Estimation of Commercial Shrimp Effort in the Gulf of Mexico

Kyle Dettloff

SEDAR87-DW-01

30 August 2023
Updated: 27 September 2023
7 November 2023
14 December 2023
2 April 2024
17 May 2024
20 August 2024

Please cite this document as:

Dettloff, K. 2024. Estimation of Commercial Shrimp Effort in the Gulf of Mexico. SEDAR87-DW-01. SEDAR, North Charleston, SC. 31 pp.

Estimation of Commercial Shrimp Effort in the Gulf of Mexico

Kyle Dettloff, *SEFSC*

July 2024

History

Electronic Logbook (ELB) devices were originally developed by LGL in 2004 as a position logging system for commercial shrimping vessels in the Gulf of Mexico, with the goal of more accurately estimating spatial patterns of trawling effort than those collected by port agents. These devices record vessel location at 10 minute intervals using GPS, and the resulting speeds are used to identify potential vessel trawling activity. From 2004-2013, data from memory chips on these devices were collected and processed by LGL, and total fleet effort was estimated using LGL code. In mid-2013, these responsibilities were transferred to NMFS, where James Primrose was responsible for data management and calculation of effort estimates using a modified version of the original LGL code.

In 2014, cellular Electronic Logbook devices (cELB) were implemented, in which positional data are automatically transmitted back to NMFS servers through the cellular network, as opposed to manual retrieval of memory chips. In early 2014, NMFS selected 500 Gulf of Mexico Shrimp Permit (SPGM) owners using a spatially stratified random sampling method weighted by landings in the prior season to participate in the cELB program. Consistent position data from devices were being received by the second quarter of 2014. An additional 100 vessels were selected to carry units in 2018. Data from these devices are stored in a Galveston Oracle database and were used to generate total fleet effort estimates using the modified LGL code through the 2019 fishing season. In the year 2020, data were received from 452 cELB devices encompassing nearly 13 million pings, 365 of which were identified to have fishing activity. A yearly breakdown of the number of vessels exhibiting ELB fishing activity between 2014 and 2022 is provided below.

Year	2014	2015	2016	2017	2018	2019	2020	2021	2022
Active ELB vessels	444	433	462	452	495	449	365	278	230

Motivation

Development of a new effort estimation method began in late-2021, after the original code was unable to be executed in a timely manner to generate 2020 estimates. Goals in developing this new method were to produce robust effort estimates with:

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

This simplified code and logic, resulting from a combination of advances in the R programming language (R Core Team 2023) and simplified and unified data sources since the inception of the ELB program, will allow code to be executed more reliably, both in the short-term and by future users.

Summary of Changes

- Effort classification
    - Distances are calculated using the Vincenty ellipsoid method (R *geosphere* package) rather than a Euclidean metric with rough fixed parameters. This results in more accurate distances that take the curvature of the earth into account.
    - A 1-minute resolution NOAA GOM bathymetric grid (R *marmap* package) is used to filter out data at depths outside of the biological ranges of target species. The default settings filter out any pings occurring at depths between 450 and 600 feet or at depths greater than 2,400 feet, though these are adjustable parameters. 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), and brown shrimp can occur to depths up to 450 feet (Renfro and Brusher 1982).
    - An updated GOM shapefile with higher resolution fathom delineations is now in use. This shapefile now encompasses the entire Gulf 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.
    - Towing activity must occur for a one-hour minimum to be classified as effort. This protects against including 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 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 (~60%). 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 effort.
- Code
    - Code for generating annual effort estimates has been greatly simplified and modernized 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 interpreted, executed, and troubleshooted, 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 a seed.

Assumptions

The following basic assumptions are required to obtain accurate estimates of total effort, given 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.
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.
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.
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 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. 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.

*Note assumptions 3-5 are required due to ELB devices being present on only a subset of the fleet, and could be eliminated with universal VMS. An evaluation of these assumptions is presented in Appendix 1.

Data Sources

All tabular data required to generate fleet-wide effort estimates are stored in the following four tables within the SEFSC Oracle database:

1. Raw cELB pings (elb.elb_data@elb_dblk)

   Complete data range from 2014-present, encompassing between approximately 12-15 million rows (10-minute pings) annually. With the termination of the 3G network in 2020, data have been uploaded to this table manually from chips beginning in 2021.

   Relevant fields include: box number, latitude, longitude, year, month, day, hour, minute, and second.

2. Vessel assignment table (elb_obj.elb_assignments)

   This table is used to assign a vessel ID (either US Coast Guard or otherwise state number) to the raw cELB data. The “serial” field in this table is used to join to the “box number” field in the cELB data. Other relevant fields include: vsbn (vessel ID), date installed, date removed, and status (current state of cELB device).

3. Trip ticket landings (scdw.v_fac_landing_prior_year@secdw_dblk)

   This is a SEFSC view of complete dealer trip ticket landings for prior years based on data in the GSMFC database. Data are refreshed on a weekly basis to accommodate any updates or corrections to existing data. Relevant fields include: landed_date, fishing_area_sub_area_num (area and subarea of landings on 1-21 GOM trip ticket grid), landing_state_fips, landing_county_fips, species_itis, gear_code, disposition_code, market_code, market_count_min, market_count_max vessel_official_number, and gutted_wgt (in this case meaning shrimp tail weight). These data are used to summarize landings according to vessel cELB status within time/area blocks.

4. Vessel permit table (scdw.dim_vessel_permit@secdw_dblk)

   This table is maintained by SERO and identifies which federal permits are associated with a particular vessel through time. This allows landings to be separated into state and federal categories based on if the vessel ID associated with the landings was in possession of a valid SPGM permit. Relevant fields include: official_number, fishery_code, permit_effective_date, permit_terminated_date, and permit_status.

5. Annual landing and gear survey table (gom_shr_permit.gear_master@secapxdv_dblk)

   This table is used to assign a net number (two or four) to federally permitted SPGM vessels based on what is reported in the Annual Landings and Gear Survey each year. Relevant variables include: year, vessel_id, and number_of_nets. Vessels reporting net values of three or five are treated as two and four, respectively, to account for vessels that may have included a try net in their reported number. Other net values are considered erroneous and data are treated as missing.

Additional static inputs include:

1. Gulf of Mexico shapefile with 1-21 trip ticket area grid and 10 and 30 fathom delineations (Courtesy Jo Williams, SEFSC)
2. Gulf of Mexico bathymetric grid (imported from *marmap* package)
3. SEAMAP Groundfish Survey data (1987-2022) containing weights of brown, pink, and white shrimp sampled on 30 minute tows by season (Summer/Fall), stat zone (1-21), depth, and time of day (D/N)

Code

All code to produce the PDF report with annual effort estimates, figures and tables is contained within the script ‘effort_scaled.R’. This code is executed through the Rmarkdown file ‘GOM_shrimp_effort_report.Rmd’. The user simply selects “Knit with Parameters” from the dropdown menu next to the Knit button in the GUI, enters the year for which estimates are desired in the Shiny interface, and clicks Knit to run the code and produce the report.

Required R packages to execute code include: ROracle, dplyr, tidyr, marmap, geosphere, sf, suncalc, mixtools, survey; ggplot2, scales (to produce figures); knitr, shiny (for report generation)

Function parameters (and defaults) are as follows:

- ping.hrs.min: minimum time in hours between pings required to attempt to classify vessel state (default 59/3600 hours, i.e., at least one minute)
- ping.hrs.max: maximum time in hours between pings at which an initial attempt will be made to classify vessel state (default 0.5 hours)
- knots.min: minimum speed in knots between pings which has the potential to be classified as towing effort. All speeds below this threshold will be considered stopped/idling (default 1.9 knots)
- knots.max: upper speed bound in knots to be considered a steaming activity rather than an erroneous data point. That is, this value cuts off the right tail of the distribution of steaming speeds (default 11.5 knots)
- tow.hours.min: sets the minimum amount of consecutive time in hours a vessel must fall in the speed profile of towing activity to be considered a true tow. To reduce false positives, speeds that otherwise fall within the range of towing activity will only be counted as effort if they meet this threshold (default 1 hour)
- trip.eff.hrs.min: sets the minimum amount of towing time in hours on a “trip” to be considered true effort. This serves as a second pass to reduce false positives that are not filtered out with tow.hours.min (default 2 hours)
- fill.hrs.max: maximum amount of time in hours between consecutive pings to attempt to fill in missing vessel activity for gaps in ELB data transmission (default 8 hours)
- trip.hrs.brk: minimum amount of time in hours that must elapse between the end of one tow and the start of the next to be considered a new “trip” (default 24 hours)

- ❖ min.transition.hrs: minimum amount of time in hours that must elapse between the end of one tow and the start of another to be considered a new tow (default ⅙ hour, i.e., 10 minutes)
- ❖ penaeid.depth.ft: sets maximum depth in feet for Penaeid species (brown, pink, white) (default 450 ft/75 fm).
- ❖ rr.depth.ft: sets minimum depth in feet for royal reds (default 600 feet/100 fm). Pings occurring at depths between penaeid.depth.ft and rr.depth.ft are excluded.
- ❖ max.depth.ft: sets maximum depth in feet, according to 1 minute GOM bathymetric grid, for vessel activity to be considered true effort. Any vessel activity occurring in deeper waters will not be considered effort (default 2,400 feet/400 fm).
- ❖ rel.net.cpue: sets relative CPUE scalar for four net vs. two net configurations (default 1.4 based on observer data)

The code to estimate effort can be broken down into two primary steps: initial classification of effort from raw cELB data and scaling of cELB effort to the total fleet. These steps are detailed below.

Effort classification

1. Data input

Raw cELB data with timestamp, latitude, and longitude (1) for a specified calendar year are first pulled into R from the Oracle database using the *ROracle* package (version 1.3-1.1). These data are joined to the vessel assignment table (2) by box number, allowing vessel IDs to be associated with the correct boxes at a given time. At this stage, data are also filtered down by box status to remove any demo, development, or test data.

2. Data filtering

Next, coordinates are converted to decimal degrees and raw pings are filtered to the extent of the Gulf of Mexico EEZ shapefile based on lat/long using the st_join function in the *sf* package, such that any data falling outside the region (e.g., South Atlantic) are removed. Any pings occurring deeper than the maximum specified depth contour (default 2,400 ft.) are removed based on the GOM bathymetric grid (*marmap* package). Additionally, any pings falling between 30 and 100 fathoms in stat zones 3-10 where Penaeid effort does not occur are assumed to be false positives and also removed. All duplicate rows (i.e., identical timestamp, lat/long, vessel) are removed in this step as well.

3. Initial speed calculation

Filtered data are next sorted in order of timestamp within each vessel ID, and the distVincetyEllipsoid function in the *geosphere* package is used to calculate distances (in nautical miles) between consecutive lat/long points. Speed in knots between consecutive pings are calculated by dividing these distances by the elapsed time in hours. To prevent unrealistic values, only consecutive pings with times between ping.hrs.min (~1 minute) and ping.hrs.max (30 minutes) are used in this initial calculation to generate a distribution of vessel speeds. The vast majority of pings occur at 10 minute intervals, so this is not an issue in most cases but serves as a safeguard to prevent extreme values from arising. Any speeds still falling above the realistic range of vessel travel (knots.max, default 11.5), likely due to bad GPS points, are also removed from the distribution.

4. Classification of vessels speeds

Vessel speeds as calculated in step 3 (above a minimum stopped/idling threshold, default 1.9 knots) tend to clearly follow a bimodal distribution, representing a mixture of the two Normal distributions that result from the distinct fishing and steaming speed footprints. Therefore, a 2-component Normal mixture model (R package *mixtools*) is fit to this distribution of speeds to help identify an optimal breakpoint between fishing and steaming activity. To reduce false positive effort classification for speeds near the start and ends of tows that may be slightly faster than those at which trawling typically occurs, the speed at 95% of this estimated cutoff value is then used as the upper bound to define fishing speeds. This model-based calculation can be useful for identifying subtle changes in fleet behavior over time as opposed to using a fixed threshold. Starting values of expected means (3, 8), standard deviations (0.5, 1.5), and densities (0.8, 0.2) are provided for the distributions of fishing and steaming speeds, respectively, ensuring that the model converges to a reasonable solution. For 2014-2020 cELB data, the resulting threshold comes out consistently very close to 3.8 knots (i.e., cutoff = 3.8 * 0.95 = ~3.6 knots), with very little variation between years (+/- 0.05 knots).

Upon classification of speeds according to the logic above, values that were unable to be accurately classified initially due to pings occurring below the minimum time threshold are filled in with the most recent known vessel state. In cases where the elapsed time between pings is greater than 30 minutes, classification of vessel state is only attempted for periods of up to a specified maximum (fill.hrs.max, default 8 hours), and only if the calculated vessel speed during that interval falls within the range of possible speeds (<11.5 knots). Reducing fill.hrs.max below 8 was observed to have very limited impact on the final result, but is a necessary parameter to handle gaps in ELB transmission. For 2014-2020, this logic typically resulted in excess of 99.8% of data points receiving a classification.

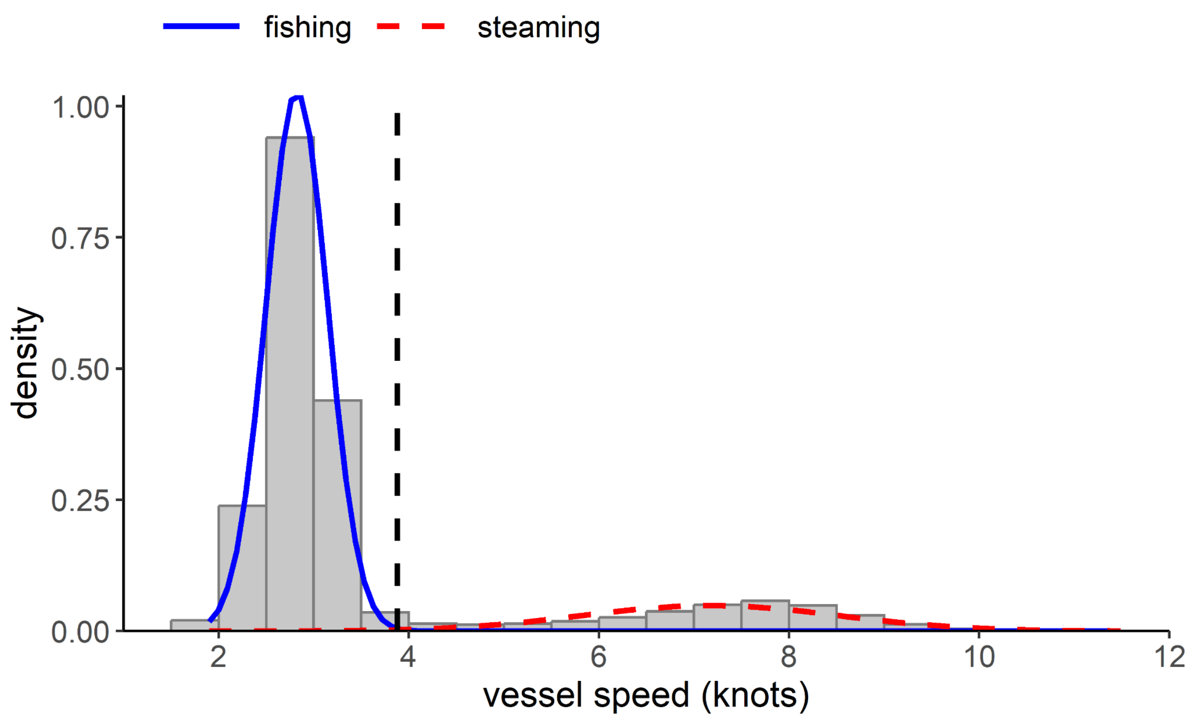


Figure 1: Distribution of cELB vessel speeds (2020) with fitted Gaussian mixture distribution.

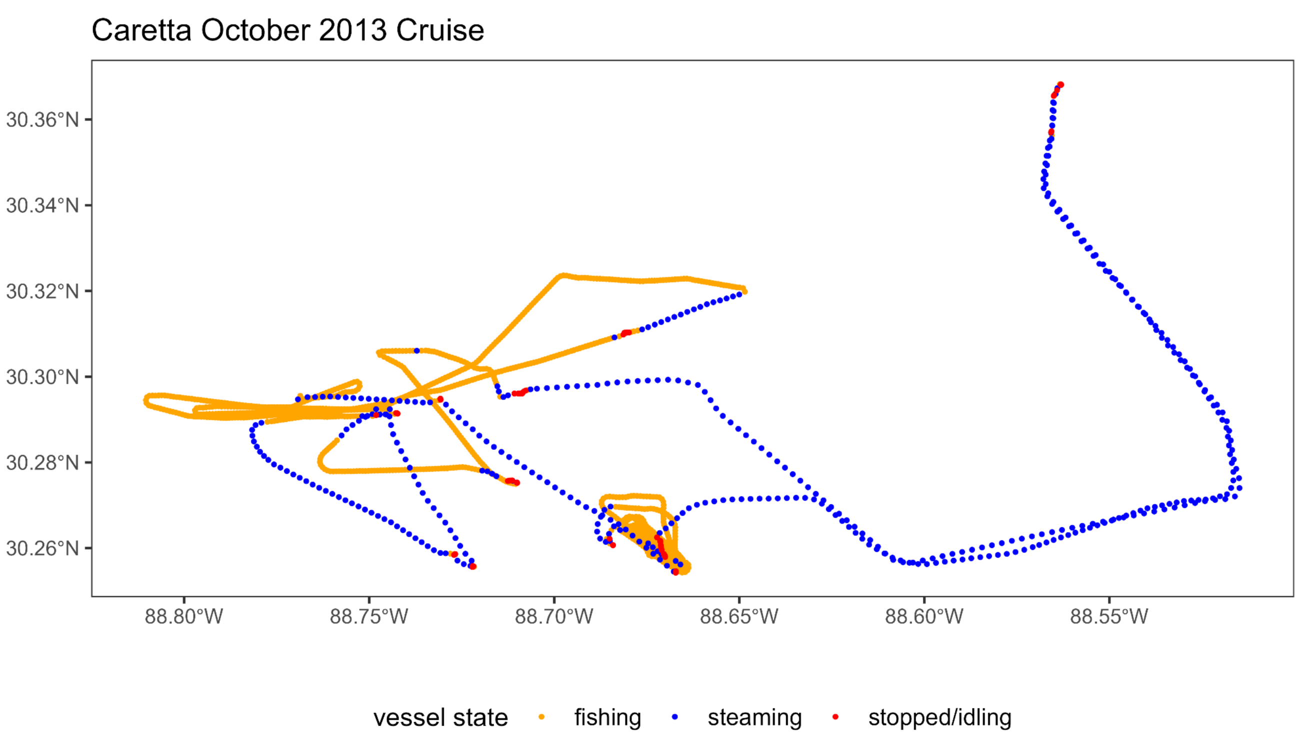


Figure 2: Example vessel track with classified vessel states.

5. Isolation of fishing activity

For the final stage in calculating total cELB effort, a series of steps are taken to remove potential false positive pings based on surrounding vessel activity. The *minimum* break time between tows, or changes between fishing and non-fishing activity, is set to 10 minutes by default (min.transition.hrs). This helps remove changes in vessel state that occur in an unrealistically short interval, likely arising from irregularities in the data (e.g., pings occurring at shorter than 10 minute intervals). A minimum transition time between tows of 10 minutes is also consistent with what is seen in the shrimp observer data. Once tows are defined in this way, only consecutive periods of activity classified as fishing effort falling at or above 1 hour (tow.hrs.min) are kept as true effort. This helps filter much of the false positives arising from vessels that happen to be traveling at towing speeds but are not truly towing. As a final pass at removing false positives, periods of effort that sum to less than a total of two hours (trip.eff.hrs.min) within periods until the duration between consecutive tows on a vessel exceeds 24 hours (trip.hrs.brk) are removed, thought to be too little effort to occur within what likely constitutes a trip. These thresholds are apparent upon examining the resulting distributions of tow times, indicating these periods classified as fishing speeds are likely artifacts of the data rather than true fishing activity.

Once effort data have been isolated from the complete cELB dataset, a spatial join with the shapefile below (Figure 3) is performed to assign a 1-21 statistical zone and depth zone (0-10, 10-30, 30+) to each ping (row).

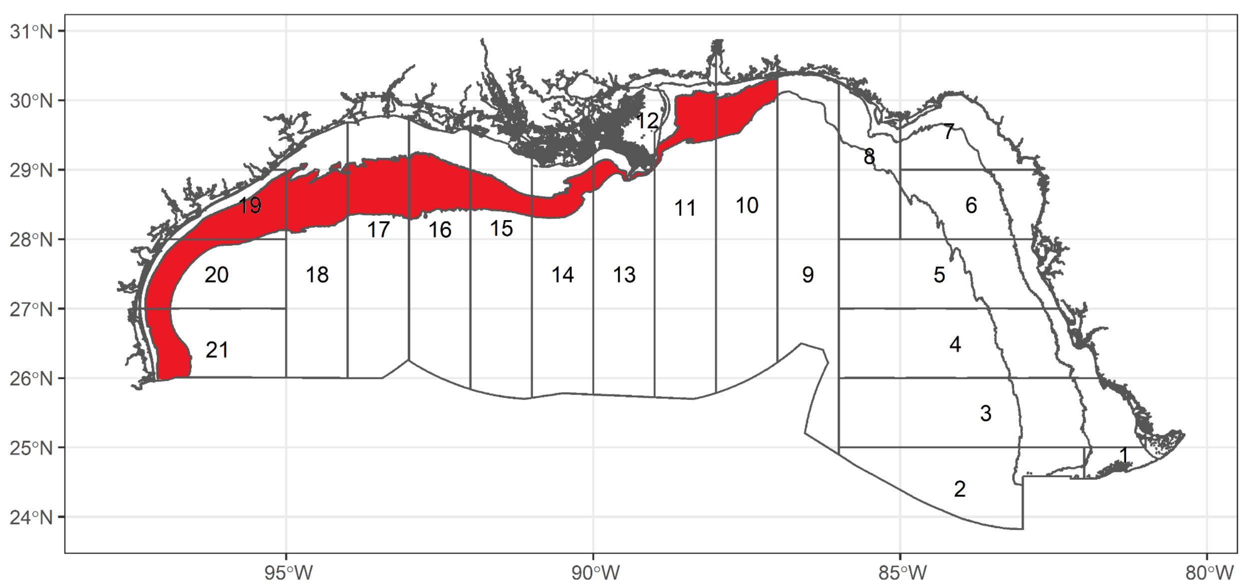


Figure 3: GOM trip ticket statistical grid, with offshore lines delineating 10 and 30 fathom boundaries. Red snapper restricted area is shaded in red.

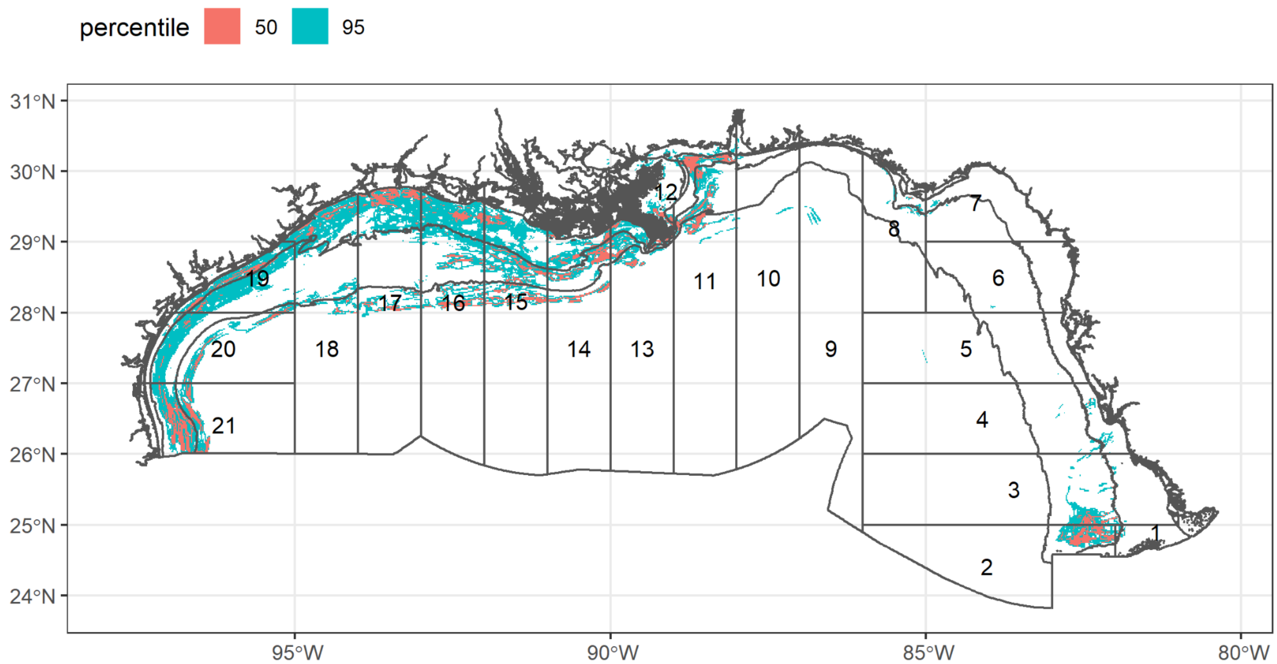


Figure 4: cELB estimated effort (2018-2020) classified by percentile (top 50% of effort falls in red areas, top 95% falls in combination of red and blue).

Scaling to total fleet

A survey-weighting approach based on aggregated trip ticket reported non-bait landings (see Data Sources, #3) is used to scale up estimated effort from vessels with cELB devices to that of the total offshore fleet. Landings have been observed to correlate particularly strongly with effort in the GOM shrimp fishery (r = 0.94), especially in relation to reef fisheries (Chollett, unpubl.).

1. Stratum definitions

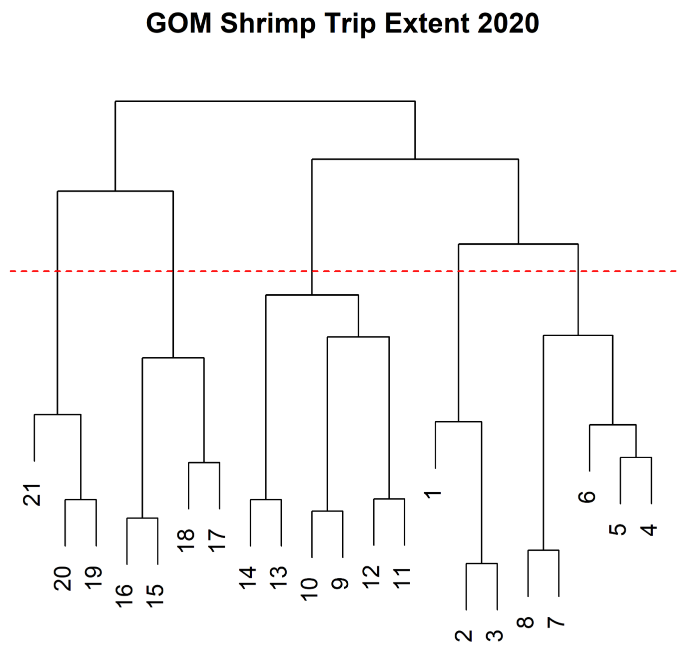


Figure 5: Dendrogram from hierarchical cluster analysis of spatial extent of trips (2020).

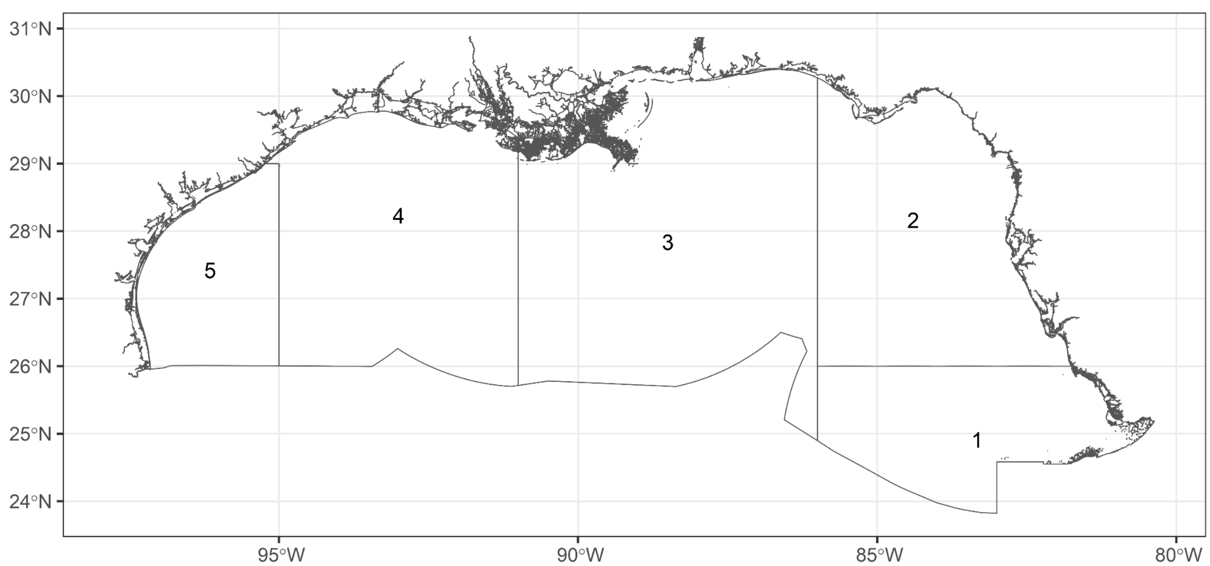


Figure 6: Map of resulting 5 aggregate areas (stat zones: 1: 1-3, 2: 4-8, 3: 9-14, 4: 15-18, 5: 19-21).

Strata are defined by the combination of quadrimester (months 1-4, 5-8, and 9-12) and area. As shrimp trips tend to be spatially extensive and span multiple statistical zones, a suitable aggregation of zones is needed to define broader areas used for scaling. To help identify these areas quantitatively, a hierarchical cluster analysis using Ward's method was performed to identify patterns in the spatial extent of trips based on an indicator matrix with trips as rows and statistical zones as columns based on cELB activity. The result for 2020 is presented in Figure 5. When the tree is cut to obtain four branches (dashed red line), the aggregations result in area 1: 1-3, area 2: 4-8, area 3: 9-14, area 4: 15-18, area 5: 19-21 (Figure 6). This results in a total of 3 quadrimester x 5 area = 15 time/area strata for Penaeids.

For royal reds, scalars are collapsed into only two area groups (1: areas 1-2 and 2: areas 3-5). Quadrimester is not included due to fishery dynamics and data limitations.

2. Species specific effort

ELB effort is first classified into one of two species groups according to depth: Penaeids (brown/pink/white) and royal reds. Any effort occurring shallower than 75 fathoms (450 feet) (depth.penaeid.ft) is considered to be directed toward Penaeids; deeper than 100 fathoms (600 feet) is considered directed toward royal reds.

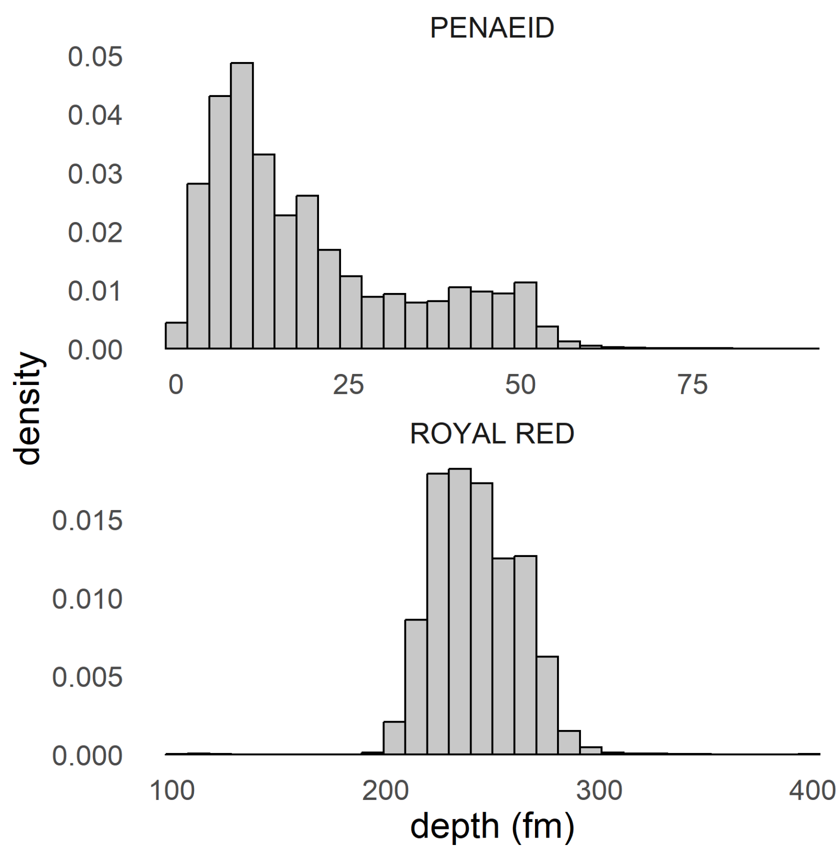


Figure 7: Distribution of 2020 ELB effort pings by depth. Note different x and y scales.

While other species (e.g., Atlantic seabob, rock shrimp, and roughneck shrimp) co-occur at depths with the three major Penaeids (brown/pink/white), they comprise only a negligible proportion of non-royal red landings (~0.4%), and are therefore not considered to be directly targeted when computing effort for the three major Penaeids. A table of total offshore landings and weight proportions for all Gulf shrimp species from 2014-2022 is given below.

ITIS CODE	COMMON NAME	TAIL WEIGHT (lbs.)	PERCENT
551570	SHRIMP, NORTHERN BROWN	312,886,616	52.2
551680	SHRIMP, NORTHERN WHITE	258,837,465	43.2
551574	SHRIMP, NORTHERN PINK	23,841,794	4.0
095966	SHRIMP, ROYAL RED	1,040,249	0.2
095750	SHRIMP, ATLANTIC SEABOB	939,506	0.2
095601	SHRIMPS, PENAEOID	887,021	0.1
096027	SHRIMPS, ROCK	865,118	0.1
096028	SHRIMP, ROCK	498,231	0.1
095647	SHRIMP, ROUGHNECK	Conf.	0

Penaeid effort (<100 fm) is then apportioned among brown, pink, and white shrimp according to the observed fishery independent catch distribution in weight of each species from SEAMAP Groundfish Survey data (1987-2022, >= 5 fm depth), such that individual species efforts are additive and sum to total effort. SEAMAP data are considered preferable to trip ticket landings data for apportioning effort due to more accurate species identification in addition to the higher spatial (stat zone) and temporal (30 minute tow) resolution of the data. This allows stratification of effort into categories known to account for significant variability in relative species abundances. These relative abundances are presented graphically in Figure 8 by season, stat zone, and time of day. Visual exploration revealed these categories to be highly explanatory within any given year, with relatively minor variation in relative distribution observed across years.

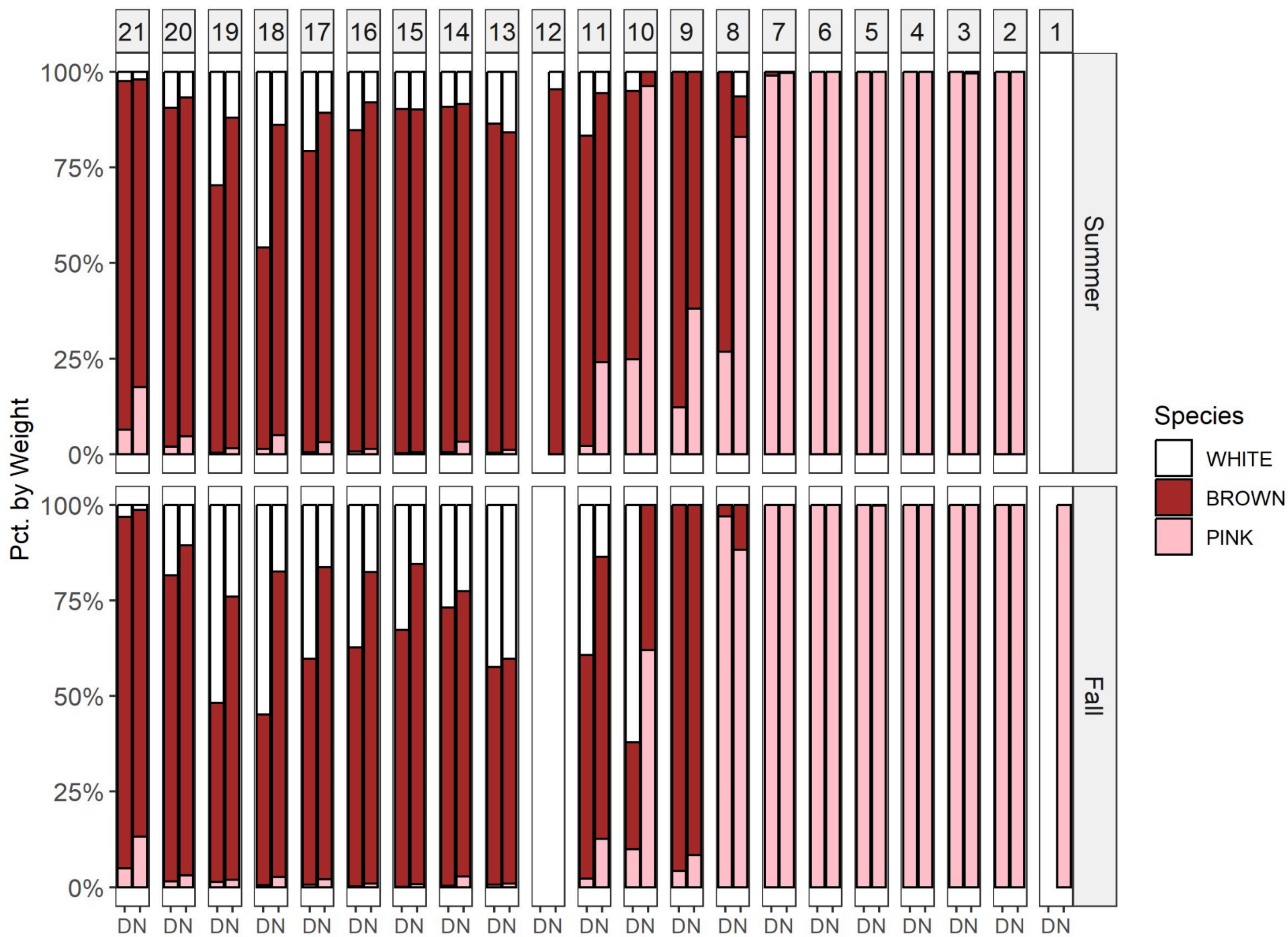


Figure 8: Distribution of GOM SEAMAP shrimp catch by season, stat zone, and time of day.

Since zones 1-7 (West FL shelf) consist of virtually 100% pink shrimp, all Penaeid effort in these zones is classified as such, and the remaining methodology outlined in this section pertains to the allocation of effort by species within zones 8-21.

Next, within zones 8-21, binomial GAM smooths were fit within season and time of day to SEAMAP tow level proportions of brown shrimp catch by depth in order to identify a suitable depth threshold beyond which Penaeid effort could be reasonably classified as directed toward browns. Figure 9 shows this threshold to be approximately 30 fathoms. Thus, any effort in zones 11-21** occurring between 30-100 fathoms is assumed to be directed toward brown shrimp (**Penaeid effort in zones 3-10 deeper than 30 fm is removed in a previous step).

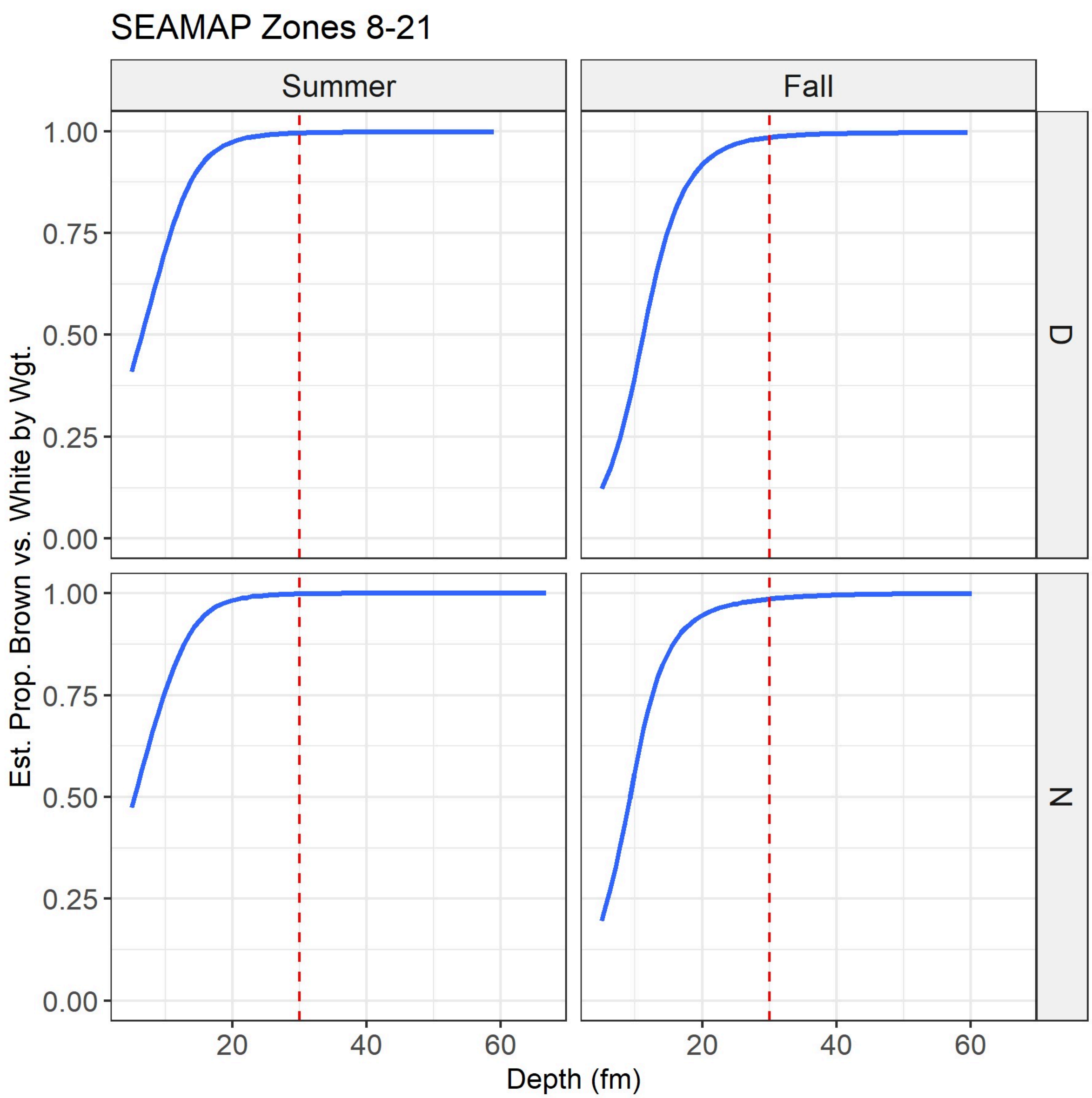


Figure 9: Expected proportion of brown shrimp catch by depth within season and time of day. Dashed vertical lines denote 30 fm threshold.

A limitation of this datasource is that SEAMAP surveys do not occur shallower than 5 fm, meaning white shrimp may potentially be underrepresented in the 0-10 fm zone.

The remaining effort in stat zones 8-21*, less than 30 fathoms (Figure 10) is proportioned by species statistically as outlined below. There was limited evidence that these proportions varied statistically over the duration of the SEAMAP time series after accounting for the main effects of season, stazone, depth, and time of day (p = 0.064, binomial GLM; 48% deviance explained).

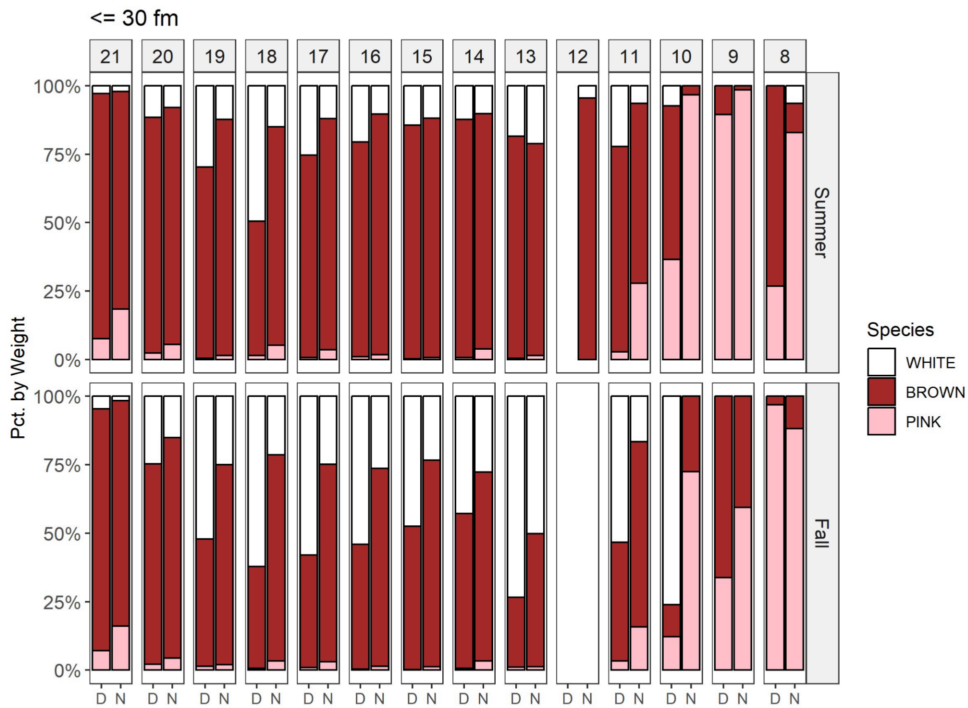


Figure 10: Distribution of SEAMAP shrimp catch less than 30 fm by season, stat zone, and time of day.

Relative weight proportions are stratified by season (Summer/Fall), stat zone (1-21*), depth zone (0-10 fm/10-30 fm), and time of day (D: sunrise-sunset/N: sunset-sunrise), such that the proportions for the three species in each of these combinations sum to one (Eq. 1).

$$p_{season/stat\ zone/depth/time/sp} = catch_{season/stat\ zone/depth/time/sp} / \sum catch_{season/stat\ zone/depth/time} \quad (1)$$

*Due to limited data in stat zones 8-10, catch in these areas is pooled into a single zone when calculating relative proportions. Likewise, due to limited data in stat zone 12, it is pooled with zone 11. If a species proportion within any stratum combination falls below 1%, it is zeroed out and the proportions for the remaining species are recalculated based on their respective total weights. This helps prevent excessively small effort estimates within strata in which a given species is not likely targeted.

Each effort ping is given a Day/Night classification as falling between sunrise and sunset based on its exact location on a given date using the ‘suncalc’ package in R.

As SEAMAP data consist of only summer and fall surveys, for purposes of joining catch proportions to effort, fall is defined as quadrimester 3 (Sep.-Dec.) and summer as quadrimester

2 (May-Aug.). In lieu of existing data, an average of the two is taken within the remaining strata and defined as winter quadrimester 1 (Jan.-Apr.) (Eq. 2), following the logic that the winter months occur between the fall and summer surveys.

$$p_{winter/stat\ zone/depth/time/sp} = (p_{fall/stat\ zone/depth/time/sp} + p_{summer/stat\ zone/depth/time/sp}) / 2 \quad (2)$$

3. Sampling weight estimation

Initial sampling weights (i.e., scaling factors) for effort in each stratum are estimated using the ratio of total offshore landings (defined by reported subarea and bottom otter trawl gear type if subarea is unknown, i.e. ‘8888’) among vessels known to be holding a federal SPGM permit to offshore landings among the subset of vessels with fishing activity identified by a cELB device in a given calendar year (Eq. 3). Landings reported by cELB vessels in a given semester of the calendar year (Jan-Jun; Jul-Dec) with no corresponding fishing activity in the same period are removed from the denominator of the calculation to avoid underestimating effort based on cELB data that may not have been received for a complete year. Finite population correction factors for variance calculations are computed similarly as the number of cELB vessels with landings over the number of total vessels with landings (Eq. 4).

$$w1_{area/quad/spg} = total\ landings_{area/quad/spg} \ / \ ELB\ landings_{area/quad/spg} \quad (3)$$

$$fpc_{area/quad/spg} = ELB\ vessels_{area/quad/spg} \ / \ total\ vessels_{area/quad/spg} \quad (4)$$

Final weights are then adjusted to account for potential CPUE differences between ELB and non-ELB vessels based on the average number of nets for each component of the offshore fleet across strata. An analysis conducted by Smith et al. 2023 (unpubl.) using shrimp observer data estimated that all else equal, CPUE is approximately 40% (1.4 times) higher for four net vessels than two net vessels on average. This means that vessels with two net configurations are expected to require 40% more towing time to catch the same amount as a vessel with a four net configuration in a given amount of towing time. If the average number of nets differs between ELB and non-ELB vessels (see Figure A6 in Appendix), this could lead to bias when using landings to scale effort if this difference is not accounted for in the scaling process.

Weight corrections are computed by taking the ratio of mean CPUE net scalars, assigned a value of 1.4 for four net vessels and a value of 1 for two net vessels, for active ELB vessels vs. non-ELB vessels reporting landings within each strata (Eq. 5).

$$c_{area/quad/spg} = (\text{mean}(net\ scalar_{ELB\ vessel}) \ / \ \text{mean}(net\ scalar_{nonELB\ vessel}))_{area/quad/spg} \quad (5)$$

The non-ELB landings in the numerator of the weight scalars are then multiplied by these values to ensure ELB and nonELB landings are on equivalent effort scales. This works out algebraically to adjusting the original weights in Eq. 3 (in this case with ELB landings defined as those coming from ELB vessels at any time of the year) according to (Eq. 6).

$w_{area/quad/spg} = w1_{area/quad/spg}\ c_{area/quad/spg} - c_{area/quad/spg} + 1$ (6)

For 2020, these values are as follows:

species	quad	area	w	fpc
PENAEID	1	1	1.99	0.469
PENAEID	1	2	11.43	0.794
PENAEID	1	3	2.40	0.696
PENAEID	1	4	2.19	0.632
PENAEID	1	5	1.91	0.642
PENAEID	2	1	2.56	0.446
PENAEID	2	2	3.75	0.574
PENAEID	2	3	2.46	0.565
PENAEID	2	4	2.08	0.594
PENAEID	2	5	2.10	0.637
PENAEID	3	1	2.55	0.383
PENAEID	3	2	3.54	0.500
PENAEID	3	3	2.97	0.542
PENAEID	3	4	2.27	0.599
PENAEID	3	5	2.00	0.622
ROYAL RED	NA	1-2	1.00	1.000
ROYAL RED	NA	3-5	1.10	0.700

4. Total effort computation

A stratified 1-stage cluster sampling design is used to obtain estimates of effort totals and variances using the sampling weights as calculated in Step 2. Estimated cELB effort for each vessel is aggregated at the quadrimester/area/species group level and the sampling weights are applied accordingly to obtain totals (Eq. 7).

$total\ federal\ fleet\ effort = \sum ELB\ effort_{area/quad/spg}\ x\ w_{area/quad/spg}$ (7)

Variance of the estimated total is calculated using the standard formula for a stratified one-stage cluster sample (Lohr 2022), with vessels representing primary sampling units. In the following formula, y represents vessel effort, w represents the sampling weight (Eq. 3), f represents the finite population correction (Eq. 4), n represents the number of PSUs (vessels) in a strata, and $_h$ denotes a given quadrimester/area stratum.

$$\hat{V}(\hat{Y}) = \sum_{h=1}^{H} \hat{V}_h(\hat{Y})$$

where, if $n_h > 1$, then

$$\hat{V}_h(\hat{Y}) = \frac{n_h(1-f_h)}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi\cdot} - \bar{y}_{h\cdot\cdot})^2$$

$$y_{hi\cdot} = \sum_{j=1}^{m_{hi}} w_{hij}\, y_{hij}$$

$$\bar{y}_{h\cdot\cdot} = \left(\sum_{i=1}^{n_h} y_{hi\cdot} \right) / n_h$$

This design is implemented in the R *survey* package (Lumley 2023) with the following code, where “cell” is a concatenation of quadrimester and area:

```
design = box_tow_effort_grp %>%
    group_by(species_grp, quadrimester, area,
            region = ifelse(StatZone %in% 10:21, "West", "East"), DepZone, VSBN) %>%
    summarise(days = sum(days)) %>%
    inner_join(sample_wgts,
       by = c("species_grp", "quadrimester" = "quadrimester_land", "area"))%>%
    mutate(cell = paste(species_grp, quad_rr, area_rr)) %>%
    svydesign(~VSBN, strata = ~cell, weights = ~w, fpc = ~fpc, data = ., check.strata = FALSE)
```

The following code can be used to produce estimates of totals and variances within specified domains (here quadrimester/area):

```
svyby(~days, ~species_grp + quad_rr + area_rr, design = ., svytotal, vartype = "var",
keep.names = FALSE)
```

Further details of statistical computations can be found here:
https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.4/statug/statug_surveymeans_details06.htm#statug.surveymeans.vartotaldetails

Confidence intervals around estimates (in units of 24 hour days) are calculated from the resulting standard errors (i.e., square root of the variances) multiplied by the desired quantile of a t-distribution with n-h degrees of freedom (number of vessels minus number of strata). Variances for brown/pink/white shrimp are divided proportionally within strata according to the estimated effort directed toward each species.

5. Offshore state landings

Penaeid landings reported in offshore waters (beyond the COLREG line) reported by vessels not in possession of a federal SPGM permit are assumed to occur in state waters (i.e, 0-3 or 0-9 nm offshore). These landings typically constitute a small percentage of the offshore total (~3.5%

in 2020). For purposes of scaling effort, these landings are allocated among depths in proportion to the area of offshore state water encompassed by each depth zone within each of the four aggregate areas (with the majority falling in the 0-10 fathom zone). For example, to allocate within a given area:

$$\textit{state landings}_{area/depth} = \textit{state landings}_{area} * \textit{state water area}_{area/depth} / \sum \textit{state water area}_{area} \quad (8)$$

As the cELB sampling frame consists only of vessels with federal SPGM permits, the total federal fleet effort and variance estimates as calculated in the previous section are adjusted upward as follows:

$$\textit{state correction} = (\textit{federal landings} + \textit{state landings}) / \textit{federal landings} \quad (9)$$

$$\textit{total effort} = \textit{total federal effort} * \textit{state correction} \quad (10)$$

$$var(\textit{total effort}) = var(\textit{total federal effort}) * \textit{state correction}^2 \quad (11)$$

6. Allocation of total landings and effort among depth zones

Since the original distribution of cELB effort data is preserved, scaled effort and landings can be allocated among depth zones and statistical zones (or other custom groupings) proportionally to the observed distribution of cELB effort. For example, to allocate landings by depth zone within each area and quadrimester:

$$\textit{landings}_{area/quad/depth/sp} = \textit{landings}_{area/quad} * \textit{effort}_{area/quad/depth/sp} / \sum \textit{effort}_{area/quad} \quad (12)$$

For allocation of landings, this assumes constant offshore CPUE among the depth zones, as depth information is not reported on trip tickets, though this assumption seems to be reasonably well met based on analyses conducted by the SEFSC Social Science Research Group.

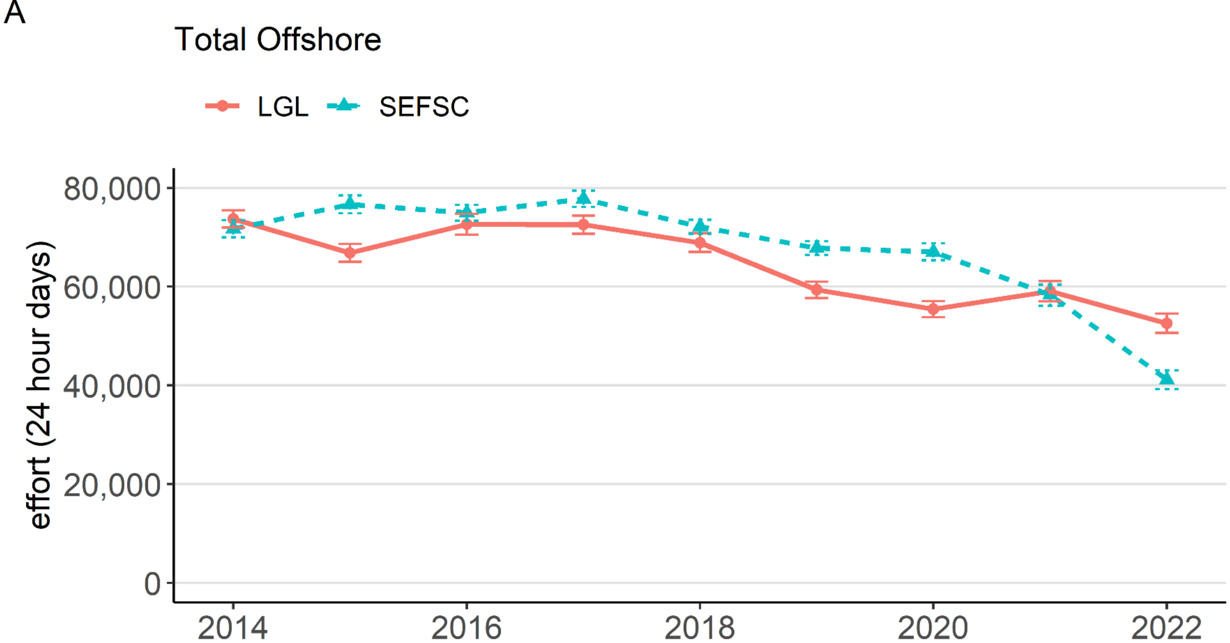
CPUE calculations can then proceed accordingly for given times/areas/depths:

$$CPUE = \sum \textit{landings}_{region/year/depth/sp} / \sum \textit{effort}_{region/year/depth/sp} \quad (13)$$

Results

Total effort estimates from 2014-2022 generally track those historically estimated by LGL. A further breakdown of estimated differences by strata is provided in Appendix 2.

A



B

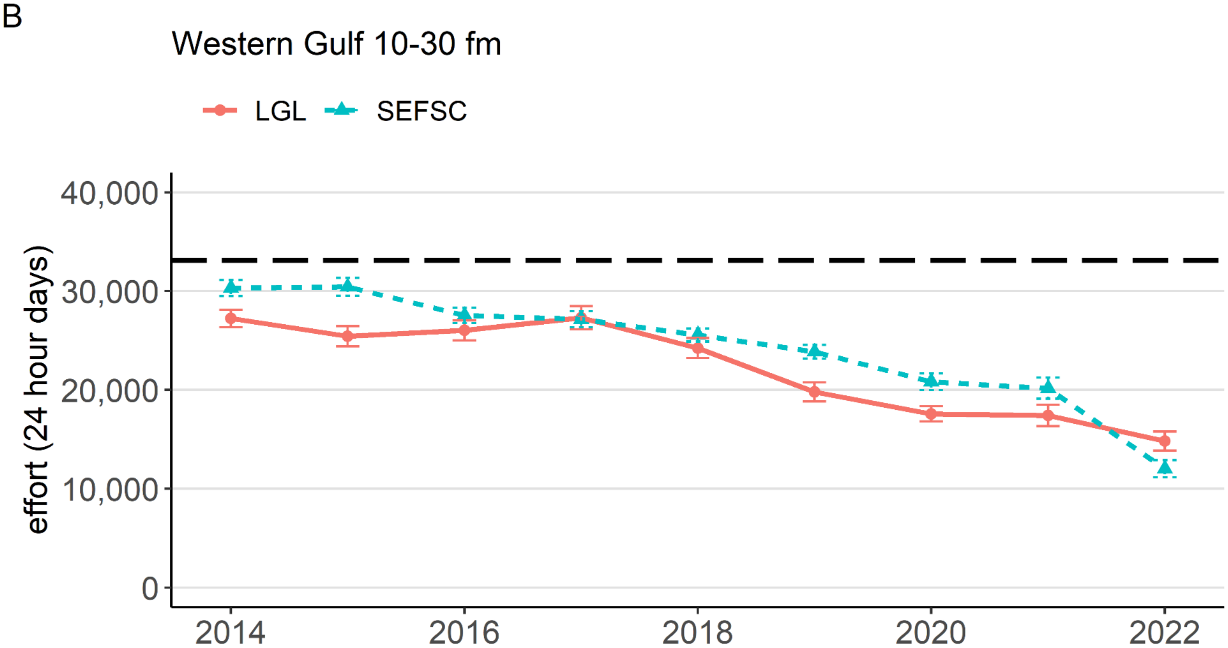


Figure 12: Comparisons of LGL and SEFSC algorithm results for total offshore effort (A) and effort within the red snapper restricted area (defined as depths between 10 and 30 fathoms in the Western GOM, areas 10-21). (B).

The following table and Figure 11 show example products that will be generated annually, including total Gulf-wide effort estimates, estimates broken down by strata, and the overall spatial distribution of effort from devices with cELB estimated effort.

Table 1: 2020 Gulf-wide Offshore Estimates

subarea	tow days	lower	upper	landings	CPUE
offshore	67,048	65,305	68,790	68,937,474	1,028

Table 2: 2020 Stratified Offshore Effort Estimates

region	depth zone	species	tow days	lower	upper	landings	CPUE
East	0-10 fm	BROWN	93	44	143	65,684	703
East	10-30 fm	BROWN	151	74	228	93,464	619
East	0-10 fm	PINK	2,460	2,186	2,733	504,063	205
East	10-30 fm	PINK	5,957	5,512	6,401	7,631,994	1,281
East	30-100 fm	PINK	1	0	1	825	1,269
East	0-10 fm	WHITE	79	32	126	57,712	732
East	100+ fm	ROYAL RED	59	32	85	52,801	899
West	0-10 fm	BROWN	7,173	6,467	7,879	8,179,384	1,140
West	10-30 fm	BROWN	17,533	16,757	18,308	18,383,386	1,049
West	30-100 fm	BROWN	14,043	13,224	14,861	14,386,638	1,024
West	0-10 fm	PINK	735	505	965	838,476	1,141
West	10-30 fm	PINK	891	715	1,067	957,957	1,075
West	0-10 fm	WHITE	15,416	14,417	16,415	15,548,776	1,009
West	10-30 fm	WHITE	2,411	2,122	2,700	2,197,446	912
West	100+ fm	ROYAL RED	47	21	72	38,868	833

Example table displaying 2020 fleet wide effort breakdown by region, depth, and species.

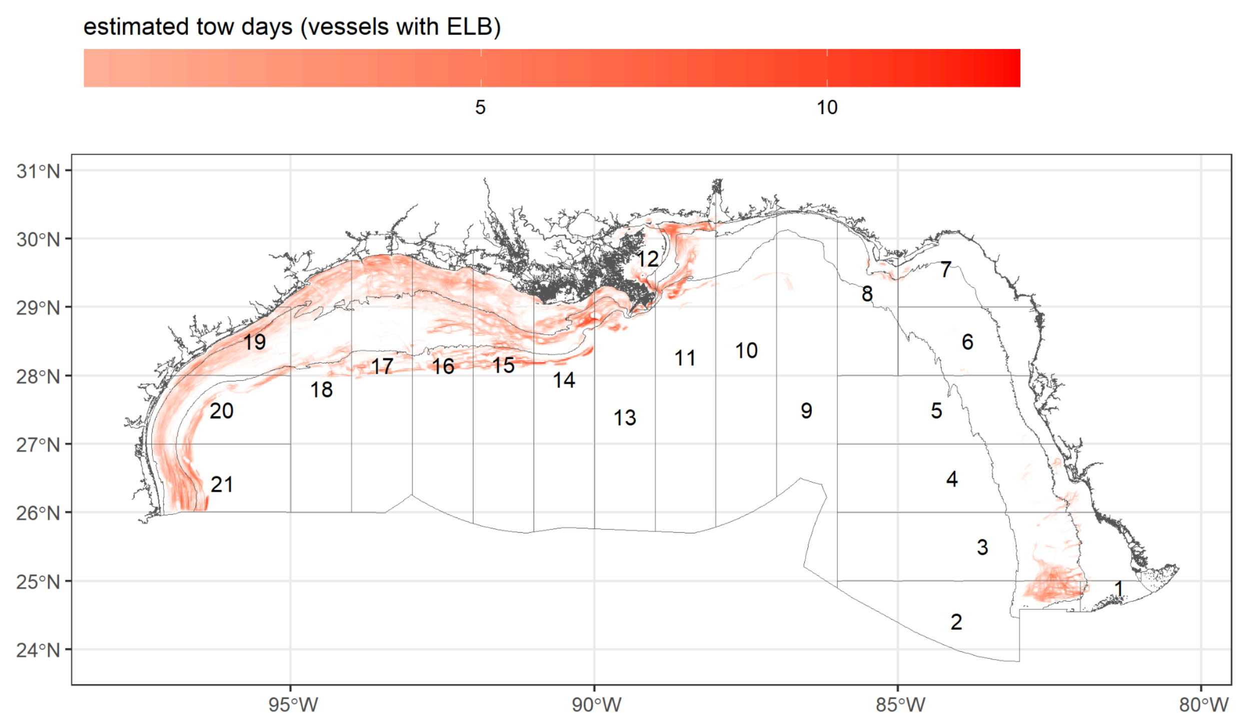


Figure 11: Heat map of ELB 24 hour tow days at 1 minute resolution (2020).

A breakdown of effort by species through time shows effort directed toward white shrimp is generally highest in shallower waters of the Northern Gulf, with brown shrimp effort dominating in deeper waters. Pink shrimp effort is primarily concentrated around the Tortugas in the 10-30 fathom zone.

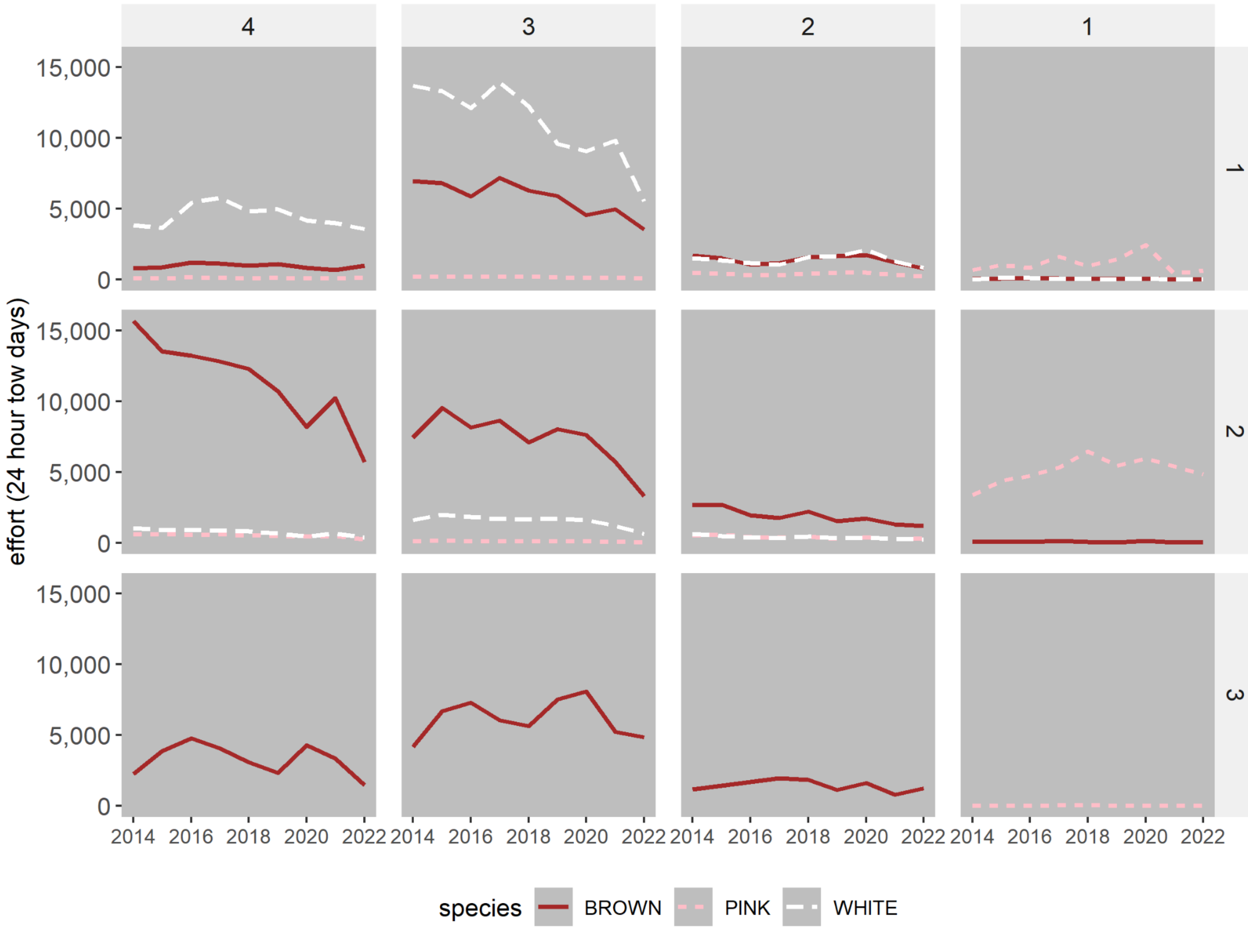


Figure 13: Time series of offshore species-specific effort estimates grouped by historical area (horizontal axis, 1 = stat zones 1-9, 2 = 10-12, 3 = 13-17, 4 = 18-21) and depth zone (vertical axis; 1 = 0-10 fm, 2 = 10-30 fm, 3 = 30+ fm).

Gulf-wide total effort trends by species remain similar back in time to 1981, with the majority of effort directed toward browns followed by whites and pinks.

Figure 14: Complete time series of offshore species-specific effort estimates, 1981-2022. Species breakdowns for 1981-2014 have been back-calculated by applying the mean annual stratum-level (quarimester/area/depth zone) species effort proportions from years 2015-2022 to the historical stratum-level total effort.

Limitations

The methods described here are only as good as the data they depend on. Incomplete ELB data (e.g., boxes not functioning or turned on for complete trips) or missing/inaccurate landings reports have the potential to create bias in estimates. Additionally, while data collection currently depends on the receipt of physical memory chips, we equally depend on all vessels with ELB units to submit complete ELB data on an annual basis in order to produce accurate estimates.

Future Work

Observer CPUE by species, area, depth, time of day, and number of nets can be used to further refine how landings are apportioned according to effort, rather than assuming constant CPUEs within the existing strata. This has implications for how inshore effort will be apportioned among the Penaeid species, since cELB and SEAMAP data are not collected inshore and the effort allocation process relies strictly on inshore landings reported on trip tickets. Additionally, use of an independent observer CPUE, when multiplied by present effort estimates, provides a means of estimating the degree of discrepancy between estimated landings and trip ticket reported landings.

Work is ongoing to calibrate historical port agent estimates of effort back in time to be comparable with the current estimation method based on years with overlapping ELB and port agent coverage (2006-2014). This may change the magnitude of historical estimates.

References

Lohr, S.L. (2022) Sampling: design and analysis. 3rd ed. CRC press.

Lumley, T. (2023) survey: analysis of complex survey samples. R package version 4.2.

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.

R Core Team. (2023) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Renfro W.C, Brusher, H.A. (1982) Seasonal abundance, size distribution, and spawning of three shrimps (*Penaeus aztecus*, *P. setiferus*, and *P. duorarum*) in the northwestern Gulf of Mexico, 1961-1962. NOAA Tech Memo NMFS-SEFC-94. NOAA, Silver Spring, MD, USA. 47 pp.

Appendix 1: Evaluation of cELB coverage

Figure A1: cELB coverage of offshore Penaeid landings by stratum over time as of November 2023. Panels represent areas and lines represent quadrimesters.

Gulf-wide offshore Penaeid landings coverage over time:

Year	2014	2015	2016	2017	2018	2019	2020	2021	2022
Proportion cELB Landings	0.529	0.528	0.575	0.573	0.614	0.584	0.489	0.361	0.330

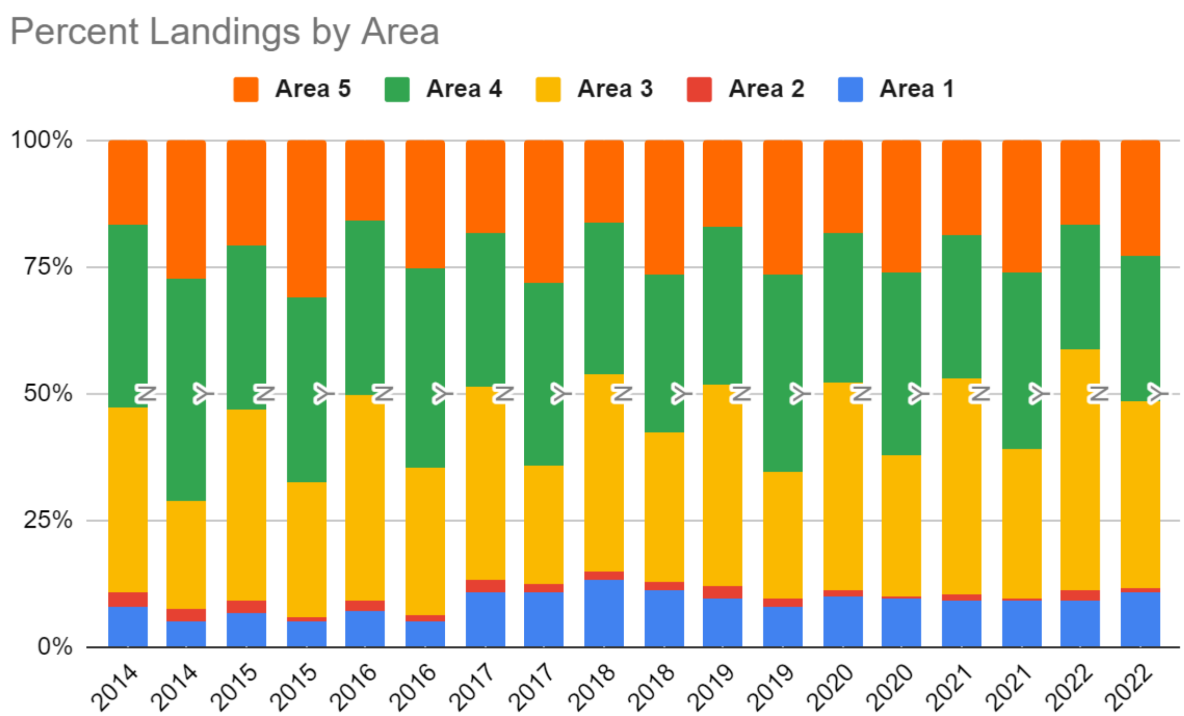


Figure A2: Percent landings by area for ELB (Y) vs. nonELB (N) vessels.

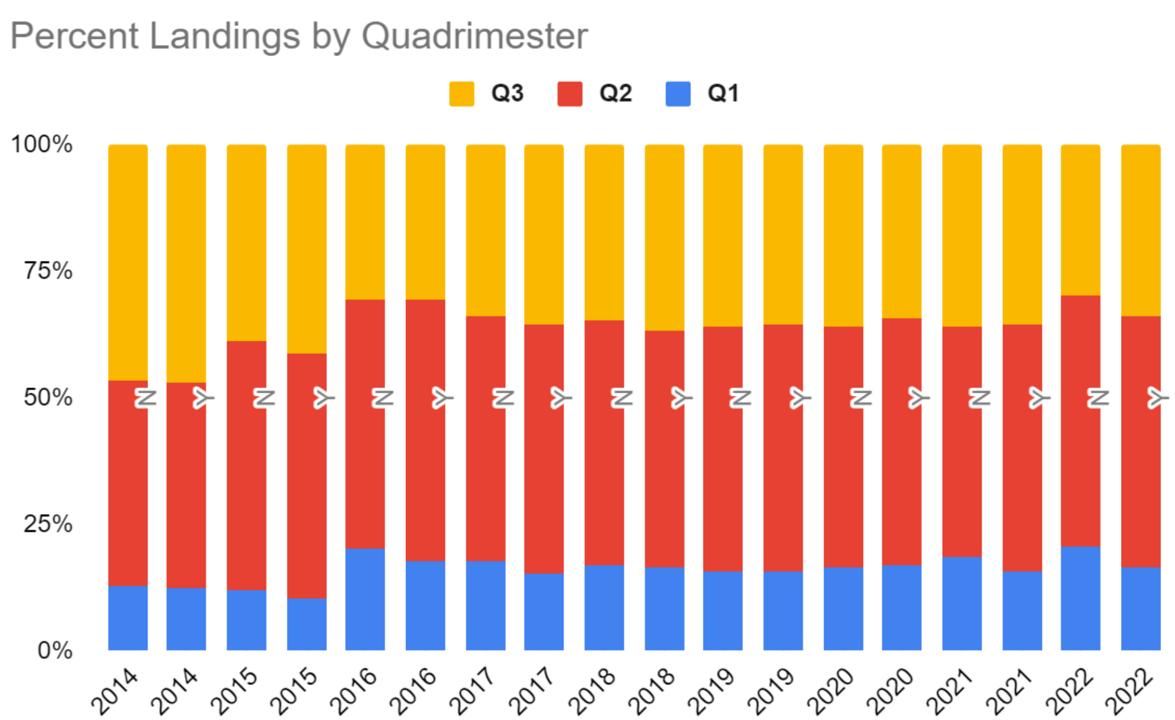


Figure A3: Percent landings by quadrimester for ELB (Y) vs. nonELB (N) vessels.

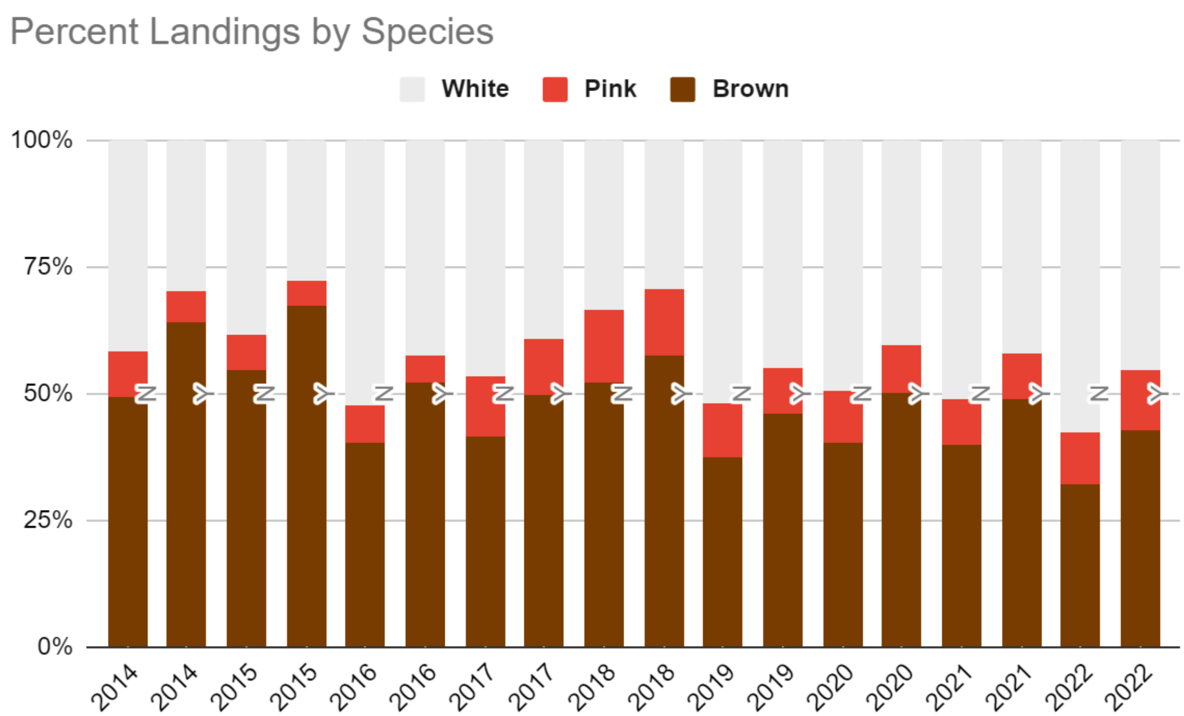


Figure A4: Percent landings by species for ELB (Y) vs. nonELB (N) vessels.

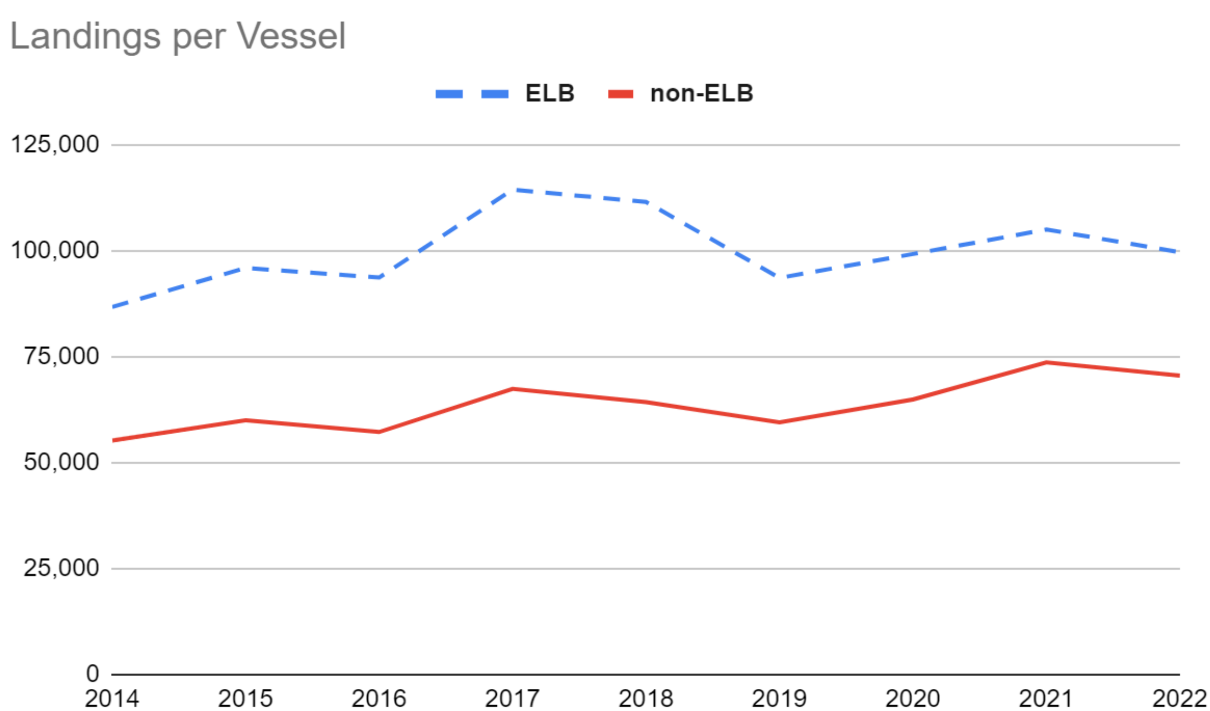


Figure A5: Average landings per vessel for ELB vs. nonELB vessels.

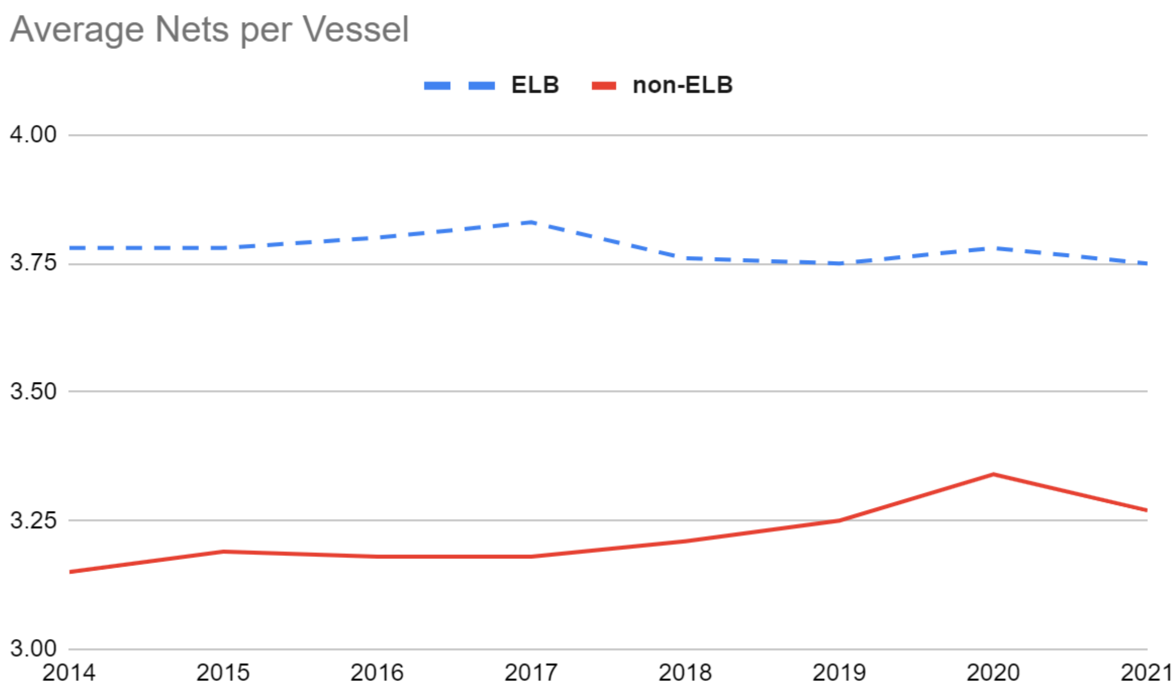


Figure A6: Average number of nets per vessel for ELB vs. nonELB vessels based on Annual Landings and Gear Survey.

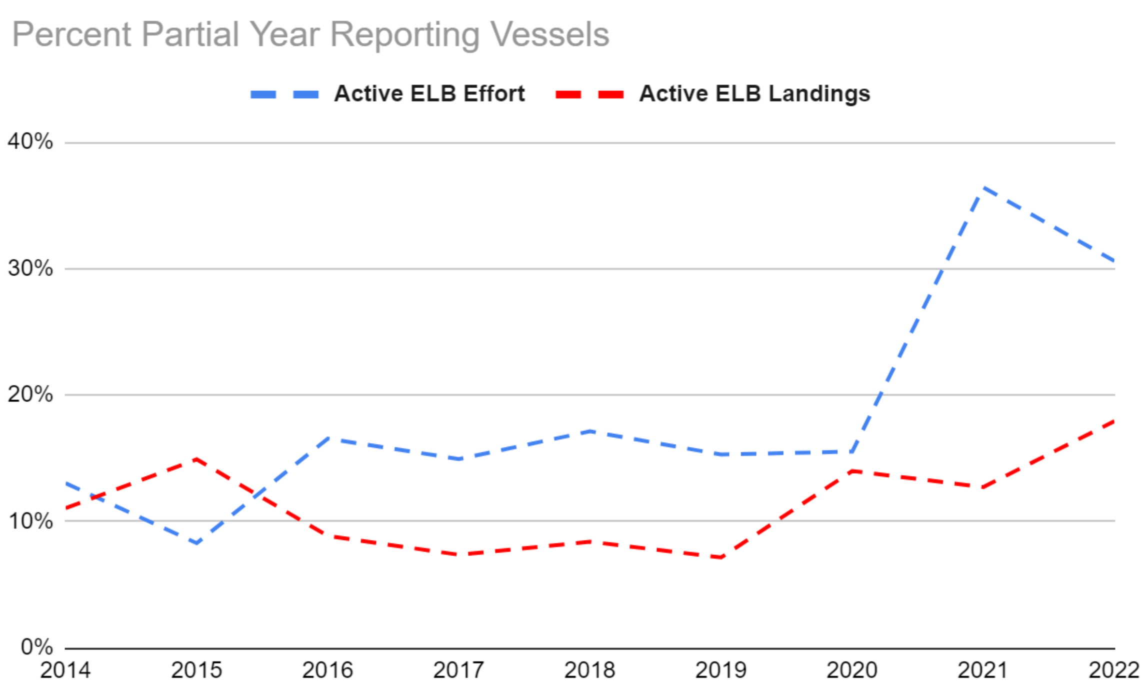


Figure A7: Proportion active ELB vessels reporting partial year (only Jan-Jun or Jul-Dec) landings and effort. Uptick in partial year effort relative to partial year landings beginning in 2021 suggests incomplete retrieval of chip data.

Appendix 2: 2006-2014 LGL Effort Calibration

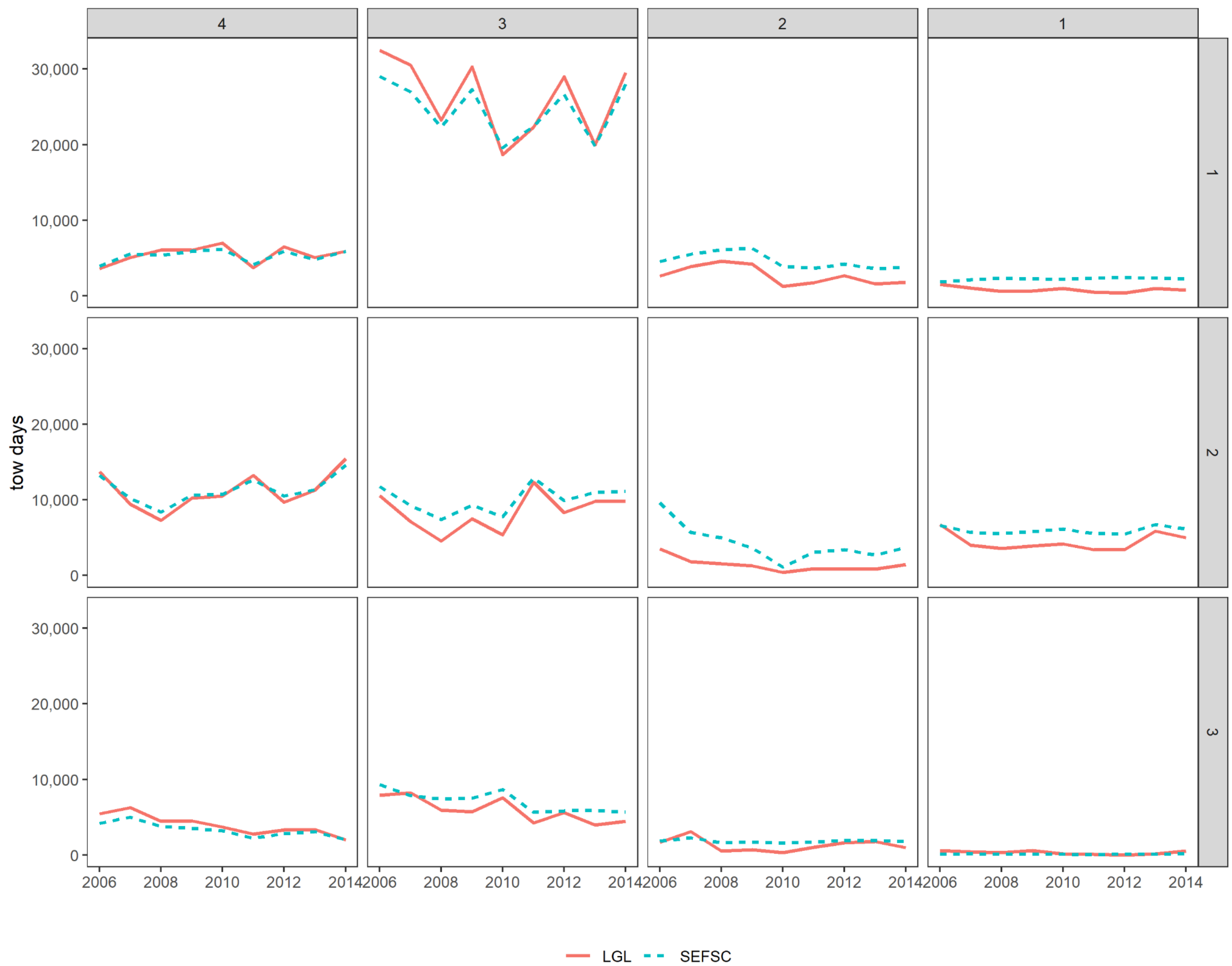


Figure A8: 2006-2014 LGL effort estimates (solid red lines) calibrated to SEFSC estimates (dashed green lines) based on linear model fit to estimates from overlapping years (2015-2020). Estimates are grouped by historical area (horizontal axis, 1 = stat zones 1-9, 2 = 10-12, 3 = 13-17, 4 = 18-21) and depth zone (vertical axis; 1 = 0-10 fm, 2 = 10-30 fm, 3 = 30+ fm).

```
formula = new_effort ~ effort * area * tri * dpz

Type III ANOVA Table:
                      Sum Sq  Df F value    Pr(>F)    
(Intercept)          3797834   1 24.6384 1.927e-06 ***
effort                983903   1  6.3830 0.0126026 *  
area                 2251426   3  4.8687 0.0029606 ** 
tri                  1761937   2  5.7153 0.0040877 ** 
dpz                  4779892   2 15.5047 7.974e-07 ***
effort:area          5711819   3 12.3518 3.088e-07 ***
effort:tri            500571   2  1.6237 0.2007523    
area:tri             3955188   6  4.2765 0.0005369 ***
effort:dpz           2161686   2  7.0119 0.0012418 ** 
area:dpz             2876695   6  3.1104 0.0067828 ** 
tri:dpz              1518251   4  2.4624 0.0478721 *  
effort:area:tri      1558420   6  1.6850 0.1286925    
effort:area:dpz      3018355   6  3.2636 0.0048739 ** 
effort:tri:dpz        817459   4  1.3258 0.2631786    
area:tri:dpz         4483717  12  2.4240 0.0067811 ** 
effort:area:tri:dpz  2347471  12  1.2691 0.2430262    
Residuals           22196604 144                      

Residual standard error: 392.6 on 144 degrees of freedom
Multiple R-squared:  0.9748,    Adjusted R-squared:  0.9624
F-statistic: 78.52 on 71 and 144 DF,  p-value: < 0.001
```