Parameter uncertainty in marine ecosystem models: what can we learn from ensemble calculations and Bayesian models?

Jerome Fiechter

University of California, Santa Cruz

PICES Meeting, Khabarovsk, 19 October 2011

Co-authors:

Jeremiah Brown[1], Williams Leeds[2], Radu Herbei[3], Ralph Milliff[1], Christopher Wikle[2], Andrew Moore[4], Thomas Powell[5] and Mevin Hooten[6]

[1] NorthWest Research Associates, Colorado Division (NWRA-CoRA)

[2] Department of Statistics, University of Missouri

[3] Department of Statistics, Ohio State University

[4] Ocean Sciences Department, University of California Santa Cruz

[5] Department of Integrative Biology, University of California Berkeley

[6] Department of Fish, Wildlife and Conservation Biology, CSU

Funding: NSF GLOBEC Program

Motivation

"Model robustness to parameter uncertainty"

Ensemble statistics

- Ensemble mean and spread vs. ensemble size
- Ensemble mean and spread vs. parameter range
- Comparison with observations (SeaWiFS)

Parameter control and variability

- Identify fundamental biological processes controlling ecosystem model solutions in space and time
- Estimate optimal parameters values and uncertainty based on available observations (satellite, *in situ*)

CGOA: Physical and Biological Properties

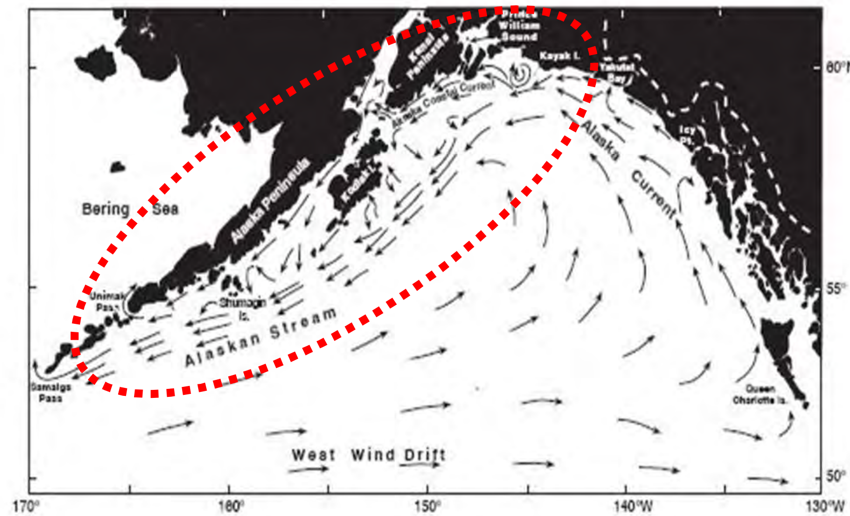
Physical Variability

- Downwelling-favorable winds
  (Stabeno et al., 2004)
- AS mesoscale variability
  (Combes and Di Lorenzo, 2007)
- Anticyclonic (Yakutat) eddies
  (Okkonen et al., 2003)

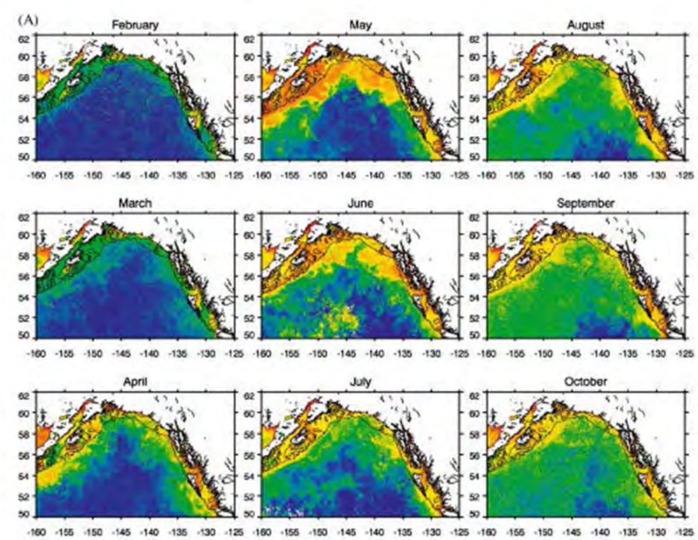
Biological Variability

- CGOA shelf: highly productive
- Subarctic Gyre: HNLC region
  (Lam et al., 2006)
- Iron limitation on phytoplankton
  (Strom et al., 2006)

860 P.J. Stabeno et al. / Continental Shelf Research 24 (2004) 859–897



P.J. Brickley, A.C. Thomas / Deep-Sea Research II 51 (2004) 229–245



CGOA: Coupled Physical-Biological Model

ROMS ocean model

- 10 km horizontal resolution
- 42 terrain-following vertical levels

Boundary/initial conditions

- Northeast Pacific (NEP) ROMS

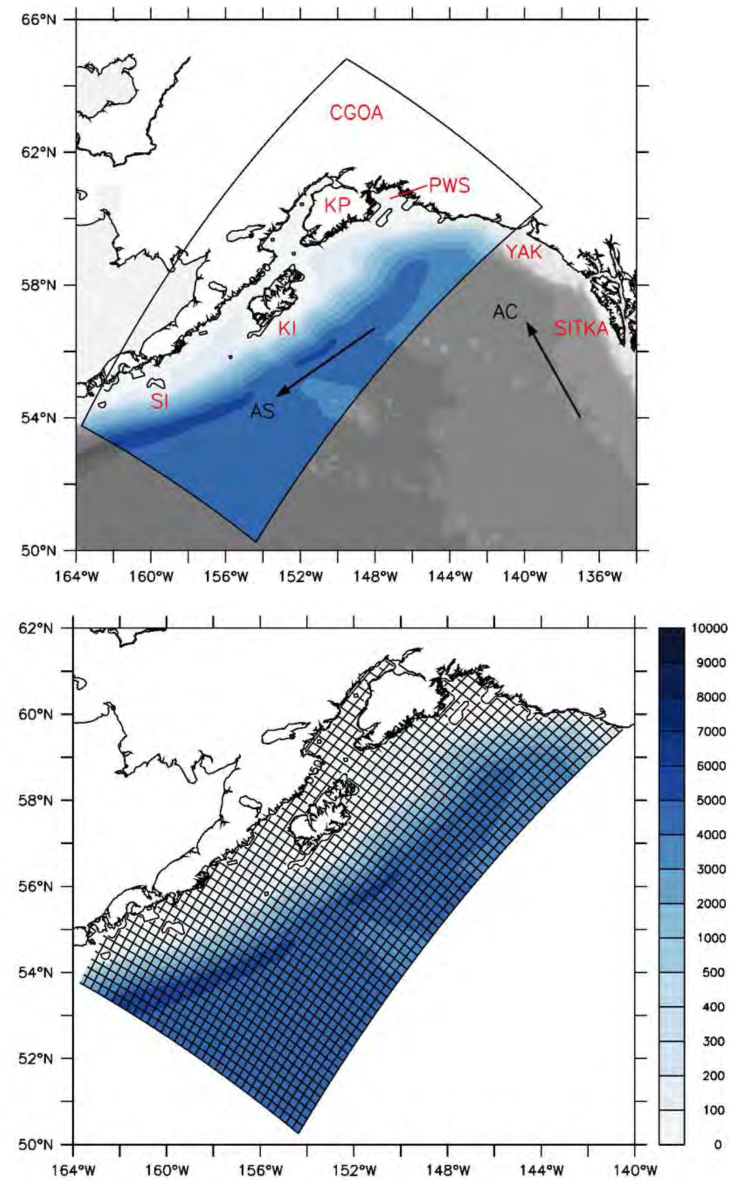
  (Curchitser et al., 2005)

Surface and river forcing

- CORE2 (Large and Yeager, 2008)
- Freshwater runoff (Royer, 1982)

4D-Var data assimilation

- Satellite SSH, SST
- In situ T, S (GLOBEC)



CGOA: Coupled Physical-Biological Model

Lower trophic level ecosystem model

- 4-component NPZD (Powell et al., 2006)
- Iron limitation (Fiechter et al., 2009)

Ensemble calculations

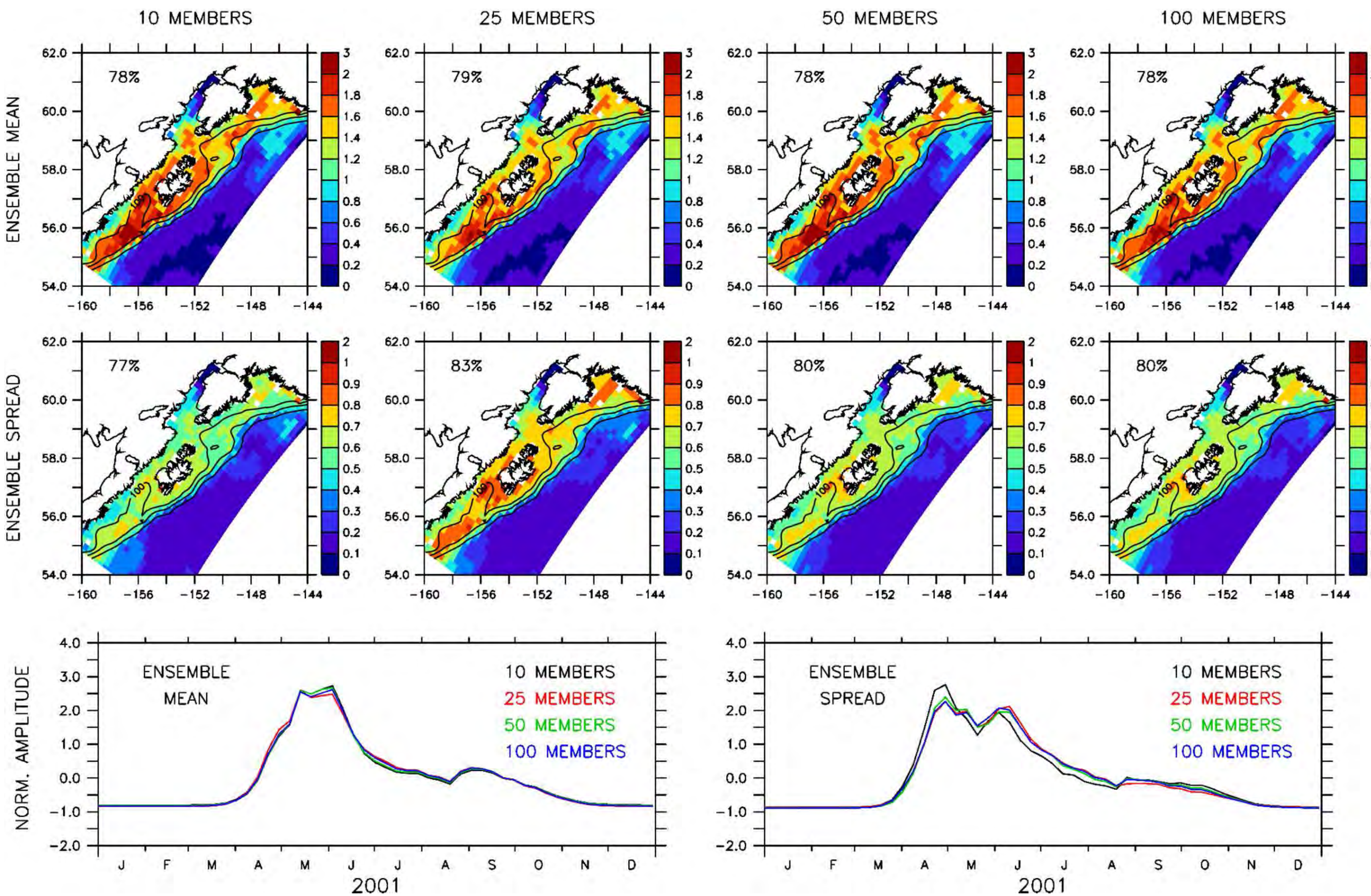
- 7 random parameters out of 17 model parameters:

  a) Phytoplankton maximum growth rate (VmNO3) and limitation by light (PhyIS), nitrogen (KNO3) and iron (KFeC)

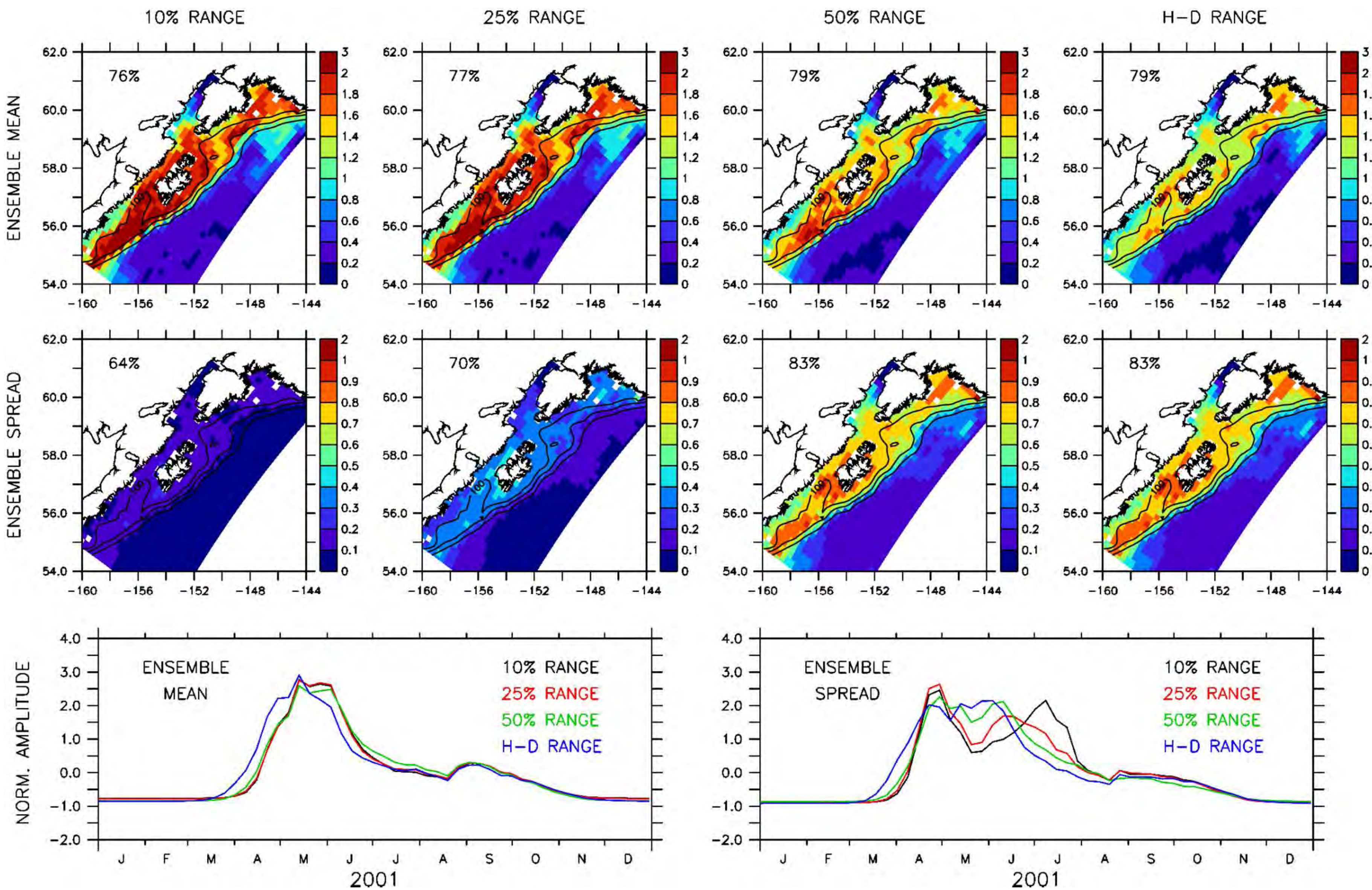
  b) Zooplankton maximum grazing rate (ZooGR)

  c) Remineralization rates for nitrogen (DetRR) and iron (FeRR)
- Parameter range: ±10%, ±25%, ±50%, and half-double
- Ensemble size: 10, 25, 50, and 100 members
- Latin Hypercube Sampling

Dependence on Ensemble Size: EOF Mode 1

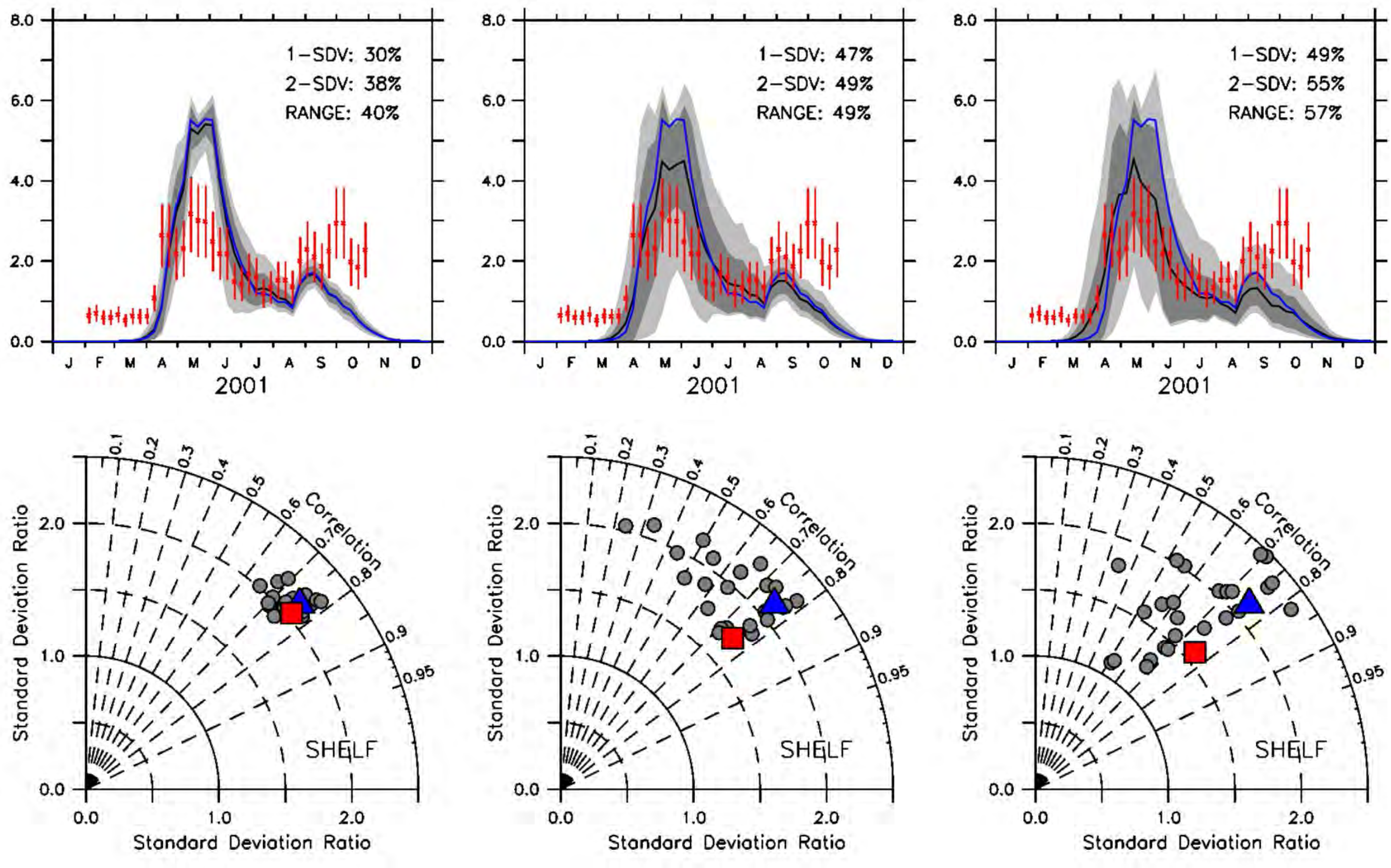


Dependence on Parameter Range: EOF Mode 1



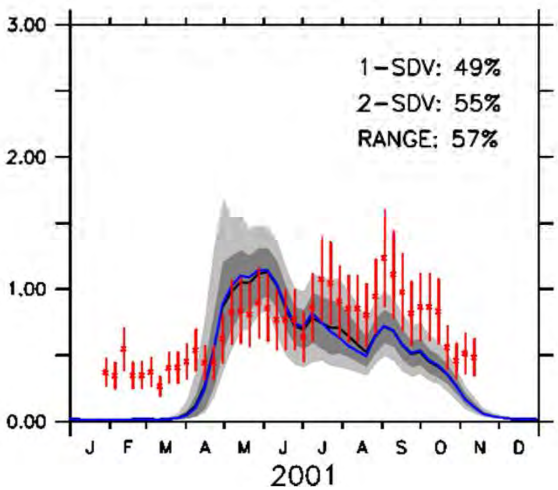
25-Member Ensembles vs. Observations: Shelf

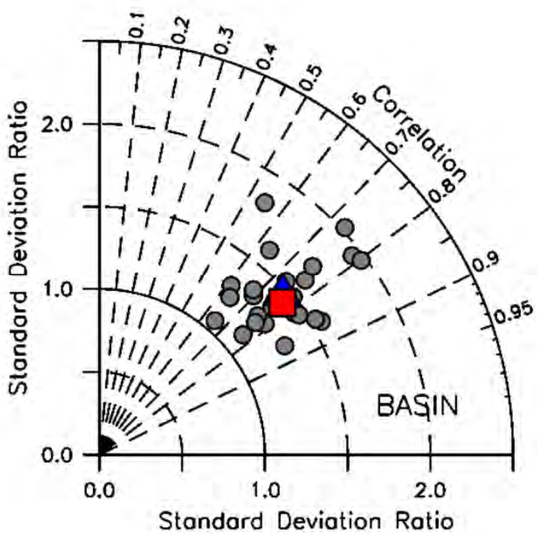
±25% Param. Range | ±50% Param. Range | H-D Param. Range



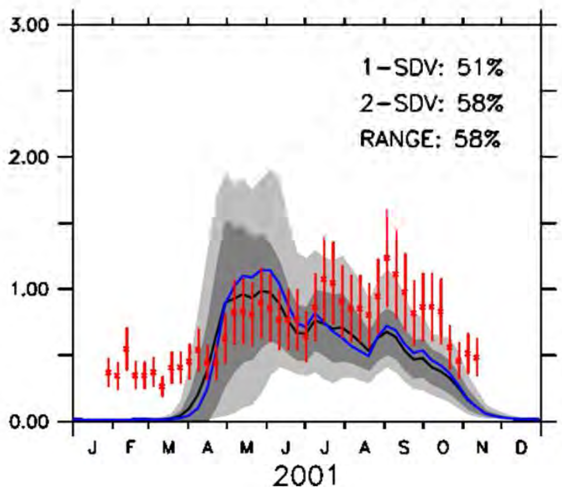
25-Member Ensembles vs. Observations: Basin

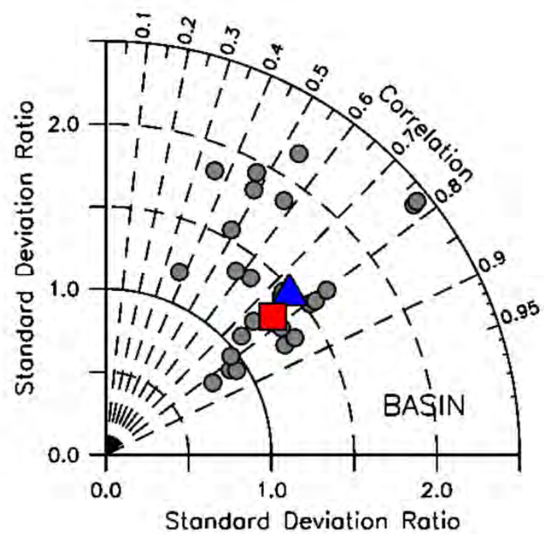
±25% Param. Range



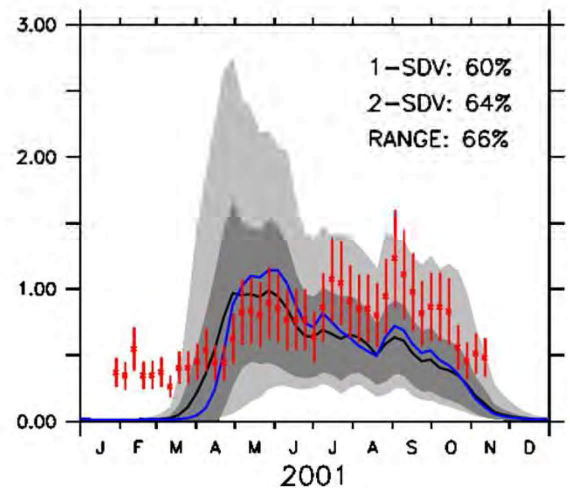


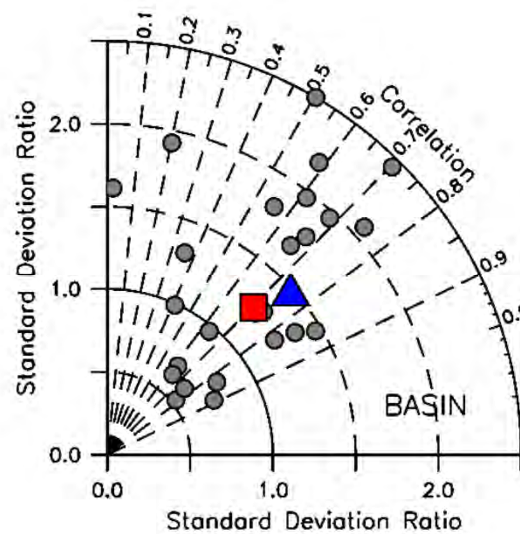
±50% Param. Range





H-D Param. Range





Parameter Control on Phytoplankton Concentrations

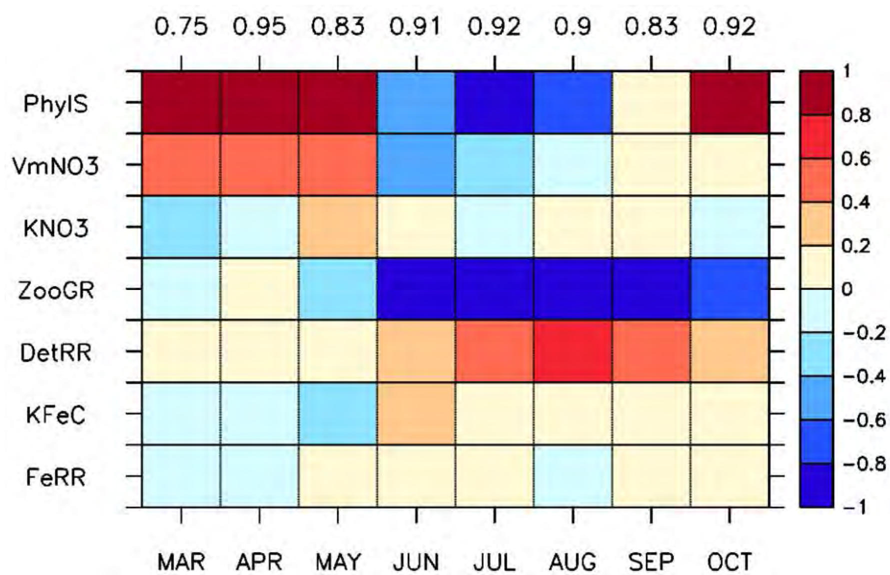
Multivariate linear regression on monthly phytoplankton concentrations:

$$P_n = a_1\theta_{1,n} + a_2\theta_{2,n} + a_3\theta_{3,n} + a_4\theta_{4,n} + a_5\theta_{5,n} + a_6\theta_{6,n} + a_7\theta_{7,n}$$

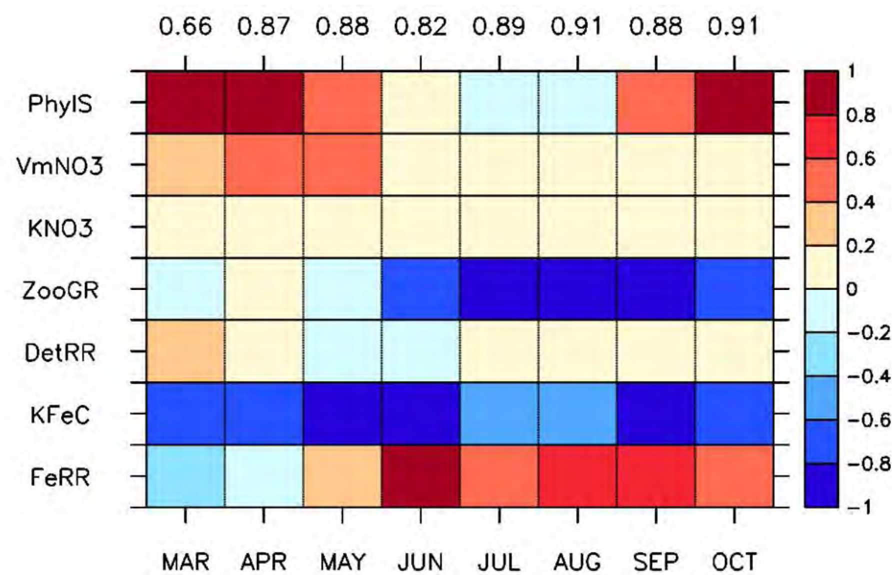
P_n = phytoplankton concentrations from n^{th} ensemble member
$\theta_{i,n}$ = i^{th} parameter value associated with n^{th} ensemble member

- a_i = regression slope for i^{th} parameter ("parameter control")

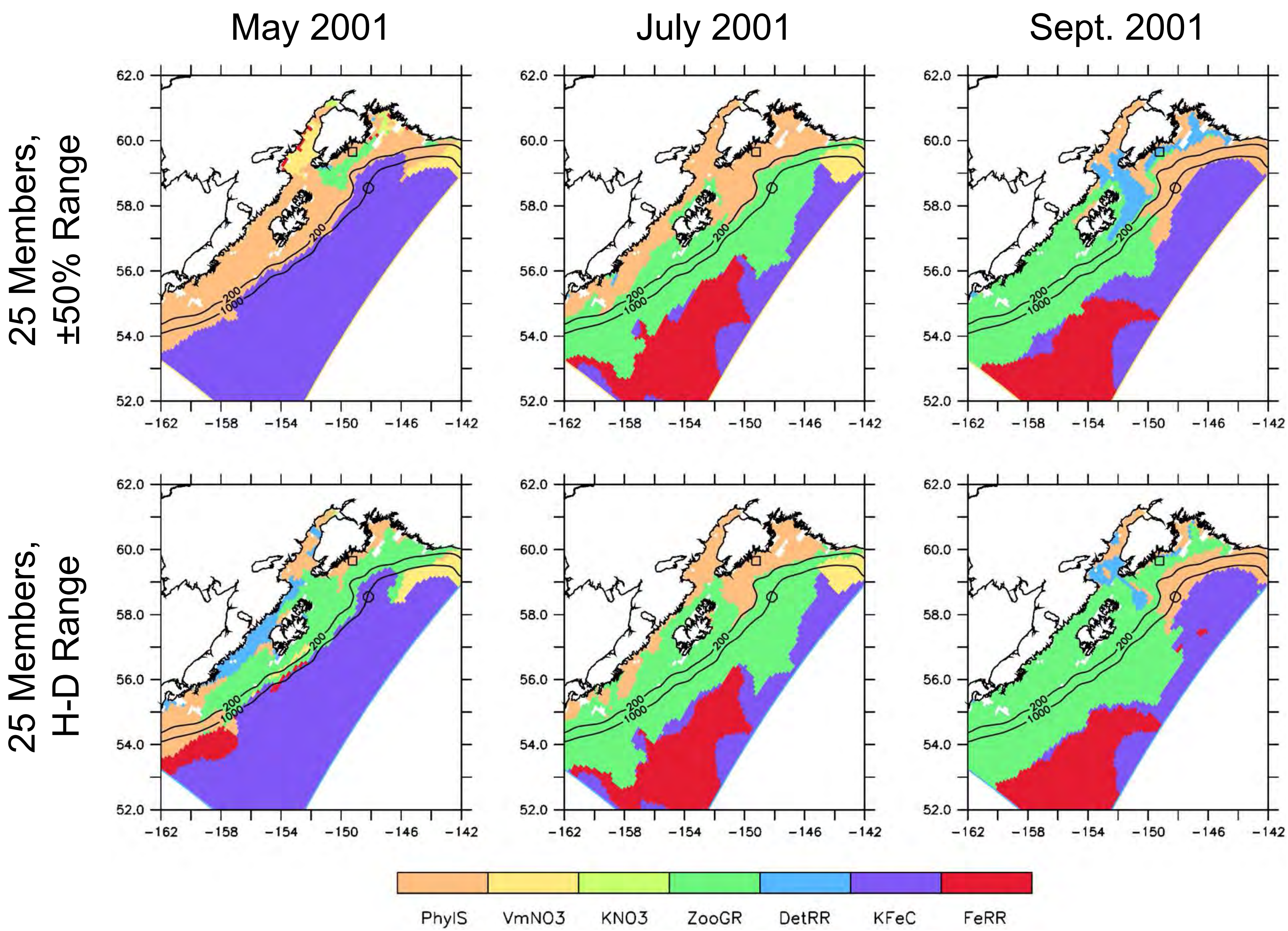
Shelf (25 members, ±50%range)



Basin (25 members, ±50% range)



Parameter Control on Phytoplankton Concentrations

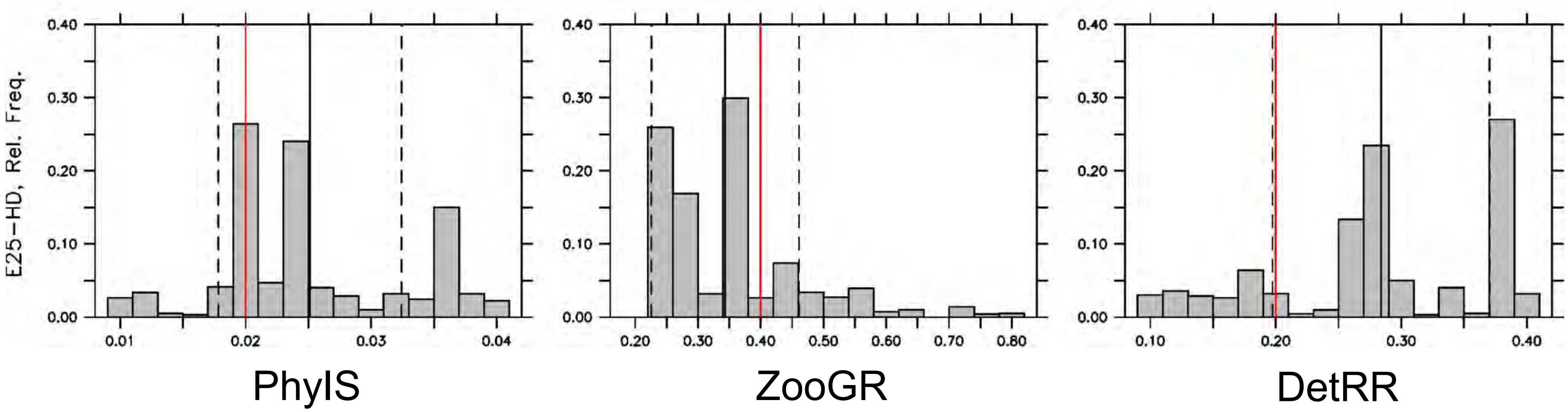


Parameter Estimation from Ensemble Members

- Parameter estimates from best ensemble members

Experiment	PhyIS	VmNO3	KNO3	ZooGR	DetRR	KFeC	FeRR
Control	0.02	0.8	1.0	0.4	0.2	16.9	0.5
Shelf best	0.029	0.55	0.81	0.42	0.12	24.79	0.61
Basin best	0.029	0.66	1.32	0.28	0.24	22.40	0.71
Domain best	0.029	0.73	0.92	0.34	0.16	21.76	0.67

- Parameter estimates from frequency histograms (shelf)



Parameter Estimation from Bayesian Model (BM)

Bayes theorem: $[\,\mathbf{X}, \theta_{\mathrm{d}}, \theta_{\mathrm{p}} \mid \mathbf{Y}\,] \propto [\,\mathbf{Y} \mid \mathbf{X}, \theta_{\mathrm{d}}\,][\,\mathbf{X} \mid \theta_{\mathrm{p}}\,][\,\theta_{\mathrm{d}}\,][\,\theta_{\mathrm{p}}\,]$

$[\,\mathbf{X}, \theta_{\mathrm{d}}, \theta_{\mathrm{p}} \mid \mathbf{Y}\,]$

Posterior Distribution (“Posterior Mean”)

- spread quantifies uncertainty (MCMC distributions)

$[\,\mathbf{Y} \mid \mathbf{X}, \theta_{\mathrm{d}}\,]$

Data Stage Distribution (“Likelihood”)

- e.g., satellite observations, in situ measurements

$[\,\mathbf{X} \mid \theta_{\mathrm{p}}\,]$

Process Model Stage Distribution (“Prior”)

- NPZD-Iron + Error Models

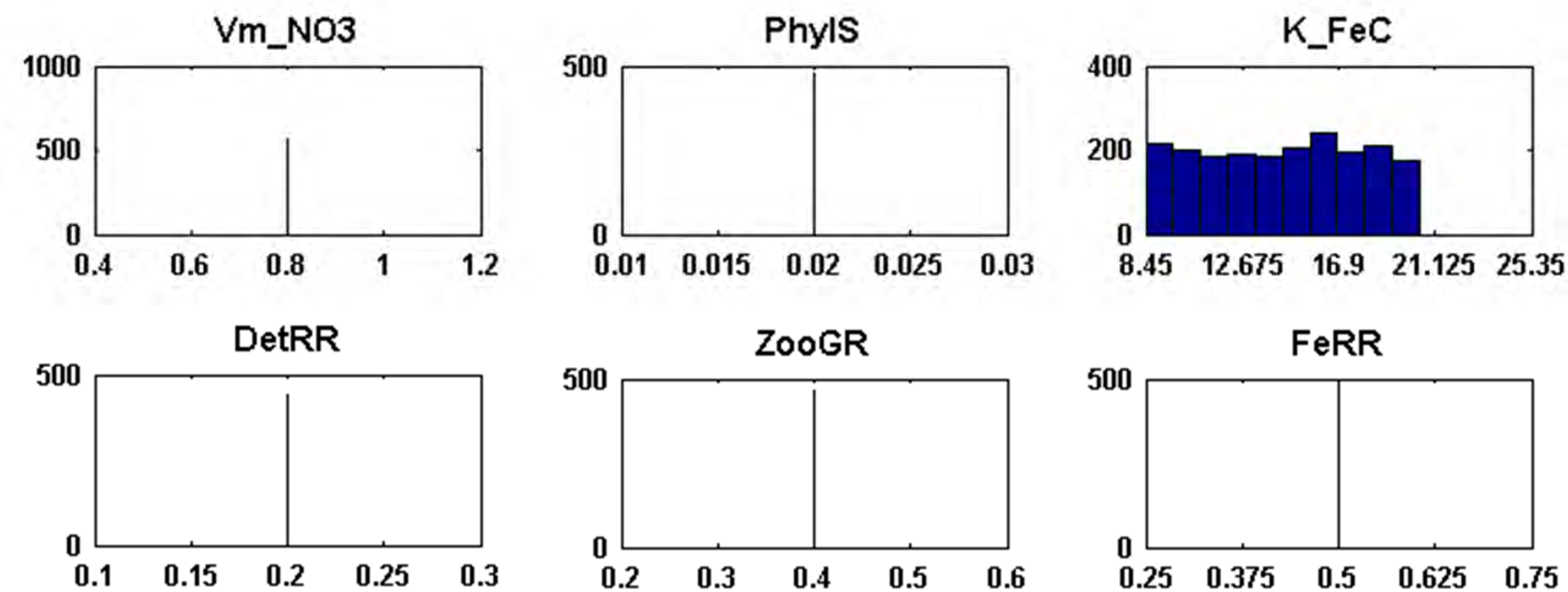
$[\,\theta_{\mathrm{d}}\,][\,\theta_{\mathrm{p}}\,]$

Parameter Distributions

- fixed vs. random parameters

1D-NPZDFe BM: Perfect Experiment, 2001

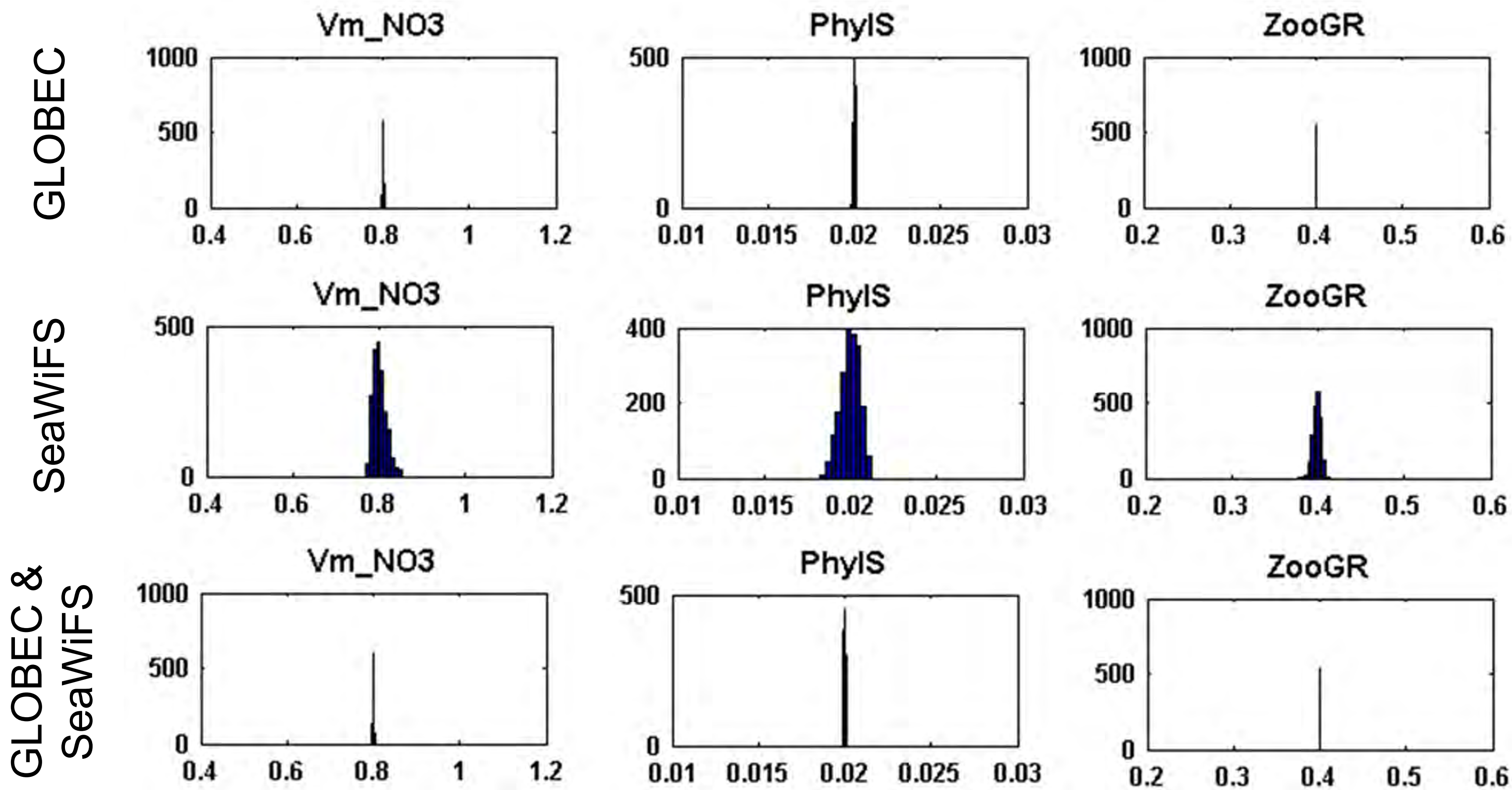
- "Observations" are generated from the 1D-NPZDFe model
- Sanity check that parameters can be recovered under "best case" scenario (all variables are known everywhere)



Parameters distribution for inner shelf along Seward line

1D-NPZDFe BM: Perfect Experiment, 2001

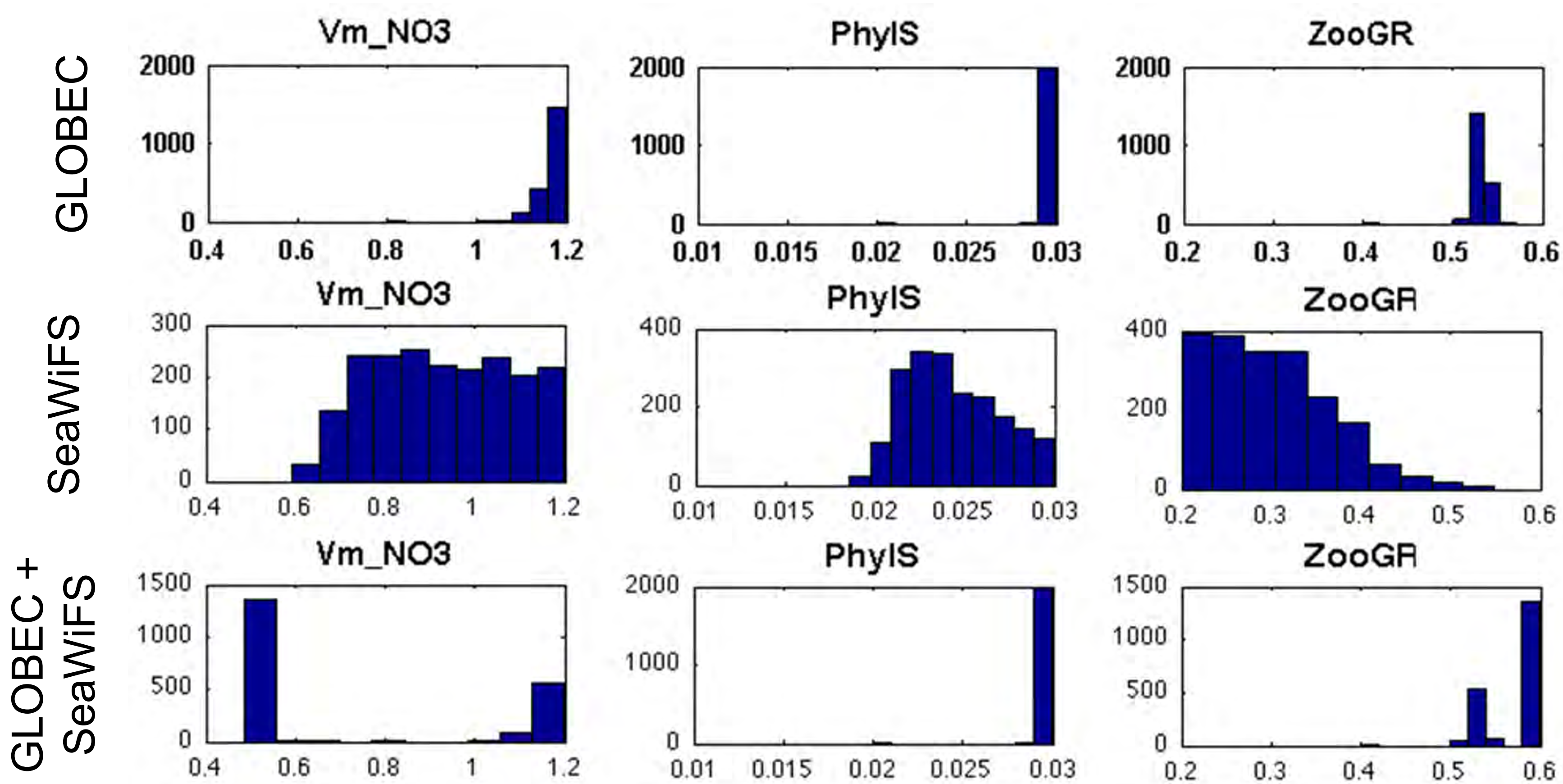
- "Perfect" data subsampled to emulate real observations (SeaWiFS Chlorophyll; GLOBEC *in situ* NO3, Chlorophyll)



Parameters distribution for inner shelf along Seward line

1D-NPZDFe BM: Real Observations, 2001

- SeaWiFS Chlorophyll; GLOBEC *in situ* NO3, Chlorophyll (SeaWiFS: daily; GLOBEC: April, May, July)



Parameters distribution for inner shelf along Seward line

Summary

Ensemble Calculations

- Ensemble statistics depend weakly on ensemble size and strongly on parameter range
- Individual ensemble members can be used to identify parameters that minimize model-data error
- Ensemble calculations can be used to identify parameters controlling variability in model solutions

Bayesian Approach

- Formal method for parameter estimations based on multi-platform observations
- May require filtering out physical variability from biological observations