

# Strawperson for comparing an alternative assessment against a previously endorsed assessment

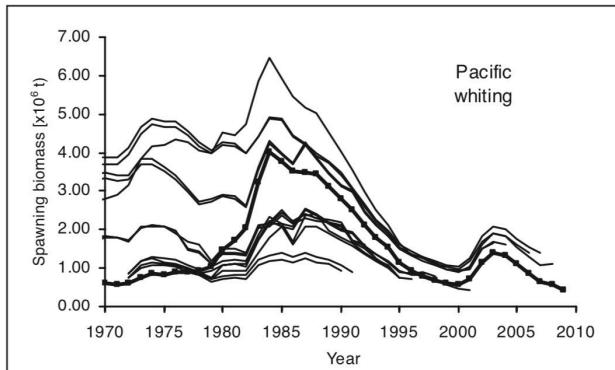
Will Satterthwaite, [will.satterthwaite@noaa.gov](mailto:will.satterthwaite@noaa.gov)

GFSC meeting 9/29/2021

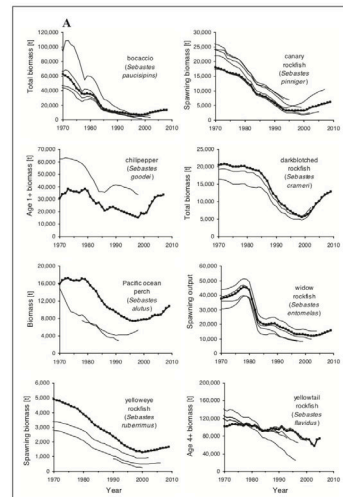
Disclaimer: I don't speak for anyone but myself, and even then I wonder sometimes.

# Some premises

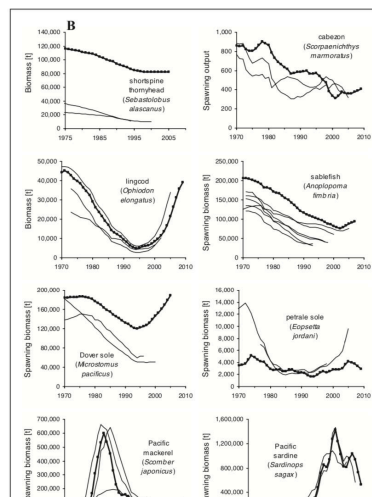
- SSC has responsibility to provide unbiased, risk & policy-neutral advice
  - GMT, GAP, public, agencies, Council, & WCRO all have roles/responsibilities too
- Assessments require numerous interacting decisions and assumptions
- ~Equally supported models can yield different results



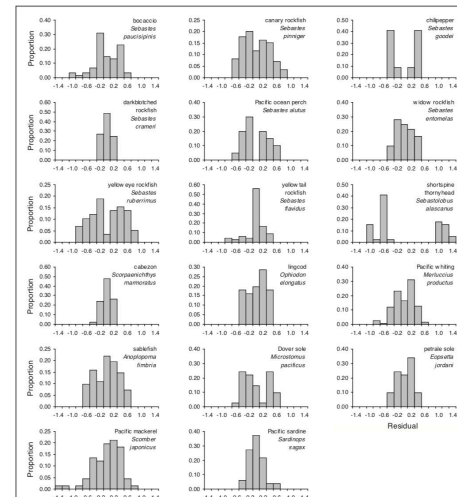
**Figure 1**  
Biomass time series for Pacific whiting (*Merluccius productus*) based on 15 historical stock assessments conducted for the Pacific Fishery Management Council. The bold line with square symbols represents the most recent stock assessment used in the meta-analysis; the other lines represent time series of abundance developed from earlier assessments.



**Figure 2**  
(A and B) Biomass time series for 10 selected groundfish and coastal pelagic species from stock assessments conducted for the Pacific Fishery Management Council on the west coast of the United States. In each panel the bold line with square symbols represents the most recent stock assessment that was completed, whereas the lighter lines represent biomass time series from earlier assessments.



**Figure 2 (continued)**



**Figure 3**  
Frequency distributions of log-scale biomass deviations for selected groundfish and coastal pelagic species in stock assessments conducted for the Pacific Fishery Management Council. Residual deviations were calculated from annual means taken from the biomass time series presented in Figure 2 (A and B).

Ralston et al. 2011  
Fish. Bull. 109:217–231  
<https://spo.nmfs.noaa.gov/content/meta-analytic-approach-quantifying-scientific-uncertainty-stock-assessments>

- Danger in working backward from desired result or cherry-picking alternatives
- Best avoided through objective, repeatable, policy-neutral criteria

# Proposed criterion for substantial difference (I)

- Step 0.1: Verify that any added data are representative and reliable, and of sufficient quality and quantity to be informative
- Step 0.2 Verify that any requested model changes are mechanistically plausible
- Step 0.3 Verify that any new models/model runs have acceptable diagnostics
- Step 1: Calculate terminal biomass  $B_{alt}$  of alternative model, for comparison with terminal biomass  $B_{end}$  of endorsed model
- Step 2: Calculate  $X = \text{abs}(\log(B_{alt}/B_{end}))$ 
  - This puts the divergence between the endorsed and alternative models on the same scale as sigma, our well-established metric of assessment uncertainty
  - Probably best to compare against sigmas based on biomass (e.g., Ralston et al 2011), not those based on OFL (i.e., values adopted in 2019) – call this  $\sigma_b$

# Proposed criterion for substantial difference (II)

- Step 3: Compare magnitude of log-scale proportional divergence between model biomass outputs ( $X$ ) to typical assessment uncertainty in biomass ( $\sigma_b$ ), then...
  - $X \leq \sigma_b$  is within the expected level of noise for accepted assessments
    - Will  $S$ 's recommendation: stick with endorsed assessment as BSIA to inform management and status
  - $\sigma_b < X \leq 2\sigma_b$  is a somewhat unexpectedly large, but not extreme, level of divergence
    - Will  $S$ 's recommendation: stick with endorsed assessment as BSIA to inform management/harvest spex
    - May not want to use to inform status (though this would contradict SSC advice in June & Sep 2021)
    - Prioritize a full assessment next cycle
  - $X > 2\sigma_b$  is a large divergence that casts serious doubt on the original assessment
    - Will  $S$ 's recommendation: reject original assessment, bearing in mind steps 0.1-0.3
    - Prioritize a full assessment next cycle
- Let's circle back to discuss specific cutoffs and responses, but first...

# Is this approach relevant & justified?

- Biomass is probably the most fundamental output of an assessment
- $\sigma_b = 0.36$  for category 1 assessments is derived from a peer-reviewed meta-analysis (Ralston et al. 2011) and essentially replicated ( $\sigma_b = 0.39$ ) in an updated analysis (Privitera-Johnson & Punt 2019)
- $\sigma_b = 0.72$  for category 2 assessments is less directly supported. Meta-analysis (Privitera-Johnson & Punt 2019) suggested  $\sigma_b = 0.51$ , but sample size was small.
- *(Approach may not be suited to category 3 assessments.)*
- Using  $\pm 1$  SD or SE is very common for characterizing spread or uncertainty around a mean or parameter estimate, using  $\pm 2$  (or 1.96) SD or SE very common for central 95% or 95% confidence interval
- Other thresholds like  $0.67\sigma$  (central 50%) or  $1.64\sigma$  (central 90%) might merit a priori consideration