Exploratory work on possible improvements to purse-seine catch data collection and estimation

Data Review Workshop, La Jolla, California, USA, 2-6 October 2023

Possible improvements to P-S data collection and catch estimation

Covered in this presentation

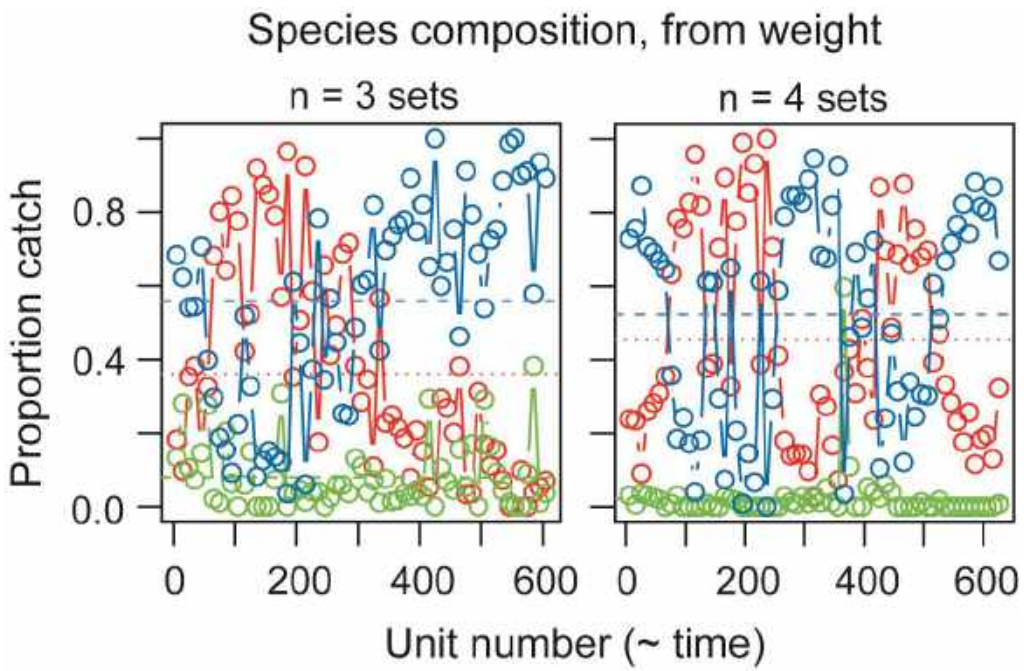
Exploratory work on possible improvements to data collection and estimation for purse-seine fleet-level catch composition.

Outline

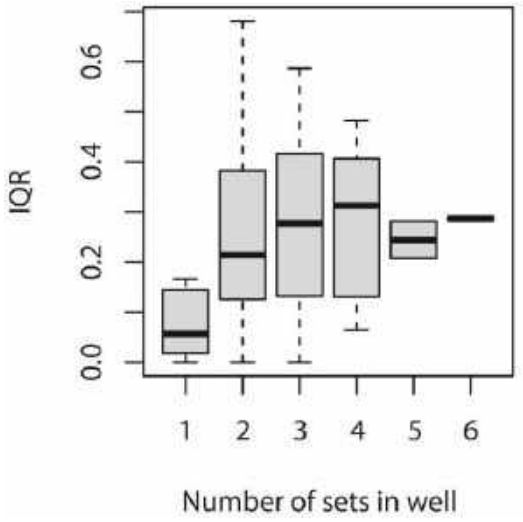
- Motivation for the work (from the Enhanced Monitoring Program)
- Background
  - Details of the regular port-sampling program for the floating-object (OBJ) fishery
- Exploratory simulation study
  - Methods
  - Results and conclusions

Motivation for the work

- Insights from Enhanced Monitoring Program*:
  - There can be considerable variability in species composition over the unloading of catch from OBJ-set wells.
  - The regular port-sampling protocol applied to these two wells would likely equate to the catch spanned by only a few units per well.
  - Pattern appears related to number of sets in the well, among other factors.
- 'New' data type available for simulation studies:
  - 'well plan' data can be constructed from various observer data sources.
- A question of interest: Could within-well variability in species composition matter when estimating fleet-level catch?



Red: BET; green: YFT; blue: SKJ



* Documents SAC-14-10, SAC-14 INF-I

Background

- Focus on the OBJ-set fishery.
- Some additional details of the regular port-sampling data collection:
  - Overall, for 2010-2019 (EPO):
    - Mostly 1 well sampled per trip, rarely more than 2 wells sampled per trip.
    - Samples mostly collected from the upper half of the well.

Number of OBJ-set wells sampled per trip (EPO; Class 1-6), by year

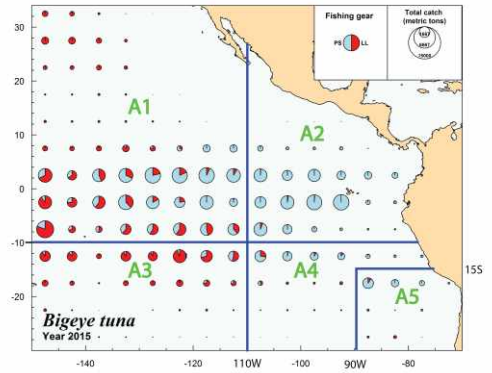
						Total trips sampled	% Trips with 1-2 wells sampled
2019							
OBJ-set wells sampled per trip	1	2	3	4			
Number of trips	181	76	15	1		273	94.10%
2018							
OBJ-set wells sampled per trip	1	2	3	5			
Number of trips	212	58	13	1		284	95.10%
2017							
OBJ-set wells sampled per trip	1	2	3	4	5		
Number of trips	179	95	23	6	2	305	89.80%

Quarter of the well from which samples were collected (EPO Class 1-6, OBJ-set wells; by row, may not sum to 1 due to rounding).

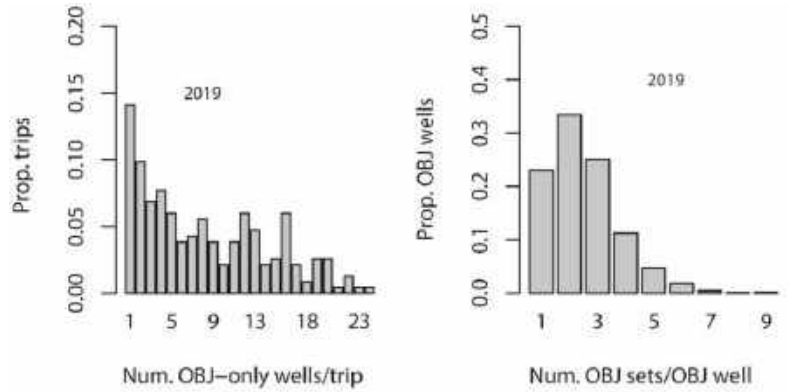
Year	1st quarter of well	2nd quarter of well	3rd quarter of well	4th quarter of well
2019	40%	40%	15%	5%
2018	42%	37%	14%	6%
2017	39%	43%	13%	6%

Background

- Going into more detail for Area A1 of previous bigeye tuna (BET) assessment (Class-6 vessels):
  - Of trips with at least 1 OBJ set, 68%, on average, had at least 1 OBJ-set well (i.e. ≥ 1 *sampleable* well).
  - Coverage of sampleable trips and wells was about 50% and 9%, respectively.



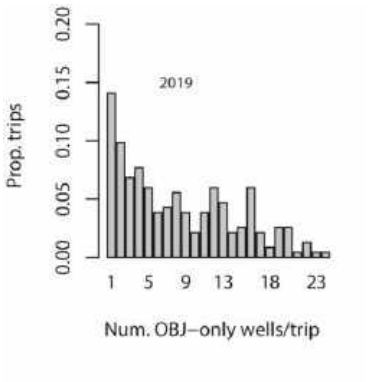
  - The distribution of sampleable OBJ-set wells per trip was right-skewed.
  - Roughly 75% of wells were loaded with catch from two or more sets.



Area A1 (Class 6)	# trips with at least 1 OBJ set	# wells with at least 1 OBJ set	# trips with at least 1 OBJ-set well	# OBJ-set wells	# trips sampled with OBJ-set wells	# OBJ-set wells sampled
2019	322	2847	234 (73%)	1911 (67%)	122 (52%)	168 (9%)
2018	297	2442	195 (66%)	1580 (65%)	93 (48%)	139 (9%)
2017	357	3266	259 (73%)	2400 (73%)	140 (54%)	236 (10%)

Exploratory simulation study: methods

- Set up
  - Uses observer data (the 'new' well plan data), 2010-2019
  - OBJ-set fishery (Class-6 vessels), area A1
- Simulation steps
  - For each year
    - Create 200 synthetic data sets from trips with at least one OBJ set in year and area.
    - For each synthetic data set, repeat the following steps 30 times to generate 30 estimates:
      - Sample fixed percent of trips (at random w/o replacement).
      - Sample fixed number of wells per trip (at random w/o replacement).
      - 'Sample' sets within a well: all sets; 1 set selected at random; the last set loaded into well.
      - Two different estimators used for the fleet proportion BET in area A1, $\hat{p}_{BET}$:
        1) An approximation to current catch estimation methodology
           - Treats wells in sample as a simple random sample from all wells in stratum
        2) Estimation based on full sampling weights*
           - The sample weight of an 'observation' in a sample is the reciprocal of the probability that the observation is included in the sample (which is a product of conditional probabilities).
           - This addresses the fact that, for example, wells in a sample may represent different numbers of wells of the population because some trips have far more sampleable wells than other trips.



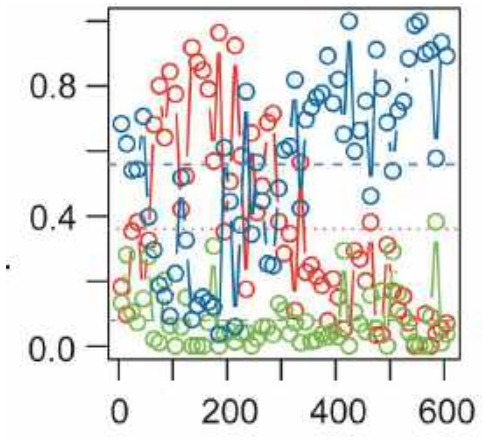
* Suggestion by Marti McCracken

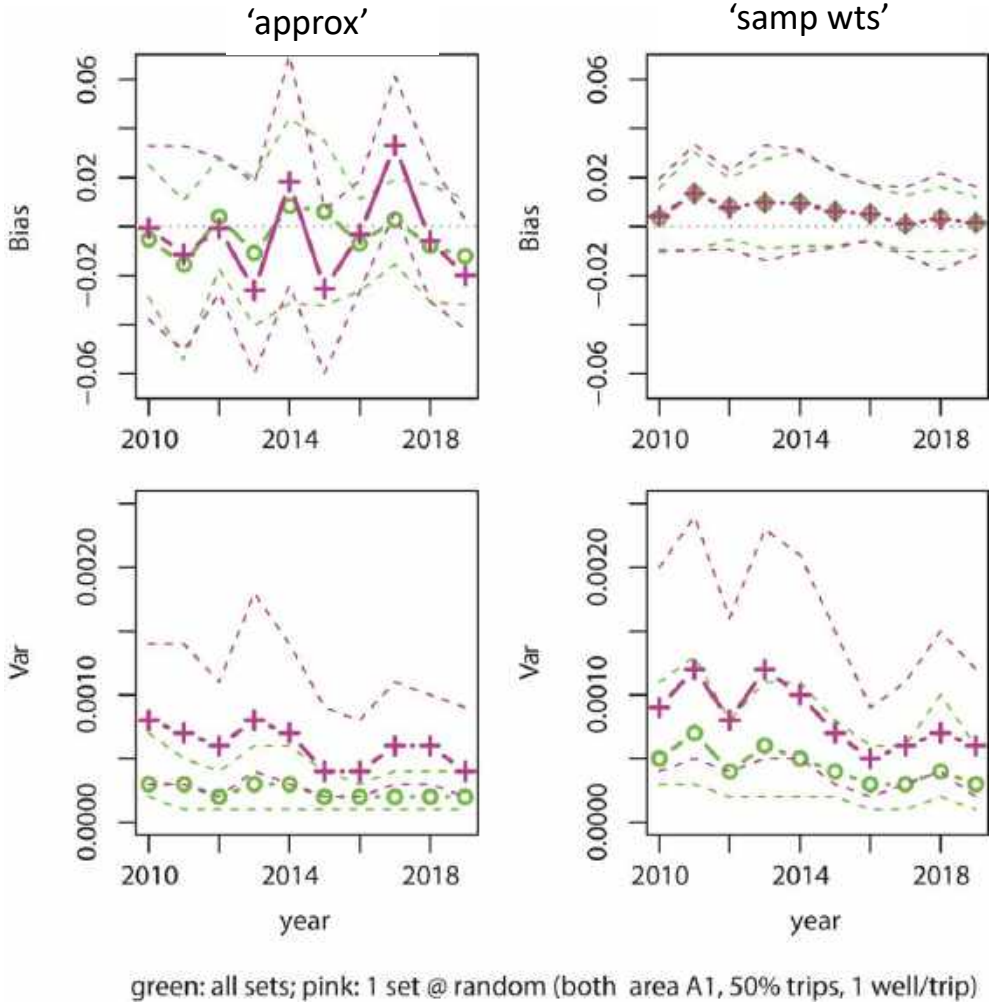
Exploratory simulation study: methods

- Summarize the simulation results for each synthetic data set x estimator:
  - Bias: $\overline{\hat{p}_{BET}} - p_{BET}$ (average of 30 estimates minus true value)
  - Variance: $\left(\frac{1}{29}\right)\sum\left(\hat{p}_{BET} - \overline{\hat{p}_{BET}}\right)^2$ (how close estimates from different samples are to each other)

Exploratory simulation study: results and conclusions

- Sampling within a well
  - 'All sets' sampling option led to somewhat better bias properties but mostly improved variance, compared to the other set-sampling options.
  - This suggests that because of limited sampling within a well, the current methodology likely inflates variance on the catch estimates, possibly resulting in a more variable catch times series.
  - To reduce variance, it is worthwhile to explore sampling protocols that sample more of a well than is done at present.

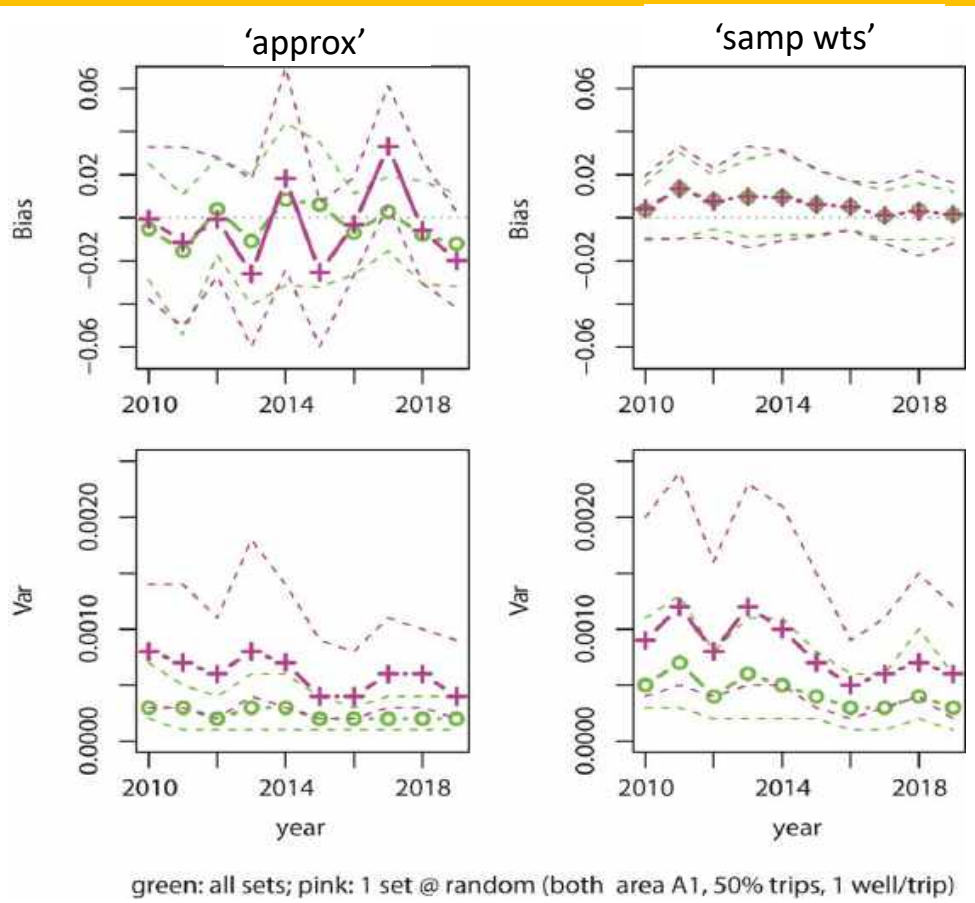




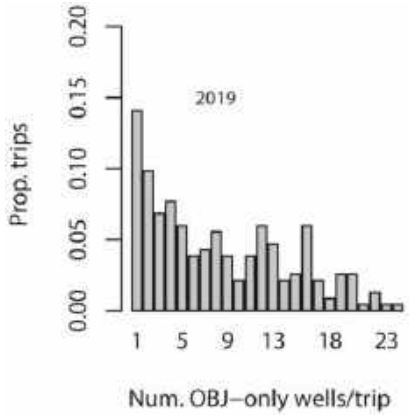
Solid lines: median, across 200 synthetic data sets
Dashed lines: extremes

Exploratory simulation study: results and conclusions

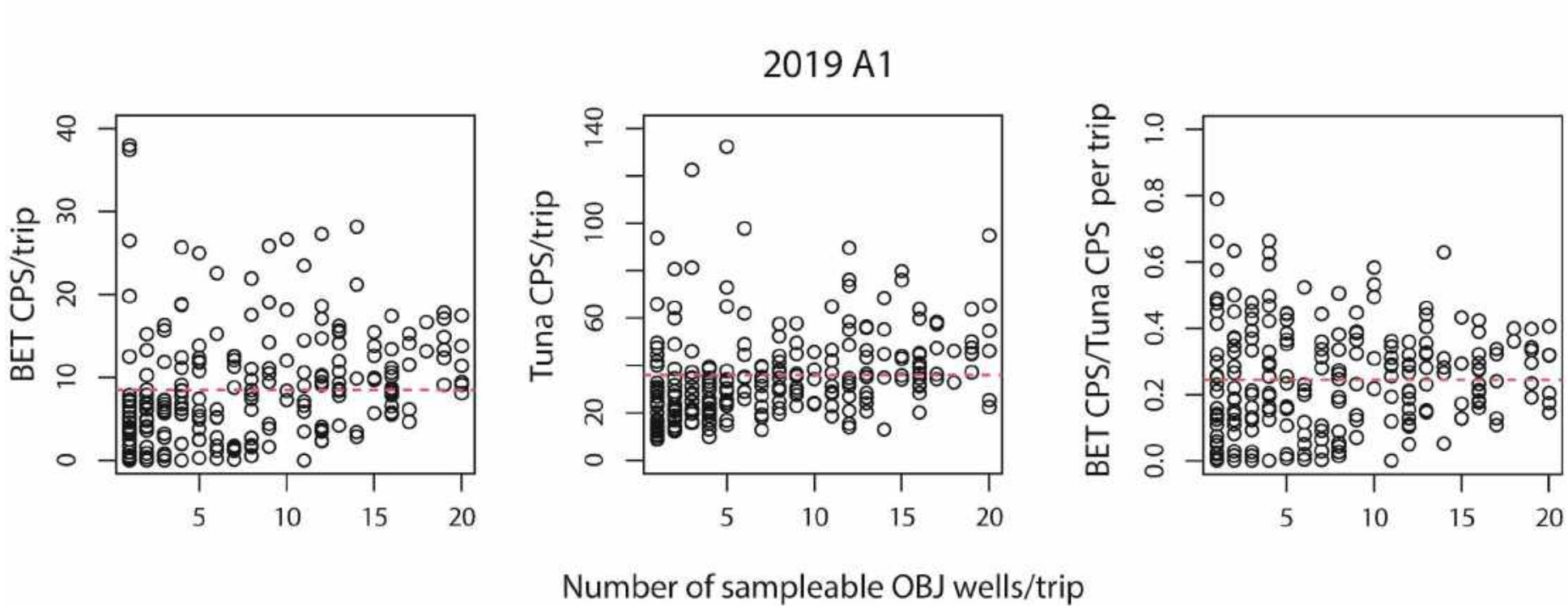
- Type of estimator
  - ‘Sample weights’ estimator had better bias properties, but it had higher variance.
  - This suggests bias may arise with the current methodology because differences among trips are not addressed in the estimation.
  - If minimizing bias is a priority, worthwhile to develop a sampling protocol for which sampling weights are known and used in the estimation (for design-based estimator).
- A workshop focusing on improving sampling designs for estimation of fleet catch should be conducted.



Solid lines: median, across 200 synthetic data sets
Dashed lines: extremes



Thank you! Questions?


2019 A1
BET CPS/trip
Tuna CPS/trip
BET CPS/Tuna CPS per trip
Number of sampleable OBJ wells/trip

Exploratory simulation study: methods

Estimator based on sampling weights

- Simulation sampling hierarchy: trips; well(s) within a trips; set(s) within a well.
- Consider the protocol where for each of the *n* trips selected, 1 well is selected per trip, and 1 set is selected per well.

$$Probability\ that\ a\ particular\ set\ is\ included\ in\ the\ sample\ = \left(\frac{n}{N}\right)\left(\frac{1}{M_i}\right)\left(\frac{1}{L_{ij}}\right)$$

n trips are selected for sampling out of *N* by SRS (*N* is the total number of *sampleable* trips in stratum), *1* is the number of wells to be sampled by SRS for the i^{th} trip, out of M_i *sampleable* wells, and *1* set is selected for sampling out of the total L_{ij} sets in the j^{th} sampleable well of trip *i*

$$Sampling\ weights\colon\ \omega_{ijk} = \left(\frac{N}{n}\right)\left(\frac{M_i}{1}\right)\left(\frac{L_{ij}}{1}\right)$$

- General form of estimator

$$\hat{p}_{\mathrm{BET}\_samp\ wts} = \frac{\sum_i \sum_j \sum_k \omega_{ijk}\, BET_{ijk}}{\sum_i \sum_j \sum_k \omega_{ijk} Tropical\ tuna_{ijk}}$$