Predictability of Future Recruitment by Parametric and Non-parametric models : Case study of G. of Alaska walleye pollock.

Yong-Woo Lee[1*]

Bernard A. Megrey[1]

S. Allen Macklin[2]

National Oceanic and Atmospheric Administration

[1]Alaska Fisheries Science Center

[2] Pacific Marine Environmental Laboratory

[*]Joint Institute for the Study of the Atmosphere and the Oceans/U. of Washington

Why Forecast Recruitment?

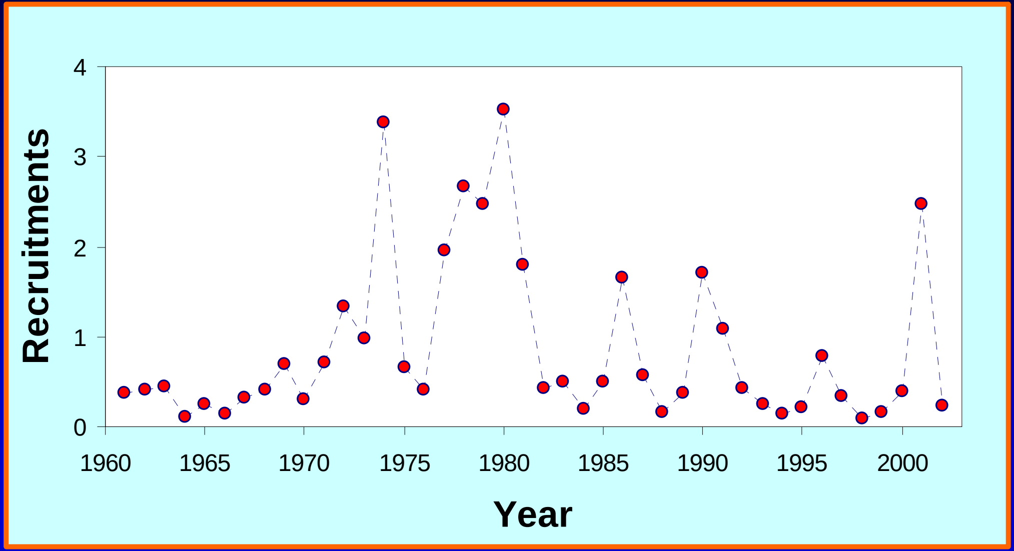
- Understand important bio-physical factors controlling the recruitment processes
- Project future stock dynamics
- Evaluate management scenarios
- Provide reference points for fishery management
- Assist commercial fisheries decision making

Problems in Forecasting

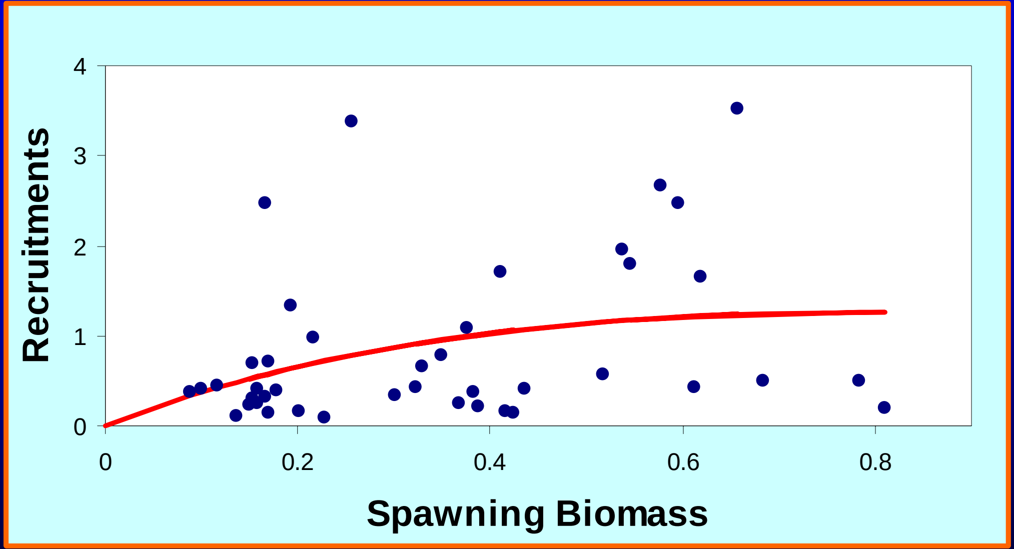
The complexity of the problem often seems beyond the capabilities of traditional statistical analysis paradigms because….

- Bio-physical relationships are inherently nonlinear
- There may be limitations in theoretical development
- Time series are too short
- Lack of degrees of freedom
- Need to partition already short time series into segments representing identified regimes

G. of A. pollock age-2 recruitments (61~02)



42 years



Ricker Model

$R = a \cdot S \cdot \exp(-b \cdot S)$

a = 4.17

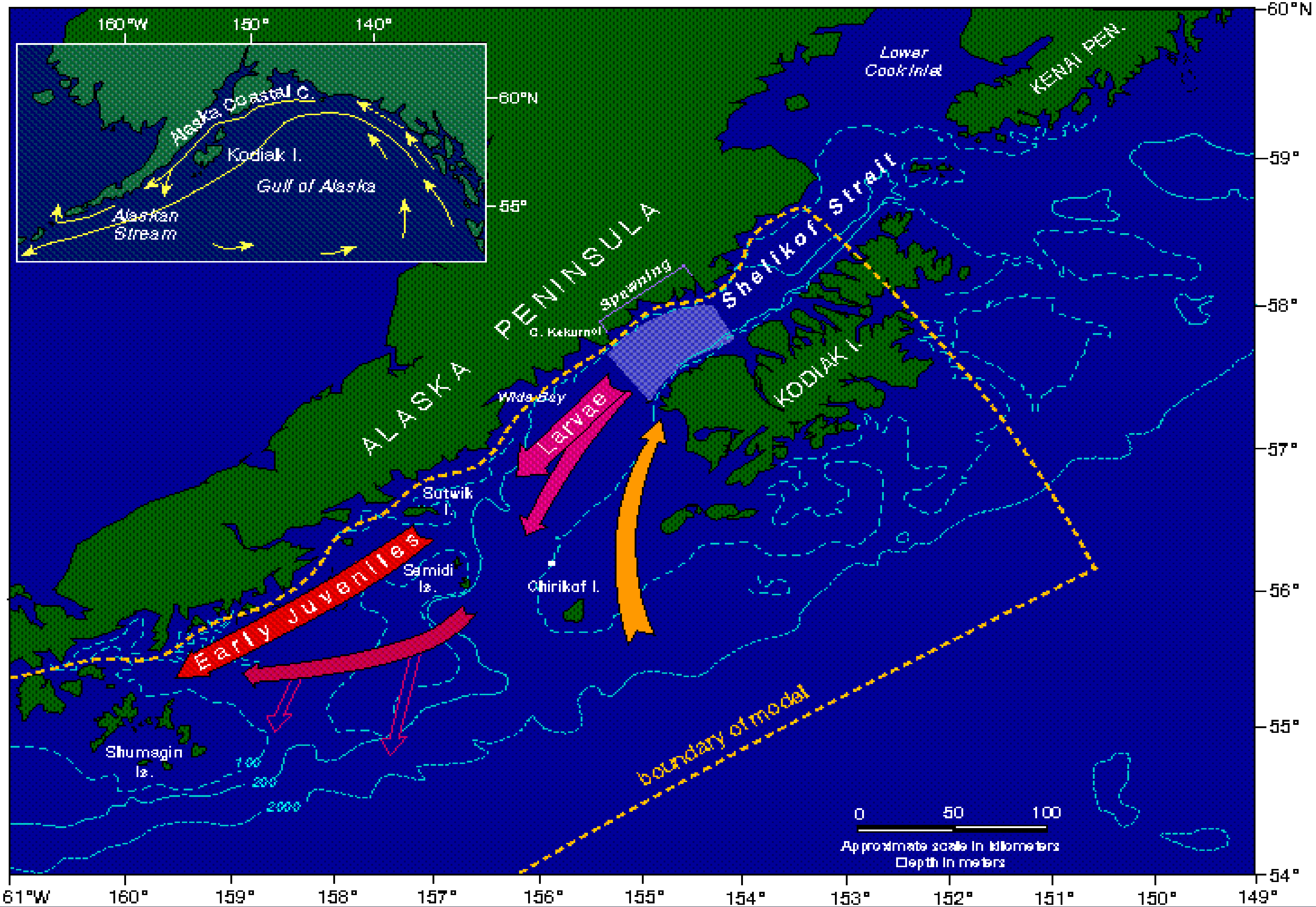
b = -1.12

r^2 = 10.4 %

Objectives

- Construct prediction models, with available environmental variables, to forecast the recruitment of Gulf of Alaska pollock.
- Test and compare several statistical methods to evaluate their ability
  - to identify recruitment-environment relationships
  - to forecast future recruitment

Early Life History of Walleye Pollock *(Theragra Chalcogramma)* Near Shelikof Strait



Examined data

Recruitment Data (response variable)

2-year old pollock, estimated from stock assess. model

Environmental Data (explanatory variables)

Annual SB + Monthly average of 6 variables (NCEP data)

- SST: Sea Surface Temperature
- WMX: Wind Mixing
- RIV: River Discharge
- NEP: North-East Pacific Pressure Index
- PDO: Pacific Decadal Oscillation Index
- SOI: Southern Oscillation Index

Examined data (cont.)

Environmental effects occur in birth year (i.e. no lags)

- recruitment and SB are <u>annual</u> data

- environmental variables are <u>monthly</u> data

- quarterly averages: ***pre-, during-, post-***spawning seasons

- resulted in total of 19 explanatory variables

2 Data Segments Partitioning (1961 ~ 2001, n=41)

- Training segment used for parameter estimation (n=35)
- Forecasting segment used for forecasting accuracy (n=6)

Tested Statistical Tools

- Multiple Linear Regression (MLR)
- Generalized Additive Models (GAM)
- Artificial Neural Network (ANN)

FISHERIES APPLICATIONS

GAM

Cury *et al*. 1995; Swartzman *et al*. 1995; Meyers *et al*. 1995; Jacobsen and MacCall 1995; Daskalov 1999

ANN

Chen and Ware 1999

Comparisons

Multiple Regression

- Conventional (good theoretical background)
- Parametric (statistical assumptions)
- Significance testing (variable selection)

GAMs & ANNs

- Innovative (computer intensive)
- Non-parametric (model-free approach)
- Flexible in function approximation

Constructing Prediction Models

- Generalized Ricker Model
  $R = \alpha \cdot S \cdot \exp(-\beta_0 S + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_p X_p)$

- 2 possible response variables: **log(R), log(R/S)**

- 2 strategies in building a prediction model
  - no prior assumption on density dependency
  - prior assumption on density dependency (SB is forced into model selection process)

- 3 statistical methods: MLR, GAM, ANN

2 responses × 2 strategies × 3 stat-methods = 12 models

Variable Selection Methods** in **MLR

Forward

-Start from empty model, F-stat-to-enter

-Sequentially entered from the most significant terms

Backward

-Start from full model, F-stat-to-remove.

-Sequentially removed from the most insignificant terms

Stepwise

-Mixture of *Forward* and *Backward*

Model Selection among All Subsets

Model-fitting Criteria

- *Mallows' **Cp** statistic*
- *Bayesian Information Criterion (**BIC**)*
- *Akaike Information Criterion (**AIC**)*

** 19 variables => 343 possible combinations of subsets.*

** each model set receives numerical score based on criterion statistic.*

MLR

- selected the variables for each model based upon the agreement of different variable selection techniques (forward, backward, stepwise, Cp, AIC, BIC)

GAM

- selected the variables based on AIC

ANN

- used the selected variables in MLR

Selected Best Prediction Models

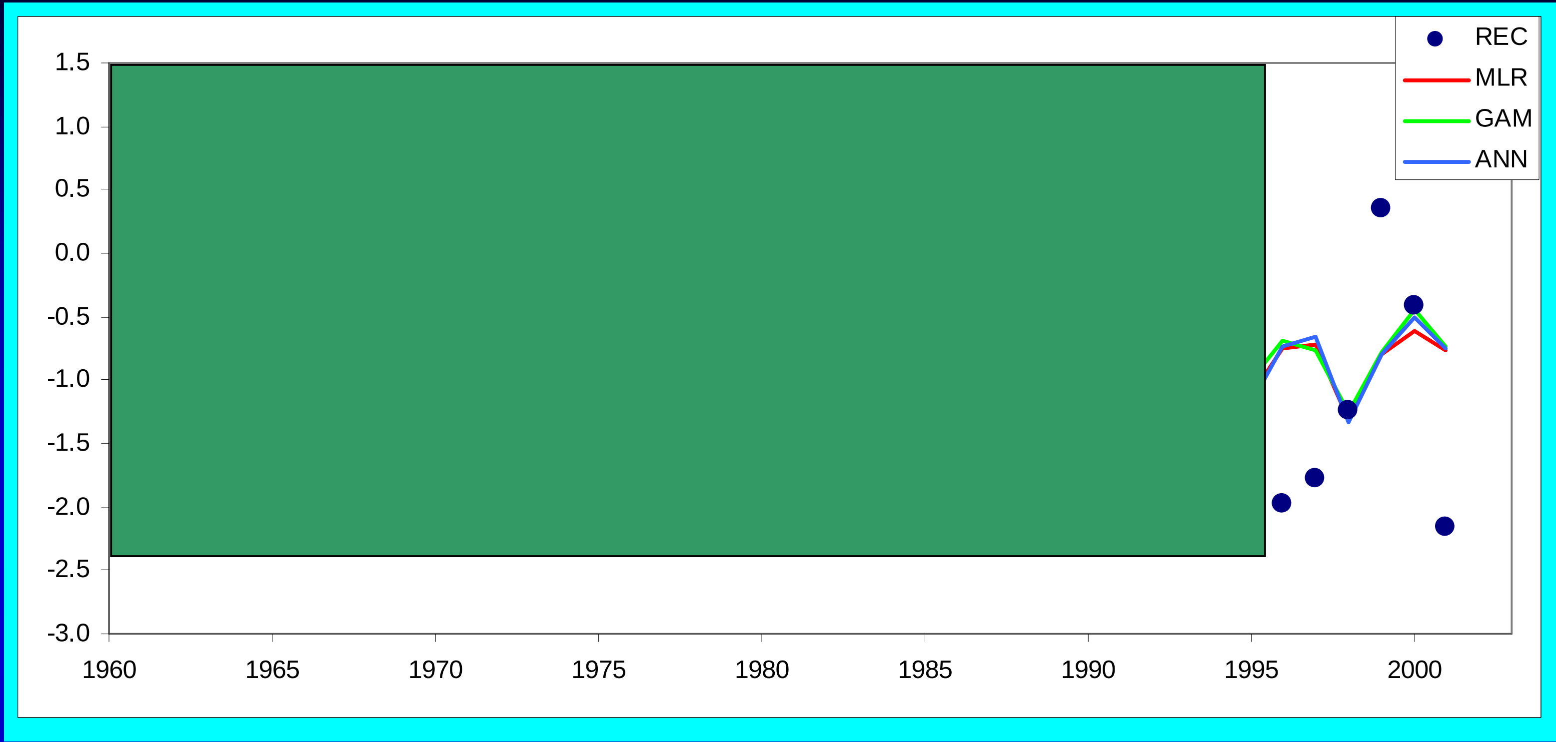
- **MLR with log(R)**
  no Prior: WMX1 + WMX3
  Prior: SB + SST1 + WMX1 + NEP1 + PDO3

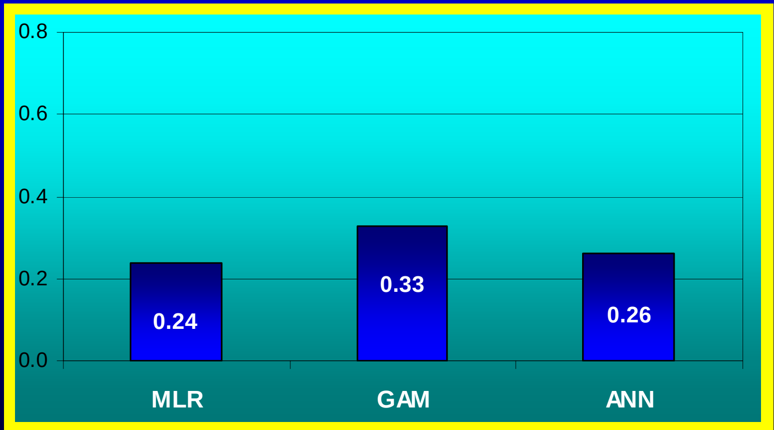
- **MLR with log(R/S)**
  no Prior: SST1 + WMX1 + WMX3 + NEP1 + PDO3
  Prior: SST1 + WMX1 + WMX3 + NEP1 + PDO3 + SB

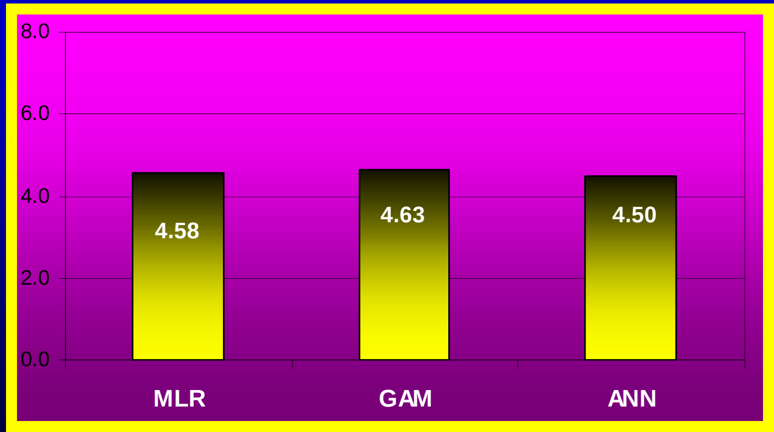
- **GAM with log(R)**
  no Prior: *S*(WMX1,df=2) + WMX3
  Prior: SB+SST1 + *S*(WMX1,df=2) + NEP1 + PDO3

- **GAM with log(R/S)**
  no Prior: *S*(SST1,df=2) + PDO3
  Prior: SB + *S*(WMX3,df=2) + PDO3

Log(R) vs Predictions (no density dependence)



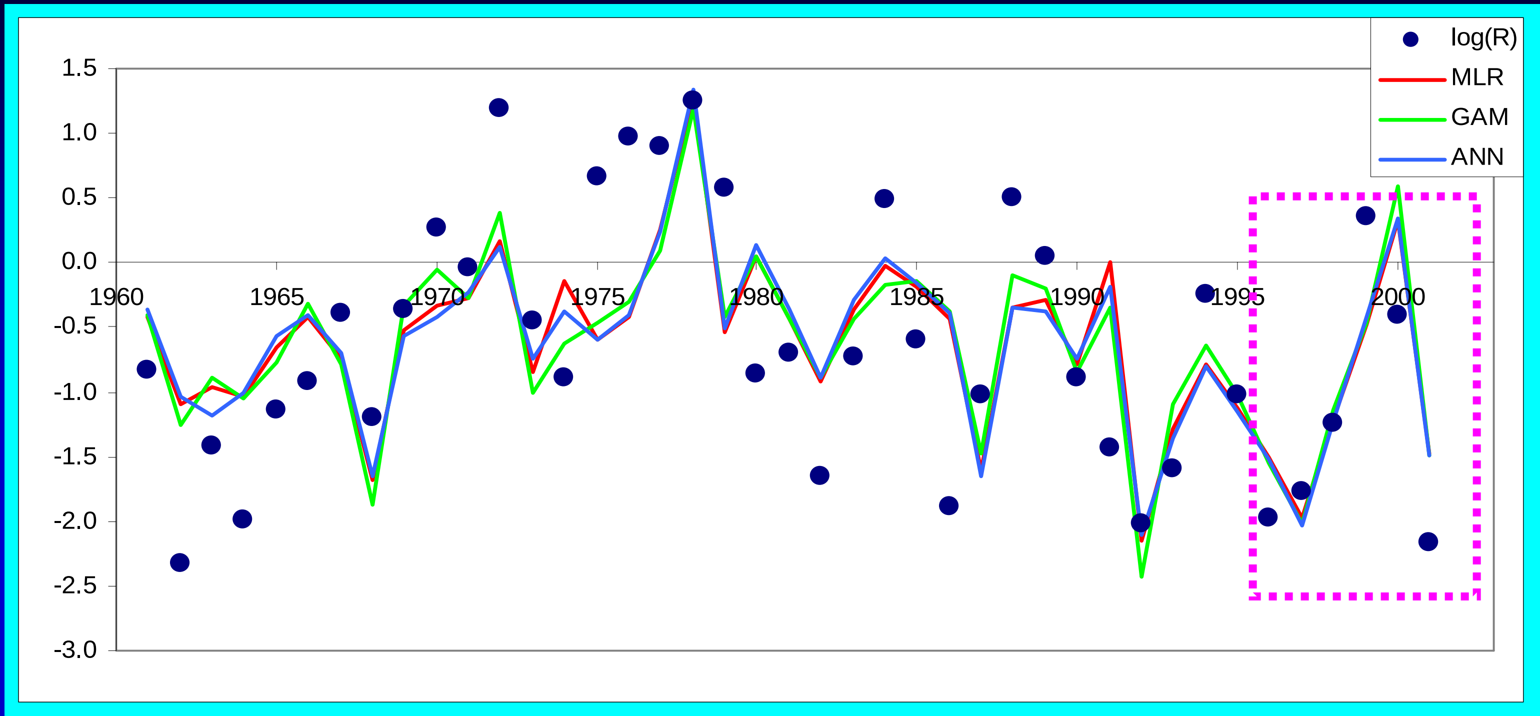


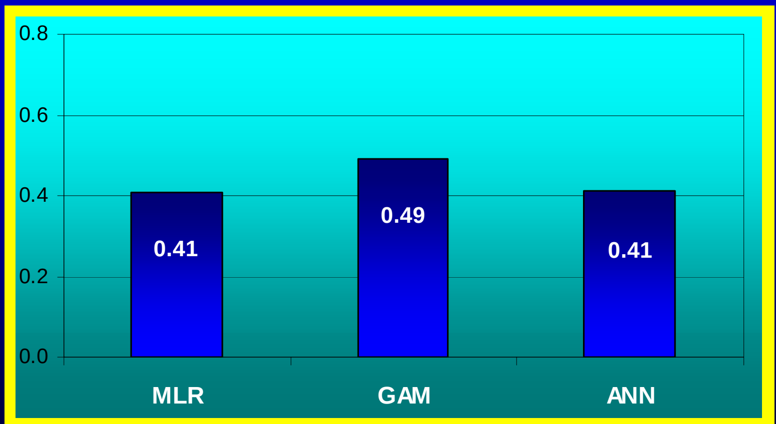


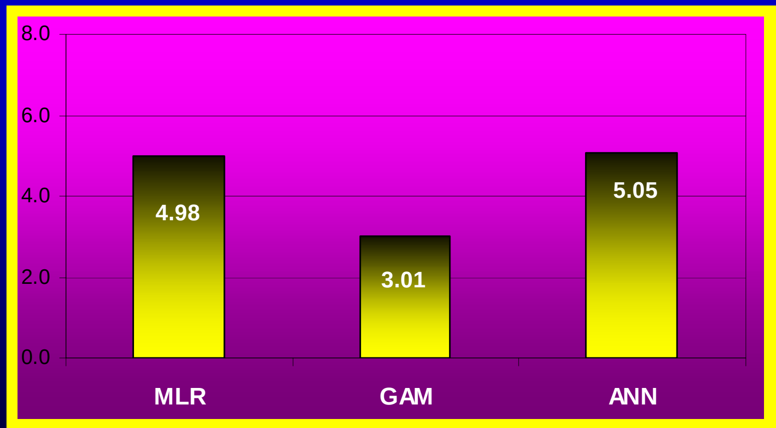
R-square for Training

MSE^{-1} for Forecasting

Log(R) vs Predictions (density dependence)







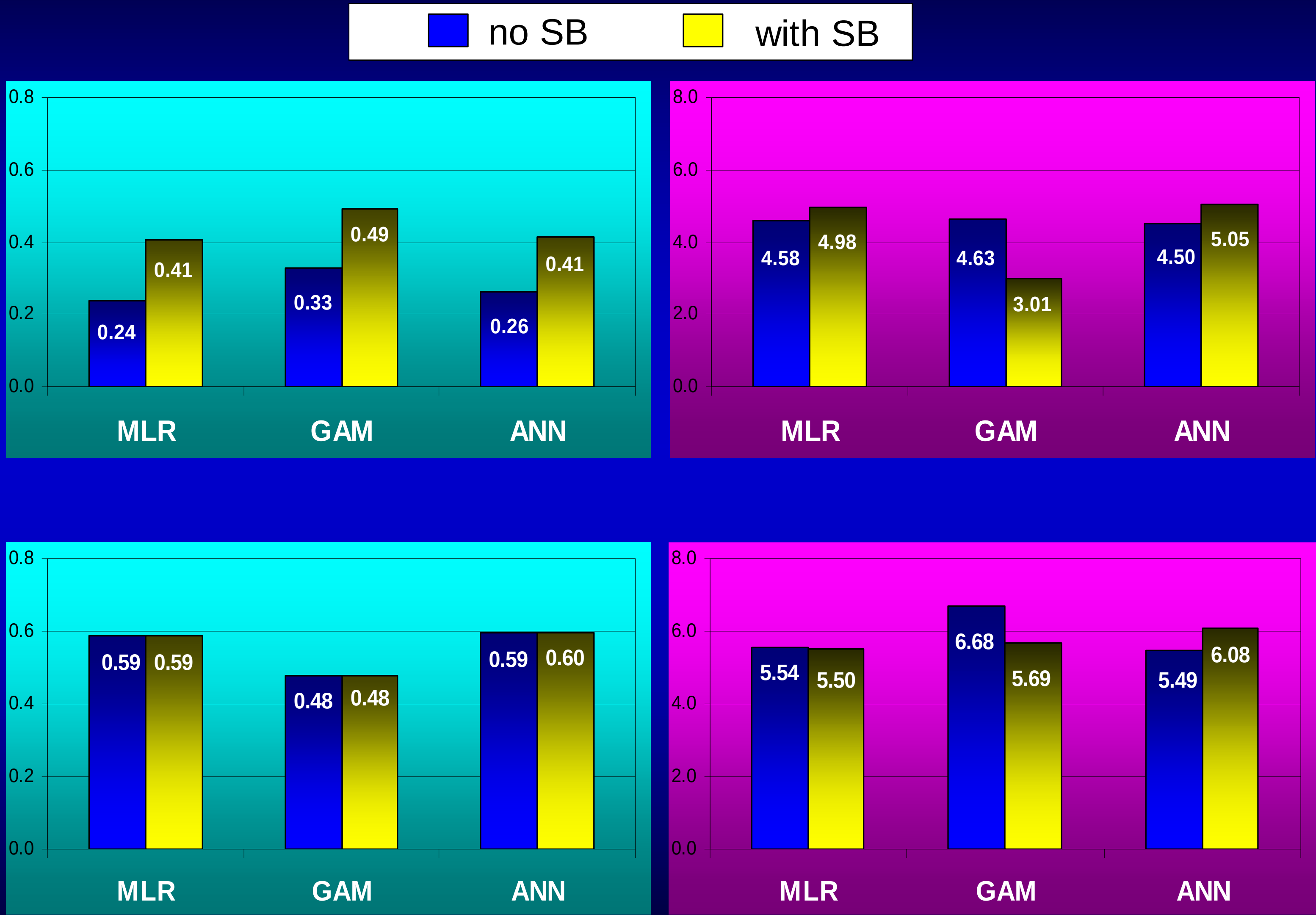
R-square for Training

MSE^{-1} for Forecasting

Summary of Model-fittings and Predictability

L(R)

L(R/S)



Training Segment (R^2) *Forecasting Segment (MSE^{-1})*

Summary of Findings

- Conditions of pre- and post-spawning seasons seem to play a big role in recruitment success: SST, WMX, NEP, PDO
- Modeling with log(R/S) performs better both for training (goodness-of-fit) and forecasting (predictability)
- There is little evidence that density-dependency is present in the model of log(R/S).
- Non-parametric methods are flexible and show promise for forecasting, thus using GAMs and ANNs together with more traditional methods should enhance analysis and forecasting.

Questions?

? ?