# Evaluation of statistical models for estimating abundance from a series of resource surveys

Paul Spencer, Grant Thompson, Jim Ianelli, and Jon Heifetz

Alaska Fisheries Science Center

# Outline

1) Why do we want to smooth survey time series?
2) Common methods for smoothing time series
3) Simulation study to evaluate performance
4) Application to Alaska data
5) Future work
6) Conclusions

# North Pacific Fishery Management Council Tier System

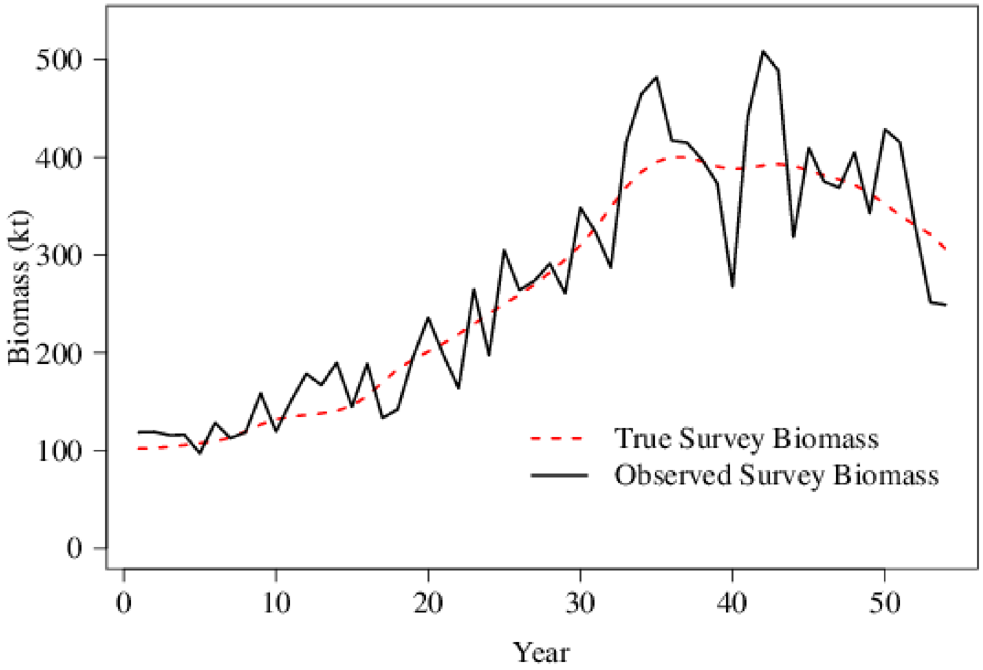
1) Information available: *Reliable point estimates of B and $B_{MSY}$ and reliable pdf of $F_{MSY}$.*
2) Information available: *Reliable point estimates of B, $B_{MSY}$, $F_{MSY}$, $F_{35\%}$, and $F_{40\%}$.*
3) Information available: *Reliable point estimates of B, $B_{40\%}$, $F_{35\%}$, and $F_{40\%}$.*
4) Information available: *Reliable point estimates of B, $F_{35\%}$, and $F_{40\%}$.*
5) Information available: *Reliable point estimates of B and natural mortality rate M.*
6) Information available: *Reliable catch history from 1978 through 1995.*

## Purpose of averaging time series

1) Estimate the biomass from survey data (Tiers 4 and 5)
2) Partition the harvest quotas within a model area, based on survey time series (Tiers 1-5)

# A signal to noise problem

1) We want to remove the observation error
2) We do not want to "smooth" the underlying "signal"
3) The last data point is the most important (for management)



# State-space representation

$$z_t = f(z_{t-1}) + a_t$$

$$y_t = g(z_t) + e_t$$

Z = Population size (unobserved)
Y = Survey index

Process and observation errors are represented by *a* and *e*, respectively

One example of special interest is the random walk model plus uncorrelated noise (RWPUN; Stockhausen and Fogarty (2007))

$$z_t = z_{t-1} + a_t$$

$$y_t = z_t + e_t$$

# Exponential smoothing

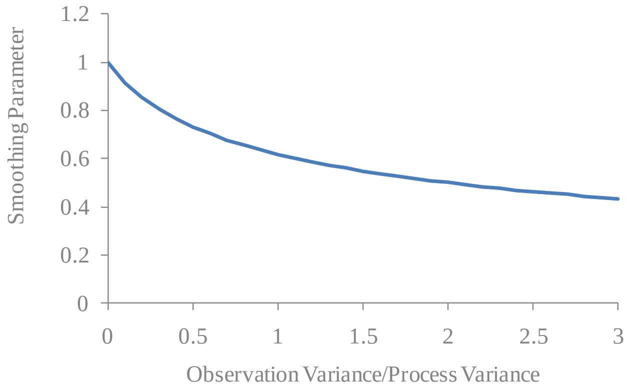
$$\hat{z}_t = (\alpha)y_t + (1-\alpha)\left[\alpha y_{t-1} + \alpha(1-\alpha)y_{t-2} + \alpha(1-\alpha)^2 y_{t-3} + \ldots\right]$$

$$\hat{z}_t = (\alpha)y_t + (1-\alpha)\hat{y}_{t-1}(1)$$

This is a Kalman Filter with constant observation error variance

For the random walk model with constant variances:

1) $\alpha$ = f(process variance/observation variance) (Pennington 1986, Thompson)
2) Exponential smoothing is the optimal forecast method (Pennington 1986)



# Random effects model

Considers the process errors as “random effects” (i.e., drawn from an underlying distribution) and integrated out of the likelihood

The state-space random walk plus noise can be formulated as a random effect model

## Differences between the Kalman filter and random effects models

1) Different statistical approaches – Bayesian updating equations vs. hierarchical random effects model
2) The random effects model can provide more flexibility with non-linear processes and non-normal error structures

# ARIMA modeling notation

ARIMA models (auto-regressive integrated moving average)

$$y_t = \alpha_1 y_{t-1} + \ldots \alpha_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-q} \ldots \beta_q \varepsilon_{t-q}$$

α – p auto-regressive parameters
β – q moving average parameters
ε – random errors

The data can be differenced d times to achieve stationarity

The structure of the ARIMA model is referred to as (p,d,q)

The random walk plus uncorrelated noise (RWPUN) is a (0,1,1) ARIMA model

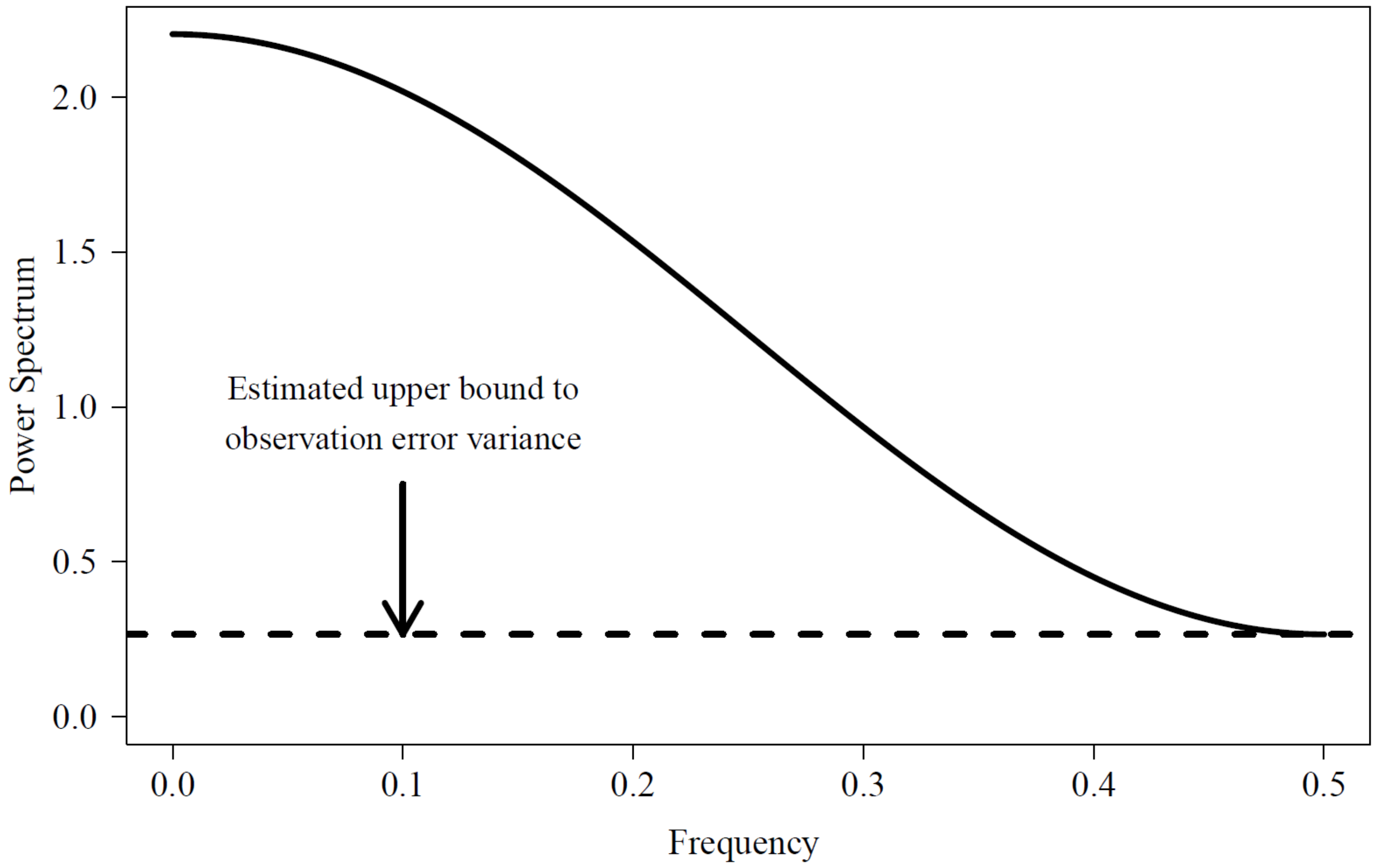
# Models where we do not assume the underlying state is a random walk

Stockhausen and Fogarty (2007) applied a smoothing procedure based on generalized ARIMA models:

1) Fit a series of candidate ARIMA models to survey data
2) Use model selection criteria to identify the best ARIMA model
3) Estimate  the power spectrum for the ARIMA process, which give an estimate of the upper bound of the observation error variance (*K**)
4) From the ARIMA parameters and K*, estimate smoothing weights to be used in a symmetric moving average

Important point – The *q* dimension we estimate for the observed data must be equal or greater than (p+d)

# Example estimation of power spectrum and K*



# Conditions for applying generalized ARIMA smoothing

1) A time series long enough to get reliable parameter estimates (Stockhausen and Fogarty (2007) suggest 40 years)

2) Estimated q >= (p+d)
3) Not white noise
4) Other (stationarity of autoregressive parameters, invertibility, variance reduction)

# Description of Simulation Study

Objective: How well does generalized ARIMA modeling compare to exponential smoothing and random effects models?

Two life-history types: Pacific ocean perch (long-lived) and walleye pollock (shorter-lived)

Recent population: Increasing, Flat, or Decreasing

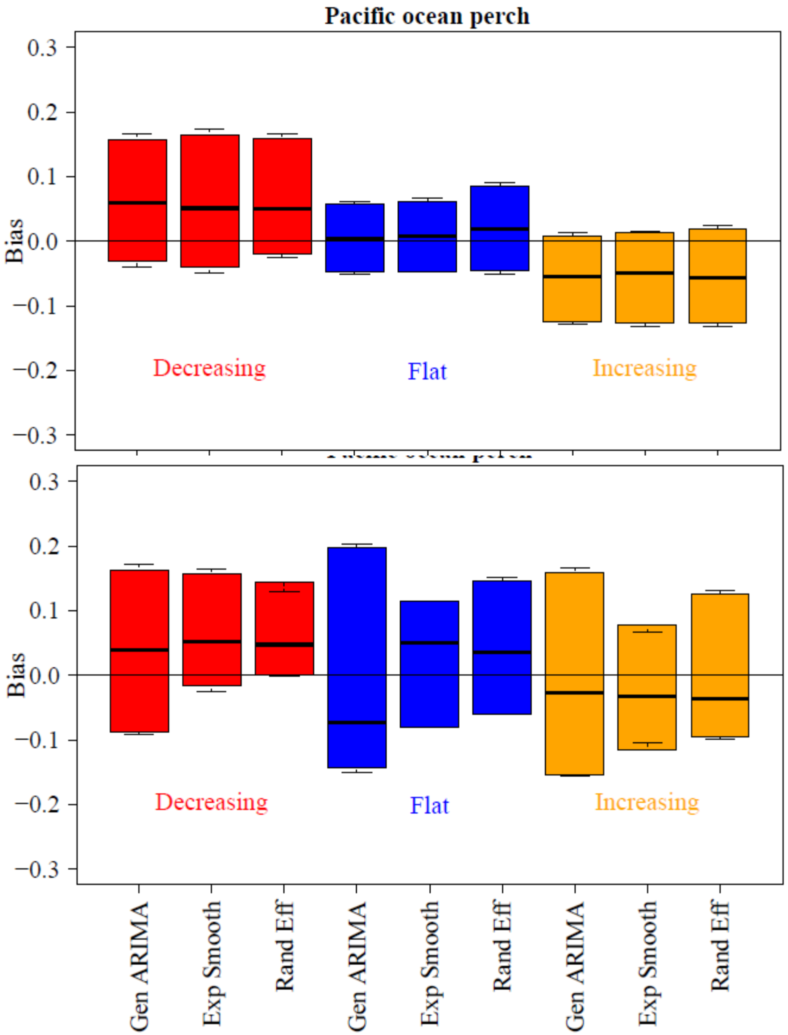
Process error – two levels of recruitment variance

Observation errors – two levels of coefficient of variation (CV) of survey biomass estimates

Three levels of survey frequency

# Classification of ARIMA model results

# Bias and variance of relative errors of recent smoothed biomass estimate



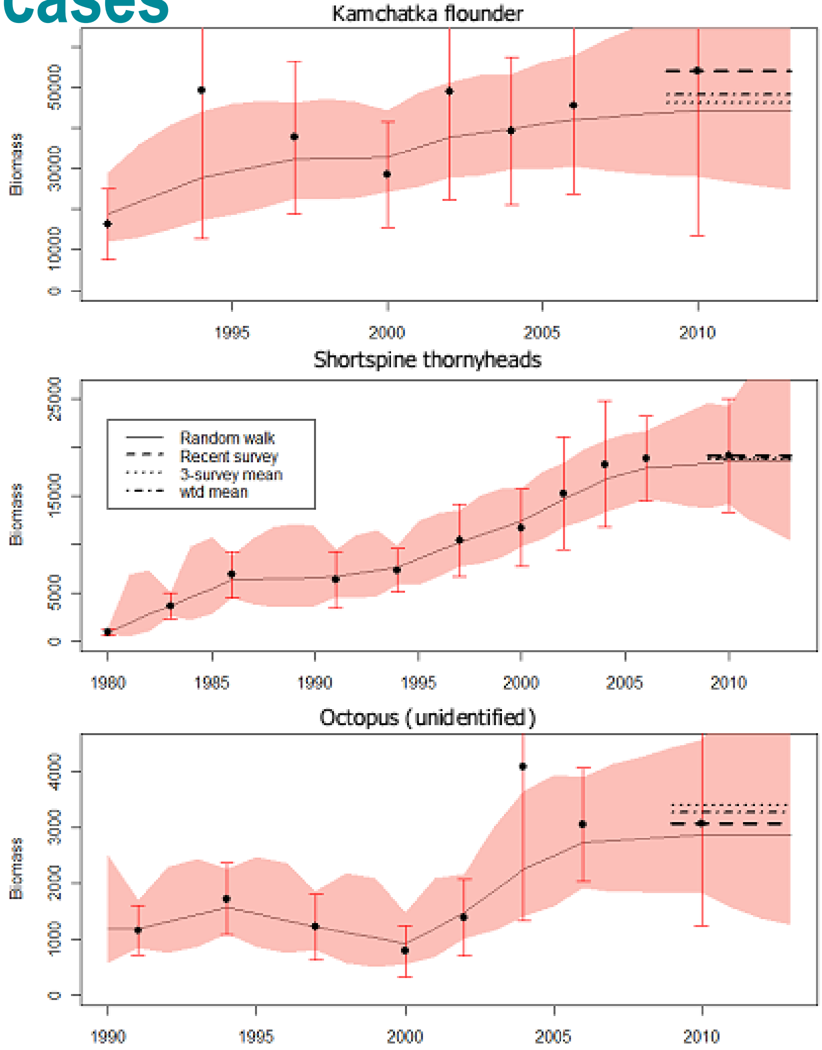

**Best ARIMA model is (0,1,1)**

Generalized ARIMA model performs about as well as exponential smoothing and random effects models

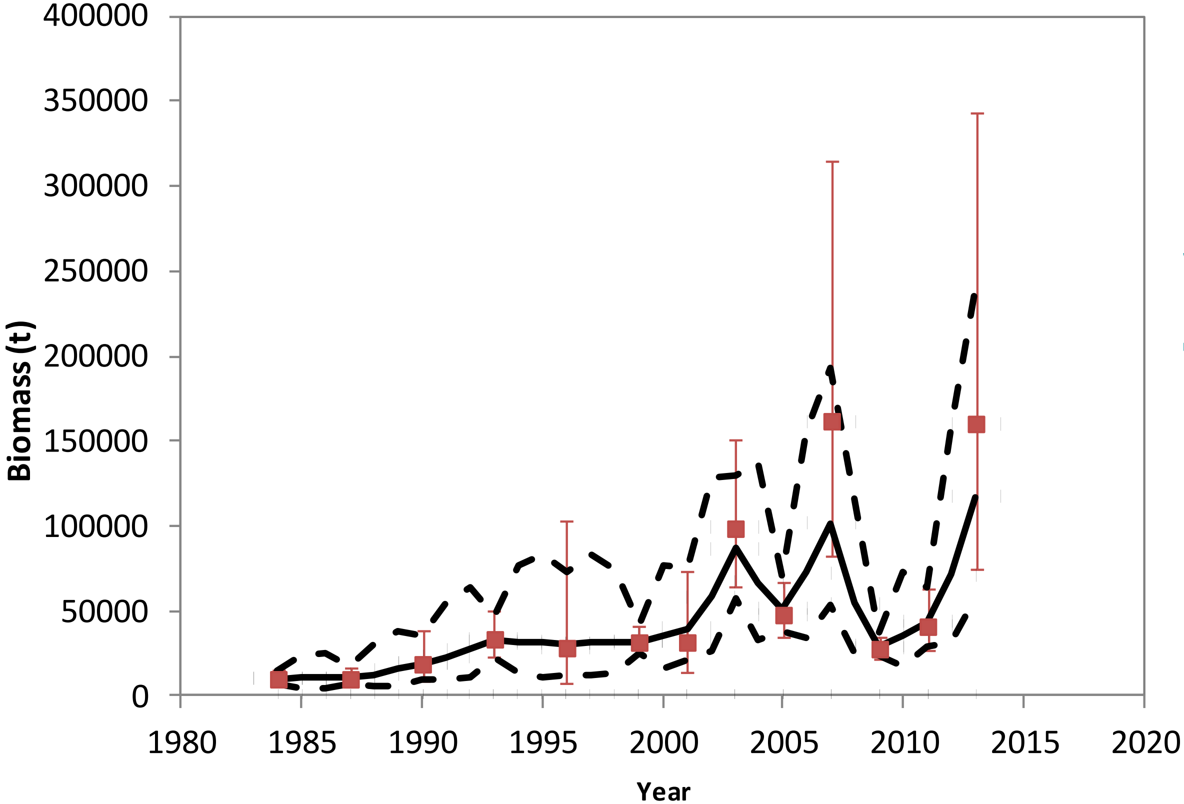
**Best ARIMA model is not (0,1,1)**

Smaller number of cases, but it appears that the generalized ARIMA modeling has greater variance

# The random effects model seems to produce reasonable fits in most cases



# Gulf of Alaska dogfish



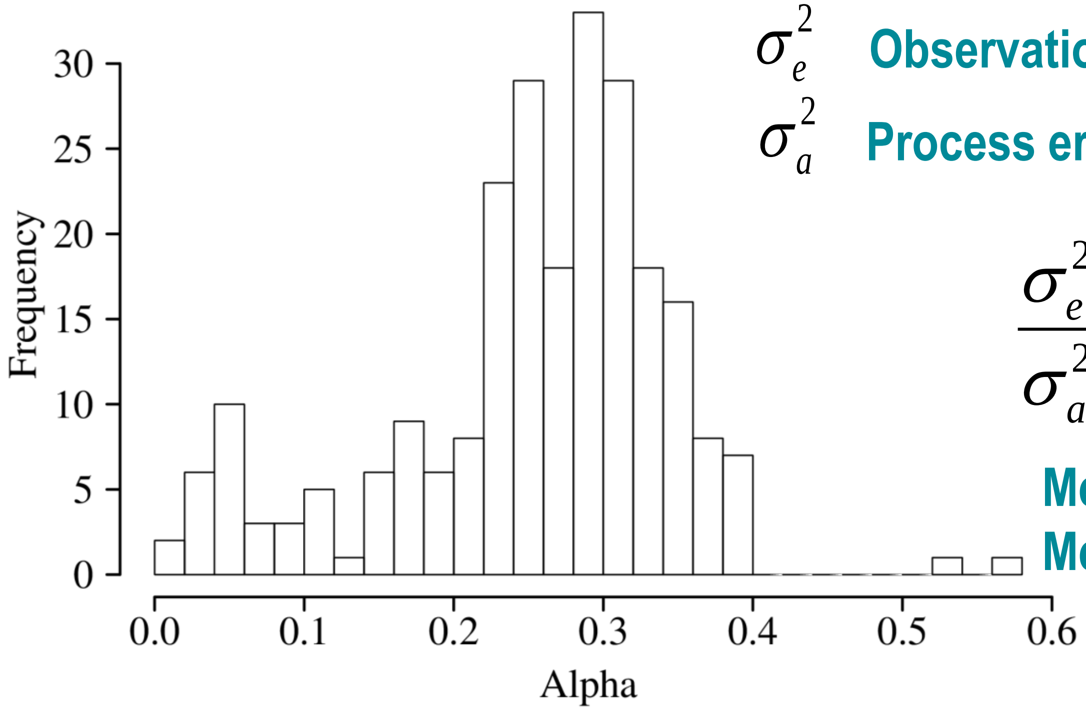

Survey biomass estimates vary widely, especially from 2003-2005

CVs of 0.22 and 0.18

Estimated log-scale process standard deviation of 0.49

# A simple exponential smoothing model can give information on the ratio of variances

$$\hat{z}_t = (\alpha)y_t + (1-\alpha)\left[\alpha y_{t-1} + \alpha(1-\alpha)y_{t-2} + \alpha(1-\alpha)^2 y_{t-3} + \ldots\right]$$




$\sigma_e^2$ Observation error variance

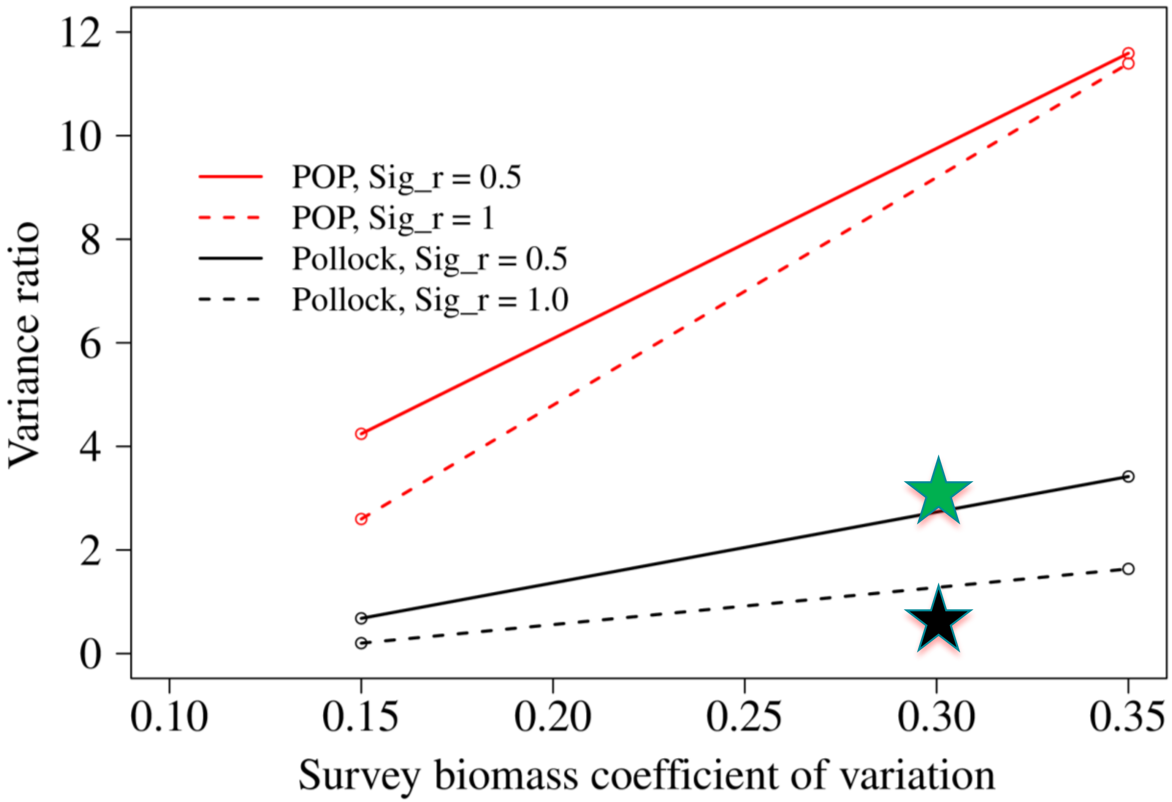
$\sigma_a^2$ Process error variance

$$\frac{\sigma_e^2}{\sigma_a^2} = \frac{1-\alpha}{\alpha^2}$$

Mean α = 0.256

Mean ratio = 11.35

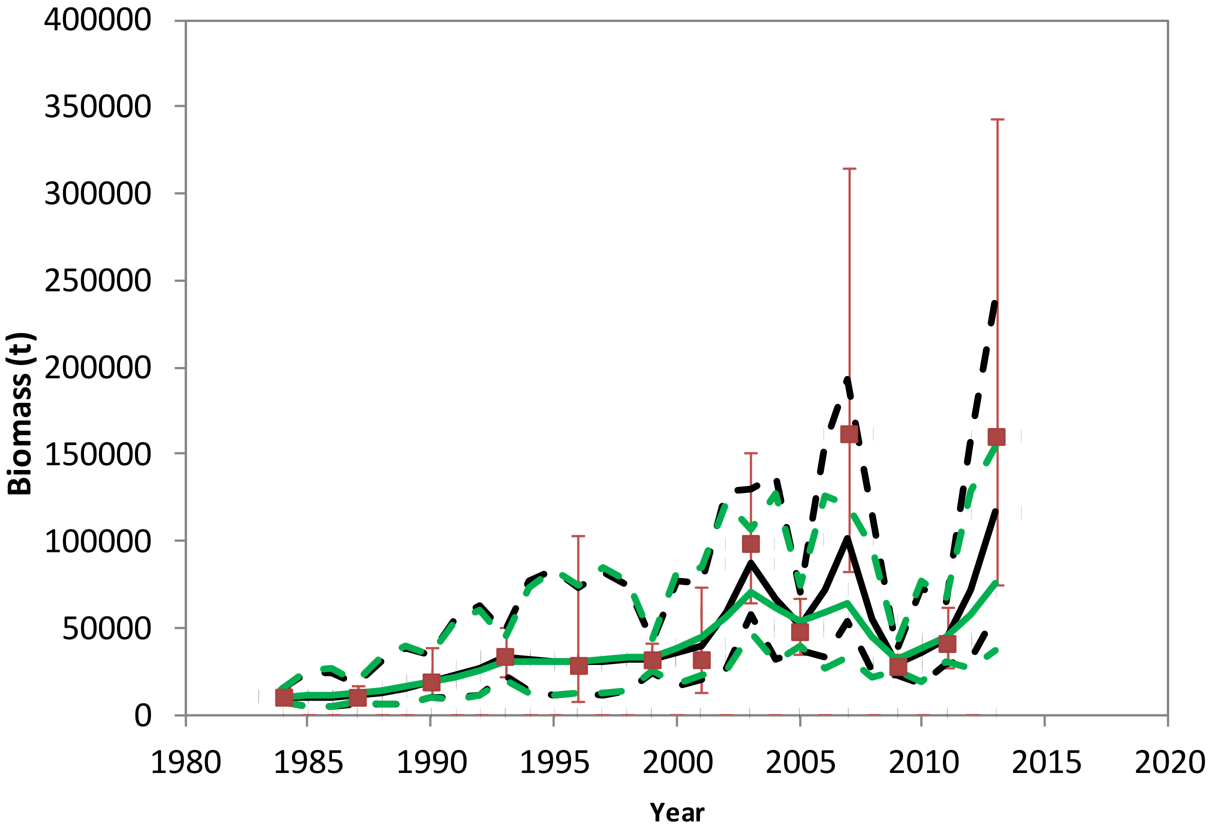
# The variance ratio is a function of stock longevity, recruitment variability, and survey variability




Used as a prior to constrain the estimate of process error standard deviations

Implied from fit to GOA dogfish

# The fit with the prior constrains the estimate of process error standard deviation, and appears more reasonable




**Fit with prior shown in green**

# Conclusions

1) The random walk model described many of our simulated datasets. For these cases, the three smoothing methods performed similarly.
2) Some cases may not be conducive to generalized ARIMA smoothing
3) Prior information on the ratio of observation to process error (either from a simple model, or knowledge of the life-history and survey process) could be used to constrain estimation in the random effects model.

# Additional work -- these methods could be used to fill in areas in which the survey was not conducted