Comparative Analysis of Statistical Tools to Identify Recruitment-Environment Relationships and Forecast Recruitment Strength

Yong-Woo Lee (AFSC, NOAA)
Bern Megrey (AFSC, NOAA)
Allen Macklin, (PMEL, NOAA)

* NOAA: National Oceanographic & Atmospheric Administration
* AFSC: Alaska Fisheries Science Center
* PMEL: Pacific Marine Environmental Laboratory
* FOCI: Fisheries Oceanography Coordinated Investigations

Recruitment Forecasting

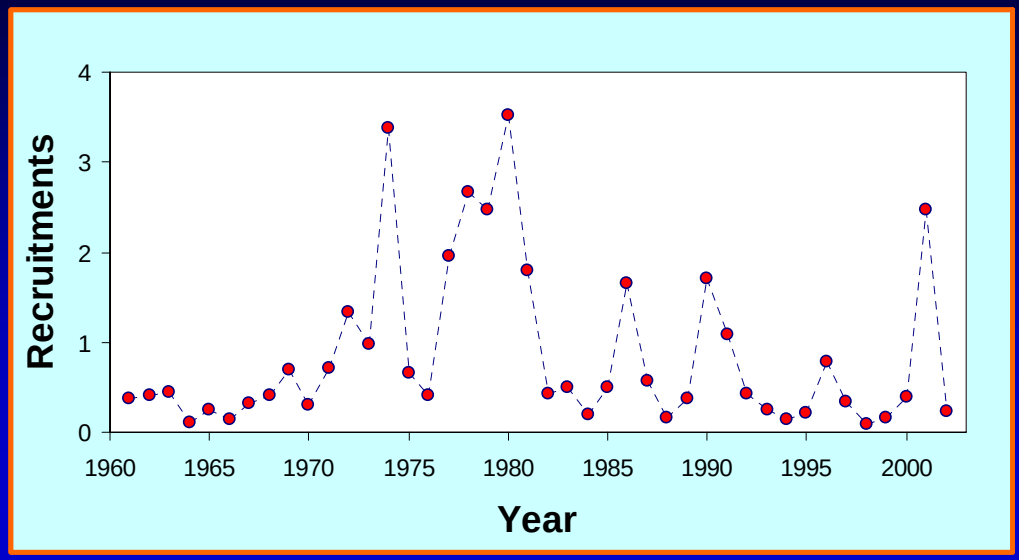
- Take recruitment estimates from the stock assessment models as the starting point.
- Attempt to relate trends in recruitment to bio-physical factors based on working conceptual model.
- GOAL: generate annual recruitment forecast.
- Why? To project the future stock dynamics & to provide reference points for fishery management.

Problems in Forecasting

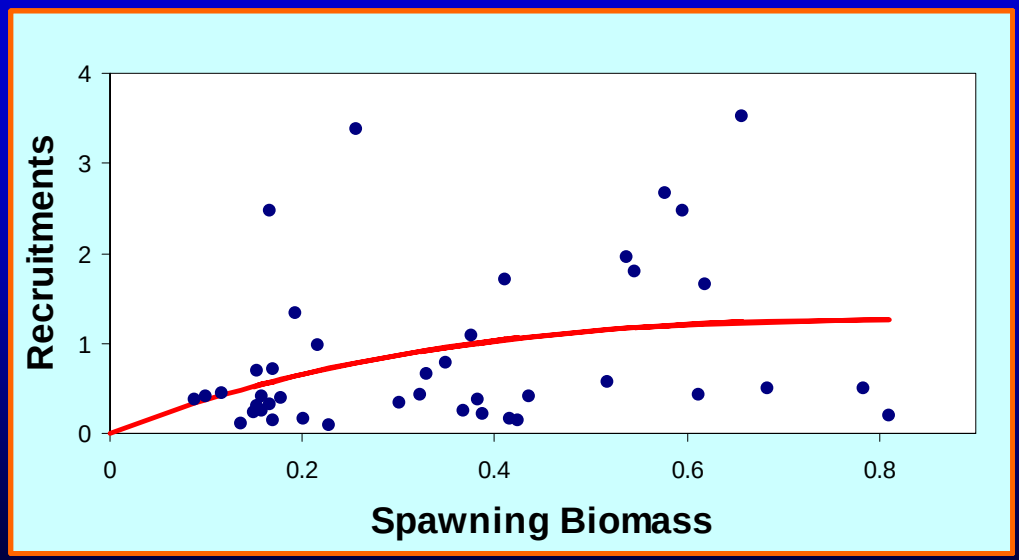
The complexity of the problem often seems beyond the capabilities of traditional statistical analysis paradigms because….

- there may be limitations in theoretical development
- inadequate length of time series
- the need to partition already short time series into segments representing identified regimes
- lack of degrees of freedom
- inability to meet required assumptions.

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 %

Reasons of Simulation

- Because we can never know the parameters and underlying relationships of actual data
- We can simulate data with known properties and different levels of measurement error

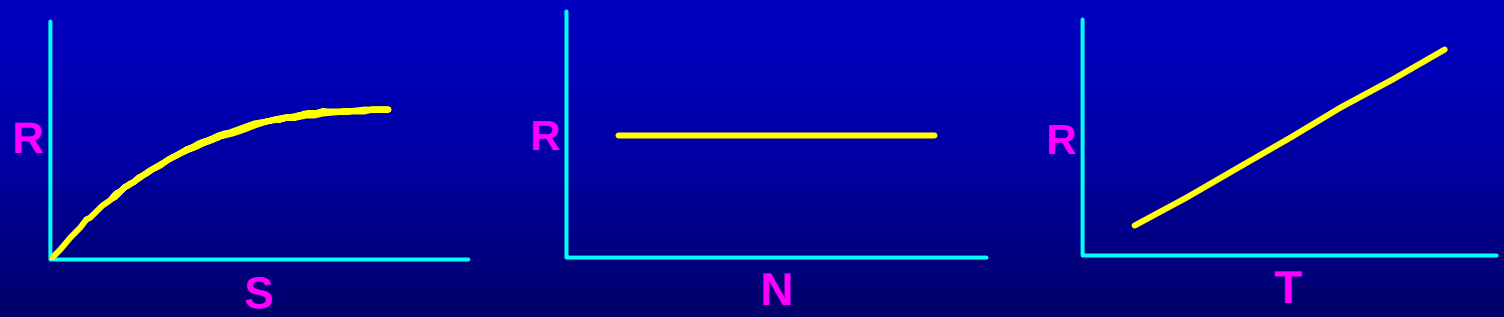
* Study Objective: to test and compare the several methods, especially their ability to forecast future recruitment.

Simulated Data with Known Properties

$$R = a \cdot S \cdot exp(-b \cdot S + c \cdot N + d \cdot T + \varepsilon)$$

R: Recruitment
S: Spawning Biomass (negative binomial distribution)
N: No relationship (lognormal dist.)
T: Temperature (normal dist.)

ε: Error, $N(0,\sigma^2)$, $\hat{\sigma}^2$ was estimated from a Ricker fit on actual data.



*** 3 Error levels: [no error] [½ $\hat{\sigma}^2$] [$\hat{\sigma}^2$]

Tested Statistical Tools

R on absolute scale (billion MT)

- Nonlinear Regression (NLR)
- Generalized Additive Models (GAM)
- Artificial Neural Network (ANN)

R on categorical scale (High, Mid, Low)

- Multinomial Logistic Regression (MLT)
- Probabilistic Neural Networks (PNN)

Parametric vs Non-parametric
Conventional vs Innovative (NNs)

Variable Testing

- Best selected GAM models

  Err = 0; R = S(2) + T(2)

  Err = 1; R = S + T(2)

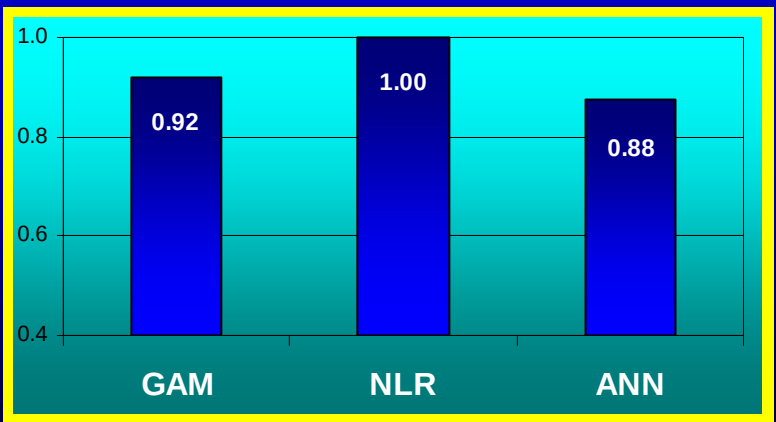
  Err = 2; R = S + N + T

- N was significant in NLR model for error level 3.

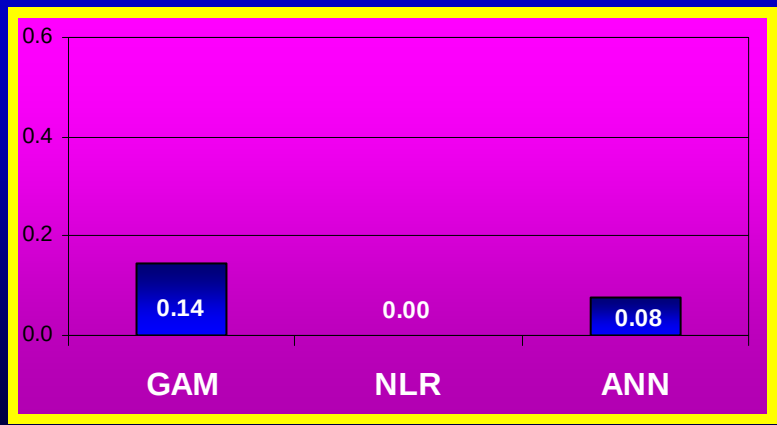
➔ It's possible to kid ourselves when dealing with highly variable data, by including unnecessary or irrelevant variables in the model as significant ones.

* For the comparisons of model performance across methods, we only used variables of SB & TEMP in the models.

Simulated vs Predicted, for Error level = 0

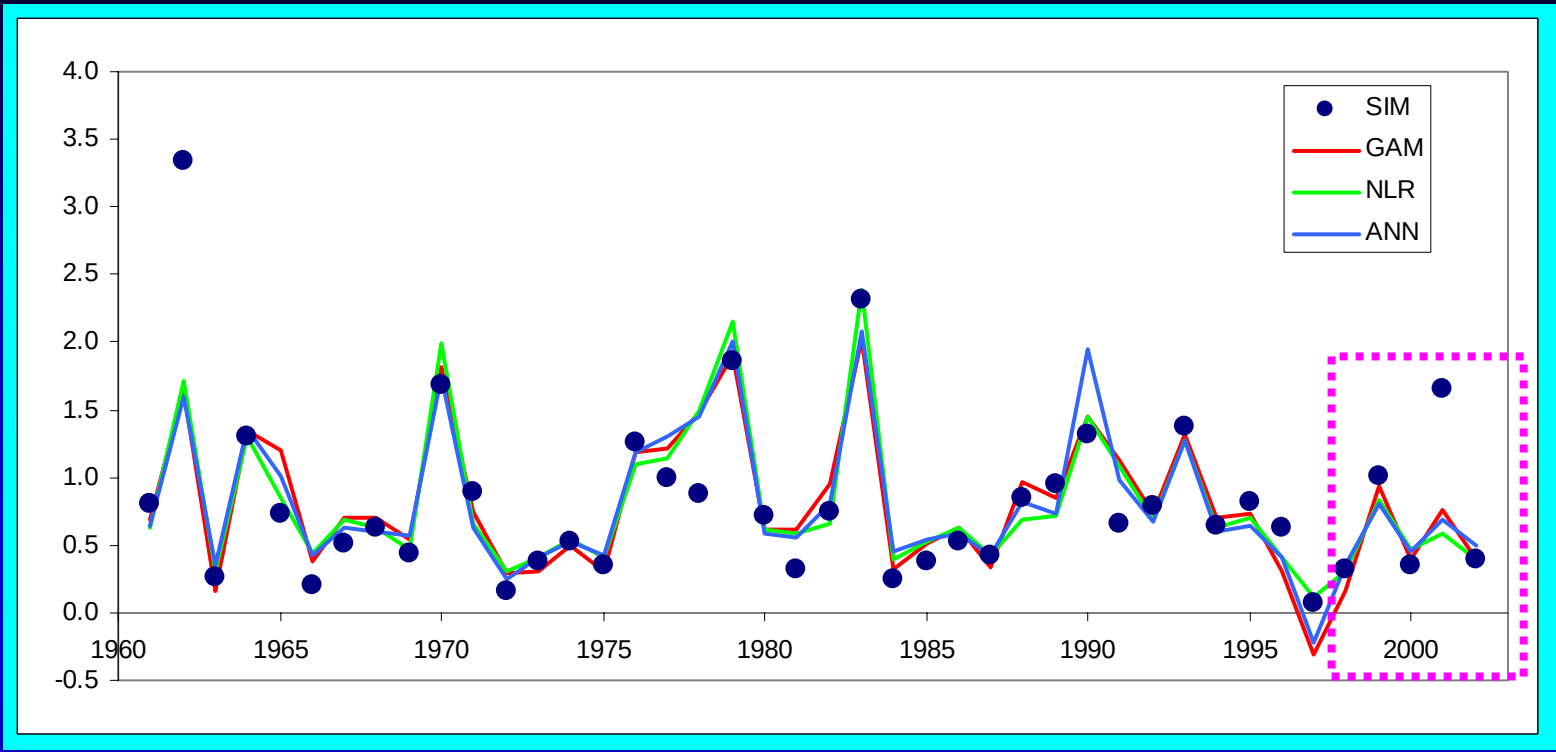


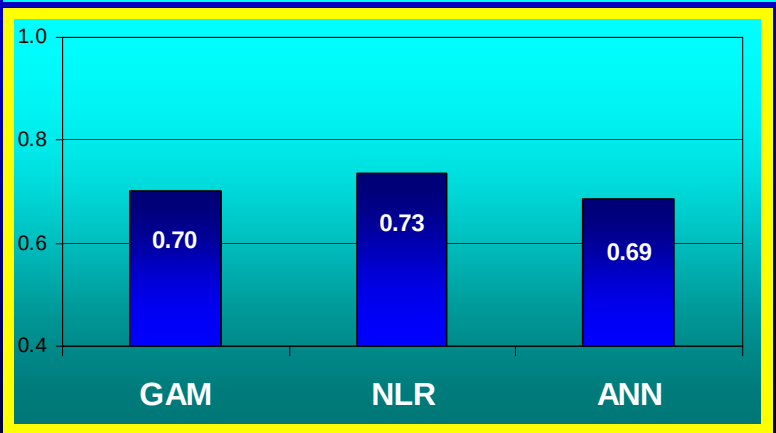
R-square for Training

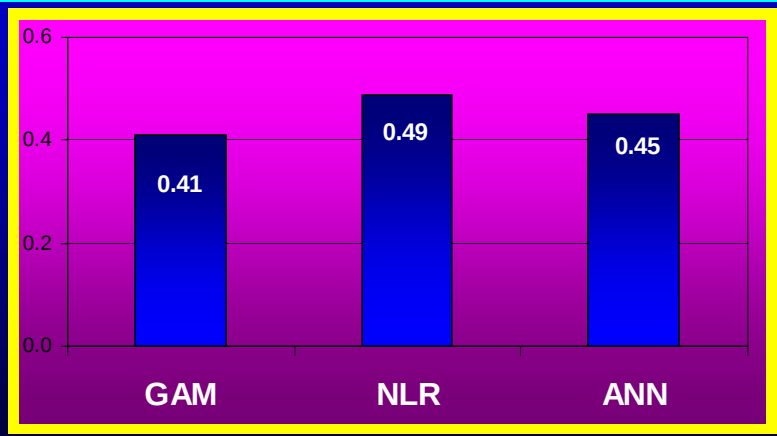


SSE for Forecasting

Simulated vs Predicted, for Error level = 1



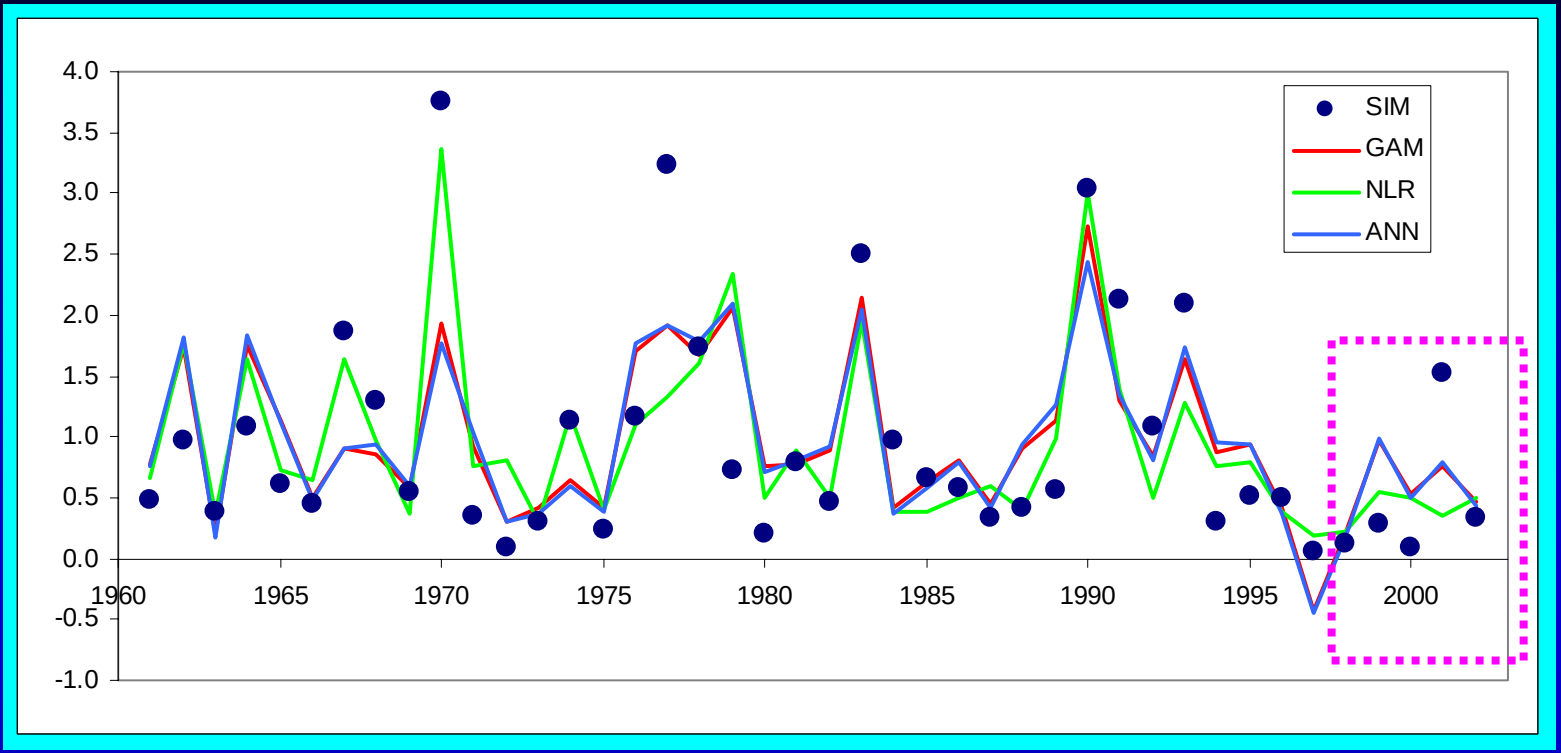


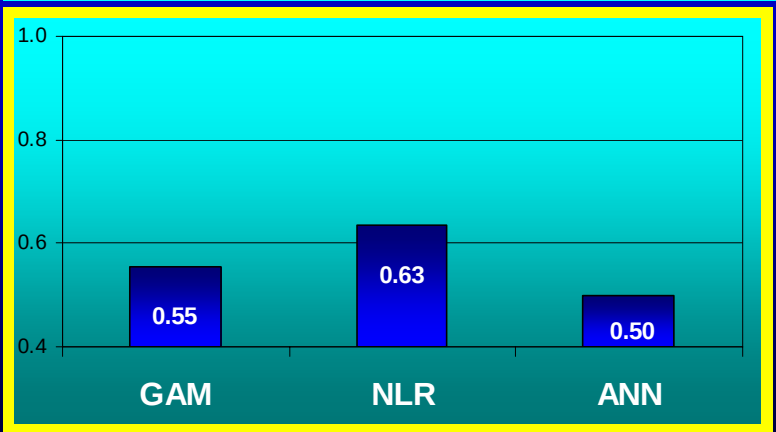


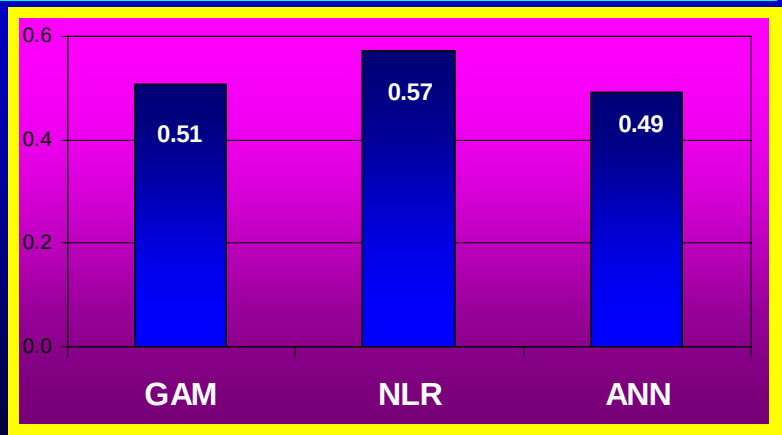
R-square for Training

SSE for Forecasting

Simulated vs Predicted, for Error level = 2



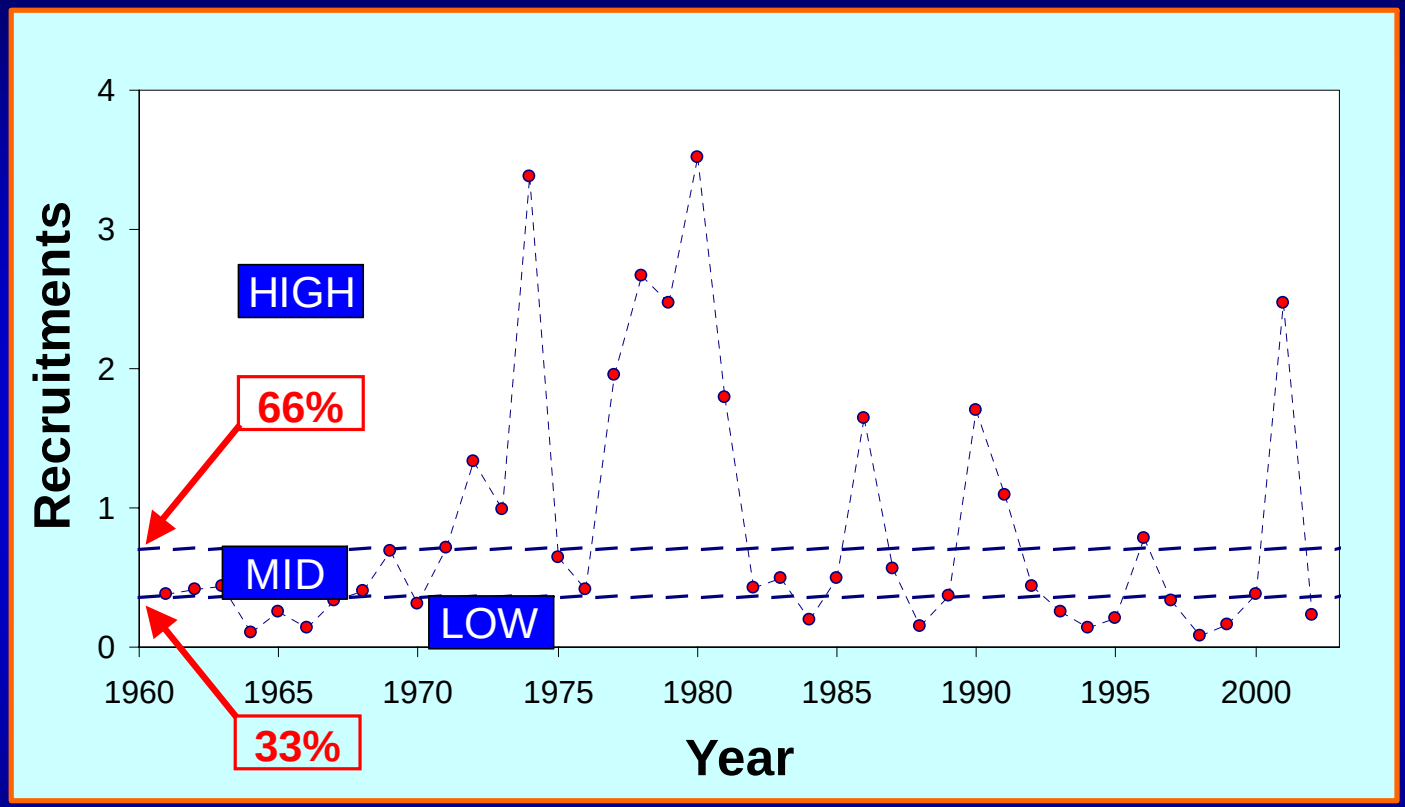




R-square for Training

SSE for Forecasting

Recruitments on Categorical Scale

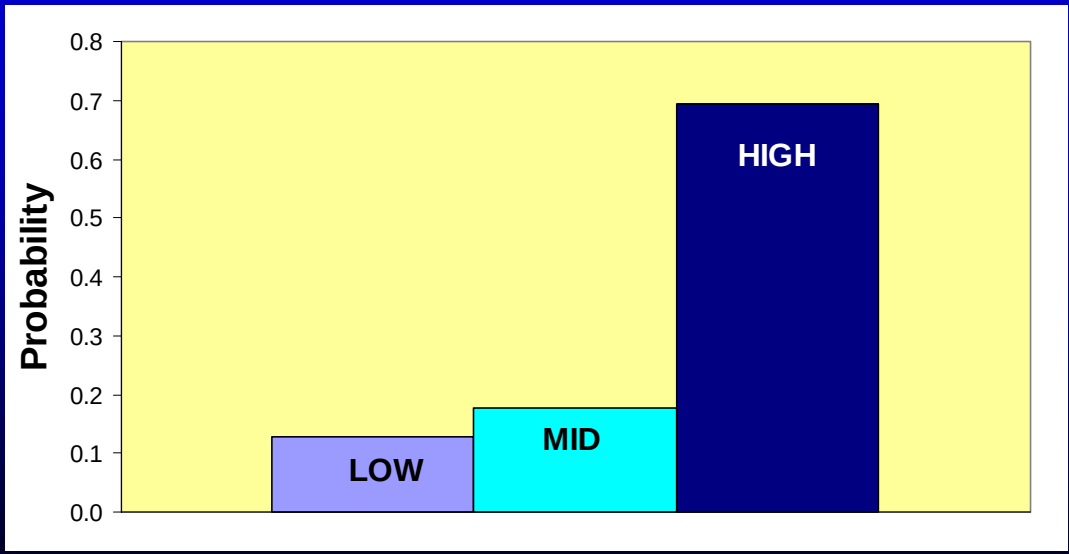


- *Multinomial Logistic Regression (MLT)*
- *Probabilistic Neural Network (PNN)*

Benefits of Using Categories

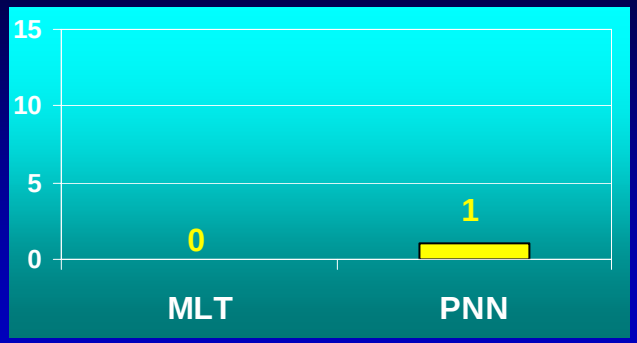
- It's easier for managers to understand the process and make decisions based on categories.
- The forecasting results provide the probability distribution for the future recruitment categories.

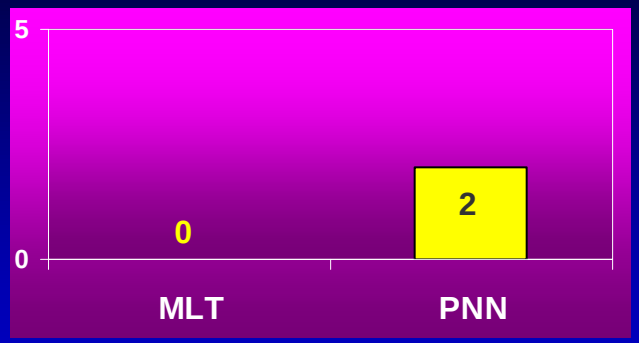
Actual = High



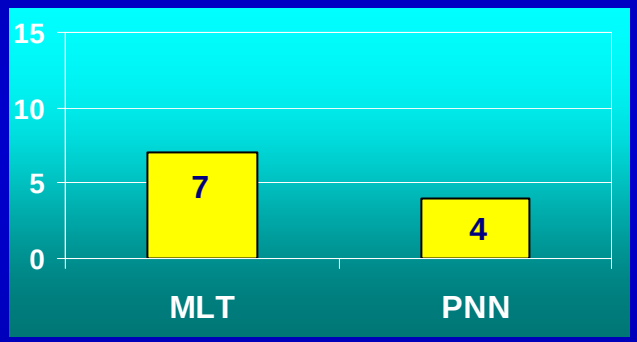
Number of Mis-classifications for Each Error Level

0

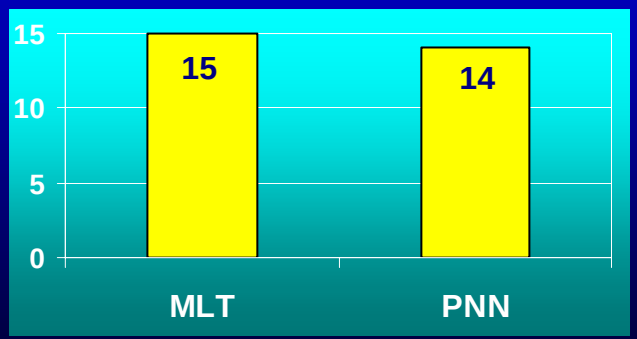


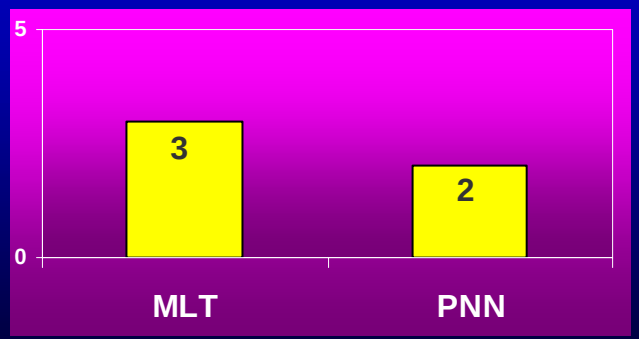


1



2





Training Segment (38) *Forecasting Segment (5)*

Summary

- Need to be cautious to deal with noisy data, because it could result in a wrong model as best.
- It appears that NNs handle noisy data better than conventional parametric methods.
- Non-parametric methods (GAM & NNs) would be better for describing the relationship & forecasting the future recruitment levels, because the real system is highly non-linear with high interactions among the variables.
- GAM is good for identifying the relationship & NNs are highly flexible and powerful for forecasting, thus utilizing them together would enhance the analysis and forecasting.

Future Research

- Extend the study to Monte Carlo simulation.
- Apply the methods to actual recruitment data to test the performance of methods & to improve our current forecasting scheme.
- Explore the different architectures of ANNs to find a suitable one for fish recruitment data.

Questions