Gingerfish

Independent Peer Review Report

Gulf of Mexico shrimp bycatch review

Prepared for the Center for Independent Experts

by

Dr. Geoff Tingley

Gingerfish Ltd, Wellington, New Zealand
email: gingerfish.ltd@gmail.com

April 2024

Table of Contents

Executive Summary

- This document is the individual CIE Reviewer Report of the review of the Gulf of Mexico shrimp bycatch estimation methodology. This was a desk review conducted during October and November 2023. This report represents the sole views of the independent CIE reviewer, Dr Geoff Tingley. Production of this report was delayed due to the reviewer catching Covid while traveling for work.
- Documents describing the historic and new methodologies were provided together with other relevant background documents. These materials were provided in a downloadable format posted on a dedicated Google Drive on the day of the prearranged call between the two CIE reviewers, the CIE program managers (NTVI) and relevant NOAA program staff. The timing of the call was determined by a potential closure of U.S. government business. All documents are listed in Appendix 1.
- Documents were provided close to the Webinar call (due to a potential government shutdown) which prevented the timely completion of Task 1 (pre-review of background documents) and limited the usefulness of Task 2 (Webinar) as described in the performance work statement (Appendix 2). These two tasks were appropriately completed during the review period, including the opportunity to seek clarification and additional information via email. Questions via email and answers were copied to all in the original meeting.
- All Terms of Reference (ToR) of the review were addressed and recommendations made where improvements were likely to be possible. The recommendations were developed and are organized by the identified ToR elements.
- The updated methodology was determined to provide improved accuracy and more robust estimates of bycatch for the shrimp fisheries than the previous approach, and is suitable for use in providing inputs to management.
- There does appear to be some additional opportunities to improve the accuracy and robustness of these analyses further.
- There was limited opportunity to explore potential bias in some data components (e.g., the function parameters and default thresholds) in this review as information provided about the data and how they were generated was limited.
- Recommendations have been developed for each ToR where considered appropriate.
- A key recommendation is that it would be appropriate to consider using a spatio-temporal modelling approach, developed over recent years and increasingly applied to support fisheries management worldwide, as a core component in the process of estimating effort and total bycatch (e.g., Integrated Nested Laplace Approximation (INLA) and Vector Autoregressive Spatio-Temporal (VAST) modelling).

Background

Shrimp trawl bycatch in the Gulf of Mexico has historically been estimated and developed as a dataset used in the assessments of multiple species of fish in the southeastern U.S. The methodology was developed specifically for red snapper, but has been applied to other species in the Gulf of Mexico, such as for king mackerel. A recent, preliminary examination of these data uncovered systematic bias in the methodology that had been used, with the estimated bycatch being consistently less than the reported bycatch by an average of about 45%. The Southeast Fisheries Science Center (SEFSC) subsequently undertook a complete overhaul of the methodology. This review is focused on the newly developed methodology. The new methodology utilizes four main data sources: effort data, observer data, a gear and landings survey dataset, and shrimp landings data.

This review of the new shrimp trawl bycatch estimation methodology was conducted as part of an independent, desk-based review for the Center for Independent Experts (CIE).

All views expressed in this report are solely those of the named, independent CIE reviewer.

Description of Review Activities

This review and reporting were scheduled to be undertaken by Dr Geoff Tingley (Gingerfish Ltd) between late September and the end of October or early November 2023. Due to a Covid infection, completion of this report was unfortunately and unavoidably delayed until early December 2023.

The supporting documents for the review were provided to the reviewers in downloadable electronic formats on a password protected Google Drive. These resources were provided a few hours in advance of the review online meeting due to a potential government shutdown, which did not provide time to read and adequately consider the material prior to the call. The provided documents included bycatch program descriptions, observer manuals and forms for different states (Texas, Louisiana and Florida), and the previous and proposed SAS code for estimating effort. All documents (but not code files) provided and used are listed in the Bibliography. Noting that, due to the timing of the provision of documents and the review online meeting, the Term of Reference (ToR) 1 of the Performance Work Statement could not be completed as intended, all documents provided were reviewed following the call at an early stage of the work.

The review (kick-off) call was conducted using the Google Meet software platform, with the CIE reviewers joining remotely from Norway and the UK, and everyone else from the east coast USA. The call was held from 13.00 to 14.00 EST on 26th September (17.00 to 18.00 GMT+1 on 26th September) 2023. The call included the two CIE reviewers, the CIE program managers (NTVI) and relevant NOAA program staff. The reviewers noted having just received the review documents and the difficulties of not being able to pre-read these before the call were discussed. Options for reviewers to correspond with NOAA staff to ask questions or seek clarifications were agreed. All parties engaged positively in the online meeting and both CIE reviewers asked questions of clarification.

The key documents describing the bycatch estimation procedure were clearly written and informative.

Information relevant to this review is presented in two appendices to this review report, as required by the ToR provided by the CIE. These are, Appendix 1 Bibliography of documents, and Appendix 2 the CIE Performance Work Statement (which includes its own annexes describing (1) the Peer review reporting requirements, and (2) the ToR for the peer review.

Some elements normally included in review material were absent from the documentation. Notably, this included sufficiently detailed comparisons and other diagnostics of the methods used. In turn, this meant that some recommendations were made without full knowledge of work or analyses that may have been done but which were unavailable to the reviewers.

The two CIE reviewers worked independently of each other and the NOAA staff, except for the few emails asking questions of clarification to NOAA staff. Reviewer questions via email were quicky and fully responded to by the NOAA team.

Summary of Findings

The review provided an in-depth understanding of the shrimp bycatch estimation processes and procedures, their development over time, and the current methodology. The bycatch team are clearly knowledgeable and competent, enabling them to deliver annual updates of bycatch estimates and an improved methodology.

Given the limited comparative analyses and diagnostics for parts of the process (especially for effort estimation), there may be more recommendations to address this than are strictly necessary based on the actual methodological performance. However, where work has already been done, surplus recommendations can easily be flagged as already addressed.

The review found that, given the quality of the principal data, the approach was an improvement of the previous method used and suitable as a basis for estimating bycatch and providing management advice, including providing improved consistency between years. The current approach does appear to somewhat underplay variation and uncertainty overall, and particularly in some of the datasets where uncertainty is known to, or would be expected to occur, as well as varying between areas and time periods.

While a number of areas for improvement or development have been identified in respect to the current approach, one area stood out as something that warranted relatively high priority. This was the exploration of the use of spatio-temporal modelling to replace some the strata and GAM-based processes currently used. It is believed that spatio-temporal modelling will provide a better understanding of the data and more robust estimates of the components and overall bycatch. These models would also likely provide a much stronger framework for addressing uncertainty, including through use of Bayesian approaches, and can be operated using the same coding language currently used, R.

Recommendations have been developed for all identified areas where improvement could be made.

Addressing the Terms of Reference (ToR) for the Peer Review

Detailed findings and recommendations are presented below, as required by the ToR for the review. Text in italics is taken directly from the requirements of the review defined in the ToR.

Shrimp bycatch estimation (U.S. eastern Gulf of Mexico).

1. ***Describe the data sources and proposed methodology.***

1a. Estimation of Effort

Initial data on effort and position for a number of shrimp fleet vessels started to be collected in 2004 and were captured on electronic logbook (ELB) devices developed by LGL, and digitally stored on memory cards until 013. The data collected for each vessel were positional data recorded about every ten minutes. These data were analysed by LGL to estimate total fishing effort. In mid-2013, the responsibility of running this program, and collecting and analysing the data passed to NMFS, when a modified version of the original LGL analytical code was used.

In 2014, NMFS introduced cellular electronic logbooks (cELBs) which were deployed on 500 shrimp vessels selected using a spatially stratified, random sampling method weighted by landings in the prior season. More details on the 2014 selection process were provided following questions from a reviewer, including the following details:

- Vessels must be associated with an active and valid SPGM permit.
- Vessels must have had landings (catch) in the previous 2 years.
- Those landings (catch) must have come from one of 9 defined geographic zones of the Gulf of Mexico.
- Within the geographic zone, vessels included two refrigeration methods (ice or freezer).

This generated 14 strata (geographical and refrigeration method), with R used to randomly select units for sampling, without replacement, using the fraction of weighted landings for the vessel to total weighted landings for all vessels. However, some details of the selection process remain missing, including the input data files and original R code.

Data from the cELBs were automatically transmitted to an Oracle database via a cellular phone network and were used to estimate total shrimp effort using modified LGL code. The proportion of vessels in the program estimated to have fished over this period was broadly constant at about 90% from 2014 to 2019, but then declined to 73% in 2020 and 55% in 2021.

The development of a new approach to effort estimation was driven by difficulties in generating effort estimates for the 2020 data. This led to a workstream and project that started in late 2021 to develop a workable replacement methodology. The objective for this project was to develop a new methodology that would produce robust effort estimates with:

- Simplified assumptions;
- Increased transparency around input parameters;
- More complete use of the cELB data; and
- Greatly simplified and streamlined R code,

This approach was supported by advances in the R programming language (R Core Team 2023) and simplified and unified data sources since the inception of the ELB program which together permit code to be executed more reliably.

Summary of Changes

Effort classification

- Distances calculated using the Vincenty ellipsoid method (R geosphere package) rather than a Euclidean metric with rough fixed parameters. This results in more accurate distance estimation that takes the curvature of the earth into account.
- A 1-minute resolution NOAA Gulf of Mexico (GoM) bathymetric grid (R marmap package) is used to define fishable depth. The default setting filters out all pings occurring at depths >2,400 feet (but this is an adjustable parameter), noting that the estimated biological depth range for royal red shrimp along the continental shelf is between 590 and 2,395 feet (Perez-Farfante and Kensley 1997).
- An updated GoM shapefile with higher resolution fathom delineations. This shapefile now encompasses the entire GoM U.S. EEZ rather than only extending to the shelf edge.
- The upper fishing speed threshold is calculated using a Gaussian mixture distribution on the observed data rather than using pre-set fixed numbers. This makes the algorithm more robust to changes in fishing dynamics that may occur over time. Starting values that correspond to the expected effort distribution are provided to protect against unrealistic classifications.
- Estimated towing activity (i.e., appropriate speed and location) must occur for a minimum of one-hour to be classified as effort. This is intended to exclude try net tows and other false positives.
- Fishery independent SEAMAP data are used to obtain species-specific effort estimates.

Scaling to total fleet

- Scaling of effort is now done using landings at aggregate combinations of time (3 quadrimesters) and area (5 zones) in a survey-design framework rather than attempting to match individual trips. This ensures all available cELB recorded effort is used to estimate spatial patterns and in the calculation of total effort, rather than only using those trips that are able to be matched to trip ticket landings (which has typically been about 60% in the past). This results in more complete use of the data and fewer assumptions related to the trip matching procedure, given no key exists to truly be able to match trips to landings.
- Separate scalars are calculated for penaeid (brown, pink, white) and royal red shrimp effort.

Computer code

- The code used to generate annual effort estimates has been greatly simplified into a single streamlined R script (<600 lines, including all data queries, figures, and comments). The result is a user-friendly product that can be more easily and reliably interpreted and executed, while requiring the user to know only a single programming language.
- All numeric decisions (thresholds, etc.) are transparent as function arguments (10) and can be modified as appropriate.
- There are no randomized components to the code. That is, results are consistent between runs without needing to set an initial seed.

Key to this approach are five assumptions described in Dettloff (2023).

1) cELB devices capture all fishing activity.
2) There is no systematic bias in effort classification, i.e., there is an equal chance of false-positives and false-negatives.
3) The spatial distribution of cELB vessels is representative of the total fleet within strata.
4) CPUE of vessels with cELBs on board is representative of the total fleet. If CPUE among cELB vessels is higher than non-cELB vessels, this would lead to an underestimation of effort, and vice versa.
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.

For assumption 2, there is some evidence to support this case from comparisons between cELB and observer recorded effort. For assumptions 1 and 3, there is some evidence for an elevated likelihood that these assumptions are not fully supported. For assumptions 4 and 5, no evidence for or against their validity is provided. All of these assumptions are important in understanding the risk to the validity of the methodology should they be violated. The identified issues in assumptions 1 and 3 need to be fully quantified in order to understand whether these could undermine the methodology, and if so, in what ways and by how much. It would be appropriate for some exploration of the robustness of assumptions 4 and 5 to be undertaken.

Full details of the evidence for or against the robustness of the five assumptions was not presented in the documentation provided; recommendations have been formulated based on what was stated. It is also noted that the advent of universal VMS would remove the necessity for checks on assumptions 3 to 5, as these would no longer be relevant.

The core data used to generate the total effort estimates are stored in four tables within the Oracle database. Additional data drawn for other sources exist as fixed sources (e.g., trip and depth grid delineations) and groundfish survey data sources. The basis of matching individual vessels to trips, permits and the cELB data is not fully described, neither are data describing the completeness of the data joining process (e.g., number or proportion of mismatches).

It is reported that, due to the withdrawal of the 3G network service in 2020, from the beginning of 2021 cELB data have been uploaded manually into the data system. There is no discussion of the possible errors that could arise in this process or of approaches being considered or in place to mitigate any data upload quality risks.

The use of standard R code in data processing is strongly supported as this will tend to reduce errors that can be generated with bespoke coding solutions.

The function parameters and default thresholds are well described. However, the reasons for choosing the particular threshold values are not given. This makes considering potential bias in some data components difficult within the context of this review. For example, the minimum time between tows used to consider the start of a new trip (trip.hrs.brk) is set at 24 hours, the implications of setting this at a different value are impossible to assess and so the presence or absence of potential biases cannot be fully ascertained.

The justification for having two distinct measures of minimum effort (i.e., tow.hours.min = 1 hour and trip.eff.hrs.min = 2 hours) and how they may interact was not fully explained in the documents provided.

The use of fixed thresholds applied to various types of data to remove unlikely high or low (extreme) parameter values in the bycatch estimation procedures is a pragmatic approach to reducing errors in the output data. This will clearly not remove all such errors, where, for example, an erroneous value lies just within the permitted data range. It would be helpful to see sensitivity tests for the chosen values, with small changes in the threshold values and reporting of numbers of accepted / rejected values and the change in average parameter values. This may have been done when these thresholds were first established but was not presented for this review. If the threshold values have not been considered for some time, it would be appropriate to review the settings as operational details of the fishery may well have changed over time and more appropriate threshold values may be warranted. The use of variable thresholds used for some parameters obviates much of the need for such review, and some parameters, such as speed, have been explored, including with some outcome reporting, with, for example 99.8% of data points being classified for speed. Precision in these estimation procedures is only one of the desirable outcomes, noting that consistency between years is probably more important that precision in any one year.

While the procedures for estimating fishing time for vessels appear to be appropriate, some checking against real vessel activity data would help to define the performance. For example, comparing model outputs against recorded activity from survey vessels and observer or camera recorded activity data from commercial vessels.

The approach to the splitting of effort between the two principal species groups (penaeids and royal red shrimps) seems well founded and base on solid data. The further splitting of effort for the three penaeid shrimps (brown, white and pink) is based on the observed fishery-independent catch distribution in weight of each species from the SEAMAP Groundfish Survey data (1987-2022) for summer and fall surveys only. With pink shrimp effort spatially distinct in the eastern Gulf off Florida, the separation focuses on the other two species taken in zones 8 to 21.

The effort directed at brown and white shrimp is estimated using the long survey timeseries of spatial, depth and temporal data (time of day and season), analysed for thresholds using binomial GAM smoothing. Limited data from stat zones 8-10 is addressed by pooling, and if the proportion of any species in a stratum falls below 1%, this is set as zero, with subsequent re-calculation. As above and given the available data, this methodology is likely to provide an adequate basis for estimation, but a more robust outcome would likely be generated using a spatio-temporal modelling approach (e.g., VAST or INLA). These alternative approaches are designed to cope with data having spatial and/or temporal gaps, such as is seen in much of the data used here, including the spatial distribution of fishing effort and also in the two surveys (summer and fall).

Scaling factors for the penaeid and royal red shrimp by time and spatial strata are calculated using the ratios of total offshore landings for cELB and non-cELB vessels.

Confidence intervals, based on 24-hour days, are calculated around the estimated values from the square root of the variances with appropriate degrees of freedom, but it is not clear why these are split proportionally for each species group rather than being calculated separately for each.

The penultimate step is to allocate a small amount (~3.5% in 2020) of total offshore state permit landings (no SPGM permit) to depth zones in proportion to the area of offshore waters in each depth zone and the four aggregate areas. As the proportion of this catch is small (<5%) and as most gets allocated to the 0–10 fathom zone, and as this is a proportional correction, it may be possible to develop a simpler correction factor at this point.

The final step permits the allocation of cELB effort, and thus landings, among the different strata (depth, time, and statistical zones). The allocation of landings requires an assumption of constant CPUE across depth zones, which is reported as reasonably supported by some studies.

Overall, the current approach appears to be sensible, practical, and pragmatic and appears to be fit for purpose in providing management advice, although adequate diagnostic checks to ensure this were not available for this desk review. It is, however, also likely that some alternative spatio-temporal approaches would provide more robust outcomes and represent better quality science to support management.

Comparison of the LGL and current (SEFSC) analyses showed broadly similar patterns of effort over 2014-2020 but with some differenced in individual years (Figure 12 of Dettloff, 2023).

While the current methodology may be adequate for the purpose of estimating total bycatch, variables are often considered separately and can be quite constrained. There clearly has been considerable thought within the team considering a range of possible improvements, seen in the response to reviewer questions, in Dettloff (2023,) and in Smith et al. (2023), and these should probably be captured together with some of the recommendations from this review. However, rather than try to further improve aspects of the current approach, it may be more appropriate to consider a different approach using, for example, one of the spatio-temporal modelling approaches that have been developed over recent years, such as Integrated Nested Laplace Approximation (INLA) or Vector Autoregressive Spatio-Temporal (VAST). These approaches would enable a much more complete consideration of the factors affecting catch and bycatch levels that are difficult to address by developing predetermined strata, and they are designed to cope with gaps in both spatial and temporal data (both present here), can incorporate Bayesian approaches, and can also be implemented in R. Using one of these approaches would not replace all of the components of the estimation procedures currently in use, and a number of these would need to be retained.

Use of spatio-temporal modelling would likely give a more robust outcome than the current approach, possibly considerably so. However, spatio-temporal modelling may be more difficult to implement than the current approach and will likely require a range of specific diagnostics to be developed.

Should spatio-temporal modelling be conducted in the future, some directly comparable analyses using the current methodology should also be prepared in order to inform end-users of the magnitude of any change and where such changes occur due to the new methodology.

Dettloff (2023) correctly notes that the methods used (as would also be the case for alternative methods) are limited by the quality of the available data, especially with regard to incomplete ELB data (e.g., incomplete trips), missing or inaccurate landings reports, and all vessels with cELB units submitting complete ELB data. No information about the potential scale or importance of the robustness of the basic data were presented.

1b. Improving Bycatch Estimation

This component draws on the outputs of the effort estimation from the ELB data cross-checked with observer catch-effort data and the data held in the annual landings and gear survey database. These are then subject to stratification and calculation of observer strata CPUE to generate estimates of shrimp and red snapper bycatch.

This work covers the waters of five states, where there are four principal data sources, (i) trip ticket landings, (ii) ELB unit, (iii) shrimp observer program, and (iv) the SEFSC annual landings and gear survey data. The last three data sources are exclusive to the federal permitted vessels.

The trip-ticket data are seafood dealer data from a long-established system. The ELB data are described above. The observer data are obtained by the SEFSC Shrimp Observer Program, where scientific observers on commercial fishing vessels record detailed information on catch and effort for a subset of trips. The SEFSC annual landings and gear survey data are collected in a mandatory (for permit holders) postal/mailback survey that has operated since 2005 and collects information on vessels and gears. This recent development of the shrimp bycatch estimation process has used the landings and gear survey data for the first time in bycatch estimation.

Estimation of bycatch is by means of a ratio-of-means estimator (Smith et al., 2023). Some alternative approaches, including spatio-temporal models may yield more robust results.

The management of the shrimp fleet is complex, with inshore and offshore components, as well as federal and state permitting systems which have different spatial coverage, and while otter trawls make up the majority gear type, other gears are also used.

The historic observer catch-effort processing is described as very complex. The new approach reports significant refinement and simplification, which would be expected to reduce or eliminate some of the data process errors associated with this type of data processing.

Concerns about the accuracy of the observer catch and effort data led to the development of detailed cross-checking of catch and effort for individual trips. Data were matched using ID and ELB data filtered to the pings between the first and last two dates of the recorded trip. This highlighted some errors in the observer data, tow date and time, for which the data processing approach was modified to handle and allow for correction. A total of 244 trips for 154 vessels were analysed for 2014-2020. Errors were approximately normally distributed about zero but with some negative skew, suspectedly due to lack of recording by some ELB units. Matches of trip tickets with observer data were done for exact reporting days and ± 4 days either way. The basis of setting the cut off at ± four days was not clear, other than if greater than this, there were multiple matches.

Notably, developing and applying parts of the new methodology enabled some significant data processing errors to be identified and corrected in, for example, the Mississippi trip-ticket identification data when stored in the GulfFIN database.

A further issue was identified, where the trip landings of shrimp were only reported as a single species while observers reported multiple species from the same trip. This highlights the type of data quality issues inherent in these types of data. This issue was addressed by cross-matching using all penaeid catches, with 374 successfully matched observer to trip-ticket data, a 62% match success. Importantly, there was no bias evident in the successfully matched data. A 62% success rate for cross-matching these data types does appear to be somewhat low.

Implementation of a unique trip identifier (UTID) for all fishing activities would considerably simplify the process of matching and cross-checking data such as trip and observer data.

Data from the SEFSC annual landings and gear survey database were analysed to better understand aspects of parameters that could affect effort, specifically including headline length (a proxy for net width) and the number of nets used during trawling (typically 2 or 4, also a proxy for net width). Some other gear parameters were also considered but eliminated based on evidence. The SEFSC annual landings and gear survey database provided fleet-level gear information that could be matched to observer reported gear information, enabling gear characteristics to be incorporated with ELB effort and landings data to estimate bycatch. These analyses used data from the time period 2014 to 2020 that were analysed as averages over this period. As a relatively minor point, data collected by user-surveys from a large number of individuals often suffer from entry errors, including errors in values entered for different years. The documents did not indicate any level of checking on the SEFSC annual landings and gear survey data for consistency, whether within or between years. It is likely that, at least, outlier checking is conducted on these data prior to inclusion in the database and there may also be between-year checks for some parameters. It would be appropriate for the types and details of any data quality checks on the SEFSC annual landings and gear survey data to be documented as part of this methodological process update and an assessment done on whether any further level of checking was desirable.

Observer recorded gear characteristics were combined with catch and effort data using GLM regression, including catch (kg), effort (hrs), a complex of year-area-depth-season, combined with head rope length, and net number (2 or 4), which provide insight into some of the key functional relationships between the variables, such as catch and effort. Most outcomes were informative, including for example, that each net in a 2-net rig caught more than one net in a 4-net rig but total catches were greater in 4-net rigs than in 2-net rigs. Also, catch showed no relationship with headline rope length. The relationships were not found to be common across all species considered.

Exploration of CPUE consistency across time and space showed both diurnal, inter-seasonal and inter-annual inconsistencies. These inconsistencies were often not obscured by averaging data. For example, seasonal differences persisted with data averaged across years.

Collectively, these analyses showed that averaging some parameters (e.g., across years) has little effect on the outcome but eased issues arising from limited observer data. There was, however, still a need to impute considerable amounts of effort data. Overall, annual observer-predicted catches were lower than the reported trip-ticket catches by an average of about 20% for the period 2014-2020, and observer-predicted catches were lower compared to trip-ticket catches across all years in areas 1, 3, and 4, but were generally higher across years in area 5.

The new methodology was used to produce estimates for red snapper bycatch in numbers, but no comparative estimates derived by the previous methodology were provided.

The detailed investigation and exploration of the new methods, described by Smith et al. (2023), to improve the reliability of bycatch estimation have clearly yielded a number of benefits. These include the ability to incorporate new datasets into the analysis to improve the estimation, identification and elimination of systematic errors in some datasets, improved testing and reporting of diagnostics for the methodology, improved standardization and cataloguing of analytical code, and improved overall accuracy.

It is, therefore, clear that the new methodology is an improvement on the historic approach and will give more robust estimates of bycatch.

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.

There has been a marked reduction in the proportion of the original 500 vessels that have been determined to have fished in the two most recent years for which vessel data are reported (i.e., in 2020 and 2021) (Dettloff, 2023). It is possible, even likely, that this level of decline either has or will lead to spatial bias in the basic data available as the loss of vessels estimated to have fished are unlikely to have mirrored the spatially stratified, random sampling used to select the vessels in 2014 (key assumption 3). If this is an issue, it may well get progressively worse should there be further decrease in the proportion of the original sample of vessels fishing. This issue does not appear to have been addressed in the documentation provided but clearly should be reviewed and any current bias corrected for, and future bias guarded against.

Information provided on another of the five key assumptions (#2) states that there is no systematic bias in classification and that this is supported by some limited analyses comparing cELB and observer classification data.

The status of the data underlying the other assumptions with respect to bias is unclear as no analyses or studies were presented. There are obvious possibilities of bias due to violation of some of the assumptions, which should be defined, and recommendations are made to evaluate the risk of this.

It is highly likely that bias exists in a number of the data components used, especially associated with the tails of distributions and especially with vessel reported data. If such bias occurs it may not occur every year, it may be consistent from year to year, or it may be inconsistent from year to year, depending on the principal drivers of the bias. The fundamental issue is whether the pattern and scale of the bias is sufficient to undermine the robustness of the estimation methodology, noting that there are already data quality issues that will tend to mask minor biases. Understanding the potential importance of bias really requires a detailed understanding of the data sources, which is beyond a review of this type. However, given the consideration of bias in the principal documentation, it may be appropriate to have a thorough in-house review of the data sources for biases that may affect outcome robustness in bycatch estimation, or at least fully document potential sources of bias.

b. Is the uncertainty adequately described? If not, make recommendations to better characterize the uncertainty.

Uncertainty in various quantities will be inherent in the data and this may or may not flow through into the bycatch estimates.

Key uncertainties are estimated, such as confidence intervals about the final effort and landings outputs. However, it is difficult to consider that uncertainty has been adequately described for all components for this review; uncertainty is only mentioned twice in the two key documents provided. However, it is likely that the relevant uncertainties are better described and understood than has been communicated to a review of this type and would have been more amenable to exploration at an in-person review.

It is noted that the focus of the program has been to deliver accuracy in bycatch estimation rather than to understand the associated uncertainty (Smith et al., 2023). Thus, while variance in some data components is considered and incorporated (e.g., observer catch rates), other data are assumed constant, when in reality that is unlikely to be true (e.g., fleet effort). The result of this is that uncertainty appears to be poorly quantified and inadequately considered in the processes currently used for some components of the overall estimation procedure.

One way to improve this position would be to develop a process to identify those datasets with uncertainties or importance in bycatch estimation, with a view to characterizing and quantifying the uncertainties. Then, a rational choice can be made of which uncertainties are most important and which of those should be incorporated into the estimation of bycatch versus which ones are important but best expressed in advice to managers outside of the estimation procedure.

3. ***Are the statistics used in the estimation appropriate, and are they used correctly?***

The statistics used in the estimation procedure described appear to have been used appropriately.

The new methodology does make wider use of estimation of variability compared to the older approach. However, some elements, such as variance estimates, are not considered for some datasets and arguably should be available, at least for use in diagnostics.

As discussed, elsewhere (e.g., ToR 1), the statistics used are a clear improvement on the previous approach and are appropriate to be used to estimate bycatch in the shrimp fishery. However, alternate approaches exist that may provide more appropriate analysis (especially with significant gaps in spatial and temporal data) and more robust estimates of bycatch. These methods include a range of relatively recently developed spatio-temporal modelling tools such as VAST or INLA, that are likely to provide some real advantages in some areas of bycatch estimation. Appropriate recommendations are made with regard to these approaches.

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.***

The updated approach estimates uncertainty for more data and parameters, and makes greater use of this in the estimation procedures, cross-checking, and in exploring variability around point estimates of, for example, bycatch. It is probable that this could be further extended, especially in regard of more detailed presentation of diagnostics for the updated methodology (see also ToR 2b).

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 issue of small sample sizes is referred to in a number of areas in the documentation, most notably around catch rates for the different species of shrimp and the spatio-temporal distribution of estimated effort. Generally, in the methodologies described, simple threshold cut-offs were applied to reduce or eliminate potential impacts of decreasing and small sample sizes.

A better understanding of the relative importance of, and how best to manage, low and declining sample sizes could and probably should be sought. This could be attempted in a variety of ways but initially a simple approach could be explored to define whether there are likely to be sample size issues and whether the current thresholds are likely to be appropriate. For example, reworking component datasets using a range of different cut-off thresholds, including some both above and below current thresholds, and then conducting some detailed comparative analyses of the results to see if using different thresholds was able to generate meaningful differences in outcomes.

Issues identified across multiple ToR

Gaps in the distributions of spatial and temporal data

Many of the datasets required in these component analyses, and eventually in the estimation of bycatch, are complex with different types of gaps in the data. This includes, for example,

- Research trawl surveys that are only conducted at certain times of year (summer and fall surveys).
- Commercial landings data by species being spatially and temporally patchy for operation and catches, with substantial areas and periods of data absence.
- Observer data on only part of the fleet that already demonstrates spatial and temporal patchiness in operation and in catches.
- ELB data devices on part of the fleet (that has incomplete overlap with observer placement) that already demonstrates spatial and temporal patchiness in operation and in catches.
- Spatial-temporal distribution of use of different gear types.

Both the previous and new approach use a complex, stepwise methodology to bring data together, ultimately toward bycatch estimation. Within these processes, dealing with gaps in data and limited data in some strata creates some challenging issues addressed by using various statistical approaches including, for example, averaging and use of GAMs. A more comprehensive approach using spatio-temporal models, which while possibly more complex to set up, is likely to better manage some of these data issues and provide more robust estimates.

Recommendations

The following are recommendations for the estimation methodology for shrimp bycatch in the southeast U.S. Gulf of Mexico shrimp fishery.

Apart from the first recommendation, that addresses an issue that was identified across a number of the ToRs, these recommendations are organized by ToR element.

Multiple ToR

- Consider using a completely different approach for at least some components of estimating effort and in estimating total bycatch. This should specifically include at least one of the spatio-temporal modelling approaches that tend to be able to handle data distributional issues better than conventional approaches. Examples to consider include Integrated Nested Laplace Approximation (INLA) and Vector Autoregressive Spatio-Temporal (VAST) modelling.

Recommendations against each ToR

ToR 1: Describe the data sources and proposed methodology.

(i) Consider developing and implementing a Unique Trip Identifier (UTID). This would greatly simplify the necessary cross-matching of fisheries data, will reduce the resources needed for replacing missing data by imputation and other means, and will improve the overall quality of the data, probably substantially.

(ii) If not already done, some appropriate success/failure metrics (e.g., numbers of records eliminated with each merge function, numbers of duplicate rows, etc.) should be reported annually for the matching process across the various data tables.

(iii) Describe the risks associated with the currently used manual upload of cELB data and define processes to manage and mitigate those risks, including finding an automated approach to uploading data.

(iv) It is probably an appropriate time to consider a full re-randomization of the vessels and strata that form the basis of the estimation of effort, including exploring whether gear-type would be a better or additional factor in the stratification than the ice/freezer factor used previously (but see also recommendations ToR 2b (iii) and (iv), below).

(v) Document as fully as possible the previous approaches for all steps used to generate the estimation of bycatch.

(vi) If and when there is any substantive future change in total bycatch estimation (e.g., re-randomization of the fleet/strata), a complete documentation of the whole procedure and archiving of code and input and output data should be implemented.

(vii) Review, document and report the basis for selecting fixed, data exclusion thresholds.

(viii) Compare individual vessel activity data (i.e., survey and/or observer reported activity from commercial vessels) with model estimates for those same vessels as check on the reliability of the model approach used. Ideally some checking should be conducted for each year.
(ix) Document the range of possible improvements that could be made to the current approach and seek to quantify the benefits of each.
(x) Review the types and scale of data quality checks on the SEFSC annual landings and gear survey data within and between years. This should be consistent with what data are relevant to bycatch estimation and the way these data are used.
(xi) Some regular basic studies of the robustness of the basic data to inform on the overall reliability of the analyses performed are warranted. This could include, for example, patterns of incompleteness in cELB data and in landings reports (especially for cELB vessels).

ToR 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.

Explore more fully and report on whether the five methodological assumptions are supported:

(i) Assumption 1: more fully explore differences in cELB and observer measures of fishing activity to quantify the scale of spatial and temporal differences so as to understand and be able to correct for any violations of the assumption.
(ii) Assumption 2: continue to periodically compare cELB classified effort with equivalent observer reported effort to provide evidence that this assumption remains unviolated.
(iii) Assumption 3: given that the proportion of the original 500 selected vessels continuing to generate effort data has recently sharply dropped, there is some likelihood that the spatial representativeness of the effort data source is no longer unbiased. The spatial and temporal distribution should be checked and reported for recent years, and in future years also as standard performance monitoring.
(iv) Assumption 3: if, for ToR 2b (iii) above, the proportion of vessels in the program estimated to fish continues to decline, or if the likelihood and scale of potential spatial bias seen in ToR 1 (iv) above is deemed sufficient to undermine the statistical robustness of the new methodology, then some specific corrective actions should be undertaken. These actions should be either to make statistical adjustments to the methodology to correct for any spatial bias in the input data or to define and implement a process for selecting additional vessels (using a statistically robust selection approach with respect to, at least, spatial distribution) to increase the fleet size of the program and ensure an appropriate, catch-weighted spatial distribution of vessels is providing data in future years. Complete replacement of the current cELB fleet may be required.
(v) Assumption 4: compare and report on spatial and temporal differences in CPUE between cELB and non-cELB vessels in such a way as to quantify any differences and allow corrections to annual effort estimation to be made as necessary.

(vi) Assumption 5: regularly monitor and report on spatial and temporal patterns (similarities and dissimilarities) in landings for vessels with and without cELB devices, so as to inform on the need for corrective actions on effort data integrity and allow corrections to annual effort estimation to be made as necessary. This should probably be done each year.

(vii) Monitoring of the integrity of the five assumptions should form part of a suite of diagnostics generated and published each year to enable all stakeholders to fully understand the scale and sources of potential errors and biases in the methodology applied.

(viii) Consider an in-house review or workshop, possibly with specialist external participants, to review the various datasets for bias at a scale or pattern that could adversely affect the estimation of bycatch in the shrimp fisheries. Such work should be focused on impacts of bias on the current procedures and any other methodological approaches under consideration for future application.

b. Is the uncertainty adequately described? If not, make recommendations to better characterize the uncertainty.

(ix) Develop a process to identify those datasets with substantial uncertainties that may be of importance in bycatch estimation, with a view to characterizing and quantifying the uncertainties. This process should explicitly distinguish between (a) uncertainties that are most important and should be incorporated into the estimation of bycatch and (b) those that are important but best expressed in advice to managers outside of the bycatch estimation procedure.

ToR 3: Are the statistics used in the estimation appropriate, and are they used correctly?

(i) See the first recommendation with respect to improving the appropriateness of the methodology of bycatch estimation.

ToR 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.

(i) Consider how to use the existing uncertainty estimates to provide improved diagnostics that can be used to demonstrate to managers the appropriateness of the methodology used currently and also to enable methodological comparisons if and when the methodology changes in future.

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?

(i) Explore the appropriateness of the cut-off thresholds where these are used, especially where these appear to be arbitrary, so as to ensure that data use is maximized within the acceptable error bounds of the desired outputs.

Appendix 1: Bibliography

Shrimp bycatch Review Documents

Dettloff, K. (2023). Estimation of Commercial Shrimp Effort in the Gulf of Mexico, Companion report, NOAA SEFSC, 22 pp.

Smith, S.G., Atkinson, S., Peterson, C., Williams, J.A., Dettloff, K. and Lowther, A. (2023). Improving Estimation of Bycatch from Shrimp Trawls in the Gulf of Mexico. 38 p.

Background Documents

Anon. (undated). Florida Trip Ticket Manual. Florida Fish and Wildlife Conservation Commission. 37 p.

DCNR. (2020). Alabama Trip Ticket Manual. State of Alabama Department of Conservation and Natural Resources, Marine Resources Division, Gulf Shores, Al. 60 p.

LDWF (2020). Louisiana Department of Wildlife and Fisheries Trip Ticket Procedures Manual. Trip Ticket Project, Po Box 80337, Baton Rouge, L. 53 p.

LGL Ecological Research Associates. (undated). Procedures Manual for Electronic Logbook (ELB) Versions 4.0-6.0. 502 p.

MDMR (2012). Trip Ticket Procedures Manual. Department of Marine Resources Office of Marine Fisheries, 1141 Bayview Avenue Biloxi, Ms.31 p.

NMFS. (2020). Observer Training Manual: Characterization of the US Gulf of Mexico and Southeastern Atlantic Otter Trawl and Bottom Reef Fish Fisheries. National Marine Fisheries Service Southeast Fisheries Science Center Galveston Laboratory. 296 p.

NMFS. (2020). Observer Training Manual: Forms. National Marine Fisheries Service Southeast Fisheries Science Center Galveston Laboratory. 49 p.

TPWD. 2019). Trip Ticket Instruction Manual (Paper Tickets) Commercial Landings Program. Texas Parks and Wildlife Department Coastal Fisheries Division. 43 p.

Other Material Provided or used

Numerous SAS code files for both the original and updated methodologies were provided.

Perez-Farfante, I., Kensley, B.F. (1997) Penaeoid and sergestoid shrimps and prawns of the world: keys and diagnoses for the families and genera. Mem. Mus. Natl. Hist. Nat. 175: 1-233.

Appendix 2: Performance Work Statement

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 standards .

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 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?