# Nonlinear Forecasts of the Gulf of Maine Zooplankton Ecosystem

Barbara A. Bailey
Department of Mathematics and Statistics
San Diego State University

Collaborator: Andrew Pershing (University of Maine and GMRI)

## Outline

- Modeling Calanus Abundance
- Nonlinear Models and Systems
- Application: Lorenz System
- Forecasts and Intervals
- Future Work

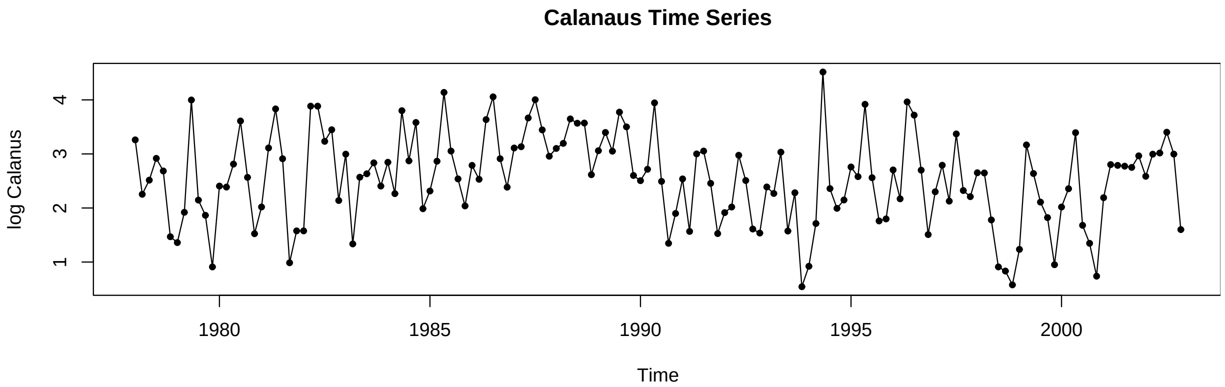
# Nonlinear Time Series Analysis of Gulf of Maine Calanus Abundance

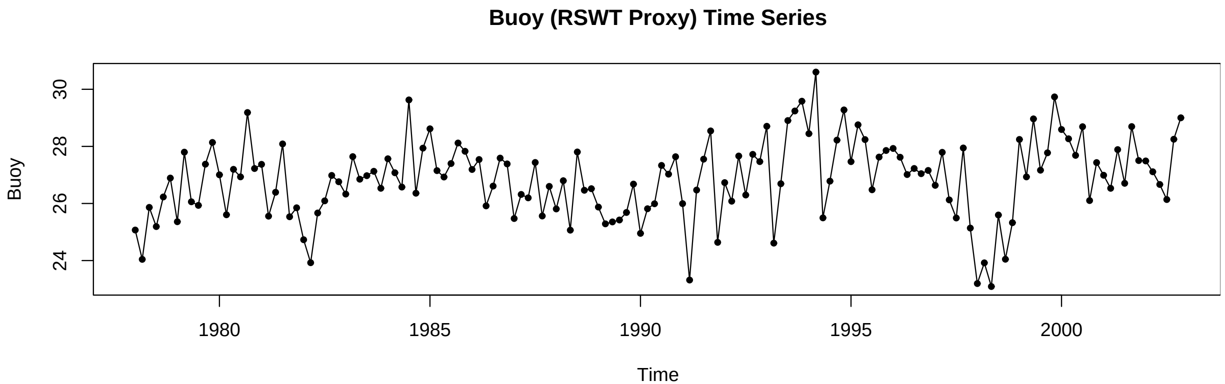
CAFE: Climate-based Assessment of Forecasting for Ecosystems in the Gulf of Maine

Goal: Understand the connection between variability and marine populations and apply this knowledge to develop ecological forecasts.

- *Calanus finmarchicus* (C) - log Calanus abundance as measured by the Continuous Plankton Recorder (CPR) surveys by NOAA Fisheries. (The major prey species of the endangered right whale.)
- Coupled Slope Water System (CSWS) - describes the system of currents and water masses between the NW Atlantic Shelf and the Gulf Stream.
- Regional Slope Water Index (RSWT) is a measure of the state of the CSWS. It is the first component of the PCA of the yearly mean temperature in 8 regions of the Gulf of Maine, Scotian Shelf, and adjacent slope waters between 150-200m.
- New Buoy Data: RSWT proxy data from Northeast Channel (near GoMOOS mooring site).
- New Herring Data: mean number per tow

# Bimonthly Data: Calanus Abundance and Buoy Data 1978-2002



# Nonlinear Models and Nonlinear Autoregressive Processes

- A nonlinear autoregressive process is a univariate time series:

$$N_t = f(N_{t-1}, N_{t-2}, \ldots, N_{t-d}; \boldsymbol{\theta}) + e_t$$

where $\{e_t\}$ is a sequence of i.i.d random variables.

- A more general model that incorporates additional state variables is given by

$$N_t = f(\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \ldots, \boldsymbol{X}_{t-d}; \boldsymbol{\theta}) + e_t$$

where $\boldsymbol{X}$ is a vector of predictors.

## Example: Lorenz System

- Coupled, nonlinear system of three first order differential equations:

$$\begin{aligned} \frac{dx}{dt} &= -s(x-y) \\ \frac{dy}{dt} &= -xz + rx - y \\ \frac{dz}{dt} &= xy - bz \end{aligned}$$

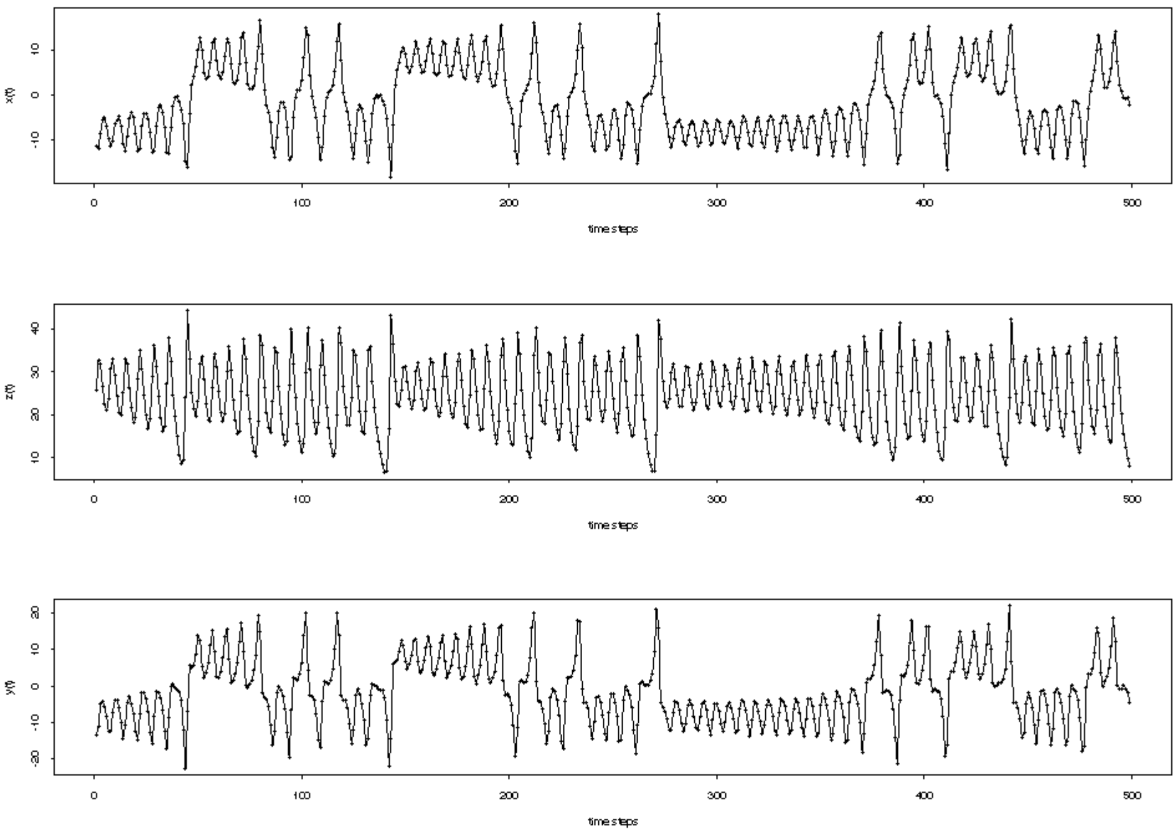
  for $s = 10, r = 28$ and $b = 8/3$ get famous "butterfly"

- Data: numerically integrate and add noise at every integration time-step.

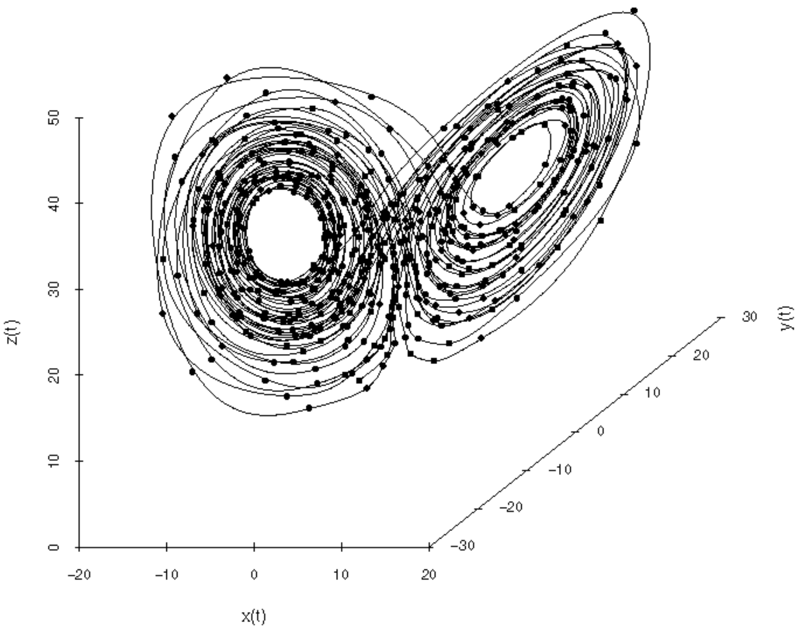
- State-Space System:

$$\begin{aligned} X_t &= F_1(X_{t-1}, Y_{t-1}, Z_{t-1}) + e_{1,t} \\ Y_t &= F_2(X_{t-1}, Y_{t-1}, Z_{t-1}) + e_{2,t} \\ Z_t &= F_3(X_{t-1}, Y_{t-1}, Z_{t-1}) + e_{3,t} \end{aligned}$$

# Lorenz Time Series, $X(t), Z(t), Y(t)$



# Phase Space for Noisy Lorenz System



## Neural Networks

- Class of nonlinear models:

$$Y = f(\boldsymbol{x}; \boldsymbol{\theta}) + e$$

- The form of the model:

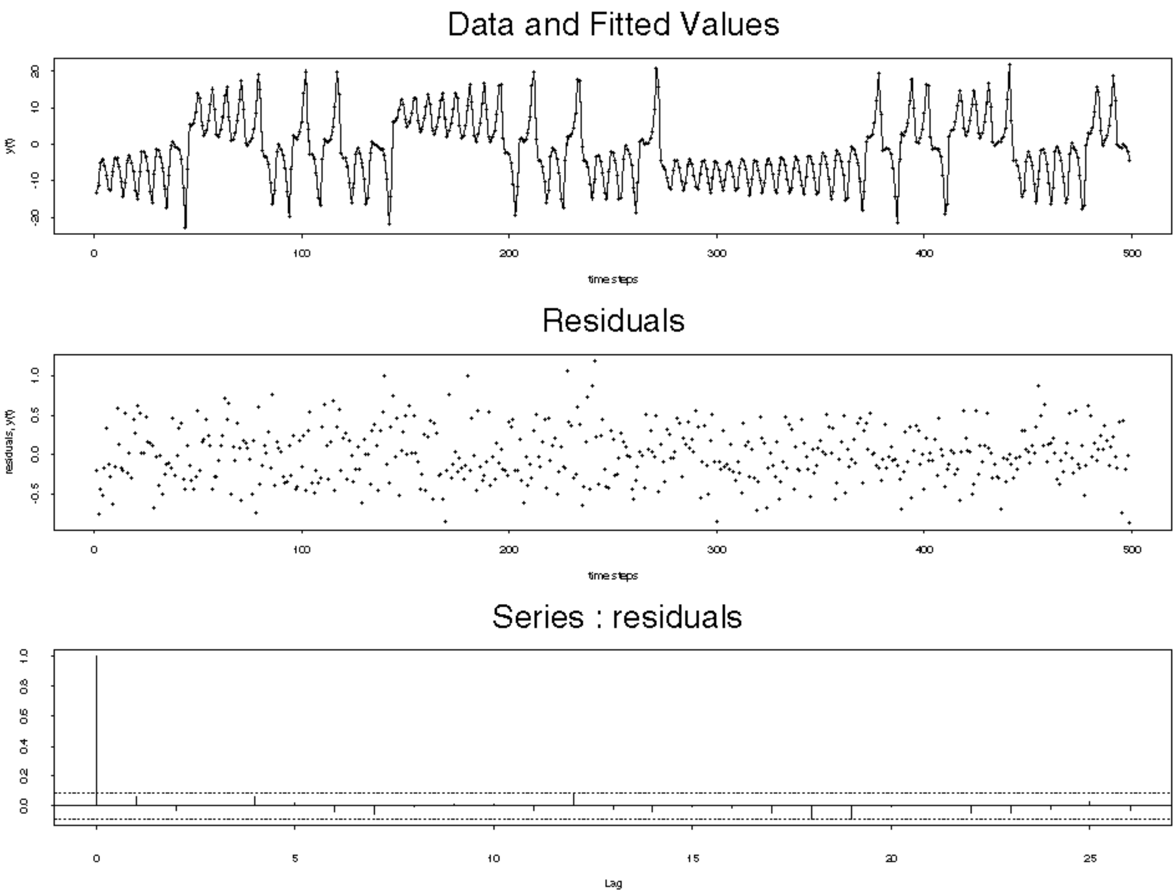
$$f(\boldsymbol{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i \varphi(\boldsymbol{x}^T \gamma_i + \mu_i)$$

  where $\varphi(u) = e^u/(1 + e^u)$

  - Net parameters are estimated by nonlinear least squares.
  - Total number of parameters is $p = 1 + k(d + 2)$ where $d$ is dimension of $\boldsymbol{x}$.
  - Complexity of the model chosen by Cross Validation:

$$V_c = \frac{\frac{1}{n} RSS}{\left(1 - p\frac{c}{n}\right)^2}$$

# Neural Network Fits to Noisy Lorenz System, Y(t)



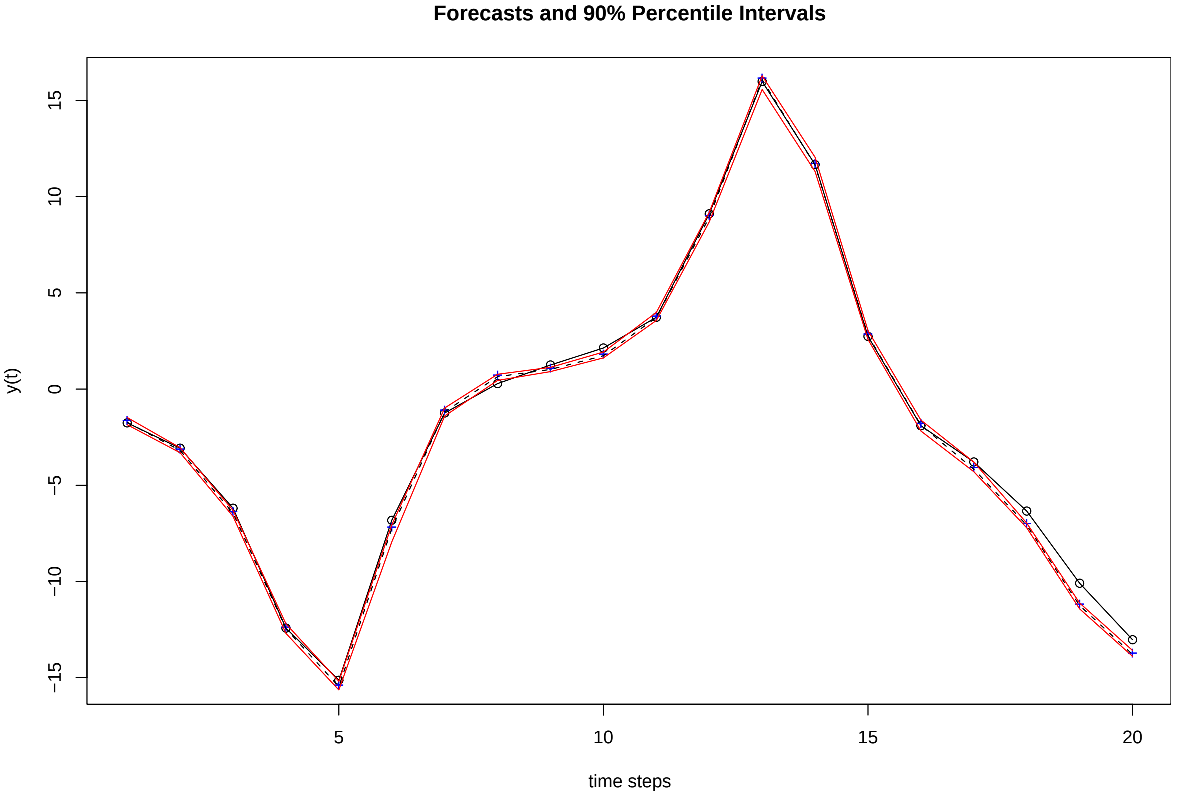
# Nonparametric Bootstrap Prediction Intervals

**Nonlinear Time Series Model**: $Y_t = f(X_{t-1}, Y_{t-1}, Z_{t-1}) + e_t$

Bootstrap "pairs" of the data. Have sample of $n$ observations: $(y_t, \boldsymbol{x}_t)$
where $\boldsymbol{x}_t = (x_{t-1}, y_{t-1}, z_{t-1})$

- Sample the $n$ data pairs with replacement to obtain $n$ bootstrap pairs with which to estimate $f(\cdot)$. Repeat to obtain B=1000 bootstrap estimates of $f$.
- Generate point forecasts as the average of the $B$ bootstrap predicted values.
- Percentile Prediction Intervals: after trimming off the top and bottom 5%, the endpoints of the interval gives the $90th$ percentile bootstrap prediction interval.

# Forecasts for Y(t)



## Nonlinear Time Series Analysis of Gulf of Maine Calanus Abundance

- *Calanus finmarchicus* (C) - log Calanus abundance as measured by the Continuous Plankton Recorder (CPR) surveys by NOAA Fisheries. (The major prey species of the endangered right whale.)
- New Buoy Data: RSWT proxy data from Northeast Channel (near GoMOOS mooring site)
- New Herring Data: mean number per tow
- **Calanus Nonlinear Time Series Model:**

  1. Data: bimonthly

  $$C_t = f(C_{t-1}, B_{t-12}, \sin(2\pi t/6), \cos(2\pi t/6)) + e_t$$

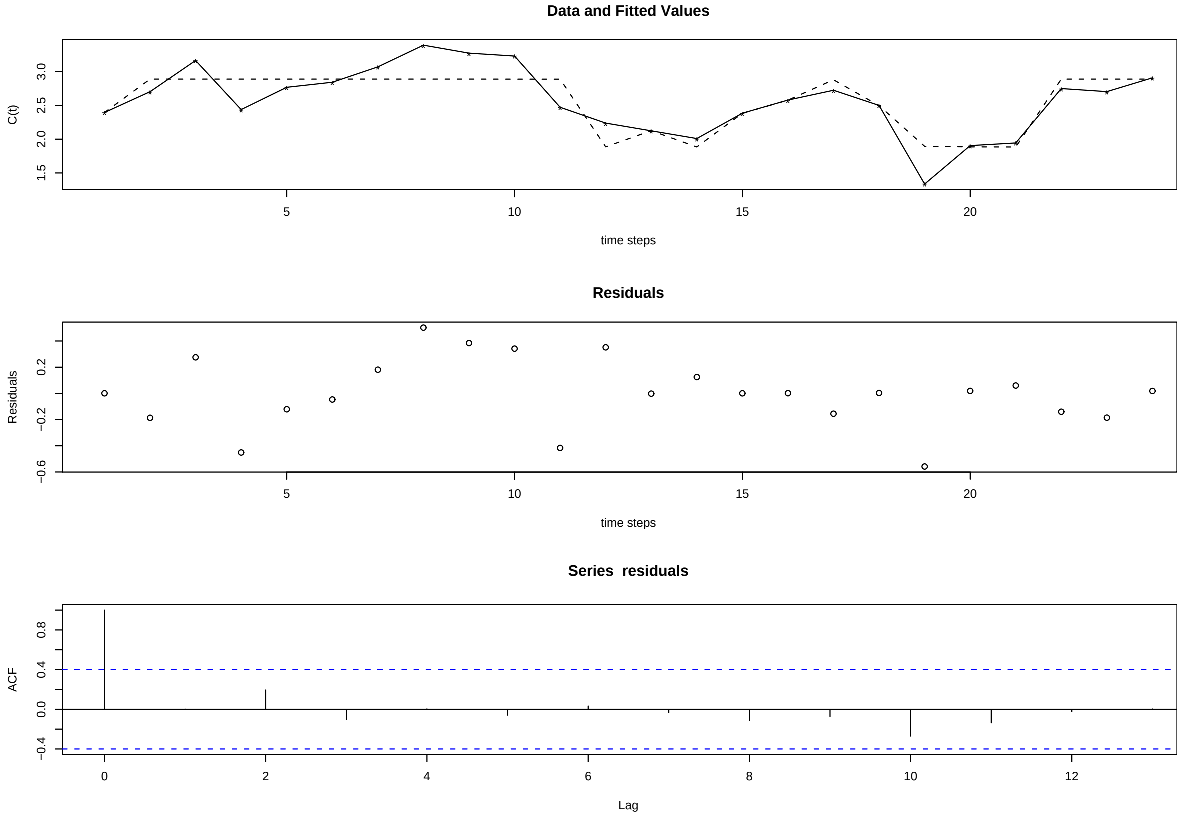
  2. Data: yearly

  $$C_t = f(C_{t-1}, B_{t-2}, H_{t-1}) + e_t$$

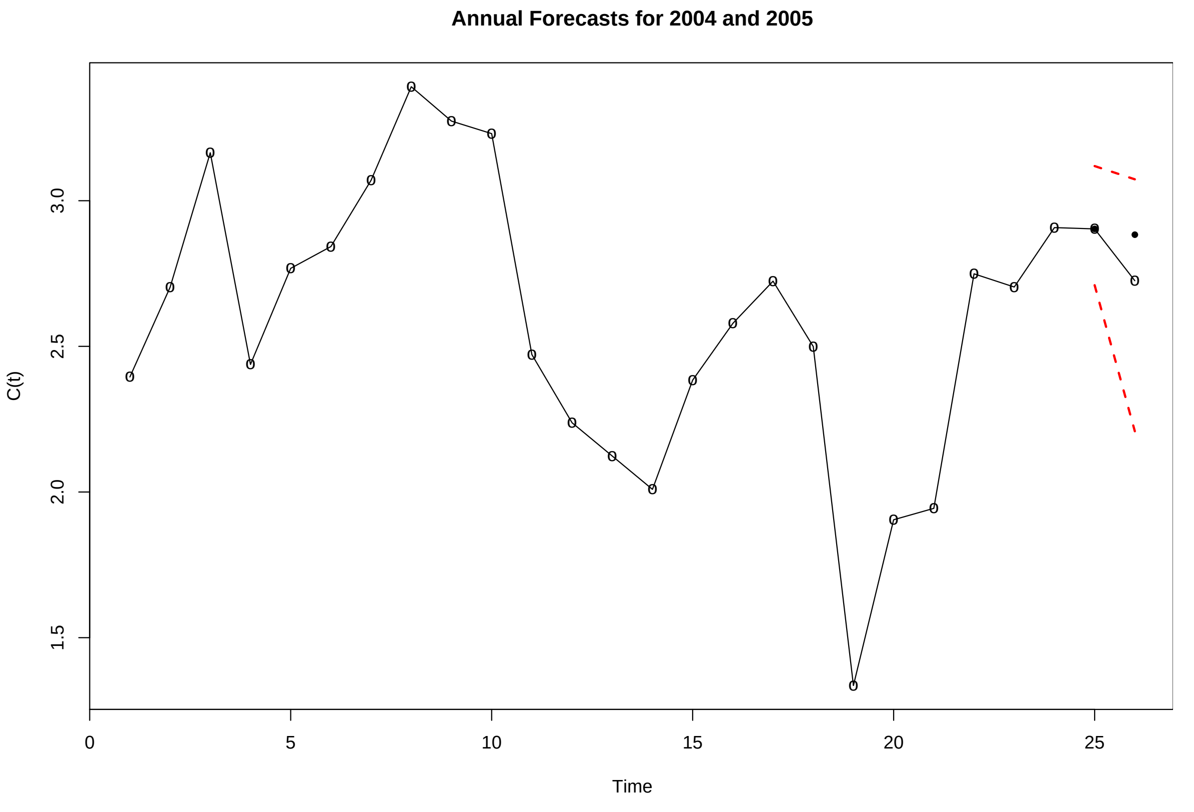
  – Fit model to data 1978-2003

  – Forecast 2004 and 2005, starting with data at time $t$ and iterating model.

# Neural Network Fits to $C_t$ for Annual Data



# Forecasts and Bootstrap Percentile Intervals



## Future Work

- Look for other important variables to include in the Calanus prediction model.
  - Variables that are suspected to interact with Calanus.
  - Regional: Eastern and Western Gulf of Main Calanus Data