

Center for Independent Experts (CIE) Independent Peer Review of the Shrimp Bycatch Estimation Methodology

By

Jon Helge Vølstad¹

April 2024

¹ Strangehagen 22, 5011 Bergen, Norway; jon.helge.volstad@gmail.com
Representing Center of Independent Experts

Executive Summary

The trawl fishery for shrimp fishery in the Gulf of Mexico (GOM) is one of the most important commercial fisheries of the US in terms of landings, economic value, and number of vessels participating. The GOM shrimp fishery also captures many non-target species discarded as bycatch. Estimates of bycatch are crucial input data for assessing many fish species in the southeastern U.S., and particularly for determining annual quotas for red snapper, a key species.

From around 2007 until 2021, estimation of bycatch has been conducted using data on CPUE from the shrimp fishery observer program, fleet effort data from vessel tracking devices, and fleet landings data from trip tickets, using data processing procedures that were developed in the mid- to late-2000s. This 'legacy code' for bycatch estimation was not well documented, and by 2021 the original developers were not available for consultation (retired, left agency, etc.) during annual updates of bycatch estimates.

In 2021, NMFS established a project team to update and quality assure the bycatch estimation procedures, and to assess possible sources of bias. A recent preliminary examination by the project team uncovered substantial systematic bias in the previous method used for bycatch estimation. Therefore, a complete overhaul of the methodology was conducted at the Southeast Fisheries Science Center (SEFSC). The new method utilizes four main data sources: effort data, observer data, a gear and landings survey, and shrimp landings data. A project team formed in 2021 implemented a cross-check procedure, estimating annual brown shrimp landings and comparing them to reported fleet landings. The results indicated a consistent 45% underestimation, revealing a systematic bias in the bycatch estimation process. The project team suspected that the bias was due to one or more of the principal data inputs for the estimation procedure: observer CPUE, fleet effort, and/or fleet landings. The origins of each input were unclear.

Shrimp trawl bycatch is estimated by combining estimates of fleetwide effort with independent estimates of catch per unit effort (CPUE) derived from the SEFSC shrimp fishery observer program in which scientific observers on commercial fishing vessels record detailed information on catch and effort for a subset of trawl hauls within a subset trips. The observer program became mandatory for federally permitted shrimp trawl vessels in 2007. The background materials provided for this review do not include a clear description of the survey design for selecting vessels and trips for the SEFSC Shrimp Fishery Observer Program.

The proposed methodology combines refined effort estimates with catch-per-unit-effort (CPUE) estimates derived from observer data within spatial and temporal post-strata. Effort scaling is now conducted using a survey-design framework that aggregates time (3 quadrimesters), area (5 zones), and three depth categories. A model-based approach was employed to explore potential stratification variables for gear characteristics and diurnal periods in the CPUE estimation. The primary data sources for catch and bycatch estimation include data on effort using positional tracking from cellular electronic logbook units (cELB) for a sample of vessels, the shrimp observer program (CPUE), and the ALG survey (gear configuration). These sources are exclusive to the federal offshore otter-trawl fleet, where most vessels using otter trawl gear possess federal fishing permits.

I commend the comprehensive analysis conducted by the project team to mitigate bias and assess potential sources of bias. The project team has used sound statistical methods given necessary assumptions driven by the available input data. In particular, the design-based and model-based analysis are conducted under the assumption of trawl hauls being primary sampling units in the effort and CPUE surveys, and that trawl hauls within post-strata are representative for the fleetwide hauls. However, I express reservations regarding the extensive post-stratification of hauls, given that the input data on effort and CPUE are derived from multi-stage sampling surveys with vessels and trips as primary sampling units. The challenge lies in accurately evaluating the annual catch, by-catch estimates, and associated variances due to low sample sizes and the necessity for imputations in post-strata with missing data. While I concur with the broader spatial post-strata groupings (regions), I propose the elimination of depth-stratification and truncation of temporal post-strata to 6-month periods, aiming to streamline the number of strata.

The systematic annual bias in brown shrimp estimates, as indicated by cross-checks, underscores issues with the fundamental representativeness assumptions in equations (1.1) and (1.3) of the bycatch estimation report. Addressing these concerns could involve adopting more robust data collection and sampling methods in the future. Additionally, exploring methods outlined in the study by Clegg et al. (2022) for assessing bias in bycatch estimates from the Norwegian reference Fleet could provide valuable insights.

Electronic Logbook (ELB) devices that recorded vessel location at 10-minute intervals using GPS and stored on memory chips were originally developed and implemented by the consulting company LGL in 2004. From 2004-2013 positional data from manual retrieval of memory chips from these ELB devices were analyzed and used to identify trawl paths based on speed calculations and to quantify distances trawled for all hauls for a subset of shrimp trawlers. In mid-2013, these responsibilities were transferred to NMFS. In 2014, tamper-proof cellular Electronic Logbook (cELB) devices were implemented, in which positional data are automatically transmitted back to NMFS servers through the cellular network. In early 2014, NMFS selected 500 Gulf of Mexico Shrimp Permit (SPGM) owners to participate in the cELB program using a spatially stratified random sampling method, based on a list-frame of vessels. Vessels with a federal shrimp permit were assigned to nine geographic areas based on the location of the port where the vessel most often landed during its most recent year of fishing activity. Vessels were selected with probabilities that were proportional to 2013 landings within 14 strata defined by geographic area crossed with vessel type (Ice or Freezer vessels).

The potential bias linked to fleetwide effort estimation, relying on a sample of vessels equipped with cELB, could be substantially mitigated by universally equipping the entire federal shrimp fleet with standard Vessel Monitoring System (VMS) units. Fixing the sample of vessels from 2014 poses a risk of decreasing representativeness over time, emphasizing the need for mandated VMS for all vessels. Mandating electronic logbook reporting of gear type and effort per haul (start and end positions) would further enhance data accuracy.

In the 2021 analysis, effort scaling from cELB vessels to the entire fleet within post-strata assumes perfect correlation between landings and effort. Given that primary sampling units

(PSUs) in the cELB program are vessels, an annual check of their representativeness is suggested by scaling up mean annual catch per vessel (from trip tickets) to all vessels (or all trip tickets). This simple design-based estimate could offer a robust comparison with reported landings from trip ticket data. If installation of cELB units for all vessels cannot be achieved, it is strongly recommended to reshuffle vessels with cELB at regular intervals to counter potential bias introduced by changes in fleet composition over time. One method could be to use sampling of vessels with partial replacement annually, for example by replacing 20% of the vessels randomly. Continuing positional data collection with cELB for only a subset of vessels necessitates reshuffling to avoid increasing bias over time, with a recommended sampling frame of vessels stratified by port, refrigeration method, and gear configuration. Simple random selection within strata, proportional to the number of vessels, ensures equal selection probabilities of vessels and is advantageous when poststratification of vessels and hauls is employed in effort data analysis.

A potential bias source in catch and bycatch estimates is linked to observer CPUE data. I believe that the main concern with respect to bias is whether the CPUE data from the observer program from a sample of trips (and a sample of hauls within trips) are representative for the fleetwide trips and hauls. In the material provided for this CIE review, the survey design for the observer program is not well documented. I could not find any description of how the annual observer trips are selected. A review of the SEFSC Shrimp Observer Program is recommended. The cross-check analyses indicate minimal bias caused by the various levels of subsampling within selected trips, but bias in vessel and trips selections cannot reliably be checked based on the available observer data. It is particularly important that the vessel and trip selection for observer sampling be unbiased. Utilizing a sampling frame based on the latest vessel list stratified by gear configuration using data from the Annual Landings and Gear survey could help enhance the program's representativeness. Addressing diurnal differences in CPUE highlights the importance of including both day and night trawling in observer programs, with randomization strategies to select day and night shifts across trips to ensure representativeness without compromising observer well-being.

Description of the Individual Reviewer's Role in the Review Activities

This CIE desk review was conducted independently by Dr. Geoff Tingley and me. Dr. Tingley is a fisheries management scientist with global fisheries experience in stock assessments, environmental impacts of fishing, and sustainability. I bring international research and management experience in quantitative fisheries biology and ecological statistics, specializing in statistical survey sampling methods. I have broad hands-on experience in the design, optimization, and analysis of fisheries-dependent and fisheries-independent surveys and monitoring programs to support stock assessments and ecosystem-based fisheries management. From 2008-2013, I chaired or co-chaired expert groups annually for the International Council for the Exploration of the Seas (ICES) to develop practical and statistically sound catch sampling surveys for commercial fisheries. I am currently a member of the ICES WGRFS and WGCATCH. My broad experience in design-based and model-based quantification of catch and bycatch from fisheries-dependent surveys of commercial and recreational fisheries, and experience with abundance index development from fisheries-independent surveys is especially relevant for this review.

The CIE reviewers were requested to provide input on the following items in the Terms of Reference.

1. Describe the data sources and proposed methodology.
2. Are the data used appropriately to estimate bycatch from the federally permitted offshore shrimp fleet?
 - a. Is there any evidence of bias? If so, provide recommendations for improvements.
 - b. Is the uncertainty adequately described? If not, make recommendations to better characterize the uncertainty.
3. Are the statistics used in the estimation appropriate, and are they used correctly
4. Describe the appropriateness of the estimation model approach (Bayesian, GLM, etc.) for the estimation of common vs. less common species assessed in the Southeast U.S. region.
 - a. Evaluate the uncertainty estimated from each approach.
 - b. Provide recommendations about a sample size or encounter rate below which a method is not recommended for use to describe the bycatch of a species?

The review started with a webinar on Google Meet on September 21, 2023. Participants included Alan Llowther (NOAA, organizer), Steven Smith (University of Miami and NOAA Affiliate), Kate Siegfried (NOAA Federal), Geoff Tingley (CIE), Jon Helge Vølstad (CIE), and Manoj Shivilani (NTVI). The Terms of Reference for this review was provided by email on September 25. The reports and background information for the review (Appendix 1) was uploaded to a Google Drive Folder and made available for this Desktop review on September 26.

Following reading of the two main reports, I sent some questions to Steven Smith by email to better understand the ELB vessel selection methodology employed in 2014. I contacted Steven Smith on October 16 requesting some more details on the methods for selecting vessels for the Electronic Logbook Program (ELB). Steven forwarded the request to Kyle Dettloff who promptly (the same day) provided a description of the statistical methodology for stratification and sample selection employed in 2004, and attached a pdf of the chapter from the ELB manual that describes the process in detail and provides the R code. It is not clear to me if NMFS employed the same methods for selecting vessels in 2014. Dettloff informed that they don't have the input data files or the specific R code used to conduct the 2014 selection exercise.

I also had two follow up questions that promptly were answered by Steven Smith:

Q1: Did you not have catches from both Ice of Freezer vessels in some of the 9 spatial strata?

A. Yes, this was the case. The ice vessels are mainly the inshore fleet, which is a small component of the federally-permitted vessels. The inshore fleet mainly uses trawl gears

other than otter trawls (butterfly nets, skimmer trawls, etc.), and fish in coastal bays/estuaries. In retrospect, probably using gear type would have been a better way to separate the inshore and offshore fleets.

Q2. So, to make sure I understand correctly: the PSUs (vessels) within each of the 14 geographical and refrigeration method strata were selected with unequal probability, proportional to size (fraction of total landings).

A. That is our understanding as well, presuming that the analyst followed the stated methodology.

I have conducted numerous CIE reviews over the years, and in recent years desk-top reviews such as this one have been the norm. I have found it difficult to conduct this review as a desktop study due to the complexity of the methods, and the relatively limited documentation on the complex input data. It would have been helpful to have had a physical meeting with the project team and leaders of the shrimp observer program.

[Summary of Findings for each ToR in which the weaknesses and strengths are described.](#)

1. Describe the data sources and proposed methodology (ToR 1).

Table 1: Data sources for effort

Data sources	Type of data	Usage
Cellular ELB (cELB)	Positional data with timestamp, latitude, and longitude; vessel location from GPS at 10-minute intervals	Estimate trawl effort (tow paths and distances per haul)
Vessel assignment table	Link vessels ID with cELB	Link vessels ID with raw cELB data
Vessel permit table	Identifies which federal permits are associated with a particular vessel through time.	Separate landings data by vessel into state and federal categories
Trip ticket landings	Date, species catch weight and price by gear, and market category, as well as the principal fishing area for the trip	Quantify annual landings by vessel cELB status and time/area blocks (post-strata).
Gulf of Mexico shapefile	1-21 trip ticket area grid and 10 and 30 fathom delineations	Area and depth post-stratification of effort and landings
NOAA Gulf of Mexico bathymetric grid	1-minute resolution spatial grid of bottom surface	Filter out data at depths too deep for shrimping activity.

Table 2: Data sources for CPUE

Data sources	Type of data	Usage
SEAMAP Groundfish survey data	Weights of brown, pink, and white shrimp from 30-minute tows by season (Summer/Fall), stat zone (1-21), depth, and time of day (D/N)	Used to obtain species-specific effort estimates based on modeling.
GOM Onboard Observer Program	Catch, by-catch, and effort data, provide general fishery bycatch characterization and catch rates for finfish species by trip, and by haul for area and target species	CPUE estimates that are independent of cELB based effort estimates
Annual Landings and gears (ALG) mail survey	Vessel-specific information on net configuration (i.e., number of trawl nets).	Standardize CPUE estimates to adjust for bias in the selection of vessel trips in the observer program. The program oversampled vessels with 4 nets, and observers sampled highest fraction of hauls during nighttime fishing

In response to concerns about bias in shrimp bycatch estimates for red snapper and other commercially important species, a 2021 project team developed a cross-check procedure. This involved applying the bycatch estimation method to assess annual brown shrimp landings, a primary focus of the shrimp fleet, and comparing these estimates with officially reported fleet landings.

The proposed new methodology for estimating bycatch is based on combining refined effort estimates with CPUE estimates of catch and bycatch from observer data within spatial and temporal post-strata with trawl hauls assumed to be primary sampling units. Scaling of effort is now done using landings at aggregate combinations of time (3 quadrimesters), area (5 zones), and three depth categories in a survey-design framework rather than attempting to match individual trips. A model-based approach was employed to investigate potential additional stratification variables for gear characteristics and diurnal period in the CPUE data from the observer program.

The main data sources used by the 2021 project team for estimating catch and bycatch were obtained from cELB units (effort) installed on approximately 500 vessels in 2014, the shrimp observer program (CPUE), and the mandated annual gear configuration (ALG) questionnaire census. Data from cellular Electronic Logbook vessel tracking devices (cELB units) were obtained from between 495 and 276 vessels annually between 2014 and 2021. These data sources were only available from the federal offshore otter-trawl fleet. Most vessels using otter trawl gear possess federal fishing permits. The sampling frame was a list of vessels (primary sampling units) that were stratified by assigning each permitted vessel to one of 9 geographic regions based on the location of the port where the vessel most often landed their catches during its most recent year of fishing activity. Vessels within each region were further stratified by the refrigeration method (Freezer or Ice) as a proxy for separating offshore and inshore vessels. A full re-randomization of the ELB vessels selection has not been conducted since 2014.

The sampling frame for selecting vessels in 2014 were established using the following steps (Kyle Dettloff, personal communication by email):

- vessel must be associated with an active and valid SPGM permit.
- vessel must have landings (catch) in previous 2 years.
- those landings (catch) must come from one of 9 geographic zones of the Gulf of Mexico
- within the geographic zone a split of vessel refrigeration method (Freezer or Ice) when applicable.

These rules above lead to the original grouping of vessels into 14 strata. Let $C_{h,i}$ be total catch by vessel i in stratum h , and let C be total catch of all vessels in the sampling frame. For each stratum, an unequal probability sample of vessels were selected without replacement, with probability of selection vessel i in stratum h :

$$p_{h,i} \propto C_{h,i}/C,$$

using landings for 2013 as proxy for strata sizes. In the documentation for the 2004 vessels selection, it was stated that ELB only were installed for 50% of the randomly selected vessels, with the remaining ELB boxes installed on vessels that volunteered. I could not find any documentation of refusal rates for vessels selected for installing cELB devices in 2014.

In the new analysis conducted by the 2021 project team, the scaling of effort from the cELB vessels to the entire fleet within post-strata is based on the assumptions of a perfect correlation between landings and effort. The 9 geographical strata described in the ELB manual are at a finer spatial scale compared to the five geographical strata used in the new effort estimation approach, but generally correspond (the 9 smaller strata can be grouped into the 5 larger strata). The 'refrigeration' designation in the vessel selection was mainly used to distinguish offshore from inshore vessels. Since the total effort (number of hauls by the general fleet) within post-strata are unknown, the project team had no alternative to using reported catches in the effort estimation. Clearly, the accuracy of the reported landings within the post-strata is difficult to assess. If VMS were installed on all vessels, then the total number of hauls within each post-stratum could be quantified, thus eliminating the potential bias from using inaccurate spatial data on catches to scale up effort from cELB vessels to the general fleet.

The National Marine Fisheries Service (NMFS) Shrimp Observer Program in the Gulf of Mexico is designed to collect important data on shrimp fisheries to support sustainable management and conservation efforts. Shrimp observers are trained individuals who work on commercial shrimp vessels to monitor fishing activities and collect data on catch composition, bycatch, and fishing practices. According to the NMFS website the observer coverage of the entire shrimp fishery in the South Atlantic is around 2% of trips.

Estimation of penaeid catch for cross-checking with trip-ticket landings conducted by the 2021 project team was restricted to the federal offshore otter trawl fleet. According to documents provided, observer sampling effort was concentrated for nighttime 4-net trawls, with sparser sampling during daytime, and sparse sampling for 2-net trawl configurations.

The project team conducted extensive analysis to standardize the CPUE data from the observer program. Fishing power factors were determined to standardize CPUE data for sparsely sampled nighttime hauls and hauls with 2-net trawl configurations relative to the well-sampled reference gear (4 nets during nighttime), and to impute observations in sparsely sampled strata. Species- and gear-specific mean CPUEs and variances were estimated using the full dataset of actual and imputed catch-effort observations.

Estimations of bycatch of red snapper were based on expanding sample-based estimates of CPUE to offshore otter trawl effort for both federal- and state-permitted vessels (Dettloff 2023). The bycatch estimation was conducted for 2014-2020 where trip-ticket databases were fully in place and reliable for all the GOM states and where cELB effort data were complete. Three of the four principal data sources for estimating catch and bycatch are exclusive to the federal fleet: cELB units, shrimp observer program, and the ALG survey. Thus, estimation of penaeid catch (Eq. 2.4) and cross-checking with trip-ticket landings was restricted to the federal offshore otter trawl fleet. The cross-check to assess representativeness of the bycatch estimation method entailed applying the bycatch estimation methodology to estimate the annual landings of brown shrimp, a primary target of the shrimp fleet, and then comparing the estimated landings with the reported fleet landings.

2. Are the data used appropriately to estimate bycatch from the federally permitted offshore shrimp fleet? (ToR 2)
 - a. Is there any evidence of bias? If so, provide recommendations for improvements (ToR 2(a)).

Effort data

I am impressed by the thorough approach to identify and estimate trawl effort by mapping the speed and paths of trawl hauls based on data from cELB units.

Comments to the stated assumptions required to obtain accurate estimates of total effort, give non-universal cELB coverage of the fleet:

- 1) cELB devices are capturing all fishing activity, and are powered on for the full extent of vessel activity per federal regulations. This assumption is generally supported, but a comparison against observer reported effort suggests cases exist when cELB devices may not be recording, which would lead to an underestimation of effort.

Comment: I agree

- 2) There is no systematic bias in effort classification. That is, there is an equal chance of false-positives and false-negatives. A comparison of cELB classified effort with observer recorded effort generally supports this assumption.

Comment: I agree

- 3) The spatial distribution of cELB vessels is representative of the total fleet within strata. There is support for this assumption given the randomized nature of the original cELB selection, however, due to changes to the fleet and vessels dropping from the sample the original selection of vessels may not remain representative.

Comment:

This is a very strong assumption that is unlikely to be fulfilled.

The sampling frame for selecting cELB vessels was a list of vessels (Primary Sampling Units - PSUs), and unequal probability sampling was employed to select a stratified sample of these vessels. The stratified sampling method ensured that cELB units were installed on a sample of vessels (PSUs) from home ports that were spread out geographically. However, the selection of vessels was based on fishing patterns in 2013, and there was no way to ensure that the vessels fished in the spatio-temporal strata used in the analysis in the following years.

In both the design-based and model-based estimation of catch and bycatch for 2014-2020, it is implicitly assumed that primary sampling units are hauls, and that hauls taken by cELB vessels form a simple random sample of fleetwide hauls within the spatio-temporal post strata defined by area, depth, and trimester. The effective sample size of cELB hauls is likely to be substantially reduced because the annual hauls were clustered within a subsample of vessels and trips. This would especially cause bias in the variance estimation.

One simple exercise to assess potential vessel selection bias would be to estimate annual mean catch of shrimp per vessel (based on all trip tickets), and then scale this mean up to all active vessels (i.e., vessels that could be connected to trip tickets) or to all trip tickets to estimate total annual catch. If the near 500 vessels (PSUs) were selected with approximately equal probability, then estimated total catch (at least for 2014) should be close to the total catch for the entire fleet in the sampling frame. If the project team has access to the selection or inclusion probability of the original sample of vessels in 2014, then a Horvitz–Thompson estimator should be used to estimate annual mean catch per vessel, accounting for the unequal selection probability for vessels. It could also be useful to use data from the annual AGS in a similar approach to assess if the sample vessels with cELB units have similar gear characteristics as the general fleet.

- 4) CPUE of vessels with cELBs on board is representative of the total fleet. This is a necessary assumption for using landings to scale up cELB effort. If CPUE among cELB vessels are higher than non-cELB vessels, this would lead to an underestimation of effort, and vice versa.

Comment:

In the observer program, multi-stage sampling is employed with trips as primary sampling units. The observers subsample hauls (secondary sampling units) within each of an annual sample of trips, and within hauls they subsample nets (tertiary sampling units), and for large catches they even subsample catches within nets. In the analysis conducted by the project

team, hauls are considered primary sampling units, and again the key assumption is that hauls sampled by observers form a simple random sample of fleetwide hauls within the spatio-temporal post strata defined by area, depth, and trimester. Fishery independent SEAMAP data were used to obtain species-specific effort estimates. Clearly, it is unlikely that the fisheries-dependent hauls have a similar distribution of targeted effort by species as the fisheries-independent hauls. The latter are presumably conducted with fixed trawl distance in randomly selected locations, while fishing vessels will likely take repeated hauls in areas with high catch rates, avoid areas with low catch rates, and are also likely to target species that yields highest income. Nevertheless, the estimates of brown shrimp catches were fairly accurate, with an estimated 20% bias as compared to reported landings.

The project team has convincingly demonstrated that the sub-sampling conducted by observers within trips yields representative data on catch by haul and catch by trip through a comparison of estimated catches with official landings for observer sampled trips that could be linked with trips tickets. Hence, the most likely source of bias in the observed data are due to the vessel and trip-selection. I could not find any documentation of the survey design for the annual sampling of trips in the shrimp fishery observer program. Ideally, annual estimates of mean catches per trip from the observer program could be scaled up to all trips, and then compared to total reported catches from trip tickets. However, since there is, to my knowledge, no census data on the total annual number of trips, it is difficult to assess potential bias in the trip selection.

I fully understand that the project team needed to conduct the analysis under the assumption that hauls were primary sampling units, since haul-based observer data are linked with independent effort data to estimate catch and bycatch. However, it should be noted that the variance estimate for CPUE is especially likely to be underestimated by ignoring that trips are PSUs. Since only a small fraction of fleetwide trips are sampled by observers, sampling with replacement can be assumed, and the variance estimation can be based on primary sampling units, ignoring the subsampling of hauls (Williams 2000).

The effective sample size of observer hauls is likely to be substantially reduced because the annual hauls were clustered within a subsample of trips. This would cause bias in the variance estimation (Nelson 2014).

- 5) Reporting of landings is similar between vessels with and without cELBs. That is, one group is no more or less likely than the other to completely and accurately report landings. This is a necessary assumption for using landings at an aggregate level to scale up cELB effort. If reporting is better among cELB vessels, this would lead to an underestimation of effort, and vice versa.

Comment:

This assumption could also be assessed by comparing total annual catch estimates based on the cELB vessels (see comment to assumption 3).

Overall comment:

The project team conducted a thorough analysis using robust statistical methods to minimize bias in catch and effort estimates. However, concerns arise due to low sample sizes and zero hauls in certain spatio-temporal strata, necessitating complex imputations.

Underestimation issues stem from representativeness assumptions for estimates of effort and CPUE in post-strata and can be mitigated with improved data collection and adjusted post-stratification based on available sample sizes. The underestimation in fleet effort is also a result of some cELB units being switched off and potential unreported shrimp catch. For fleet effort, advocating the adoption of standard VMS units over ELB units is strongly recommended for better reliability and coverage. This would yield census data on number of hauls and eliminate the need for using landings data as a proxy for effort.

Issues with observer sampling can be addressed by refining the allocation strategy considering diurnal period, gear configuration, and area-depth-season strata.

Pooling years within strata and imputing catch rates for sparsely sampled gears may contribute to an overall underestimation of catch rates. The divergence from the original sampling design is attributed to factors rendering it unsuitable for scaling to current fleet effort.

To address these issues, 9 geographic strata are aggregated into five broader areas, incorporating spatially explicit effort and landings. This method relies on equal efficiency assumptions between ELB and non-ELB vessels, using annual landings as a more reliable scaling variable than a snapshot from 2013.

Concerns about bias in CPUE focus on the representativeness of the observer program data. The survey design lacks documentation and randomized selection of vessels/trips with documented refusal rates is crucial for assessing representativeness. Design-based estimation assumes hauls as primary sampling units, but two-stage cluster sampling estimators would be more appropriate, considering observers are placed on a sample of trips throughout the year.

The project team has conducted a very comprehensive analysis using sound statistical methods to minimize bias in estimates of catch and effort. My main concern with the analysis is that the large number of post-strata results in low sample sizes of hauls in many strata, and even zero hauls in some strata and, hence require complex imputations.

Information provided by Kyle Dettloff, SEFSC, identified two aspects of fleet effort that may have resulted in underestimation:

- First, the cross-check evaluation between trip-level observer and ELB effort (Fig. 3.1) found that the ELB units were switched off or not functioning for a portion of the matched trips, and thus were not capturing the complete trip effort. This would have led to a lower value of ELB effort relative to ELB catch in Eq. (1.2), resulting in an underestimate of fleet effort.

- The second aspect concerns the level of unreported shrimp catch. Anecdotal evidence indicates that not all the shrimp catch is sold to licensed dealers for a variety of reasons (personal consumption by the crew, given to relatives and friends, sold directly to consumers, etc.). It is plausible that the vessels equipped with ELB units may have a higher reporting rate in the trip-ticket system compared to non-ELB vessels. This would have led to a lower value of fleet catch relative to ELB catch in Eq. (1.2), also resulting in an underestimate of fleet effort.

The net overall effect of pooling years within strata and of imputing catch rates for sparsely sampled gears, however, may have been an underestimation of catch rates.

The strata used in the original selection were based on lists of vessels, and the spatial and temporal distribution of landings and effort by vessel were only known at the end of the year. The probabilities of selecting vessels by strata were proportional to the fraction of annual landings by strata, based on a snapshot in time of the landings that occurred in 2013.

To address these issues, the current post "strata" used by the project team for scaling are aggregated into the 5 broader areas that better satisfy the requirement of spatially explicit effort and landings for trips that typically occur over very large areas, a seasonal component is included to better address the fact that different vessels behave differently throughout the year, and the aggregated landings within each year (theoretically a census) are used as the ancillary variable to scale ELB effort to the total fleet. This still relies on the assumption of equal efficiency (CPUE) between ELB and non-ELB vessels, but landings within each year are likely a much more reliable way to directly scale the corresponding annual ELB effort rather than using a snapshot of landings from 2013.

CPUE data

One of the main concerns with respect to bias in catch and bycatch estimates is whether the CPUE from the observer program from sampled trips (and sampled hauls within trips) are representative for the fleetwide trips and hauls. In the material provided for this CIE review, the survey design for the observer program is not well documented. I could not find any description of how the annual observer trips are selected. Randomized selection of vessels/trips with documentation of realized trips and refusal rates is important for judging the representativeness of the observe data.

The design-based estimation of mean CPUE from observer data assumes that hauls are primary sampling units. It would be more appropriate to use two-stage cluster sampling estimators of mean CPUE and associated variance since observers are placed on a sample of trips throughout the year.

- b. Is the uncertainty adequately described? If not, make recommendations to better characterize the uncertainty (ToR 2b).

Total catch or bycatch \hat{C}_h of a species within a stratum h is estimated by (equation 2.4) as the products of the respective estimates of total effort, \hat{E}_h , and catch rates, \hat{R}_h

$$\hat{C}_h = \hat{E}_h \times \hat{R}_h$$

with a simplified variance estimator (equation 2.5) under the assumption of zero variance in effort,

$$var(\hat{C}_h) = \hat{E}_h^2 \times var(\hat{R}_h).$$

A more appropriate method is to employ a variance estimator of the product (Goodman 1960; Anderson 1988, p. 343) where also the variance in effort estimates is accounted for:

$$var(\hat{C}_h) = \hat{E}_h^2 \times var(\hat{R}_h) + \hat{R}_h^2 \times var(\hat{E}_h) - var(\hat{E}_h) \times var(\hat{R}_h)$$

3. Are the statistics used in the estimation appropriate, and are they used correctly (ToR 3)?

I am not sure if I understand this question. I believe that the project team has used available input data appropriately given their assumptions, and I agree with the choice of design-based estimators of catch and bycatch that combines independent effort and CPUE estimates. Especially, I think the use of the annual ALG data on fleetwide gear configurations for bias correction is very important. As already pointed out, I do have reservations about the extensive fine-scale spatio-temporal post-stratification of hauls given the clustering caused by vessels being PSUs in the effort surveys, and trips being PSUs in the observer program.

4. Describe the appropriateness of the estimation model approach (Bayesian, GLM, etc.) for the estimation of common vs. less common species assessed in the Southeast U.S. region (ToR 4).

Trawl gear characteristics and diurnal period (daytime, nighttime trawling) were evaluated as potential additional stratification factors to control variation in catch rates (CPUE from the observer program) of penaeid shrimp species using GLM. Predicted catches were generally higher for 2-net vs. 4-net trawls within a diurnal period.

The compound probability function regression approach used to predict catch for species, where the CPUE data were zero-inflated, seems reasonable. This method is designed to handle a scenario where you have presence-absence data (modeled using logistic regression) and, when the condition is present, you are interested in modeling the quantity or catch (modeled using a standard GLM regression). The combination of these two regression models forms a compound PDF regression approach for predicting catch. The categorical space-time blocking covariate (area-depth-season) helps account for spatial and temporal heterogeneity in the data, allowing for a more nuanced analysis.

- a. Evaluate the uncertainty estimated from each approach (ToR 4a).

The standard errors (measures of precision) of the predicted catch based on the compound probability function regressions are not provided. The variance estimator of the catch

based on a product (Goodman 1960; Anderson 1988, p. 343) may not be appropriate here since the two data sources on occurrence (p) or catch when present (u) are not independent. The standard errors for occurrence and catch when present are also likely to be biased downwards, since the effective sample sizes of hauls likely are reduced because of the clustering by vessel and trips.

- b. Provide recommendations about a sample size or encounter rate below which a method is not recommended for use to describe the bycatch of a species (ToR 4b).

It is difficult to provide a recommendation on sample sizes due to the extensive post-stratification and since the precision values of the predicted catch are not provided. I would recommend a pragmatic approach here. One rule of thumb could be to ensure that sample sizes are sufficient to secure annual data from at least two hauls with effort and CPUE data for all post-strata.

Conclusions and Recommendations in accordance with the ToRs

I commend the extensive analysis conducted by the project team to minimize bias and to assess potential sources of bias. The design-based and model-based methods for estimating bycatch are reasonable given the assumptions, but the estimates of precision are biased or lacking. I especially have concerns about the extensive post-stratification of hauls given that the input data on effort and CPUE are based on multi-stage sampling with vessels and trips as primary sampling units, with limited sample sizes annually. It is very difficult to evaluate the accuracy of the annual catch and bycatch estimates and associated variances due to low sample sizes and the need for imputations for post-strata with missing data. I agree with the larger groupings of spatial post-strata (regions), but perhaps the depth-stratification could be eliminated, and the temporal post-strata cut to 6-months periods to reduce the number of strata.

The current estimates of precision for the design-based method are generally biased downwards since they do not account for uncertainty in the effort estimates, and do not account for the actual clustering of hauls within vessels and trips. I recommend that the variance estimator for a product be employed for catch and bycatch to account for the variance in the effort estimates. For the model-based method, no estimates of precision are provided for the predicted catch or bycatch. In theory, estimates could be based on bootstrapping, but this will be problematic given the complex analysis with the large number of post-strata and multiple computations.

The systematic annual bias in estimates of brown shrimp revealed by the cross-check suggests that the underestimation issues relate to the fundamental assumptions of representativeness in Eqs. (1.1) and (1.3) in the bycatch estimation report, which can be addressed with more robust data collection/sampling methods in the future. Methods for assessing bias in bycatch estimates from the Norwegian reference Fleet (Clegg et al. 2022) might also be useful for assessing bias.

The potential bias related to fleetwide effort estimation based on a sample of vessels equipped with cELB could largely be eliminated by equipping the entire federal shrimp fleet with more standard vessel monitoring system (VMS) units, as required for all other federally permitted fishing vessels in the Southeastern U.S. region. By fixing the sample of vessels from 2014, there is clearly a risk that the group of vessels with cELB units becomes less representative over time. For example, this group of vessels may have a different technological creep than the general fleet. Also, the cross-check evaluation between trip-level observer and cELB effort found that the cELB units were switched off or not functioning for a portion of the matched trips, and thus were not capturing the complete trip effort. Mandated VMS for all vessels would also eliminate these sources of bias. Also, it would be ideal to mandate electronic logbook reporting of gear type (2-nets or 4-nets) and effort (start and end position) per haul. As evident in the reports provided for this review, the estimation of effort based on cELB data from only a subset of vessels, expanded to the fleet using landings data from trip tickets, is requiring extensive statistical analysis and the estimates are subject to potential biases that are difficult to quantify.

Since the primary sampling units (PSUs) in the cELB program are vessels, it could be beneficial to check their representativeness annually by scaling up mean annual catch per vessel from trip tickets to all vessels (or trip tickets) and compare these of total landings to the respective annual reported landings from trip ticket data. This simple design-based estimate relies only on the selection or inclusion probabilities for vessels and does not require imputations due to post-strata with missing data. This simple comparison can potentially reveal issues with the post-stratification, and should be particularly robust for 2014 since the sampling frame of vessels were based on 2013 data on vessels and landings. Clearly, even if the vessels were representative of the general fleet in the year they were selected, changes in fleet composition by strata over time would likely introduce bias in estimates based on a fixed group of cELB vessels.

If the collection of positional data continues with cELB installed for only a subset of vessels, I recommend that the vessels equipped by such units be reshuffled at regular intervals (ideally annually). It would be recommended to continue to use a sampling frame where vessels are stratified by port (area), and refrigeration method, but it could also be efficient to classify vessels by gear configuration (e.g., 2-nets, and 4-nets) using data from the most recent AGL survey. Such a sampling frame based on a list of vessels is well defined, and likely to cover the target population of active vessels well, assuming stability in the gear configuration and a recent vessel list from one year to the next. The coverage of the frame can easily be assessed annually. A simple random selection of vessels, without replacement, within strata, with sample sizes allocated proportional to the number of vessels in each stratum, is recommended. This simple approach would result in equal selection probability of vessels across strata. This is advantageous if poststratification is employed in the analysis of effort data. Also, even if vessels were selected with known inclusion probabilities in 2014, the fleet composition will likely change over time such that the fixed set of cELB vessels will be less representative for the fleet over time. The use of landings by spatial area to scale up effort from the cELB vessels is also problematic since this assumes perfect correlation between landings and effort.

I suspect a main source of bias in estimates of catch and bycatch is related to the observer CPUE data. For observer data, the cross-check analyses indicated that effort and catch observations were unbiased at the net-, tow-, and trip-levels. Cross-checking to detect vessel and trip selection bias could not be achieved based on the available data since there are not unique trip ids for the general fleet annually. I recommend that a review of the NEFSC Shrimp Observer Program be conducted, with particular focus on the survey design for selections vessels and trips. The selection of vessels and trips in the observer program could benefit from using a sampling frame based on the most recent vessel list and coupled with gear configuration data from the ALG survey. This would help ensure representative data from vessels trawling with 2-nets and 4-nets. Diurnal differences in CPUE also suggests that it is important to cover day and night trawling, requiring two observers, or alternatively (and more cost effective) that night or day shifts be alternated from trip to trip by randomization. Randomization of day or night shifts within a trip will not be possible in practice since this would be detrimental to sleep patterns for observers.

References

- Anderson, O.D., 1988. Kendall's Advanced Theory of Statistics, Volume 1: Distribution Theory.
- Clegg, T. L., Fuglebakk, E., Ono, K., Vølstad, J. H., and Nedreaas, K. 2022. A simulation approach to assessing bias in a fisheries self-sampling programme. – ICES Journal of Marine Science, 79:76-87
- Nelson, G. A. 2014. Cluster sampling: a pervasive, yet little recognized survey design in fisheries research. Transactions of the American Fisheries Society, 143(4), 926-938.
- Goodman, L.A. 1960. "On the exact variance of products." *Journal of the American statistical association* 55, no. 292: 708-713.
- Williams, R. L. 2000. A note on robust variance estimation for cluster-correlated data. *Biometrics*, 56(2), 645-646

Appendix 1: Bibliography of materials provided for review

The reports and supplementary background information made available as pdf files through a Google Drive folder included the

- Report “Estimation of Commercial Shrimp Effort in the Gulf of Mexico” by Kyle Dettloff, SEFSC, August 2023
- Report “Improving Estimation of Bycatch from Shrimp Trawls in the Gulf of Mexico” by Steven G. Smith, Sarina Atkinson, Cheston Peterson, Jo Anne Williams, Kyle Dettloff, and Alan Lowther, September 2023
- Supplementary folder with SAS Legacy Code, SAS Refined Code, and tables of analysis results.

Appendix 2: A copy of the CIE Performance Work Statement (PWS)

Performance Work Statement (PWS)
National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

Desk Review of the Shrimp Bycatch Estimation Methodology

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards².

Scope

Shrimp trawl bycatch is an estimated data set that is used in the assessments of multiple fish species in the southeastern U.S. The methodology was developed specifically for red snapper but has been applied to other species such as king mackerel in the Gulf of Mexico. A recent preliminary examination uncovered systematic bias in the previous method. Therefore, a complete overhaul of the methodology was conducted at the Southeast Fisheries Science Center, and this process is meant to review those results. The new method utilizes four main data sources: effort data, observer data, a gear and landings survey, and shrimp landings data.

Requirements

NMFS requests that two CIE reviewers conduct a peer review of the scientific information and framework of this new methodology based on the Terms of Reference (TORs) referenced

² https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/memoranda/2005/m05-03.pdf

listed Annex 2. The specified format and contents of the individual peer review reports are found in Annex 1. Each reviewer should have working knowledge and recent experience in a fishery survey sampling (independent or dependent) as well as abundance index development (e.g. Generalized Linear Models (GLMs), Zero-Inflated Negative Binomial (ZINB) Models, etc.).

Tasks Each CIE reviewer shall complete the following tasks in accordance with the PWS and Schedule of Milestones and Deliverables herein.

Task 1. Pre-review Background Documents: At least two weeks before the peer review, the NMFS Project Contact will make all the necessary information and reports available electronically for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE on where to send documents. The CIE reviewer shall read all documents in preparation for the peer review.

Task 2. Webinar: Approximately two weeks after the CIE reviewers receive the pre-review documents, they will participate in a webinar with the NMFS Project Contact and appropriate staff to address any clarifications that the reviewers may need regarding the TORs or the review process. The NMFS Project Contact will provide the information for the arrangements for this webinar.

Task 3. Desk Review: Each CIE reviewer shall conduct the independent peer review in accordance with this PWS and TORs and shall not serve in any other role unless specified herein. Modifications to the PWS and TORs cannot be made during the peer review, and any PWS or TORs modifications prior to the peer review shall be approved by the Contracting Officer's Representative (COR) and the CIE contractor.

Task 4. Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with this PWS. Each CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. Each CIE reviewer shall complete the independent peer review addressing each TOR as described in **Annex 2**.

Place of Performance

Each CIE reviewer shall conduct an independent peer review as a desk review, therefore no travel is required. Work will be conducted at other than the government facility (i.e. off-site)

Period of Performance

The period of performance shall be from the time of award through November 2023. The CIE reviewers' duties shall not exceed 10 days to complete all required tasks.

Schedule of Milestones and Deliverables

The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Schedule	Milestones and Deliverables
Within two weeks of award	Contractor selects and confirms reviewers
No later than two weeks prior to the review	Contractor provides the pre-review documents to the reviewers
September – October 2023	Each reviewer conducts an independent peer review as a desk review
Within two weeks after review	Reviewers submit draft peer-review reports to the contractor for quality assurance and review
Within three weeks of receiving draft reports	Contractor submits independent Peer-Review reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards: (1) The reports shall be completed in accordance with the required formatting and content; (2) The reports shall address each TOR as specified; and (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Confidentiality and Data Privacy

This contract may require that services contractors have access to Privacy Information. Services contractors are responsible for maintaining the confidentiality of all subjects and materials and may be required to sign and adhere to a Non-disclosure Agreement (NDA).

Government Furnished Resources

The Government will provide all necessary information, data and documents to the Contractor for work required under this contract.

Travel

Travel is not expected or authorized for this task order.

Project Contact:

Kate Siegfried
 Supervisory Research Mathematical Statistician
 NOAA/NMFS/SEFSC
kate.siegfried@noaa.gov

Annex 1: Peer Review Report Requirements

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is adequate.
2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

Appendix 1: Bibliography of materials provided for review

Appendix 2: A copy of the CIE Statement of Work

Annex 2: Terms of Reference (TOR) for the Peer Review

The reviewers will provide input on the following questions.

1. Describe the data sources and proposed methodology.
2. Are the data used appropriately to estimate bycatch from the federally permitted offshore shrimp fleet?
 - a. Is there any evidence of bias? If so, provide recommendations for improvements.
 - b. Is the uncertainty adequately described? If not, make recommendations to better characterize the uncertainty.
3. Are the statistics used in the estimation appropriate, and are they used correctly?
4. Describe the appropriateness of the estimation model approach (Bayesian, GLM, etc.) for the estimation of common vs. less common species assessed in the Southeast U.S. region.
 - a. Evaluate the uncertainty estimated from each approach.
 - b. Provide recommendations about a sample size or encounter rate below which a method is not recommended for use to describe the bycatch of a species?