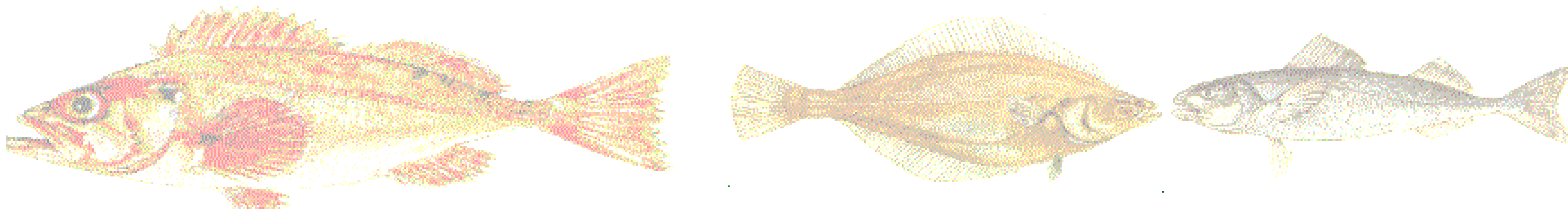
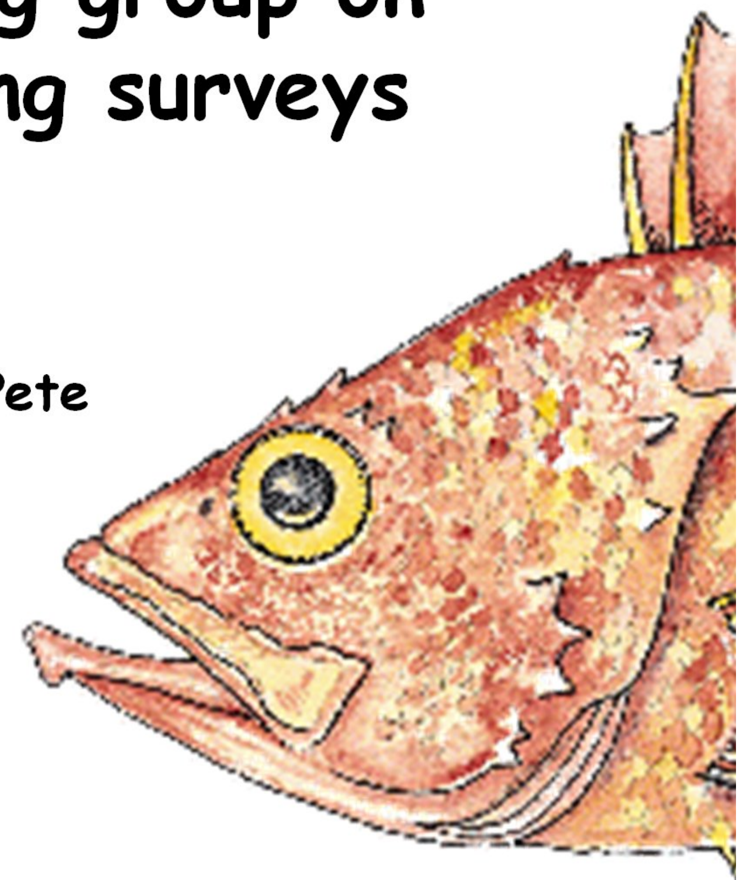
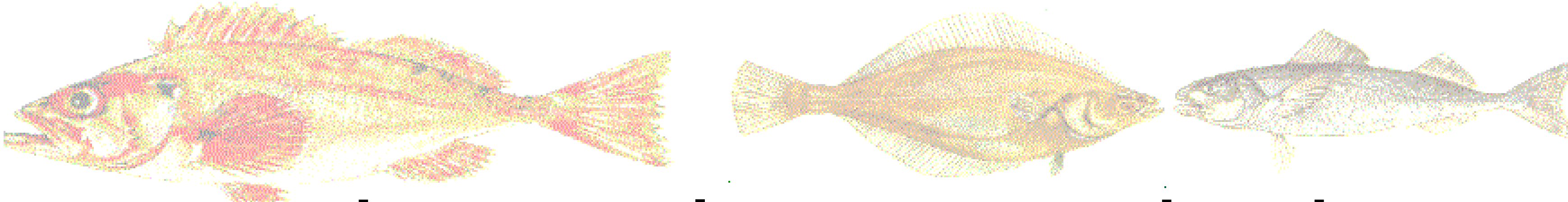


Report of the working group on methods for averaging surveys

Paul, Grant, Jon, Jim, Pete





Tasks: evaluate methods

1. To produce “reliable” biomass estimates (mainly Tier 5)

2. Use of surveys for apportionment

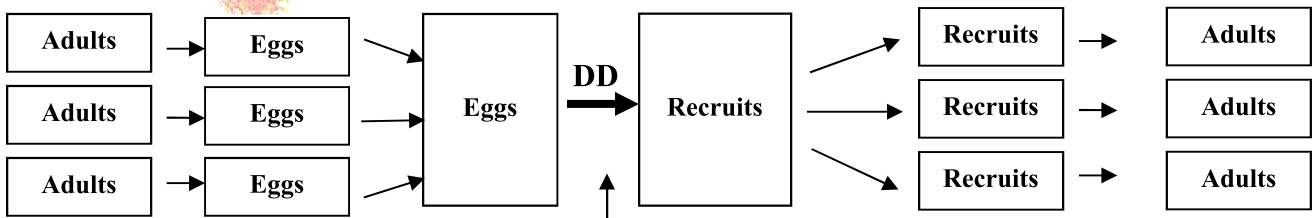
3. Help with incomplete surveys

Outline

- Simulation study
  - Simulated spatial survey data
  - Evaluation of estimated proportions
  - Evaluation of method for accounting for missing data
- Application to real data
  - Methods for constraining unrealistic estimates of process error variance
- Conclusions/ Recommendations
- Next Steps

Simple Spatial Recruitment Models (modified from Ralston and O'Farrell, 2008)

Intracohort density-dependence



'global' recruitment variability

'local' recruitment variability

$$R_a = \hat{R}_a e^{(\varepsilon_g + \varepsilon_a)}$$

'global' recruitment standard deviation (σ_g) 0.12 and 0.24
'local' recruitment standard deviation (σ_a) 0.48 and 0.96

Also, the global recruitment errors may be autocorrelated:

$$\varepsilon_g = \rho\varepsilon_{g-1} + \left(\sigma_g \sqrt{(1-\rho^2)}\right)\tau \qquad \tau \sim N(0,1)$$

Autocorrelation coefficient (ρ) set to 0 and 0.3

Simulation approach

Survey CV (within each subarea):

lognormal distribution, 0.25 and 0.6

Natural mortality (*M*):

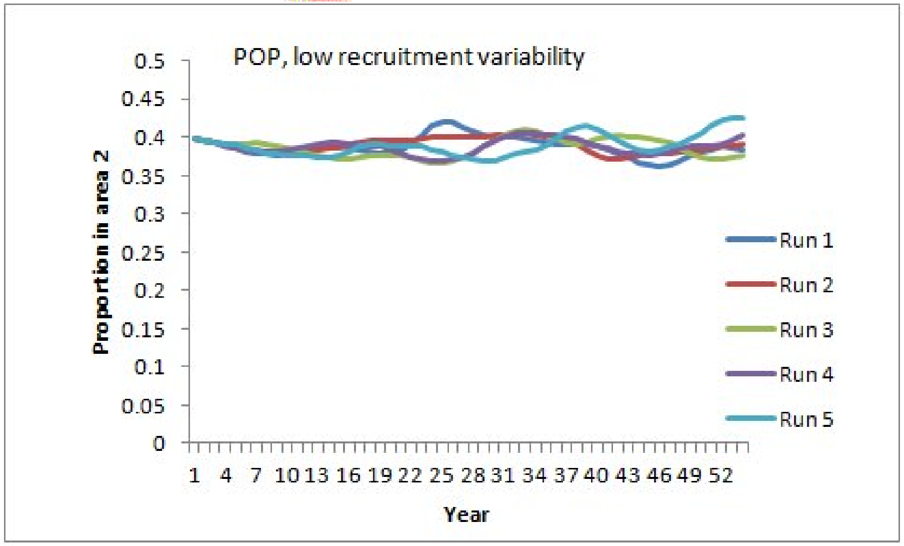
0.06 (POP) and 0.30 (pollock)

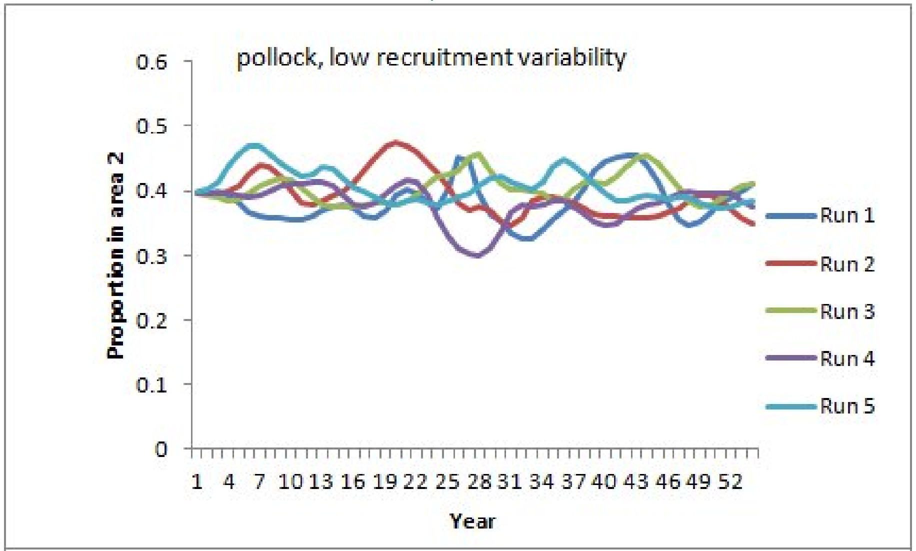
Two levels of adult movement

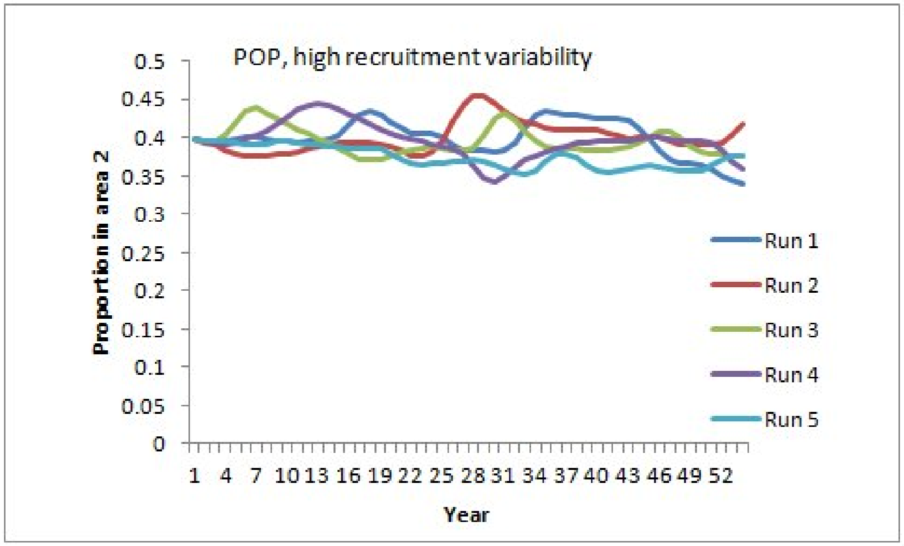
Trend in biomass :

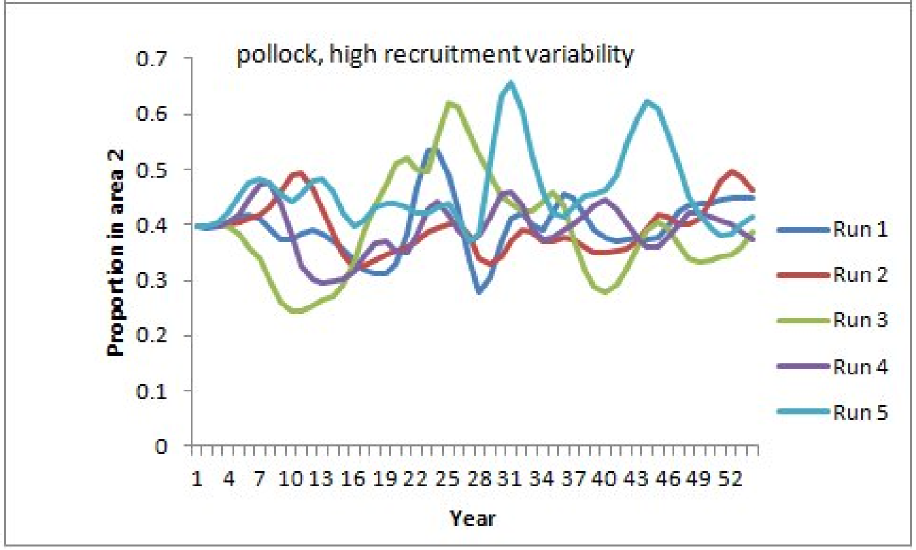
1) increasing, then decreasing
2) decreasing, then increasing
3) constant

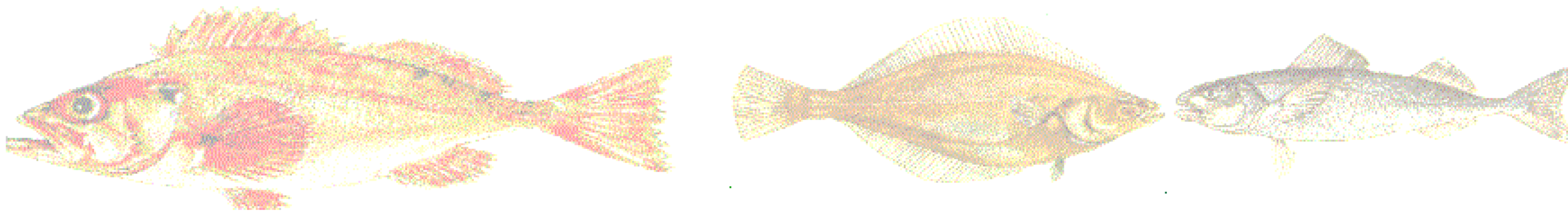
Some example simulated populations





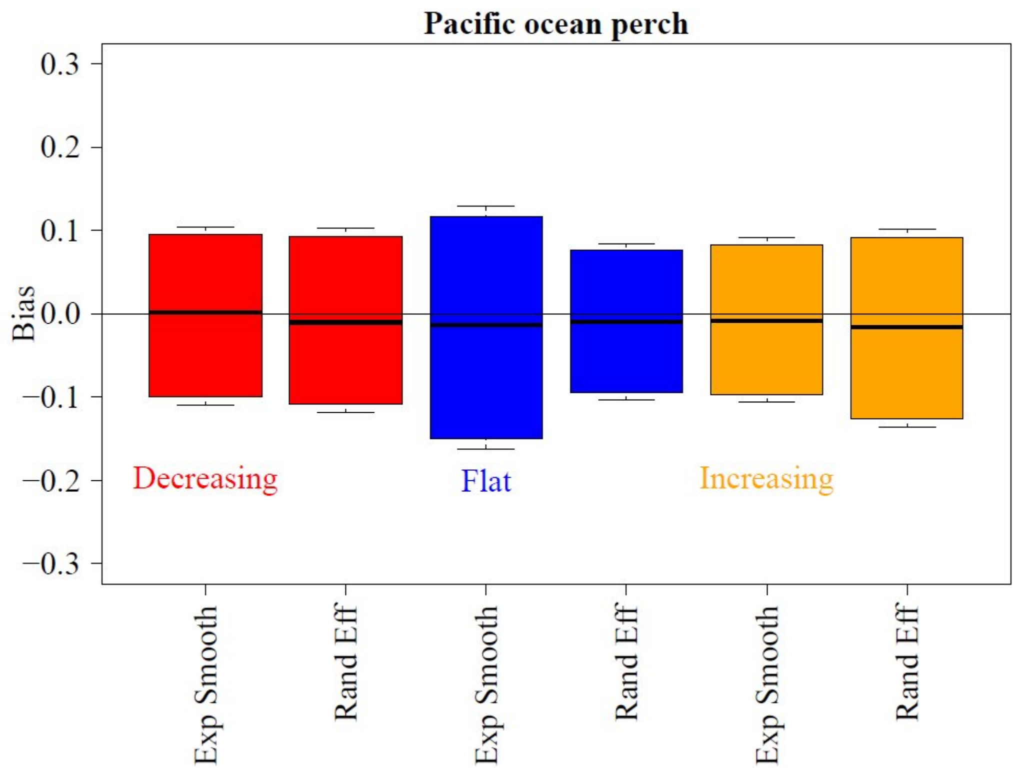




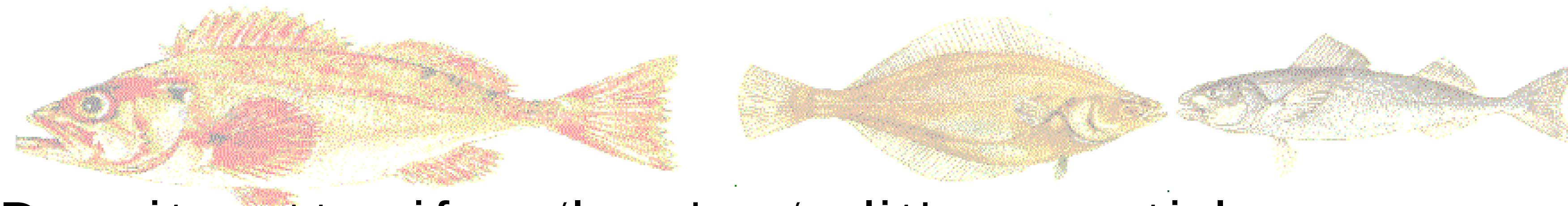


Is there bias in the estimated distribution among the subareas?

Fit the smoother to the observed data from each subarea, and use the smoothed results to compute the estimated proportions.



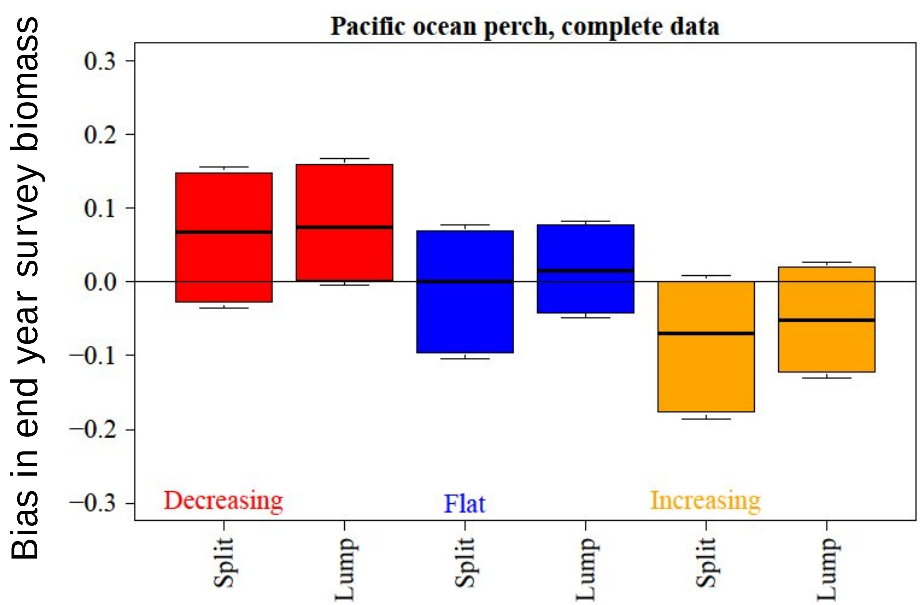
Bias = $(\hat{p} - p)/p$



Does it matter if we ‘lump’ or ‘split’ our spatial survey data?

Lump – Fit a single smoother to the sum of the spatial survey estimates

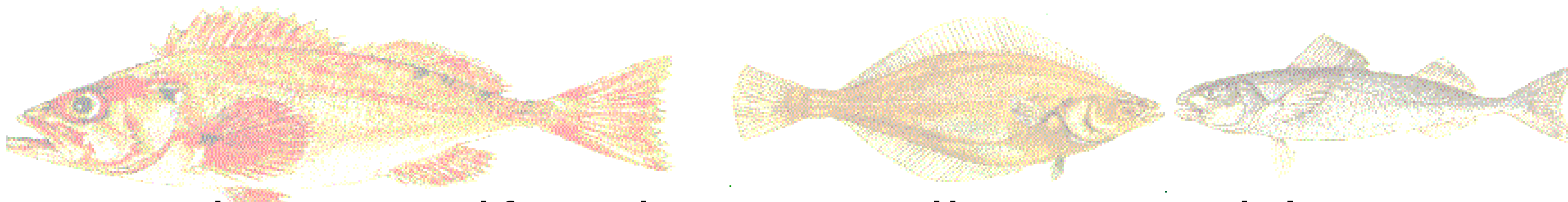
Split – Fit the smoother to data in each spatial area, then combine the smoothed results across areas



Bias = $(\hat{S}_{y=54} - S_{y=54})/S_{y=54}$

Random effects model was used

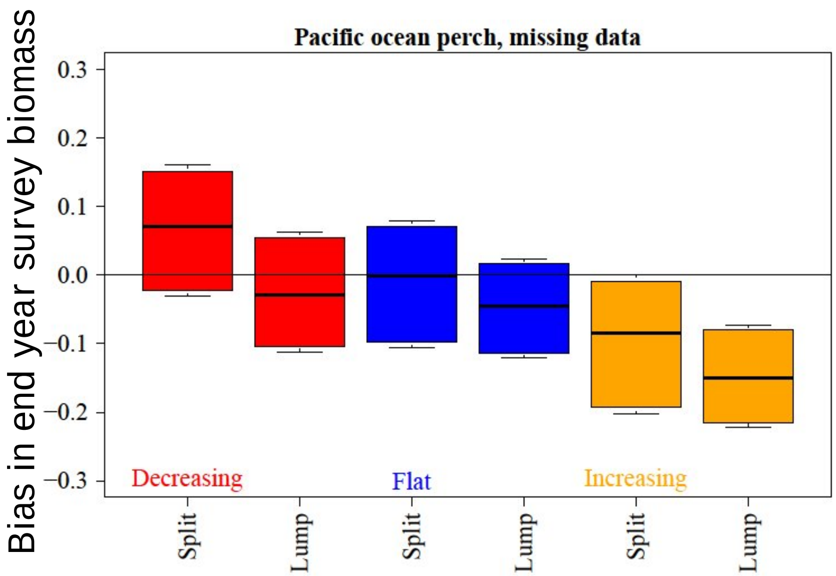
When we do not have missing data, the two methods give similar results



Does it matter if we ‘lump’ or ‘split’ our spatial survey data when we have missing data in some areas?

Modeling of missing survey biomass estimates

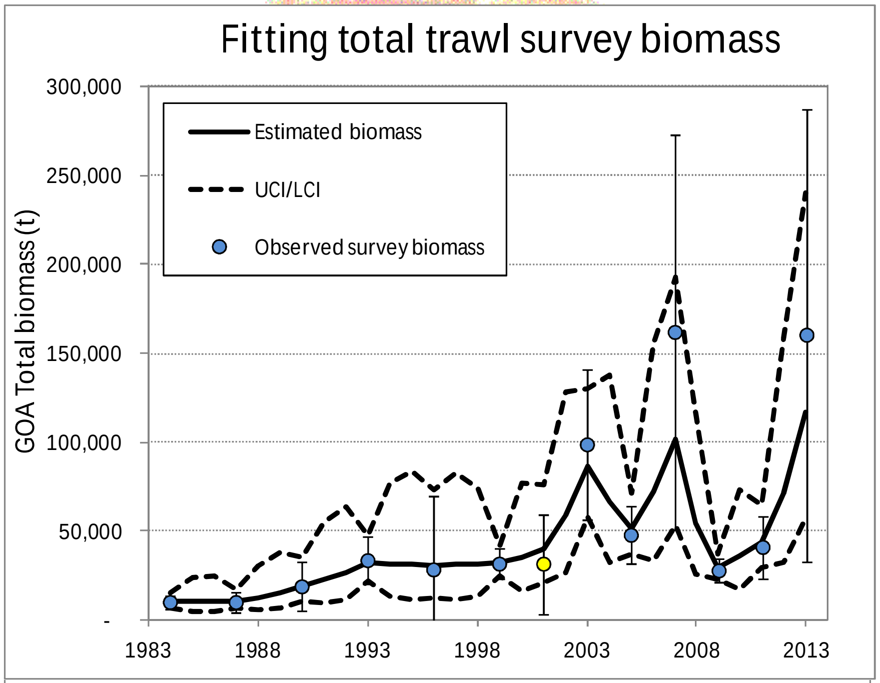
	Year 51	Year 52	Year 53	Year 54
Total	393	284	292	362
Area 1	178	136	96	107
Area 2	99	147	196	129
Area 3	116			126



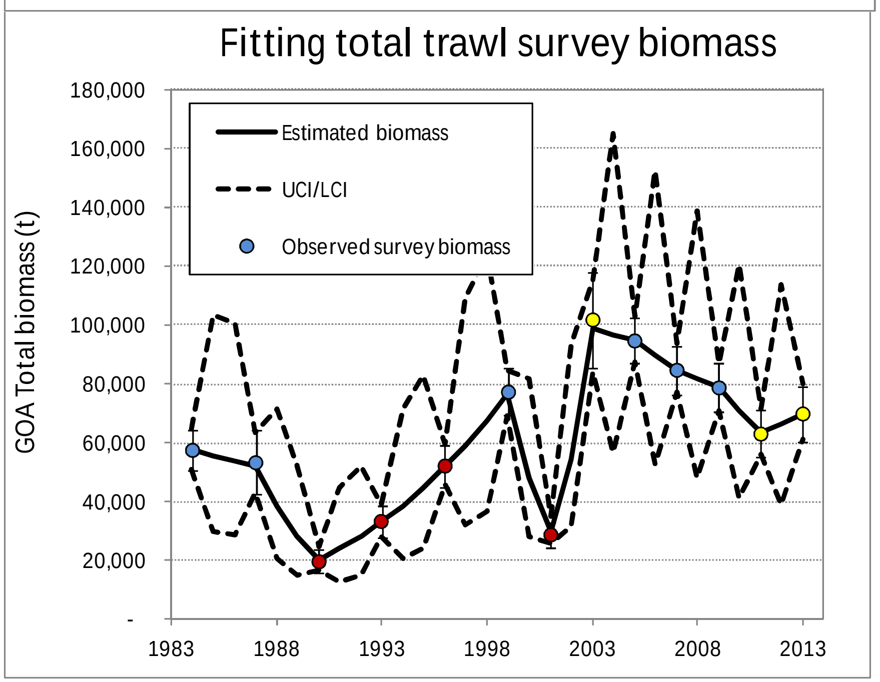
As expected, not accounting for missing data will induce a negative bias in our results.

Applying the random effects smoother by area ‘fills in’ the missing data.

When survey CVs are low, smoothing diminished



GOA dogfish
Survey biomass estimates vary widely
especially from 2003-2005;
CVs 0.22 and 0.18

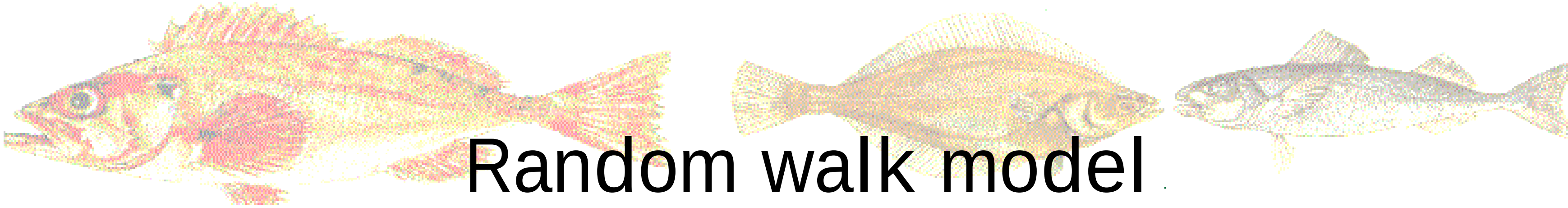


GOA shortspine thornyheads
CV averages 7%

Possible remedies

Constraining estimated process error

- May relate to species longevity
- Could include temporal variability in catchability and selectivity
- Results from exponential smoothing could be used to develop a prior for the ratio of observation error to process error



Random walk model with observation error

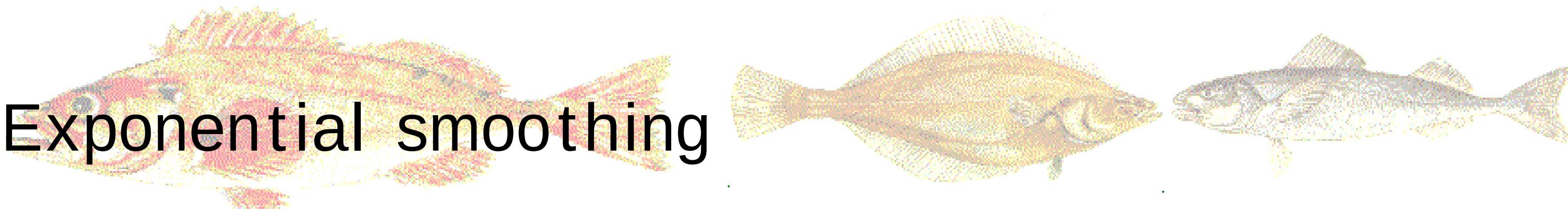
$$z_t = z_{t-1} + a_t$$

$$y_t = z_t + e_t$$

z = Population size (unobserved)

Y = Survey index

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

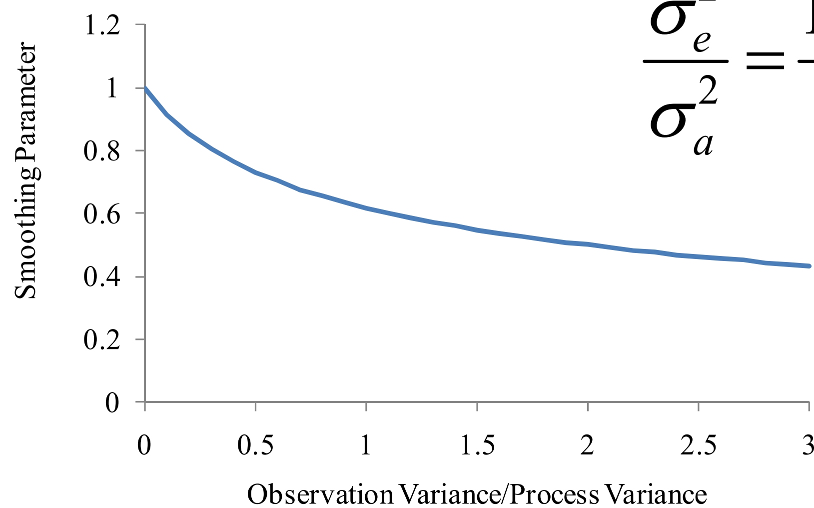


Exponential smoothing

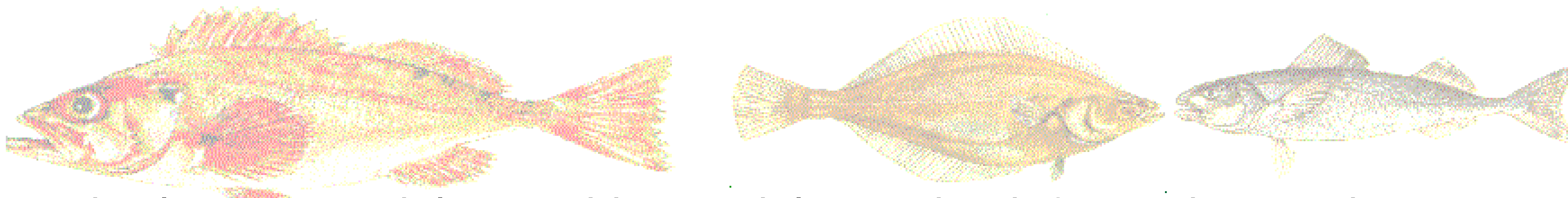
$$\hat{z}_t = (\alpha) y_t + (1-\alpha)\left[\alpha y_{t-1} + \alpha(1-\alpha) y_{t-2} + \alpha(1-\alpha)^2 y_{t-3} + \ldots\right]$$

For the random walk model with constant variances:

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

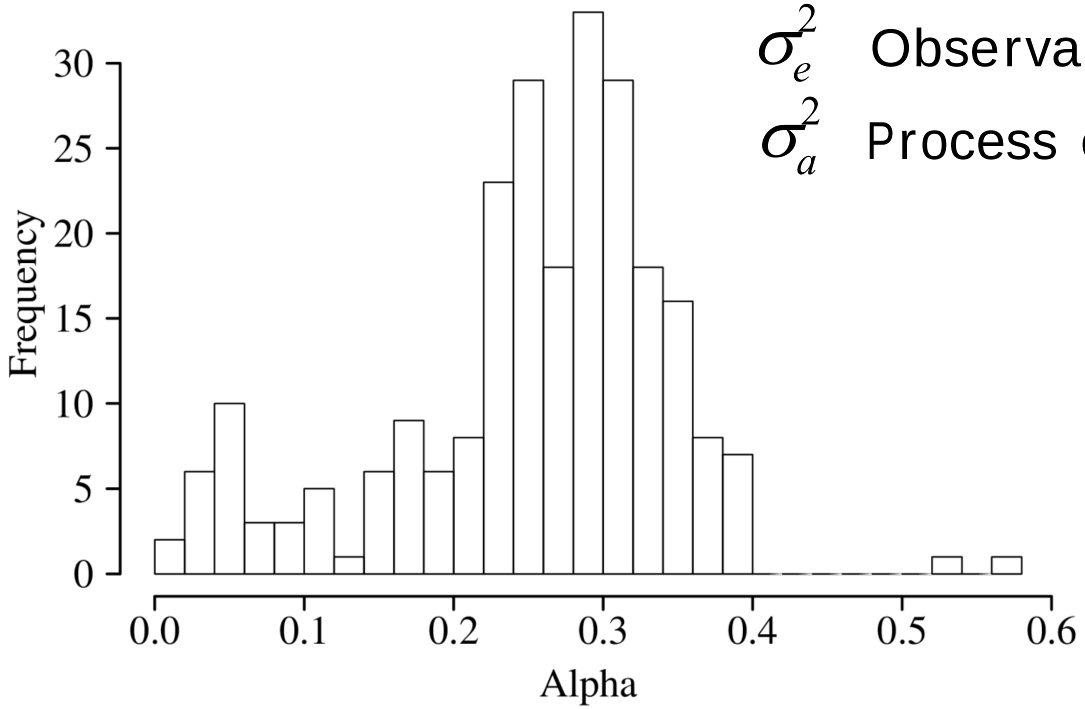


$$\frac{\sigma_e^2}{\sigma_a^2} = \frac{1-\alpha}{\alpha^2}$$



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

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



σ_e^2 Observation error variance

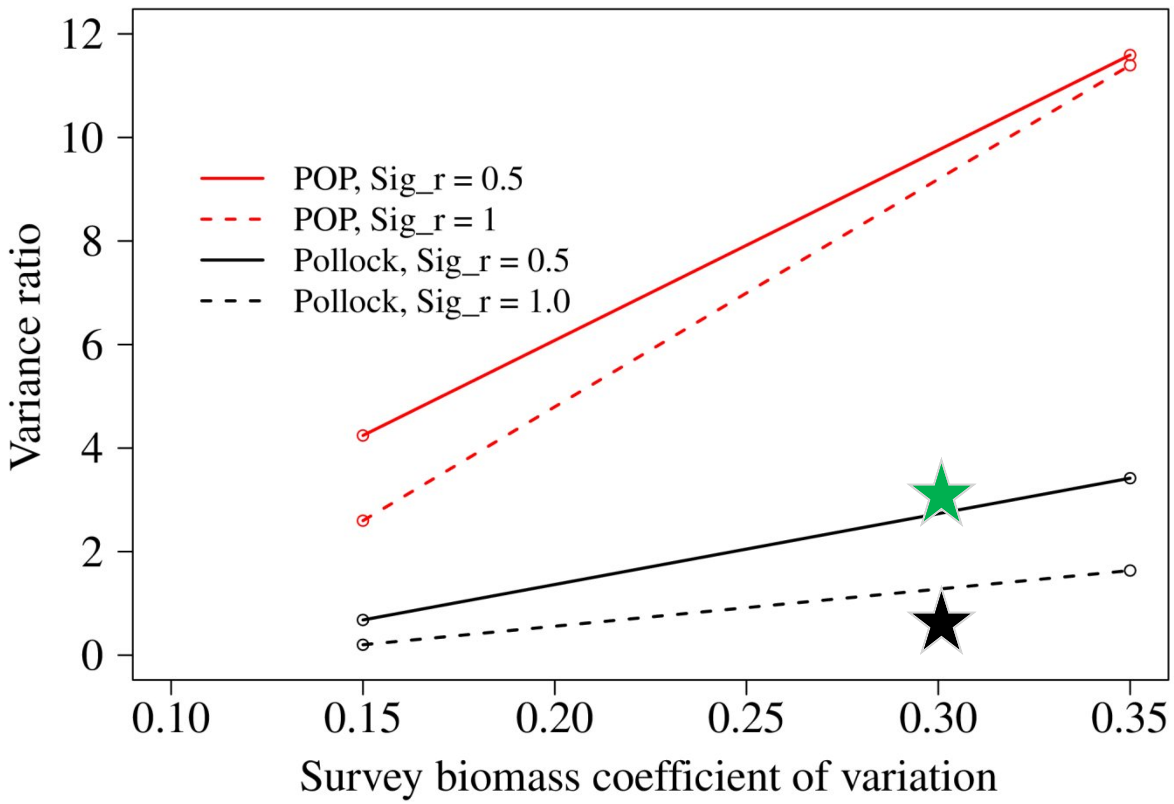
σ_a^2 Process error variance

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

Mean $\alpha = 0.256$

Mean ratio = 11.35

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

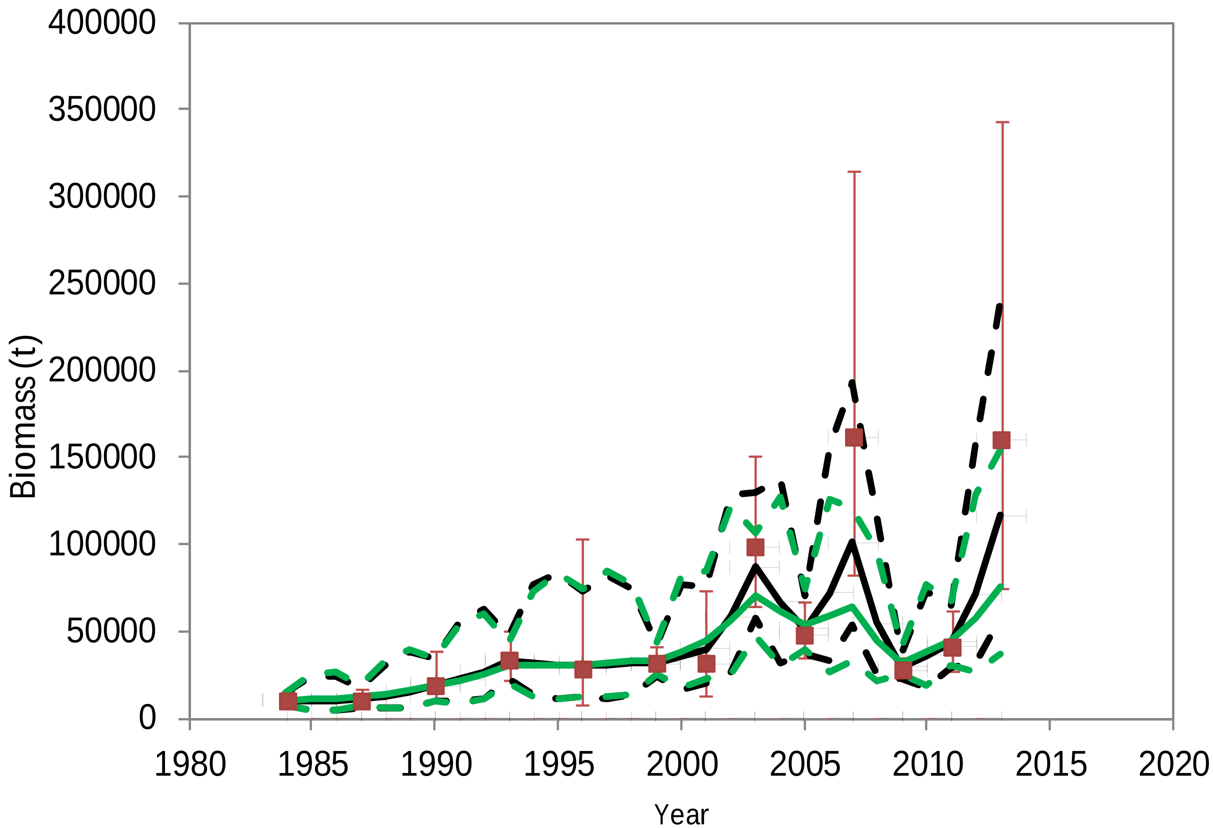


$$\frac{\sigma_e^2}{\sigma_a^2}$$

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

Implied from fit to GOA dogfish

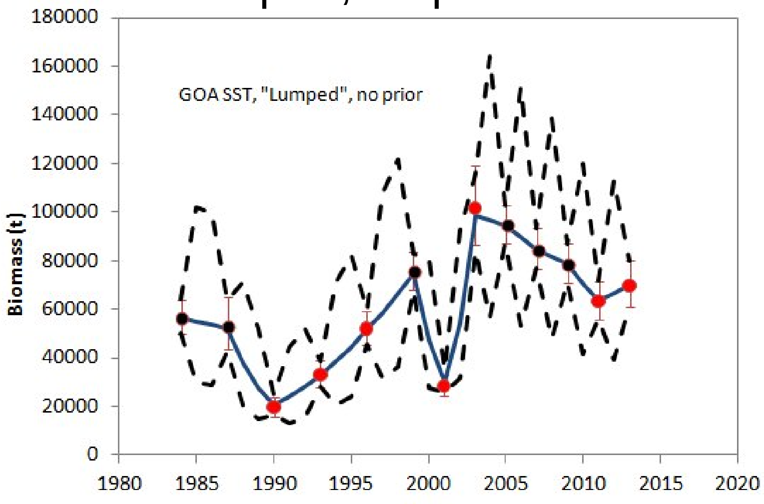
The fit with the prior constrains the estimate of process error standard deviation, and appears more reasonable



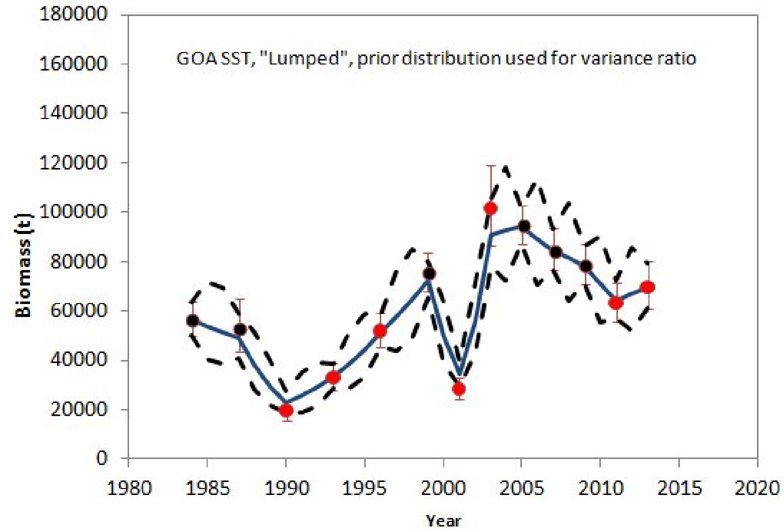
Fit with prior shown in green

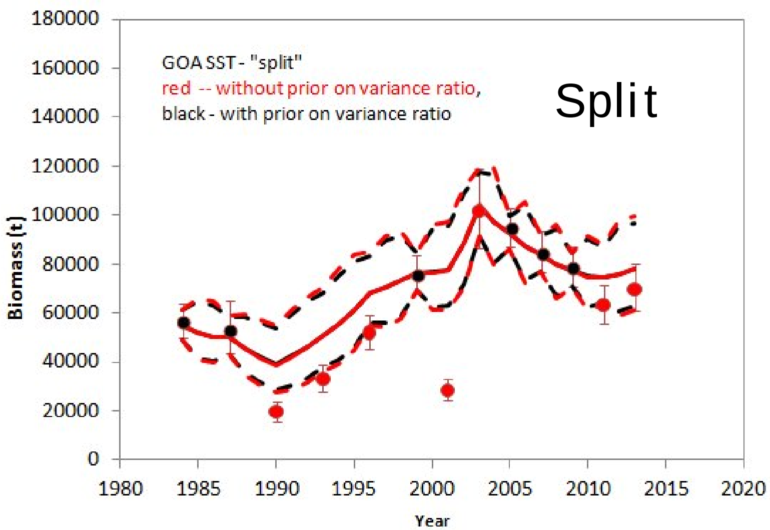
GOA Shortspine Thornyhead -- also has problem of missing data strongly affecting the results.

Lumped, no prior



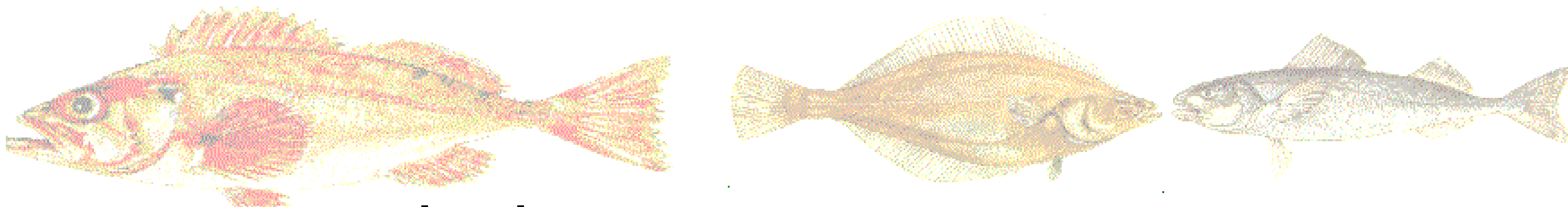
Lumped, with prior





Red data points are years in which some areas/depths were not sampled

In 2001, a large portion of the population was not sampled.



Recommendations

Obtaining survey biomass estimate

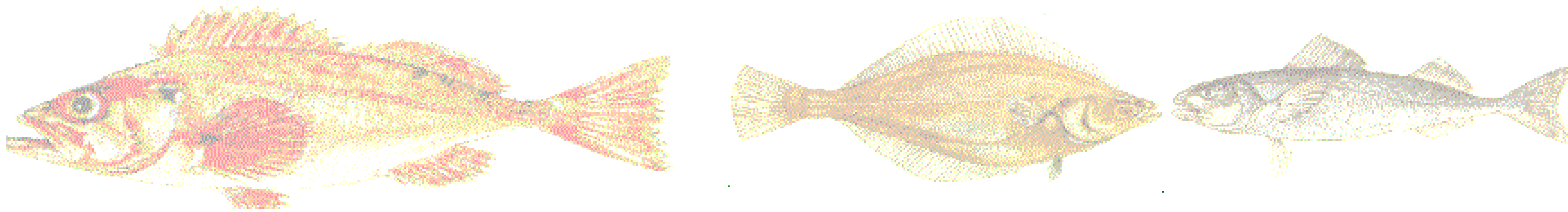
1) If the areas/depths sampled are consistent between years, apply the random effects model to the sum of the survey biomass estimates from the areas/depths.

2) If the areas/depths sampled differ between years, apply the random effects model separately to survey biomass estimates from each subarea, and sum the smoothed results.

Obtaining subarea proportions

1) Obtain subarea proportions by applying the random effects model separately to survey biomass estimates from each subarea.

Note: For multispecies complexes, combined the survey biomass estimates across the component species within the complex.



Next steps

Further developments to model code will address

1) Fitting many subareas simultaneously (with option to estimate a single process error variance across the areas).

2) Use of prior distribution on the ratio of observation error variance to process error variance.

3) Use of multiple surveys.