



NOAA
FISHERIES

**Southeast
Fisheries
Science Center**

Weight Estimation Methods

Gulf of Mexico Fishery Management Council

April 13, 2021



Marine Recreational Information Program Access Point Angler Intercept Survey Average Weight Procedure



References

General

- David Haziza, J.N.K. Rao. (2010) [Variance Estimation in Two-Stage Cluster Sampling under Imputation for Missing Data.](#) *Journal of Statistical Theory and Practice* 4:4, pages 827-844.

NOAA Fisheries Website

- Recreational Fishing Data Glossary: [Weight Data](#)
- [Survey Statistics Overview](#)
- [Survey Design and Statistical Methods for Estimation of Recreational Fisheries Catch and Effort](#)

Workshops

- [South Atlantic Council SSC Workshop](#) (August 2019)
- [Gulf Council SSC MRIP Workshop](#) (August 2020)
- SEDAR 67: [Sample Size Sensitivity Analysis for MRIP Weights](#)

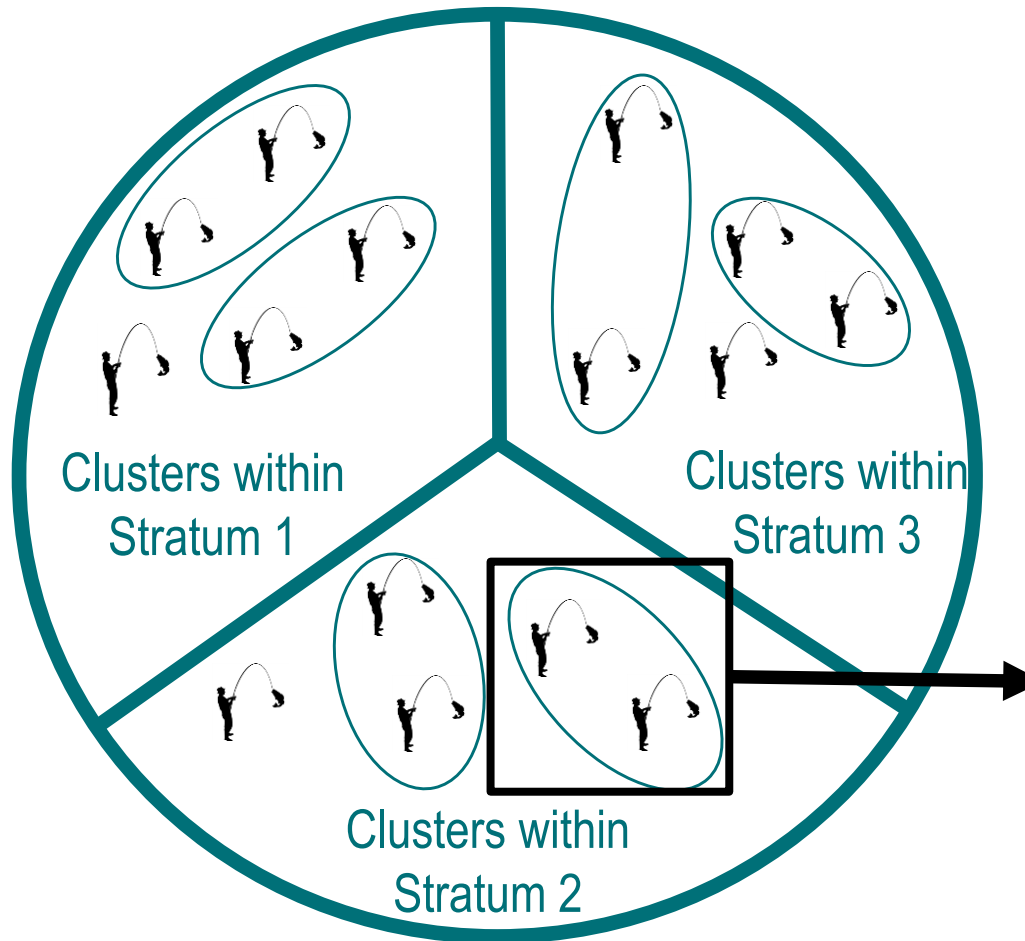
AP AIS Overview

- In-person **interviews of anglers** intercepted at **public fishing access sites**.
- Sample frame derived from NOAA Fisheries' Public Fishing Access Site Register.
- Data collected continuously, used to **estimate catch rates** and trip characteristics for two-month waves.



AP AIS Design

- Stratified, clustered **multi-stage design**.



Primary Stage Unit (PSU):
Site Cluster-Day-Time Interval

Secondary SU:
Sample Duration (time spent
sampling each site in a
cluster)

Tertiary SU:
Angler Trips

Quaternary SU:
Catch

APAIS Sample Weights

PSU – Stage I Sample Weight ¹

$$w_1 = \frac{1}{\text{inclusion probability}}$$

Sample Duration – Stage II Sample Weight

$$w_2 = \frac{\text{length of assigned time interval}}{\text{time spent subsampling at a site within a cluster}}$$

Angler-Trip – Stage III Sample Weight

$$w_3 = \frac{\text{Total angler trips observed at a site}}{\text{Total angler trips intercepted at a site}}$$

Catch – Stage IV Sample Weight

$$w_4 = \frac{\text{Total number of fish harvested by an angler trip}}{\text{Number of fish sampled from that angler trip}}$$

APAIS Catch Rate Estimates

Mean catch per angler trip is calculated as a **domain estimate**, defined by year, wave, region, state, fishing mode, area fished (inland, nearshore, offshore), species, and catch type.

Final sample weight ($w_1 * w_2 * w_3$ if estimating numbers of fish, or $w_1 * w_2 * w_3 * w_4$ if estimating total landed weight) for angler trip in domain d

Number of individuals OR weight (lbs or kg) of fish caught on angler trip in domain d

Catch rate
in domain
 d

$$\hat{\bar{y}}_d = \frac{\sum w_d y_d}{\sum w_d}$$

This is a standard weighted mean estimator used in survey statistics (e.g., SAS Institute Inc., 2016).



Catch Information Collected by APAIS

Type A

Observed Harvest

Type B1

Unobserved Harvest

Type B2

Releases

Used to estimate:

- Total landings (in numbers of fish)
- Total landings (in lbs or kg)
- **Mean fish weights and lengths**
- Length frequencies

Used to estimate:

- Total released alive (in numbers of fish)

Total Catch (Landings + Releases in Numbers of Fish)



Missing Data

Potential scenarios:

- Missing length **or** weight data
- Missing length **and** weight data



Missing Length or Weight Data

Fill length (L) or weight (W) data with standard length-weight relationship models:

$$L = \left(\frac{W}{a} \right)^{\frac{1}{b}}$$

Shape parameter for the body type of the fish species

Scaling coefficient for the weight at length of the fish species

$$W = aL^b$$

Missing Length *and* Weight Data

Fill up to five length and weight observations for each species on a given trip using a combination of **hot and cold deck imputation**:

Hot Deck

Missing values are replaced by **values from a similar unit in the same dataset**

(e.g., species lengths and/or weights from the same year, wave, and state)

Cold Deck

Missing values are replaced by **values from a similar unit in a different dataset**

(e.g., species lengths and/or weights from the same sub-region, but the previous year)

Length *and* Weight Imputation

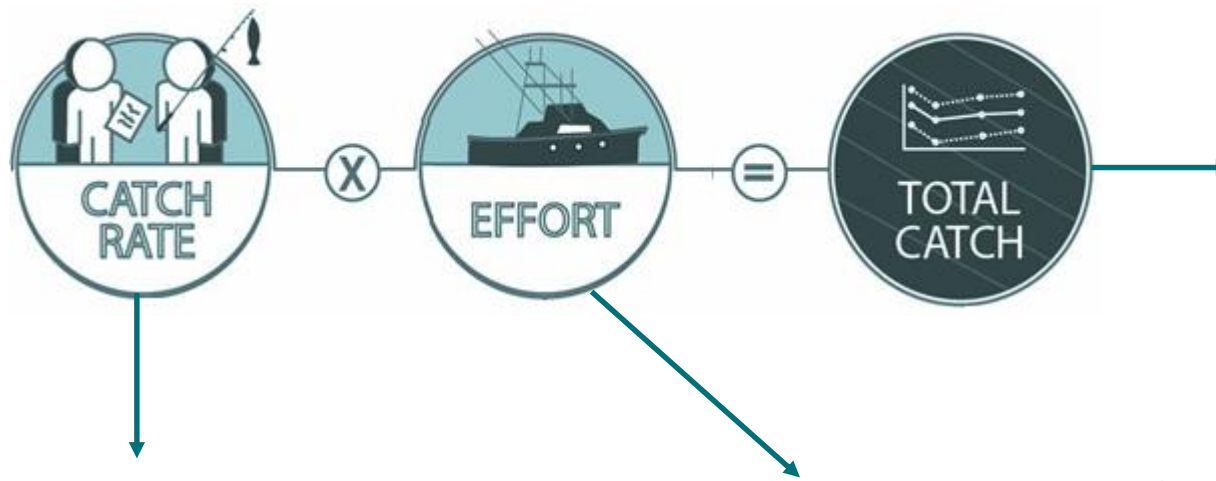
- Five rounds of attempted imputation. Each round starts with the most similar data and proceeds to less similar data.
 - Most imputations are completed within three rounds.
 - **Imputed data are always from the same species and sub-region.**
 - Sample weights are not factored into imputation to avoid introducing unknown biases into the data.

Length *and* Weight Imputation

Round	Imputation Cell	Min. Number of Completed Observations Required
1	Current year, wave, sub-region, state, mode, area fished, species	10
2	Current year, half-year (waves 1-3 or 4-6), sub-region, state, mode, species	5
3	Current + most recent prior year, wave, sub-region, state, mode, area fished, species	5
4	Current + most recent prior year, sub-region, state, mode, species	5
5	Current + most recent prior year, sub-region, species	1

If the minimum number of observations are not met, imputation proceeds to the next round. If no imputations can occur after five rounds, no additional attempts are made to fill in missing data.

Catch Estimation



- Observed Landings (Type A)
- Unobserved Landings (B1)
- Releases (B2)
- Total Landings (A+B1)
- Landings + Releases (A+B1+B2)

Weighted APAIS Catch Rate

- Includes three sample weighting components.
- Calculated using standard weighted mean estimator.

Weighted FES Effort

- Includes three sample weighting components.
- Calculated using standard weighted total estimator.
- **From APAIS:** Adjustment factor to account for out-of-state angler trips/
- **From APAIS:** Partitioned by area fished (inland, nearshore, offshore).

Total Catch Estimates

Total Catch
Wave

Weighted FES Effort
*Adjusted for out-of-state trips,
partitioned by area fished*

$$\hat{Y}_d = \hat{T}^* \hat{y}_d$$

Weighted
APAIS Catch
Rate Estimate

Total
Catch
Annual

$$\hat{Y}_D = \sum \hat{Y}_d$$

Summary

- Imputation is an established practice for large-scale surveys, wherein replacement values are used when data are missing for a given cell (usually due to item non-response).
- NOAA Fisheries' Marine Recreational Information Program uses a conservative imputation approach to assign average weight to an estimation domain for which **both** lengths *and* weights are missing.
- There are limitations to any imputation approach:
 - To help inform decisions related to data use, NOAA Fisheries is working on metrics to show the relative contribution of imputed data to final estimates. (See red grouper tables provided as Appendix.)
 - The agency continues to evaluate potential improvements to current methods.
- For 2020, imputations will include 2018 and 2019 data.

SEFSC weight estimation



SEFSC weight estimation

- MRIP provides species specific catch estimates by stratum (species/year/wave/sub-region/state/mode/area) in numbers of fish
- Corresponding weight measurements are not always available due to sampling constraints or incomplete self-reporting.
- Recreational landings estimates were historically provided in numbers of fish for stock assessments.
- When management measures, such as ACL monitoring, began requiring estimates in weight on a routine basis, a standard methodology was developed to estimate missing recreational weights (SEDAR 22-DW-16)

SEFSC weight estimation

- Ensures consistent weight estimation methodology across all years
- General methodology documented in SEDAR 32-DW-02; used to estimate all general recreational weight estimates
 - calculate average weights by strata
 - strata hierarchy: species, region, year, state, mode, wave, and area fished.
- Currently use new MRIP size datasets, including weights imputed by the survey program- since November 2018
- Change in minimum sample size from 30 to 15- Fall 2019 (SEDAR67-WP-06)

SEFSC weight estimation

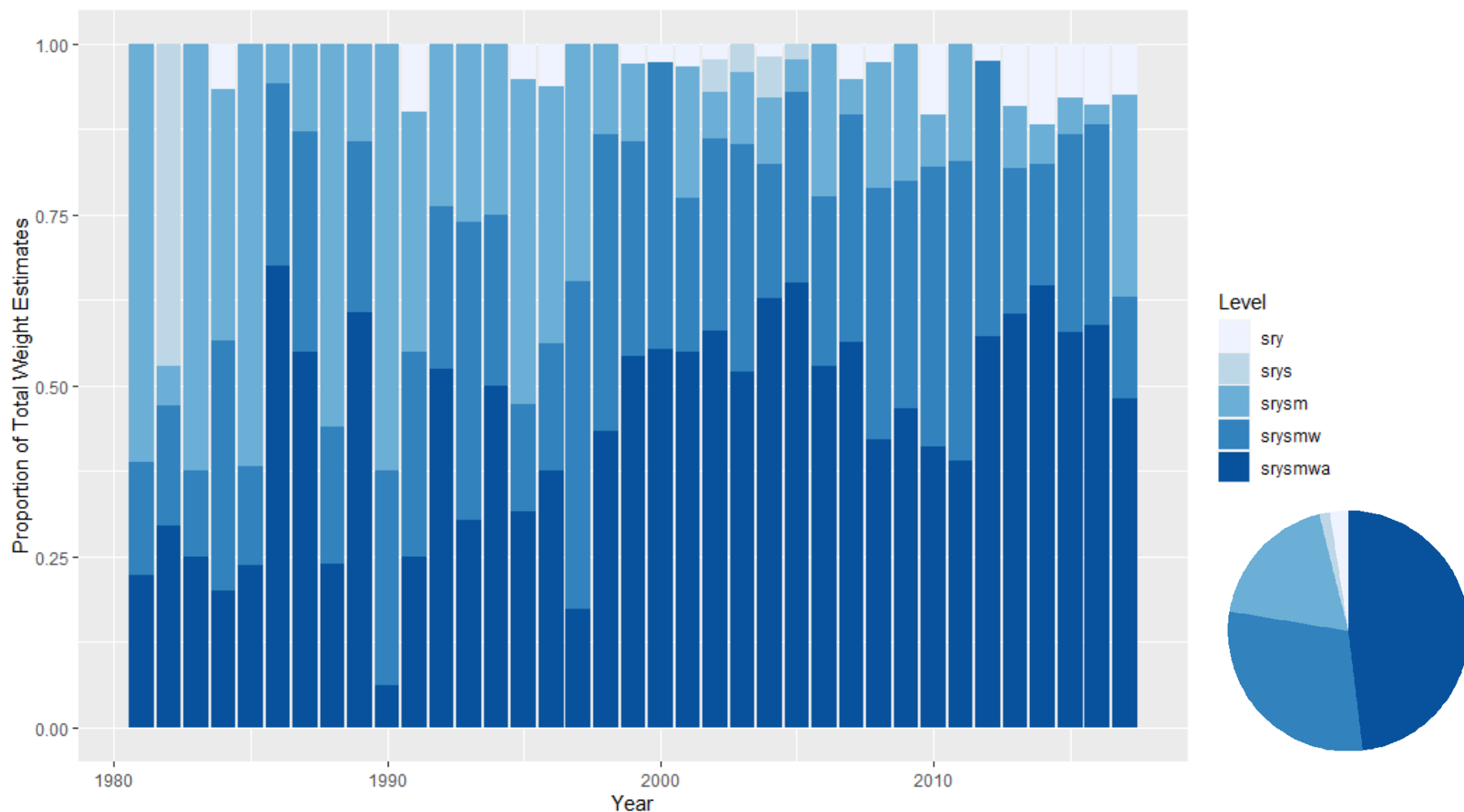
- Step 1 – Calculate average weights from intercept data (WGT) for each i strata
 - strata hierarchy: species, region, year, state, mode, wave, and area fished

$$\overline{WGT}_{i(strata)} = \frac{\sum_{n=1}^N WGT_n}{N}$$

- Step 2 – Identify strata that meets minimum sample size threshold (15 fish)
- Step 3 – Apply appropriate average weight to convert estimate of landings-in-numbers ($AB1$) to landings-in-weight
 - Landings-in-number provided at finest (area fished) strata

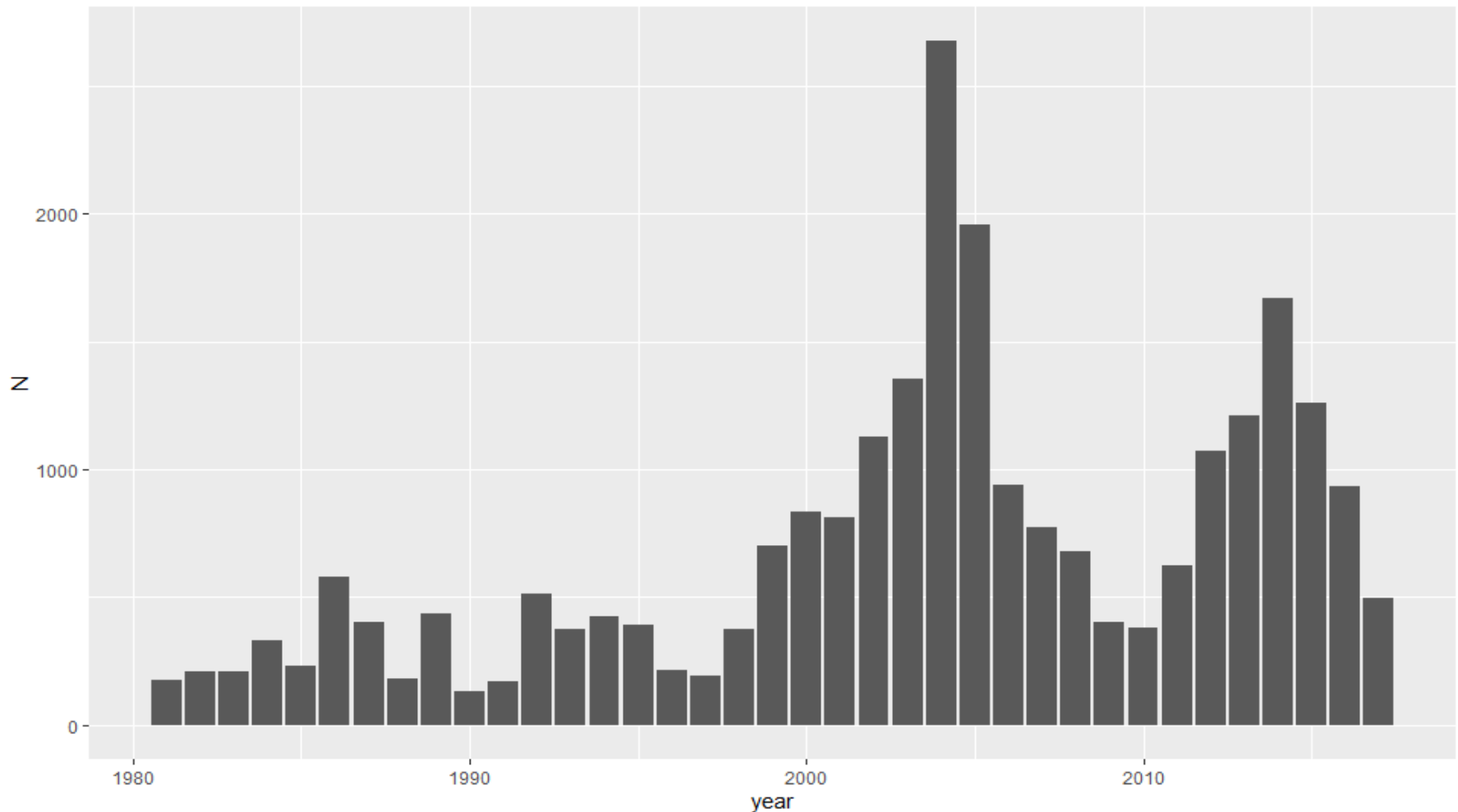
$$AB1(lbs)_i = AB1_i * \overline{WGT}_i$$

Proportion of MRIP weight estimates by strata for Gulf of Mexico Red Grouper



MRIP sample sizes by year (weights)

SEDAR 61 Gulf Red Grouper



SEDAR 61 Gulf Red Grouper model weight estimation

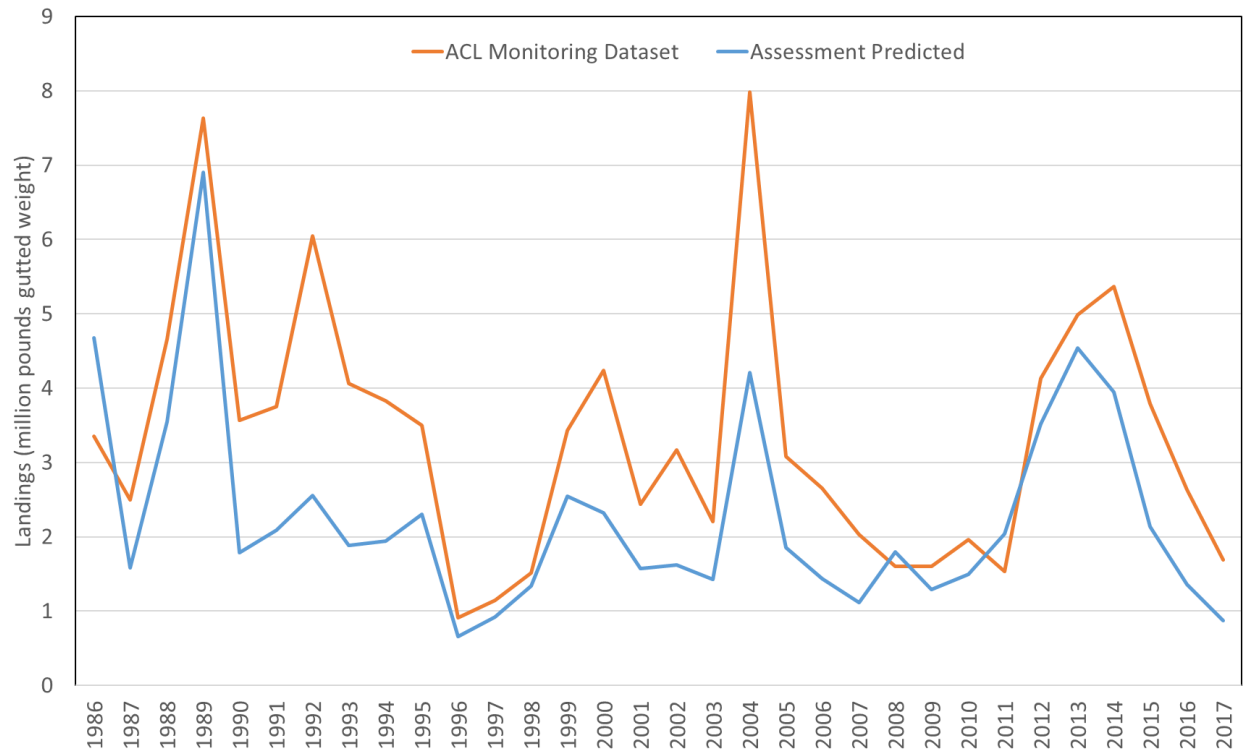


Background

- Data provided in November 2018 did not include recreational landings in weight (not reviewed at the DW/AW) due to the following factors:
 - Severe backlog of assessments due to the new release of MRIP FES/APAIS adjusted estimates back in time
 - Significant updates to SEFSC data processing required as a result of the new data
 - SEDAR 42 assessment model used recreational landings in number of fish. Since SEDAR 61 was a standard assessment, recreational landings in weights were not prioritized

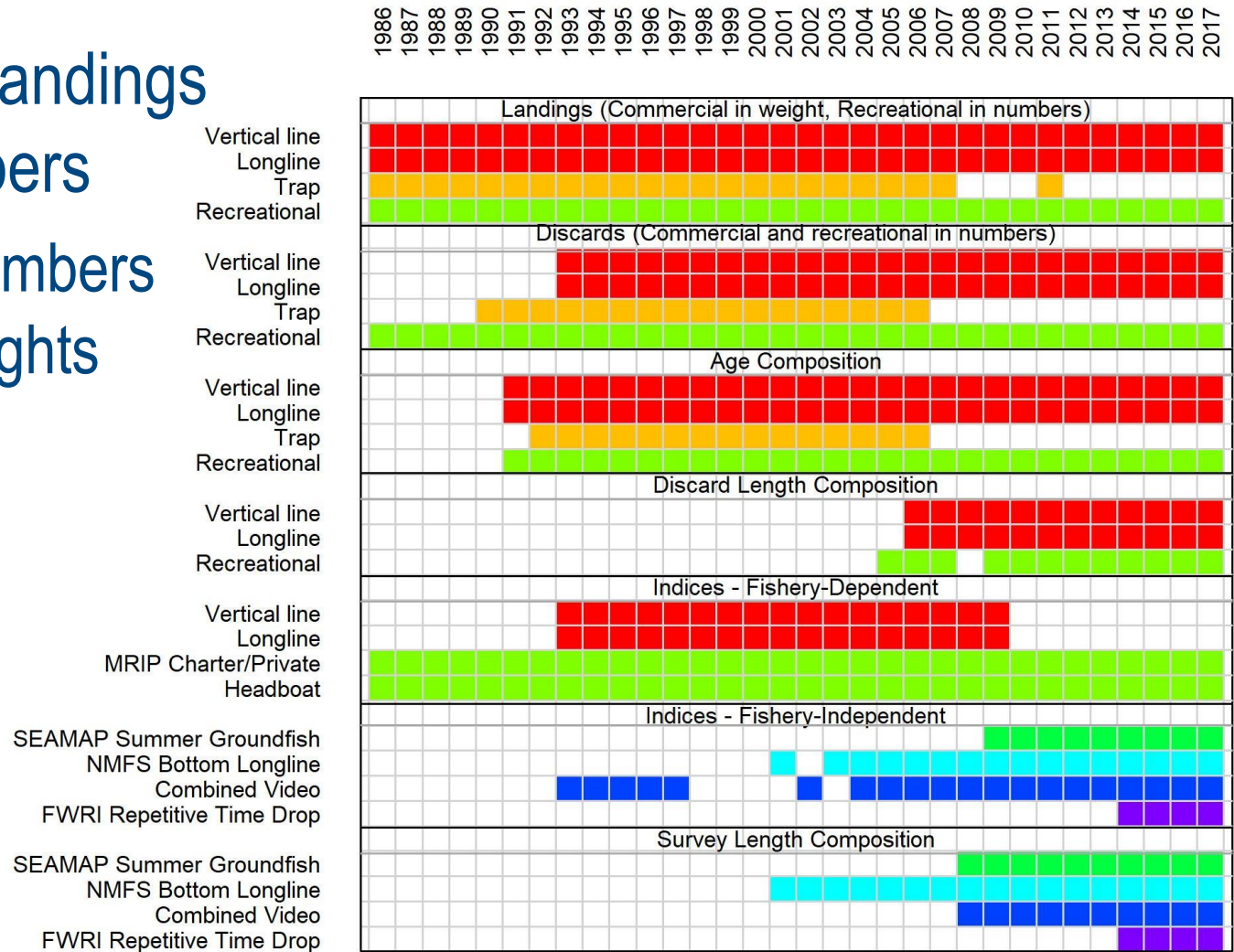
Why do predicted landings from the SEDAR61 assessment model differ from landings estimates in the ACL monitoring dataset?

1. Input data
2. Uncertainty assumed in assessment
3. Differences in weight estimation



1. Input Data

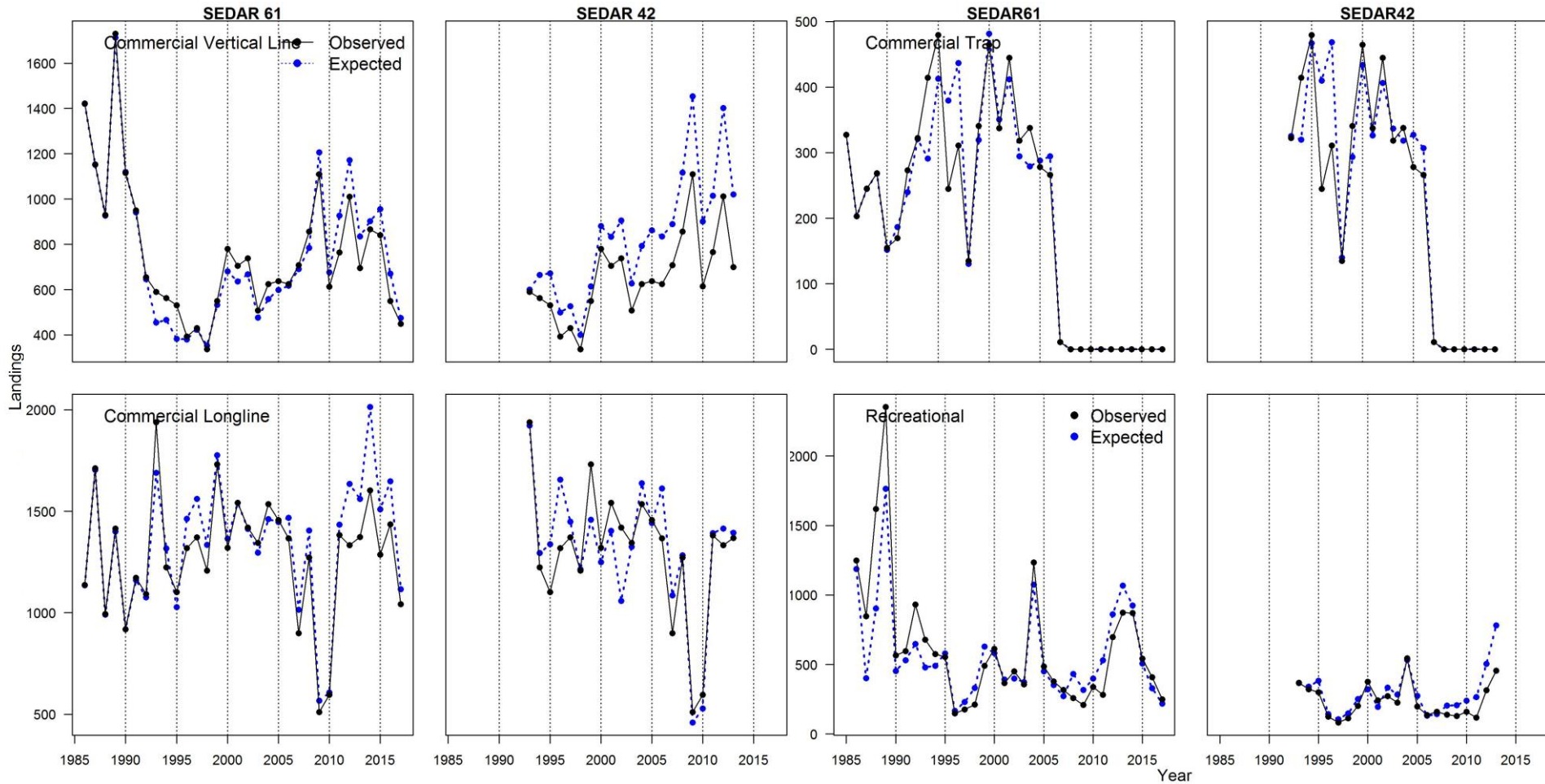
- Recreational landings input as numbers
 - Fitting to numbers and not weights



2. Uncertainty assumed in the assessment

- Assessment model assumes greater uncertainty in recreational landings compared to commercial landings
 - Commercial CV = 0.15
 - Recreational CV = 0.3
 - The model has more flexibility and is also fitting to a variety of other data sources beyond just landings (discards, compositions, indices)
- Results in predicted landings (in numbers) that are not identical to input landings (in numbers)

Fit to Data: Landings (slide 42 in Sep 2019 pres)



Commercial (metric tons): $CV = 0.15$ Recreational (1000s of Fish): $CV = 0.3$

3. Differences in weight estimation

- Assessment model uses the length-weight equation to convert predicted landings (in number) into weight units
- MRIP estimates are based on stratified observations of average weight (SEDAR 67-WP-06), but no measure of uncertainty is provided with weights.

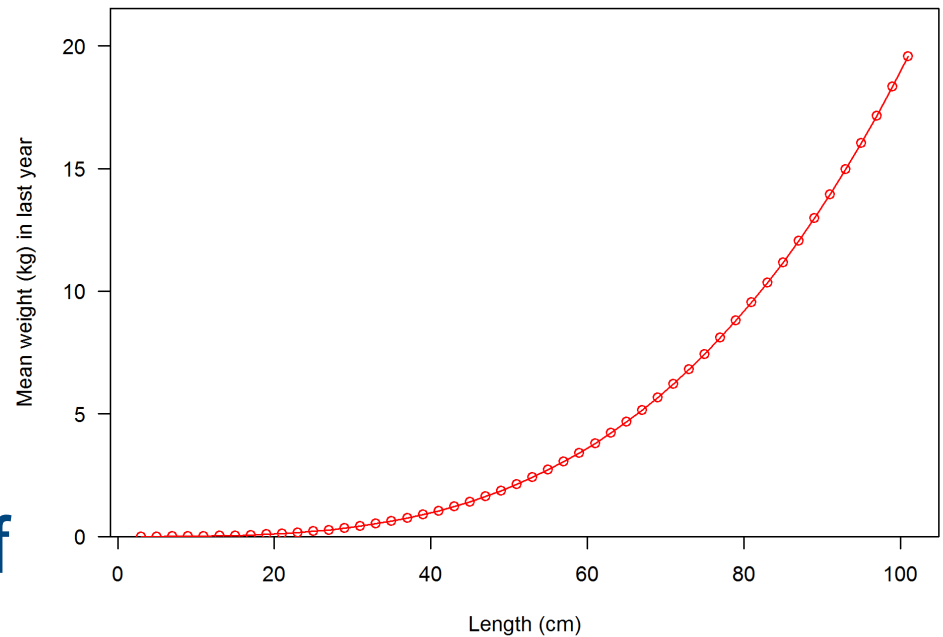


Figure 4.2 in SEDAR61 report

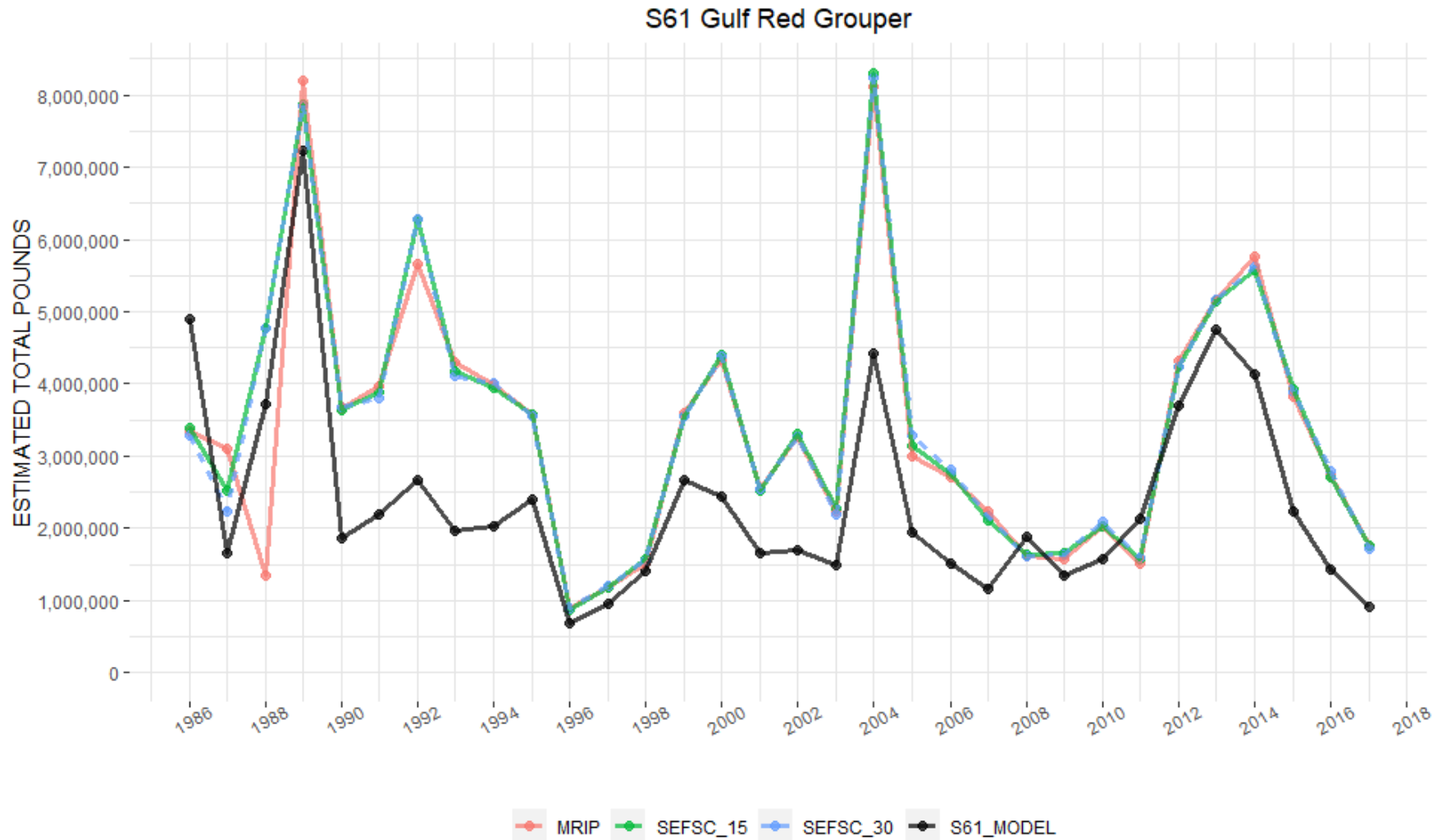
Research track on Gulf scamp (SEDAR68)

- Why have we used recreational landings in numbers?
 - Numbers have been the most reliable measure in past assessments
- Weight estimates are now consistently provided for MRIP landings (but not discards)
 - Can look into modeling recreational landings in weights instead of numbers during ongoing research track assessments
 - Will require an assumption about CV around these estimates

Weight Estimation Method Comparison



Weight Estimation Method Comparison



Summary

- Red Grouper are well sampled throughout the time series with 78% of the total SEFSC weight estimates using average weights from the finest two strata levels
 - 48% at the species, region, year, state, mode, wave, area level
 - 30% at the species, region, year, state, mode, wave level
- Difference in landings estimates in weight between those predicted by the stock assessment model and those estimated by the SEFSC weight estimation procedure and MRIP is a function of how recreational landings are input (numbers) and treated (e.g., with uncertainty) in the stock assessment

Questions

