Shore-based and boat-based fishing surveys in Guam, the CNMI, and American Samoa: survey design, expansion algorithm, and a case study

Shore-based and boat-based fishing surveys in Guam, the CNMI, and American Samoa: survey design, expansion algorithm, and a case study

Hongguang Ma [1a], Toby Matthews [2a], Marc Nadon [1,3], and Felipe Carvalho [1]

[1] Pacific Islands Fisheries Science Center
    National Marine Fisheries Service
    1845 Wasp Boulevard
    Honolulu, HI 96818
[2] University of Hawaii at Manoa
    Hawaii Institute of Marine Biology
    PO Box 1346
    Kaneohe, Hawaii 96744

[3] Joint Institute for Marine and Atmospheric
    Research
    University of Hawaii
    1000 Pope Road
    Honolulu, Hawaii 96822

[a] These authors contributed equally as lead
    authors to this work.

U.S. Department of Commerce
Gina Raimondo, Secretary

National Oceanic and Atmospheric Administration
Richard W. Spinrad, Ph.D., NOAA Administrator

National Marine Fisheries Service
Janet Coit, Assistant Administrator for Fisheries

Table of Contents

List of Tables

List of Figures

Executive Summary

This document describes annual catch and effort estimations from shore-based and boat-based fishing surveys in Guam, the Commonwealth of the Northern Mariana Islands (CNMI), and American Samoa. The catch and effort estimates in these three United States Pacific island territories were historically produced with a series of expansion scripts written in Visual FoxPro (VFP). Recently, scripts in the programming language R were developed to replace and improve upon the VFP scripts. This report describes the current survey design and expansion methods, including some recent modifications incorporated in the R scripts.

The annual catch is estimated as the product of catch-per-unit-effort (CPUE, from the catch rate survey) and annual fishing effort (from the fishing effort survey). The shore-based effort and catch rate surveys each utilize a "roving survey" design. During the roving catch rate survey, encountered fishers are interviewed to gather data on fishing methods, hours fished, and fish caught. In the roving effort survey, accessible shorelines are visited to record active fishing events, characterizing fishing methods and gear counts. The shore-based survey is stratified by day type (weekday vs. weekend) and shift (different periods in a day) in all three territories. The shore-based CPUE and effort estimates, and thus catch estimates, are made separately for each fishing method.

The boat-based survey is mainly an access point survey by design. Catch rate and effort surveys are conducted at major ports, and the surveys are stratified by day type (in all three territories) and port (except for American Samoa). The catch rate and effort estimates are made separately for different fishing methods and charter statuses (charter fishing vs non-charter fishing). As for the shore-based survey, total catch is then estimated as the product of CPUE and fishing effort.

We use the boat-based survey in Guam as a case study to describe how effort and catch rate estimates for different fishing methods at different ports are combined to estimate total catch. Trolling and bottomfishing are the most common fishing methods on Guam, and trolling accounts for 80% of the boat-based catch. Non-charter fishing dominates the catch, contributing approximately 90% and 95% of the total catch for trolling and bottomfishing, respectively. The three sampled ports on Guam account for 90% of the total catch. Interview pooling is used when insufficient interviews are available for an estimation domain, but was needed for no more than approximately 10% of the non-charter trolling and bottomfishing domains at the three sampled ports. The non-charter bottomfishing method was used to detail the expansion steps for producing the total catch for a highly targeted deep bottomfish species.

Sensitivity analyses were conducted to evaluate the impact of the interview pooling algorithm and selection of representative ports on catch estimates. There is little difference in the total catch across four selected pooling algorithms or four different representative port selection scenarios. This was unsurprising given that interview pooling is rarely used for the estimation domains with the greatest catch and unsampled ports, which are affected by the selection of representative ports, contribute the minority (~10%) of the total catch. Interview pooling choices and representative port selections had a greater impact on species-level catch than total catch, but impacts were still minor for the species most caught by trolling and bottomfishing.

Introduction

Small-scale nearshore fisheries in the United States territories of Guam, the Commonwealth of the Northern Mariana Islands (CNMI), and American Samoa target reef, bottom, and nearshore pelagic species. These fisheries are of fundamental importance to subsistence fishers, critical in local cultural traditions, and provide a source of income and recreation to the local population (Dalzell et al. 1996).

Guam is the largest and southernmost of the Mariana Islands (Figure 1). It has a potential coral reef ecosystem habitat area above 100 fathoms (183 m) depth of approximately 276 km^2, including offshore banks. Of this area, 203 km^2 are associated with the island of Guam directly. Prior to European arrival, inhabitants possessed sailing canoes that allowed fishing of nearshore and offshore banks (Allen and Bartram 2008). Fishing around Guam continues to be important to the subsistence needs of the local population, preserving history and identity, and maintaining cultural practices (Allen and Bartram 2008). For Guam's boat-based fisheries, trolling and bottomfishing are the most important fishing methods in recent decades. Trolling accounts for most boat-based trips and the major pelagic landings, including skipjack tuna, mahimahi, yellowfin tuna, wahoo, and blue marlin (Myers 1993). In 2019 there were an estimated 472 boats involved in Guam's pelagic fishery. Since many fishermen sell a portion of their catch, it is difficult to distinguish recreational, subsistence, and commercial fishers (WPRFMC 2020a). Guam's bottomfish fishery can be divided into shallow- and deep-water components according to the targeted depth and species. The shallow-water component (< 150 m) targets emperors, trevallies, snappers, and groupers. The deep-water component targets snappers of the *Pristipomoides* and *Etelis* genera (Myers 1993). The contemporary fishing methods for Guam's inshore fisheries include hook and line, net fishing, spear fishing, hook and gaff, and other methods (Hensley et al. 1993). The most popular inshore fishing method is hook and line.

The CNMI consists of the entirety of the Mariana island chain, excluding Guam and its southern banks, and extends approximately 500 nm in a north-south direction. It is paralleled by a chain of seamounts about 150 nm to the west. Most of the fishing activity occurs around the population centers of Rota, Tinian, and Saipan islands and extends north to the Zealandia Bank, approximately 120 nm north of Saipan. The CNMI has a long history of fishing, with evidence dating back 3,000 years. The recent fisheries developments during the German occupation (1899-1914), Japanese occupation (1914-1945), and the US military occupation (during and after World War II) have been summarized in the recent Mariana Archipelago Fishery Ecosystem Plan (WPRFMC 2020b). As in other Pacific Islands territories, fishing has deep traditions and cultural significance (Hospital and Beavers 2014).

American Samoa is the only U.S. territory south of the equator. Prior to European contact, indigenous fishers of the Samoan Islands fished for subsistence from canoes using pearl shell hooks and sennit lines. By the 1950s, the Samoa fleet had adopted small boats equipped with outboard engines and fished with steel hooks and monofilament lines, but fishing remained mainly a subsistence activity. Surveys conducted in the late 1960s by the American Samoa Office of Marine Resources revealed substantial deep bottomfish resources around the island of Tutuila, and by the early 1970s a small commercial fishery was established. In an attempt to develop local fisheries, two subsidized boat building programs, the dory program in the 1970s

and the alia program in the 1980s, provided fishers with low-cost vessels. In 1982, a fisheries development project aimed at exporting high-priced deep-water snappers to Hawaii resulted in a notable increase in bottomfish landings and revenue (WPRFMC 2020c). A small fishery for bottomfish was developed as a result of these government-funded projects (Craig et al. 1993). In 1988, a decline in the bottomfish fishery occurred as many skilled and full-time commercial fishermen converted to fishing for pelagic species. The bottomfish fishery was recently further impacted by the 2009 tsunami (WPRFMC 2020c). Currently, the largest domestic fishery in American Samoa is the longline fishery (including alia boats) followed by trolling. Pelagic fishing in American Samoa is mostly for commercial purposes (WPRFMC 2020a).

Boat-based and shore-based fishing surveys are conducted by local fisheries agencies in each territory: the Guam Department of Agriculture, Division of Aquatic and Wildlife Resources (DAWR); the CNMI Department of Lands and Natural Resources, Division of Fish and Wildlife (DFW); and the American Samoa Government Department of Marine and Wildlife Resources (DMWR). Data management and programming support for the fishing surveys are provided by NOAA's Western Pacific Fishery Information Network (WPacFIN, Hamm 1993). WPacFIN support began in American Samoa in 1981 and was implemented in Guam and the CNMI shortly thereafter. The survey data have been used to estimate catch-per-unit-effort (CPUE) and annual fishing effort to derive annual catch in each region. The statistical method used to compute these estimates is referred to as the expansion algorithm.

Catch estimates are provided by the territorial agencies in their fiscal year reports and in calendar year reports to the various plan teams and committees under the Western Pacific Regional Fisheries Management Council ("the Council"). The Council and NOAA Fisheries evaluate estimated landings with respect to Annual Catch Limits established under the Revised Magnuson-Stevens Fisheries Management Act for the U.S. Pacific Islands Region (PIR). Both raw and expanded data are shared with WPacFIN for a variety of uses, including annual reports on U.S. PIR fisheries (e.g. Fisheries of the U.S. (NMFS 2021)) and Annual Stock Assessment and Fishery Evaluation Reports (WPRFMC 2020a-c). More specifically, these data have been used for reef fish and bottomfish stock assessments (e.g., Langseth et al. 2019; Nadon 2019).

The purpose of this report is to describe the shore- and boat-based survey designs and present the expansion algorithms used to estimate total catch in Guam, the CNMI, and American Samoa. There are three component objectives, which are covered sequentially in the next three sections:

- Present the survey design and data expansion for the boat- and shore-based surveys in each territory
- Describe empirical survey data and detail its use to produce the total catch and species-level catch estimates, using the Guam boat-based survey and a bottomfish species as an example
- Analyze the sensitivity of the expansion algorithm to two central assumptions, again using the Guam boat-based survey as an example.

The following paragraphs provide further specifics of what is included in each section and how the objectives together fulfill the purpose of the report.

First, the statistical design of the fishing surveys and the mathematical expansion of survey data to produce total catch estimates are presented. While these components are largely consistent across the boat- and shore-based surveys in each territory, local differences exist, particularly in the implementation of the general survey design. The computation of total effort, average CPUE, and total catch are described, along with their variances. The section also explains how to perform these expansions using the R programming language, and presents the theoretical basis and practical implementation of the fishing surveys and their associated expansion algorithms.

Next, a case study is used to connect the theory of the previous section to its practical application. The Guam boat-based survey is selected for the case study as the longest-standing and most significant survey by total catch. Empirical data from the survey is detailed and the expansion algorithm is applied to the data step-by-step to produce a species-level catch estimate for Onaga (*Etelis coruscans)*, a highly targeted deep bottomfish species. This provides a detailed account of data limitations and corresponding steps the expansion algorithm takes to address them.

While the expansion algorithm is taken as a rigid formulation in the previous two sections, in practice, there are situations where data must be borrowed to account for incomplete survey coverage and limitations in data availability. These algorithmic decisions can be difficult to rigorously justify and the sensitivity of expansion results to two central decisions is analyzed: 1) the borrowing of CPUE data when insufficient catch data has been collected; and 2) the use of data from sampled ports to estimate CPUE and fishing effort from unsampled ports. Alternate scenarios are considered, and the sensitivity of total catch and species-level catch estimates to these selections are analyzed.

Taken together, these sections investigate the expansion algorithm at a level of detail from broad to specific and a range of domains from theoretical to applied. A final discussion synthesizes the findings and results from previous sections and provide some recommendations for future research.

Fishing surveys: survey design and expansion

The shore-based and boat-based fishing surveys are administrated by the Guam DAWR, the CNMI DFW, and American Samoa DMWR. Shore fishing is distinguished from boat fishing by its origination from the shoreline and the absence of a boat unless only used to transport equipment, such as a small boat launched from the beach to carry a gill net (Oram et al. 2014).

General description of fishing surveys

The shore-based effort and catch rate surveys utilize a roving survey design to collect fishing effort and catch rate information. During the roving effort survey, a surveyor drives across a pre-determined circuit along different sections of accessible coastlines and counts the fishers and gears engaged in fishing (Amesbury et al. 1991). In CNMI and American Samoa the catch rate surveys are conducted jointly with the effort survey. Due to the greater length of the shore-based survey route in Guam, catch rate surveys are conducted separately from the effort surveys. During the roving catch rate survey, fishers encountered along a designated survey route are interviewed to gather data on fishing methods used, hours fished, and fish caught or released. Surveyors identify fish to the species level and measure their lengths and weights, when possible.

The boat-based survey is primarily focused on boat facility access points. Catch rate and effort surveys are conducted at major ports, collectively referred to as sampled ports. In American Samoa, no additional measures are taken to estimate activity outside of the sampled ports since it is believed to be minimal, but in the other two territories fishing boat trailers are counted across all boat launching areas to estimate fishing activity that otherwise may not have been encountered by surveyors. In Guam, an island-wide roving survey conducted separately from the access point surveys is used to count fishing boat trailers (Myers 1993; Jasper et al. 2016). In the CNMI, additional surveys (similar to roving surveys) are conducted during the access point survey (at a sampled port) to count boat trailers at all public boat launching areas.

Shore-based fishing survey

Survey design

Roving surveys are used to estimate CPUE and fishing effort (measured in total gear hours) of shore fishing. The roving surveys for catch rate and effort are conducted on separate survey assignments (i.e., on different survey days) in Guam, but are conducted on the same assignments in the CNMI and American Samoa with runs of effort and catch rate surveys in alternating order. The roving surveys are stratified by day type (weekday and weekend/holiday) and shift (day and night for Guam, and shorter time intervals for the CNMI and American Samoa). In Guam the catch rate survey is additionally stratified by region, and in American Samoa both catch rate and effort surveys are stratified by route (a segment of the coastline). A pre-specified number of survey assignments are randomly selected for each stratum. An island-wide aerial survey, stratified by day type, is also conducted in Guam to estimate the proportion of fishing effort from areas not covered by the ground-based roving effort surveys.

Shore fishing effort is measured in gear hours. During each roving effort survey, all easily accessible shorelines in Guam are visited and all fishing activities are recorded in terms of the fishing method and number of gears deployed. Instantaneous gear counts from day and night shifts are assumed to be representative of daytime and nighttime fishing at any hours during their

corresponding periods (06:00 to 18:00 for daytime fishing and 18:00 to 02:00 for nighttime fishing; Amesbury et al., 1991). The gear counts during daytime are adjusted for the areas not covered by the ground-based roving survey, with a ratio estimated using the aerial survey data. The aerial surveys are conducted one weekday and one weekend day per month, when possible. During these surveys, an airplane flies along the shoreline in a clockwise direction and surveys begin and end at a fixed location.

In the CNMI, there are two 6-hour day shifts and two 6-hour night shifts fully covering the 24 hours of a day. A pre-determined number of survey assignments are randomly selected for each of the four shifts. During a 6-hour survey assignment, typically, three survey runs can be completed, e.g., an effort survey being followed by a catch rate survey, which travels in the opposite direction, and then followed by another effort survey. The instantaneous gear counts per survey run during a shift represent the number of gears present on the survey route at any hours during the 6-hour shift.

In American Samoa, the survey shifts have not been consistent over time. Gear counts are tallied per 2-hour time interval, and the instantaneous gear counts during a survey run (taking about 1 hour to complete) represent the number of gears present at any moment during the corresponding 2-hour time interval.

Catch interviews are conducted with fishermen when they are done with fishing or still in the process of fishing. The data collected include: 1) catch composition identified at the finest taxonomic level possible, 2) length and/or weight measurements, 3) the number of gears used, and 4) hours fished. The interview data are used to calculate CPUE.

Catch is estimated as the product of CPUE and fishing effort, separately by fishing method. In Guam, catch and effort estimates for "hook and line" fishing are produced for region, day type, and shift combinations, while region is excluded for other fishing methods. This is due to the greater availability of catch data for "hook and line" fishing versus other methods. In American Samoa and the CNMI, catch and effort estimates are produced for route, day type, and shift combinations for all fishing methods. In this report, fishing method, day type, shift, and region/route combinations (used for catch rate and effort estimations) are called "estimation domains". A domain is any subpopulation of interest for producing estimates (Breidt et al. 2012). An estimation domain may or may not be a "stratum", which is a subpopulation that is identified prior to sampling. For instance, the shore-based surveys in three PIR territories are stratified by day type and shift and combinations of day type and shift are strata. However, fishing gear types are generally not known until surveys are conducted. Estimation domains associated with different fishing methods or other non-stratifying factors are not strata. The terms *strata* and *stratum* are only used in this document for the combination of stratifying factors for survey design.

The shore-based surveys in Guam, the CNMI, and American Samoa are carried out by the local natural resource agencies. Survey frequencies are determined by available personnel, which sometimes leads to less optimal statistical rigor. Survey days are chosen randomly and stratified between weekdays and weekends, with holidays grouped with weekends. Specific shifts, regions (or routes), and starting points for each survey day are chosen randomly but at pre-determined frequencies over broader scheduling periods. Table 1 provides the frequency and times of shore-

based catch rate and effort surveys, along with other implementation details, for each territory. Information in the table reflects the current or recent survey implementation, as shift times and survey regions have changed over time in response to local agency needs and available resources.

Supplemental notes on the available resources, scheduling process, survey coverage, and on-the-ground implementation for the Guam, CNMI, and American Samoa shore-based surveys are provided in Appendix B.

Survey expansion methodology

While CPUE is computed similarly across territories, the computation of fishing effort exhibits some differences according to the survey design and scheduling for each territory.

In Guam, fishing effort is computed for each combination of fishing method, day type, and survey shift. Furthermore, fishing effort estimates for the "hook and line" method are further separated by region because of the greater availability of catch and effort data. The annual fishing effort in gear hours (ghr) is computed as the product of average gear counts ($\bar{g}$) across survey days, number of calendar days (n_c) in a year for the day type, and fishing hours per day (12 hours for day shift and 8 hours for night shift, EQ 1&2). The term g_i is the number of gears counted in an effort survey and n_s is the number of survey days for a day type. Daytime gear counts are also adjusted for the proportion of fishing effort missed by the ground-based roving survey using adjustment factor p_2 (EQ 1 for daytime). The spatial adjustment p_2 is the ratio of fishing effort in un-sampled areas to the effort in sampled areas based on the aerial survey. It is assumed that the fishing activities observed during the survey time (finishing around noon during day shift and finishing around midnight during night shift) are representative of those during daytime fishing hours (0600 – 1800 hours) or nighttime fishing hours (1800 - 0200 hours) (Amesbury et al., 1991). There is no spatial adjustment for nighttime fishing since aerial survey data are only available for daytime (EQ 2 for nighttime). The symbols used for expansion equations are also listed and defined in Table 2.

(1) $$ghr = (1 + p_2)\, \bar{g} \cdot n_c \cdot 12$$

where $\bar{g} = \frac{\sum_{i=1}^{n_s} g_i}{n_s}$ and g_i is the number of gears in a domain on a sample day

(2) $$ghr = \bar{g} \cdot n_c \cdot 8$$

Variance for $\bar{g}$ is calculated as:

(3) $$var(\bar{g}) = \frac{\sum_{i=1}^{n_s}(g_i - \frac{\sum_{i=1}^{n_s} g_i}{n_s})^2}{n_s(n_s - 1)} = \frac{\sum_{i=1}^{n_s} g_i^2 - \frac{(\sum_{i=1}^{n_s} g_i)^2}{n_s}}{n_s(n_s - 1)}$$

In EQ 3, $\sum_{i=1}^{n_s} g_i^2$ is the sum (across sample days) of the squared gear counts in a region (for "hook and line") or the sum of squared gear counts in all regions (for other fishing methods) across sample days. The variance describes variation among sample days for gear counts from a region (for "hook and line") or from all regions combined. Only fishing effort from "hook and line" is estimated at a region level.

Fishing durations at day/night times (12 hours and 8 hours, respectively) and p_2 are assumed to be constant for fishing effort variance estimation. For a new variable $y = cx$, $var(y) = c^2 var(x)$ when c is a constant. Variance for fishing effort island-wide during daytime is thus calculated as:

(4) $\qquad var(ghr) = var(\bar{g})((1 + p_2) \cdot n_c \cdot 12)^2$

In the CNMI, there are four 6-hour shifts (two during the day and two at night) and effort estimation is made for each shift. Each shift is assumed to represent six fishing hours. There is only one survey route (and hence no separation into regions) and there is no p_2 adjustment for spatial under-coverage. Otherwise, expanded effort values are computed similarly to EQ 1-4.

In American Samoa, the effort estimation is made for each 2-hour time interval (EQ 5). In addition, an adjustment (p_1) is made for temporal under-coverage. The temporal adjustment p_1 is calculated as the number of 2-hour time intervals with effort surveys during day or night, divided by the total number of time intervals within the period of interest).

(5) $\qquad ghr = \frac{\bar{g} \cdot n_c \cdot 2}{p1}$

(6) $\qquad var(ghr) = var(\bar{g}) \cdot (\frac{n_c \cdot 2}{p1})^2$

Mean CPUE (*cpue_ghr*) is calculated as the sum of catch weight ($\sum w$) from all interviews in a given estimation domain, divided by the sum of gear hours ($\sum ghr$) from those interviews (EQ 7).

(7) $\qquad cpue\_ghr = \frac{\sum w}{\sum ghr}$

In most cases, the domains for CPUE estimation are the same as those for effort estimation. However, the estimation domain for CPUE in American Samoa is day or night (encompassing six 2-hour time intervals), while the effort is estimated for individual 2-hour time intervals. When there are fewer than three interviews in an estimation domain, interviews from different day types (weekday and weekend days) are pooled for CPUE estimation. For the CNMI, two 6-hour shifts within day or night are pooled before pooling across day types. If there are still fewer than three interviews after day type pooling, interviews from other years are pooled.

Since $\sum w / \sum ghr = (\sum w/m)/(\sum ghr/m) = \bar{w}/\overline{ghr}$ where m is the number of interviews, *cpue_ghr* is therefore the same as the ratio of mean catch weight to mean gear hours. The variance of a ratio of $\bar{x}/\bar{y}$ is calculated based on Wolter (2010):

(8) $\qquad var\left(\frac{\bar{x}}{\bar{y}}\right) = \frac{\bar{x}^2}{\bar{y}^2}\left(\frac{var(\bar{x})}{\bar{x}^2} + \frac{var(\bar{y})}{\bar{y}^2} - 2\frac{cov(\bar{x},\bar{y})}{\bar{x}\cdot\bar{y}}\right)$

Total catch (*w_tot*) is the product of fishing effort (*ghr*) and CPUE (*cpue_ghr*) in an estimation domain (EQ 9).

(9) $\qquad w\_tot = ghr \cdot cpue\_ghr$

It is assumed that the estimates from effort and catch rate surveys are independent. The variance for a product of two independent variables (*ghr* and *cpue_ghr*) is calculated based on Goodman (1960, EQ 5 of the paper):

(10) $var(w\_tot) = var(ghr \cdot cpue\_ghr) = var(ghr) \cdot (cpue\_ghr)^2 + var(cpue\_ghr) \cdot (ghr)^2 - var(ghr) \cdot var(cpue\_ghr)$

In VFP scripts, the last term in EQ 10 carried a positive sign, which would overestimate the variance of the total catch. The catch of individual species (*w_tot_sp*) is generated by multiplying the total catch within each estimation domain by the proportion of the species (by weight) in interviews from the domain. When there are fewer than three interviews in a year for the domain, interviews are pooled, as for CPUE estimation, for the proportion estimation.

(11) $w\_tot\_sp = w\_tot \cdot \frac{c\_sp}{\sum c\_sp}$

Where $\frac{c\_sp}{\sum c\_sp}$ is the percent composition of the species within the catch.

Boat-based fishing survey

Survey design

The boat-based access point survey is used to estimate CPUE (as catch per boat trip) and fishing effort (measured in boat trips) at sampled ports in Guam, the CNMI, and American Samoa. Except in American Samoa, additional information is collected to also estimate fishing effort originating in areas outside of the sampled ports, which are collectively referred to as the unsampled ports. Catch rate and effort surveys are conducted on the same survey assignments and the surveys are stratified by day type and port in Guam and the CNMI. In American Samoa, all four sampled ports are covered on a survey assignment. During the access point survey, catch interviews are conducted for CPUE estimation and boats leaving and returning are recorded in a separate boat-log form for fishing effort estimation. In the CNMI, boat trailers at all sampled ports and unsampled ports are tallied on a survey day when catch interviews are conducted at a sampled port.

During a catch interview, catch is identified to species or group for each fishing method used (including number of gears and hours fished) on a boat trip. When possible, individual fish are measured for length and/or weight. On the same survey assignment, boat log data are collected to obtain the number of boat trips originating at the port with their fishing methods and other trip information that are needed for fishing effort estimation (see Figure A. 1 and Figure A. 2 for the boat log form and interview form from Guam DAWR). At ports with berthed boats such as Agana in Guam and Smiling Cove in the CNMI, boat slip presence/absence maps are also used to assist in the fishing effort data collection. Charter trips are identified in both catch interview and boat log data.

In Guam, three sampled ports are covered for catch interviews (Figure 2): Agana Boat Basin (Agana), Agat Marina (Agat), and Merizo Pier (Merizo). A roving survey is also conducted to estimate the proportion of fishing activity outside of the three sampled ports. During this survey, boat trailers are enumerated at all public boat launching facilities. The trailer roving survey is

conducted simultaneously with the roving effort survey for shore-based fishing since the shore-based survey route passes all of the launch facilities. The surveys are stratified by day type and the trailer count data are collected during two separate shifts (morning and evening) on a scheduled survey day. The starting location follows the same protocol as the shore-based survey (i.e., from a randomly selected location).

In the CNMI, catch interviews are conducted at Smiling Cove, Fishing Base, and Sugar Dock (Figure 3). On a sample day, the catch rate survey is assigned to one of the sampled ports and boat trailers are monitored at all sampled ports and unsampled ports for boat trip estimation. The proportion of trips by fishing method is estimated based on the boat-log data at the sampled ports. Each of the three sampled ports or three unsampled ports is an estimation domain.

In American Samoa, catch interviews and effort surveys are conducted at four ports: Pago Pago, Fagatogo, Utulei, and Faga'alu (Figure 4). No supplemental surveys are conducted to measure fishing activity outside of these sampled ports.

The boat-based surveys in Guam, the CNMI, and American Samoa are carried out by the local natural resource agencies. Survey frequencies are determined by available personnel and other logistics. Survey days are chosen randomly and stratified between weekdays and weekends (with holidays grouped with weekends) and by port (with the exception of American Samoa). The survey days are selected at pre-determined frequencies over broader scheduling periods. Table 3 provides the frequency and times of boat-based catch rate and effort surveys, along with other implementation details, for each territory. Information in the table reflects the current or recent survey implementation, as shift times and survey regions have changed over time.

Supplemental notes on the available resources, scheduling process, survey coverage, and on-the-ground implementation for the Guam, CNMI, and American Samoa boat-based surveys are provided in Appendix C.

Survey expansion methodology

CPUE and fishing effort are estimated by combinations of fishing method, day type, and charter status (estimation domains) in American Samoa. In Guam and the CNMI, ports also contribute to estimation domains. In the CNMI, charter boats are further separated into six-pack charter boats and head boats.

Catch per trip (*cpue_t*) is calculated for each domain as the sum of total catch weight from that domain's interviews divided by number of such interviews (EQ 12). In equations 12 and 13, w_{ij} is the catch weight from interview j in domain i and n_i is the total number of interviews in domain i. The variance of catch weight among interviews in domain i is denoted as *var(w_i)* (Table 4). The CPUE (*cpue_t*) for unsampled ports is assumed to be equal to the overall CPUE at the representative sampled ports.

$$(12) \qquad cpue\_t = \frac{\sum_{j=1}^{n_i} w_{ij}}{n_i}$$

$$(13) \qquad var(cpue\_t) = \frac{var(w_i)}{n_i} = \frac{\sum_{j=1}^{ni}(w_{ij} - \frac{\sum_{j=1}^{ni} w_{ij}}{n_i})^2}{n_i\,(n_i-1)} = \frac{\sum_{j=1}^{ni} w_{ij}^2 - \frac{(\sum_{j=1}^{ni} w_{ij})^2}{n_i}}{n_i\,(n_i-1)}$$

When there are fewer than three interviews in an estimation domain, interviews from different day types (weekday and weekend) are pooled for CPUE estimation. For Guam and the CNMI, some ports and similar fishing methods can also be pooled, in a prespecified order. If there are still fewer than three interviews after these pooling steps, interviews from other years are pooled.

In Guam, the observed number of fishing trips within a known method (n_t) is adjusted for the trips recorded on the boat logs with unknown fishing status (whether or not the trip conducted fishing is unknown, i.e. counts of U for column "Fish? (Y/N/U)" on the boat log form (Figure A. 1)). For some fishing trips, it is known that fishing occurred, but the specific fishing method cannot be identified. The trips with unknown fishing status are assumed to have the same proportion of fishing trips as those with known fishing status. The fishing trips with unknown fishing method are assumed to have the same distribution of fishing methods as those trips with known fishing method. The observed number of trips is also adjusted for fishing activity outside of the two sampling shifts, namely the duration from midnight to the start of the morning shift and a portion of the early afternoon between the morning and afternoon shifts. These adjustment factors are called a_1 (for unknown fishing status), a_2 (for unknown fishing method), and p_1 (for temporal undercoverage). While values for a_1 and a_2 are estimated from survey data, fixed values of p_1 for each estimation domain are based on expert opinion.

$$(14) \qquad n\_t_{adj} = \frac{n\_t \cdot a_1 \cdot a_2}{p_1}$$

In addition to the trips with fishing status known (n_t_fkn), the boat log data include trips with fishing status unknown (n_t_fukn) (i.e., it was unknown whether fishing happened on these boat trips). All fishing trips recorded on the boat log are adjusted as:

$$(15) \qquad a_1 = 1 + \frac{n\_t\_fukn}{n\_t\_fkn} = \frac{n\_t\_fkn + n\_t\_fukn}{n\_t\_fkn}$$

The quantity n_t is the number of fishing trips with a known fishing method recorded in the boat log. The boat log data also include fishing trips with fishing method status unknown (n_t_mukn). The adjusted trips for a specific method with an additional allocation of fishing trips from unknown method status are as follows.

$$(16) \qquad a_2 = 1 + \frac{n\_t\_mukn}{\sum n\_t} = \frac{n\_t\_mukn + \sum n\_t}{\sum n\_t}$$

In Guam, the fishing effort outside of the three sampled ports is estimated as a proportion (p_2) of the combined fishing effort at representative sampled ports, based on the proportion of trailer counts from the roving survey at unsampled ports relative to the representative ports. The sampled ports selected as representative ports vary depending on years and the expression of p_2 in recent years is shown in EQ 23 in the next section.

In American Samoa, the adjustment factors for n_t include a_2 (adjustment for unknown fishing method), p_1 (adjustment for temporal undercoverage), and p_2 (adjustment for unsampled ports).

The values for a_2 are estimated based on boat log data, while values for p_1 and p_2 are assigned based on expert knowledge.

For Guam and American Samoa, the adjusted number of fishing trips per sample day ($\bar{t}$) is then estimated as:

(17) $\qquad \bar{t} = \frac{n\_t_{adj}}{n_s}$

For the CNMI, fishing boat trailer counts are used for total boat trips and boat log data contribute to the distribution of fishing methods (EQ 18). In EQ 18, n_{tot} is the total number of boat trailers (over all sample days n_s) at a port, n_t the number of fishing trip of a specific fishing method (from boat log data) and $\sum n\_t$ the number of fishing trips over all fishing methods. Each of the six sampled and unsampled ports is an estimation domain.

(18) $\qquad \bar{t} = \frac{n_{tot} \cdot \frac{n\_t}{\sum n\_t}}{n_s}$

The adjusted effort per sample day is expanded to the annual fishing effort by the number of weekday or weekend calendar days in the year (EQ 19).

(19) $\qquad t\_tot = \bar{t} \cdot n_c$

The variance of average fishing effort per sample day is the variance of effort among sample days divided by the number of sample days (EQ 20). The variance of the annual fishing effort is scaled up by the number of calendar days (n_c^2, EQ 21).

(20) $\qquad var(\bar{t}) = \frac{\sum_{i=1}^{n_s} t_i^2 - \frac{(\sum t_i)^2}{n_s}}{n_s(n_s-1)}$

where t_i is the number of adjusted fishing trips in a domain on a sample day i

(21) $\qquad var(t\_tot) = var(\bar{t}) \cdot n_c^2$

For each estimation domain, the total catch (w_tot) is the product of the CPUE ($cpue\_t$) and the total trips (t_tot, EQ 22).

(22) $\qquad w\_tot = cpue\_t \cdot t\_tot$

The variance of total catch is calculated as the variance of a product of two independent variables (EQ 10).

The catch of individual species is generated by multiplying the total catch within each estimation domain by the percentage of the species (by weight) in interviews from the domain (see EQ 11). When there are fewer than 3 interviews in a year for the domain, interviews are pooled, as for CPUE estimation, for the proportion estimation.

In the VFP expansion scripts, the variance estimates for species-specific catch were not available. We have developed a method to estimate the variance of species-specific catch by using species-specific CPUE. The CPUE of an individual species is estimated using the catch weight of that species in interviews. Then, the variance of catch for that species is calculated according to the values of species-specific CPUE, total fishing effort, and the variances of species-specific CPUE and total effort (EQ 10).

Expansion script

The expansion algorithm is coded in the programming language R (Crawley 2007), which is widely used as a statistical software and data analysis tool.

This section details the file setup, instructions for use, and R code organization specifically for the Guam boat-based expansion algorithm. The expansion algorithms for CNMI and American Samoa are very similar, with slight differences mainly in the data format; therefore, they were not included in this report.

To ensure transparency, the various components of the R code that implement the Guam boat-based expansion algorithm can be found in Appendices D, E, and F. In addition, all the files (R code and survey data) necessary to reproduce the Guam boat-based expansion results presented in this report can be found in a folder located at:

https://drive.google.com/drive/folders/1V7q8ealStGSE8s945Of_3FHEcOgFQ8Xn?usp=sharing

The folder also contains a "readme" file describing the steps to run the expansion.

To further facilitate the implementation and review of the Guam boat-based expansion algorithm, the R code presented in Appendices D, E, and F has been turned into an R package **expalg.cie** (https://pifscstockassessments.github.io/expalg.cie/). The R package **expalg.cie** can be easily installed (see instructions on the GitHub web page) and also used to reproduce the expansion results presented in the Guam case study of this report.

File setup

The expansion code comprises three R files:

- *guam BB.R*: the runnable file that contains user inputs and produces expansion data products (Appendix D)
- *guam BB expansion.R*: a support file that provides functions for computing expanded values (Appendix E)
- *guam BB interview pooling.R*: a support file that provides a function for aggregating interview data (Appendix F)

The expansion also requires eight CSV files:

- *CIE_sample_days_allyears.csv*: sample day counts
- *CIE_bl_allyears.csv*: boat log information
- *CIE_p1_allyears.csv*: temporal adjustment factors for survey coverage

- *CIE_days_allyears.csv*: calendar day counts
- *CIE_interviews_raw.csv*: catch rate survey information, including total catch across species
- *CIE_catch_raw.csv*: species-level catch information from catch rate surveys
- *CIE_iwc_allyears.csv*: island-wide trailer counts from the roving survey
- *CIE_reference_raw.csv*: fishing method-level CPUE reference values for total catch across all species

All R and CSV files should be placed in the same file directory. Note that in practice, data are loaded from a MySQL database instead of CSV files; CSV files were chosen for the external review to simplify data access for the reviewers. The working directory should be updated at the top of "*guam BB.R*" so that the other two R files can be sourced and, later, the CSV files loaded:

```
setwd("/Users/Toby/Desktop/Guam BBS/")
source("guam BB expansion.R")
source("guam BB interview pooling.R")
```

The expansion code was built in R version 3.6.0 and requires the *dplyr* package to be installed prior to use. When data input is from a MySQL database, the *RMySQL* package is also required. If not already available on your system, install the *dplyr* package by running:

```
install.packages("dplyr")
```

Running the scripts

There are four inputs available to the user, located near the bottom of "*guam BB.R*":

- *start_year*: the first year to run the expansion for
- *end_year*: the final year to run the expansion for
- *pool_f*: either true (T) or false (F), representing whether to pool interviews when three or fewer interviews are available
- *species*: either NA if the expansion should be run for all species, or a vector of species keys (as specified in "CIE_species.csv") if the expansion should only be run for specific species

For example, to run the expansion for all species during 2018 without interview pooling, the inputs would be specified as:

```
start_year = 2018
end_year = 2018
pool_f = F
species = NA
```

Running the entire script in "*guam BB.R*" would then produce two data frames holding the expansion results:

- *expansion*: total catch by estimation domain, along with other intermediary values and corresponding variances
- *species_composition*: total catch by species and estimation domain, along with other intermediary values and corresponding variances

As another example, to obtain skipjack tuna (species key 41606) catches during 2017-2018 with interview pooling, the inputs would be specified as:

```
start_year = 2017
end_year = 2018
pool_f = T
species = c(41606)
```

In this case, the *expansion* data frame would only contain the total catch for skipjack tuna by estimation domain and the *species_composition* data frame would not be used. The CPUE estimates (and the variance) in the data frame *expansion* would be species-specific (for skipjack tuna only). If two or more species are listed in the species input, *expansion* will represent the total species-level catch aggregated across those species, and *species_composition* would still not be used.

Upon running the script, two data frames will be added to the global environment (*expansion* and *species_composition*). These can be processed or exported by the user.

Code organization

Functions

Throughout the following discussion, the term "grouping variable" is used to describe a categorical variable with levels estimated separately during the expansion process. Expansions are conducted at the domain level, where one domain represents the combination of one level for each grouping variable. For example, for the Guam boat-based expansion, the grouping variables are fishing method, type of day, port, and charter status, and an estimation domain would represent a single level of each of those variables (e.g., trolling, weekdays, Agana, non-charter).

As mentioned above, the expansion code is contained in three R files. "*guam BB.R*" is the main file that users interact with. It contains user inputs that specify the year(s) and type of expansion to run. CSV files are then parsed and the required data is aggregated into data structures (e.g. arrays). These data structures are passed to functions in "*guam BB expansion.R*", which perform the actual expansion calculations and return output data frames to "*guam BB.R*". In situations where few interviews (i.e. fewer than 3) are available, "*guam BB expansion.R*" will use a supporting function in "*guam BB interview pooling.R*" to gather similar interviews before calculations are made. The relationships among these files and functions are illustrated in Figure 5 and further explained below.

The file "*guam BB.R*" contains a single function, *run_expansion*, that is called upon running the entire script. This function takes a year and *pool_f*, as defined above, as inputs to specify which values to extract from the CSV data files. The function generates data structures needed to compute the expansion, and in turn, calls the function *df_expansion* in "*guam BB expansion.R*".

The file "*guam BB expansion.R*" contains three main functions that are called, in a nested fashion:

- *df_expansion*: This function is called by "*guam BB.R*", which computes the expansion for a given year. This function iterates through the fishing methods and calls *df_method_expansion* for each one, before aggregating those results and returning two data frames, one for total catch and the other for species-level catch.
- *df_method_expansion*: This function is called by *df_expansion* and computes the expansion for a specific fishing method during a given year. This function iterates through the remaining grouping variables (i.e. type of day, port, and charter status) and calls *calc_df* for each combination. In return, it receives expanded values, from which it computes the species composition and returns both.
- *calc_df*: This function is called by *df_method_expansion* and computes the expansion for a specific domain. This function contains most of the actual computation needed to produce expanded values and returns a data frame containing several expanded values and corresponding variances.

Lastly, "*guam BB interview pooling.R*" contains a single function, *pool_interviews*,that takes data structures and variables defining the estimation domain and returns the interviews to use when computing the expansion for that domain, as well as a string describing the level of pooling that was used. This function is called by *df_method_expansion* and *calc_df* in "*guam BB expansion.R*" when fewer than three interviews are available for a domain and the user has specified that pooling should be used (i.e., the input *pool_f* is defined as true).

Data Structures
While not necessary when only running the expansion code, if one wants to read through the code, it is necessary to understand the implemented data structures. Since expansion computations are performed separately for each combination of levels for these grouping variables (i.e., each possible estimation domain), it is logical to store the values needed for the expansion in arrays with dimensions corresponding to each grouping variable and entries within each dimension for the corresponding levels. Then, each element within the array represents a single domain. For example, boat log data is stored in a 4-dimensional array, and each dimension has length equal to the number of levels for the corresponding grouping variable.

```
bl = array(0, c(length(ports_refer), length(days_refer), length(methods_refer), length(c
harters_refer)))
```

Above, the array holding boat log counts is initialized as a 4-dimensional array with dimensions corresponding to port, type of day, fishing method, and charter status, respectively. The length of each dimension is specified by the corresponding reference vector. These vectors contain each

valid level for the corresponding grouping variable. In most cases, these are predefined, as is the case for port (Agana, Agat, and Merizo), type of day (weekday or weekend/holiday), and charter status (charter or non-charter). However, for fishing method there are a large number of potential values recognized in the database, many of which are not encountered during a given year due to their scarcity. In this case, to minimize storage space the reference vector only includes those fishing methods that were actually encountered.

```
methods_refer <<- sort(unique(c(interviews_current_year$METHOD_FK, bl_current_year$METHO
D_FK, p1_current_year$METHOD_FK)), decreasing = FALSE)
```

Above, the fishing method reference vector only contains those fishing method keys that occurred in the catch rate surveys, boat logs, or temporal adjustment factors for the year being analyzed. Due to this realignment of vector indices, if one wants to know the fishing method corresponding to level "x" in one of the arrays, the database key will not be "x" and must instead be computed as:

```
methods_refer[x]
```

and if one wants to know which level to find a fishing method key "y" in one of the arrays, it must be computed as:

```
match(y, methods_refer)
```

Case Study - Guam

In this section, we selected the Guam boat-based survey to provide a hands-on demonstration of how survey data are collected and how total catch, total effort, and species-level catch are estimated (Figure 6). The section concludes with a bottomfishing expansion example that tracks the expansion algorithm step-by-step from the survey data to species-level catch estimates.

Survey data

Catch interviews

Surveyors conduct catch interviews at three sampled ports on Guam: Agana, Agat, and Merizo. Table A.1-3 include the number of interviews in each estimation domain (port/day type/fishing method/charter status combinations) at each of the three ports. Interview pooling is employed during the catch expansion when the number of interviews in a domain is fewer than three. Trolling and bottomfishing are the most common boat-based fishing methods and correspondingly have the highest interview counts. Non-charter domains almost always have more interviews than charter domains, so non-charter domains are discussed first for each of the three sampled ports.

At Agana (1982-2019), there are more than three interviews for all non-charter estimation domains for the two most common fishing methods (Figure 7 and Table A.1). However, the number of interviews conducted at Agana is not always sufficient for other fishing methods. For spear/snorkel, around 15% of the estimation domains have fewer than three interviews. Approximately 35% and 45% of the estimation domains have fewer than three interviews for spear/scuba and "atulai night light," respectively. For other fishing methods, the majority of the domains have fewer than three interviews, and interview pooling is frequently needed for CPUE and species composition estimation.

At Agat (1995-2019), there are more than three interviews in all non-charter domains for trolling (Figure 7 and Table A.2). For bottomfishing, 10% of estimation domains have fewer than three interviews (all during weekday, see Table A.2). For spear/snorkel, there are fewer than three interviews in 38% of the estimation domains. For other specific methods, there are fewer than three interviews in the majority of the domains.

At Merizo (1989-2019), there are fewer than three interviews in about 10% of the non-charter estimation domains for trolling and bottomfishing (Figure 7 and Table A.3). Spear/snorkel and gillnet both have 46% of the domains with fewer than three interviews. There are also some interviews for cast net and spincasting at Merizo, but the majority of the domains have fewer than three interviews.

Most of the charter interviews are from trolling and bottomfishing (Figure 8). For trolling, there are more than three interviews in all domains at Agana, while 38% of the domains at Agat have fewer than three interviews. For bottomfishing there are fewer than three interviews in approximately 20% and 50% of the estimation domains at Agana and Agat, respectively. At Merizo, there are only 22 charter interviews in total from 1982 to 2019, and nearly all domains have fewer than three interviews.

Fishing effort

The boat trips and the number of boat trailers presented in Figure 9 and Figure 10 are from the raw survey data (prior to expansion), including boat log data (from the access point survey) and trailer counts (from the roving survey). Twice as many access point surveys are assigned to Agana than Agat or Merizo. Thus vessel trips are standardized by the number of sample days at individual ports (Figure 9). The number of boat trips is larger at Agana and Agat than at Merizo. The number of charter trips at Merizo is minimal.

Boat trailers are counted during the roving survey. On a roving survey, all ports (sampled and unsampled) are covered (i.e., equal sampling probability for all ports). The trailer counts shown in Figure 10 are trailer counts per sample day for individual years and are indicative of different fishing intensities among ports. Since 2000 there are fewer boat trailers at unsampled ports than at Merizo and Merizo has the lowest number of trailers among the three sampled ports (Figure 10).

Catch-per-unit-effort

Here, we use the CPUE (kg per trip) for trolling and bottomfishing to provide examples of typical patterns observed among the three sampled ports (Figure 11) and between weekday and weekend (Figure A. 3). For trolling, there are no apparent differences in CPUE between Agana and Agat (Figure 11 (a) and (b)). The trolling CPUE for non-charter fishing at Merizo has a larger fluctuation among years, probably due to a smaller number of interviews available for the estimation, and the CPUE (non-charter) at this port in recent years appears lower than those at Agana and Agat (Figure 11 (a)). For bottomfishing, the CPUE at Merizo was lower than at the other two ports prior to 2010 (Figure 11 (c)). The CPUE appears higher at Agana than at Agat in most years for charter bottomfishing (Figure 11 (d)), while the CPUE for charter trolling is more similar between the two ports (Figure 11 (b)).

The CPUE for non-charter trolling appears to be higher during weekdays than during weekends at Agana and Agat (Figure A. 3 (a)). However, the difference in the CPUE between weekday and weekend is not apparent for bottomfishing and all charter fishing at these two ports (Figure A. 3 (b-d)).

Expanded fishing effort and catch

Expanded fishing effort

Trolling is the most common fishing method at Agana and Agat, with annual trips ranging from 1,000 to 6,000 (Figure 12 (a)). Bottomfishing is the second most common fishing method at Agana and Agat, with annual trips ranging from 500 to 3,000 (Figure 12 (c)).

At Merizo, bottomfishing, trolling, and snorkel spearfishing are all similarly common (Figure 12). The annual trip estimates for gillnet fishing are ~ 500 at Merizo prior to 2000. Prior to 2000, unsampled ports (combined) have similar numbers of annual trips as one of the sampled ports, for most fishing methods. Since 2000, the number of fishing trips from unsampled ports has decreased (~10% of all bottomfishing trips and less than 10% of all trolling trips in the past 10 years).

Prior to 2000, the number of charter trolling trips at Agana is similar in magnitude to the non-charter trips (Figure 12 (a) and (b)). The number of charter bottomfishing trips at Agat is also similar in magnitude to the non-charter trips at the same port prior to 2000 (Figure 12 (c) and (d)). Though both charter and non-charter trips experience a decline around 2000, the decline is much sharper for charter trips. Contrary to non-charter trolling trips, charter trolling trips have not bounced back since then. The decrease in the number of tourists (especially Japanese visitors) due to the economic slowdown likely contributed to the drastic change at that time. The number of bottomfishing charter trips is larger at Agat than at Agana (Figure 12 (d)), while the number of trolling charter trips is larger at Agana (Figure 12 (b)). The number of charter trips for all methods is small at Merizo, and there are no charter trips at unsampled ports.

Expanded catch

Trolling represents most of the catch from boat-based fishing in Guam. The annual catch estimates at Agana and Agat are ~100,000 kg for non-charter trolling since surveys began in the 1980's and have increased in recent years (Figure 13 (a)).

The annual catch from bottomfishing has decreased across ports over time (Figure 13 (c)). In recent years Agana appears to contribute the most bottomfish catch, with Agat and Merizo contributing similarly to each other. In the past 10 years, catch from unsampled ports accounts for 7% of the total catch for trolling and 9% of the total catch for bottomfishing.

The catch from "atulai night light" fishing method decreases substantially around the year 2000, though the catch has rebounded slightly since 2010 (Figure 13 (e)). For spearfishing/scuba, pulses of high catch (> 20,000 kg) occur in several years at Agat (Figure 13 (g)), with the high catch in 1996 and 2000 coinciding with the highest fishing effort (Figure 12 (g)). The catch from spearfishing/snorkel is of the same order of magnitude across sampled ports (Figure 13 (f)). For gillnet, the catch is often the highest at Merizo (Figure 13 (h)).

Since 1995, charter trolling catch has come mainly from Agana and has followed a downward trend (Figure 13 (b)). The catch prior to 2000 is comparable to the non-charter trolling catch at the port. The catch estimate from charter bottomfishing ranges from 3,000 to 5,000 kg at Agat in the first 5 years after the survey was initiated in 1995, but drops sharply in 2000 and remains low (Figure 13 (d)). The charter bottomfishing catch is significantly less than the catch from the non-charter sector (Figure 13 (c)), accounting for 5% of the total bottomfishing catch over the past ten years.

Expansion example

The expansion algorithm can be difficult to follow when presented theoretically. Here a practical example is provided to explain the algorithm step-by-step. Equations from the survey design and expansion section are referenced and explained where applicable. Assumptions are noted to provide a more comprehensive view of the expansion algorithm than can be gleaned from equations alone. This section first presents the data used in this example and then presents the steps used to derive average daily effort, average CPUE, and expanded catch and effort values for this specific example.

Example data

The previous section presented a detailed overview of the Guam boat-based survey data. In this section, we further focus our expansion example on non-charter bottomfishing in 2016, at the species level, using Onaga (*Etelis coruscans*), a highly targeted species by deep bottomfishing. This results in eight estimation domains, representing combinations of the type of day (weekday or weekend) and port (each of the three sampled ports and the combined unsampled ports). Bottomfishing was selected as it is a common fishing method with generally good data availability, and non-charter fishing was selected to include catch estimation for unsampled ports. Furthermore, the year 2016 was selected because the estimation domains have abundant interviews (and thus do not require interview pooling) in more recent years. The total effort, total catch, and species-level catch for Onaga are derived for these estimation domains.

Average effort

The first step in obtaining total fishing effort is to estimate the number of fishing trips per sample day, adjusting for known biases arising from survey implementation. This computation relies on several values from the effort survey data:

- n_s: number of sample days
- n_t: number of observed fishing trips with known fishing method (e.g., bottomfishing)
- n_t_mukn: number of observed fishing trips with unknown fishing method
- $\sum n\_t$: number of observed fishing trips across fishing methods
- n_t_fkn: number of observed trips with known fishing status
- n_t_fukn: number of observed trips with unknown fishing status
- p_1: temporal adjustment factor to account for trips outside of survey shifts

Through the process of computing the average effort, several intermediate values are also computed before arriving at the final desired quantity, $\bar{t}$:

- a_1: adjustment factor to account for trips with unknown fishing status
- a_2: adjustment factor to account for trips with unknown fishing method
- p_2: spatial adjustment factor to account for trips outside of the sampled ports
- n_t_{adj}: adjusted number of fishing trips
- $\bar{t}$: adjusted number of fishing trips per sample day

Values for each of the aforementioned survey and computed variables are provided in Table 5, and the following text explains the required computations.

Four adjustments are made to the observed fishing trip counts (n_t), corresponding to a_1, a_2, p_1, and p_2. These adjustments address the following biases:

- a_1: Surveyors may observe a boat departing or returning but be unable to discern whether the boat is fishing. In this case, the trip cannot be attributed to any estimation domain and will bias average effort downward if not accounted for.

- a_2: Surveyors may observe a boat departing or returning from a fishing trip, but be unable to discern what fishing method is used. In this case, the trip cannot be attributed to any estimation domain and will bias average effort downward if not accounted for.
- p_1: Survey shifts only cover about 14 hours on a given sample day, missing several hours in the early afternoon and late night. Any trips that both depart and return during these non-surveyed hours will not be logged, biasing average effort downward.
- p_2: Fishing trips are logged only at three main ports, and trips are known to originate from other areas. Data are not available to compute the expanded catch from these unsampled ports without using an adjustment factor and will otherwise bias average effort downward.

Equation 15 provides the formula used to adjust for trips of unknown fishing or non-fishing activity:

$$(15) \qquad a_1 = 1 + \frac{n\_t\_fukn}{n\_t\_fkn} = \frac{n\_t\_fkn + n\_t\_fukn}{n\_t\_fkn}$$

Equation 16 is of a similar form and is used to adjust for trips of unknown fishing method:

$$(16) \qquad a_2 = 1 + \frac{n\_t\_mukn}{\sum n\_t} = \frac{n\_t\_mukn + \sum n\_t}{\sum n\_t}$$

In both cases, when the number of observed fishing trips is multiplied by the adjustment factor, we are effectively assuming that the distribution of unknown trips matches the distribution of known trips. In other words, the proportion of trips that are fishing (for a_1) and the distribution of fishing methods (for a_2) are the same in known and unknown trips. Note that both a_1 and a_2 are set to a value of 1 for the unsampled ports (Table 5). By doing this, we are assuming that surveyors are able to identify trailers used for fishing with certainty and that the distribution of fishing methods is the same at sampled as unsampled ports.

Unlike the other adjustment factors, p_1 is based on the expert opinion of survey supervisors, who estimate the proportion of daily fishing trips observed departing or returning during the shift hours. For non-charter fishing trips on Guam, values ranging from 0.75 to 1 are assumed for p_1, depending on the fishing methods.

Lastly, p_2 is computed using trailer counts across all ocean access points collected during the roving survey. This value is computed as the ratio of trailer counts at unsampled ports to representative ports (defined as Agat and Merizo for recent years, as fishing activity at unsampled ports is believed to be more similar to fishing at these two ports than at Agana):

$$(23) \qquad p_2 = \frac{trailer\ count_{unsampled}}{trailer\ count_{Agat} + trailer\ count_{Merizo}}$$

Before using p_2 in effort computations, we first need to define n_t for the unsampled ports. For these estimation domains, n_t is defined as the sum of n_t for the similar estimation domains of Agat and Merizo. The p_2 parameter is then multiplied by n_t to estimate the number of trips from the unsampled ports. In doing this, we are assuming that the relative number of fishing trailers at

port areas directly reflects the relative amount of all fishing activity at those ports. Lastly, note that p_2 is defined to be 1 for all sampled ports.

With the adjustment factors in place, we can compute the adjusted number of fishing trips using Equation 14, altering this equation slightly to include the term p_2 which is specific to Guam boat-based surveys:

$$(24) \qquad n\_t_{adj} = n\_t \cdot a_1 \cdot a_2 \ / \ p_1 \cdot p_2$$

Based on their design, a_1, a_2, and p_2 must be multiplied, while p_1 must be divided. We now have the adjusted number of fishing trips over the year's sample days. To obtain our desired metric of the number of fishing trips per sample day, we simply divide by the number of sample days, n_s, as in Equation 17:

$$(17) \qquad \bar{t} = \frac{n\_t_{adj}}{n_s}$$

Again, the unsampled ports require a separate definition of n_s. In this case, n_s is defined as the maximum of n_s for the similar estimation domains of Agat and Merizo, which in practice have identical or very similar numbers of sample days.

Average catch-per-unit-effort

The next step in obtaining total catch estimates is to estimate the average CPUE, though insufficient data availability can complicate the process. These computations rely only on information gathered from interviews with a single initial value of n_i, the number of interviews conducted. Through the process of computing the average CPUE, several intermediate values are also computed before arriving at the desired average catch-per-unit-effort, *cpue_t*:

- *n_i_pooled*: number of interviews after pooling
- *w_i*: catch weight from interviews
- *cpue_t*: average catch-per-unit-effort from interviews, expressed as catch per trip

Due to the complexity of the interview pooling algorithm, additional values that may be required during the pooling process depending on the estimation domain are not included in the above list. Instead, values for each of the survey and computed values listed above are provided in Table 6, and the following text walks through the required computations.

The expansion algorithm requires a minimum of three interviews for each estimation domain. While a single interview could provide a CPUE estimate and two interviews could provide a variance estimate, the baseline is set at three interviews to minimize the impact of outlier interviews (e.g. interviews for trips with a particularly large catch) on the CPUE estimate. Since the interview pooling process uses information from other estimation domains to supplement any available information from the estimation domain of interest, there is an implicit assumption that information from these external estimation domains benefits the CPUE estimate more through increasing the interview sample size than it detracts by introducing information from external domains with potentially different true CPUE values.

22

The interview pooling algorithm progressively accumulates interviews from estimation domains that are thought to be similar to the estimation domain of interest. For the Guam boat-based survey, the algorithm borrows from estimation domains in the following order, stopping when it has accumulated three or more interviews:

1. Borrow from the other type of day
2. For "mix spearfishing": borrow from "snorkel spearfishing" and/or "SCUBA spearfishing"
3. According to the port:
   a) If the port is Agana, borrow from Agat
   b) If the port is Agat, borrow from Agana
   c) If the port is Merizo and the fishing method is trolling or "atulai night light", borrow from Agana and Agat
   d) If unsampled ports, borrow from Agana, Agat, and Merizo
4. Borrow from previous years, beginning with the most recent year

These steps are cumulative in the sense that once levels of a grouping variable have been pooled, they are kept in all following steps. For example, in step 2, estimation domains for "snorkel spearfishing" and "SCUBA spearfishing" from both types of day would be used, since step 1 pools across type of day. If the four steps do not provide three or more interviews, any pooled interviews will be used to compute the species composition and a reference CPUE value will be used to compute the total catch aggregated across all species. A reference table consisting of method-specific CPUE values estimated based on expert opinion is used in such instances.

In our current example, only the estimation domain for Merizo during weekdays has fewer than three interviews (Table 6). In this particular case, we proceed to the first step of the interview pooling algorithm and see that the estimation domain differing in type of day (Merizo during weekends) has 13 interviews. This gives us an accumulated 15 interviews, which is greater than the minimum requirement of three interviews. Thus, we exit the interview pooling algorithm and set *n_i_pooled* to 15 interviews. All other estimation domains do not require interview pooling, so for these *n_i_pooled* equals *n_i*.

Now that each estimation domain has at least three interviews, we sum the catch weight across pooled interviews for each domain to compute *w_i*. The average CPUE, *cpue_t*, is then computed as the catch weight divided by the number of trips, with each interview representing one trip:

$$(25) \qquad cpue\_t = \frac{w\_i}{n\_i\_pooled}$$

Since we are using limited interview information to estimate CPUE for an estimation domain, we are assuming that the interviews collected are representative of fishing activity within that estimation domain as a whole. While the grouping variables account for some potentially significant sources of variance in fishing activity, others still remain, such as time of day and species targeting within a fishing method. Bias arising from shift scheduling and small interview counts, respectively, may give rise to these additional sources of variation.

For unsampled ports, all interviews from the representative ports (defined as Agat and Merizo for recent years) are used. If this still provides fewer than three interviews, interview pooling proceeds as above, with an implicit first step to pool Agat and Merizo together. In this example, interview pooling is not required for either unsampled port estimation domain and the accumulated interviews from the representative ports can be used to compute *cpue_t*.

Expanded values

Having computed the average effort and CPUE, we can now compute the expanded values of total effort and total catch. Beyond previously computed values, these computations require only one additional value n_c, the number of calendar days for a day type, from which the total effort (*t_tot*) and total catch (*w_tot*) can be computed. To compute species-level catch (*w_tot_species*), we also need the proportional catch weight of that species from interviews (*w_i_species*). Values for each of these surveys, previously computed and newly computed values, are provided in Table 7, and the following text explains the required computations.

Equation 19 provides the formula used to compute total effort, measured in fishing trips, from average effort per sample day and the number of calendar days for a day type:

$$(19) \qquad t\_tot = \bar{t} \cdot n_c$$

Since catch is estimated as the product of effort and CPUE, it is straightforward to compute the total catch at this point. Equation 22 provides the formula used to compute total catch:

$$(22) \qquad w\_tot = cpue\_t \cdot t\_tot$$

Species-level catch is computed by allocating the total catch across species according to the species composition in interviews. Thus, it is assumed that the species composition in interviews reflects the species composition of the entire catch. This implicitly assumes the interviews collected are representative of all fishing trips, as assumed for earlier computations. It may be problematic if some species are mainly caught at night, as night fishing trips tend to be undersampled with the current survey shifts.

In this example, Onaga is only caught in interviews for two of the six sampled port estimation domains. This is reflected in the values of *w_i_species*, which are computed by summing the catch of the species of interest across the pooled interviews for each estimation domain. The proportional catch of Onaga is obtained by dividing *w_i_species* by the total catch in interviews for each estimation domain (*w_i*), multiplied by the total catch (*w_tot*):

$$(26) \qquad w\_tot\_species = w\_tot \cdot \frac{w\_i\_species}{w\_i}$$

As expected, this gives zero species-level catch for the estimation domains without interviews containing Onaga. In this specific case, there may have been many interviews without the species present because it is only caught by deep bottomfishing rather than shallow bottomfishing. If some estimation domains are dominated by shallow bottomfishing (e.g., the port is not situated near deep bottomfishing grounds, or a typically-longer deep bottomfishing trip is unlikely to occur on a weekday), they would be expected to record very little or even zero catch of Onaga.

The interview species composition is only as reliable as the survey's ability to encounter single instances of different classes of fishing trips within an estimation domain with equal probability.

Though not derived in this example, variance estimates are produced by the expansion algorithm for all of the expanded values and some of the intermediate values. Of particular note is the variance estimate for the species-level catch. For Onaga, the coefficient of variation (CV) of the species-level catch is about 58%, reflecting low precision in the estimate. Figure 14 illustrates Onaga catch and catch uncertainty over time. Two aspects are important: 1) As mentioned, species-level catch estimates generally have low precision; and 2) Species-level catch is zero in some years simply because the species did not appear in any interviews that year, though it was inevitably caught outside of those interviews. This imprecision is less of an issue for the total catch estimate. Even at the aggregated species level representing deep bottomfish species (Figure 15) or shallow bottomfish species (Figure 16) the CV estimate is reduced to about 30%.

Sensitivity analyses

The expansion algorithm requires certain assumptions to compensate for survey limitations, including incomplete survey coverage and insufficient quantities of survey data. The degree to which these assumptions influence the total and species-level catch is largely unknown. While some of these limitations are addressed through simple expert-informed correction factors, others require complex algorithmic decisions. In the current section, the sensitivity of expansion results for the Guam boat-based survey to two assumptions is analyzed:

- The choice of pooling algorithm used to borrow interviews when an estimation domain has fewer than three interviews available
- The selection of representative ports to provide proxy CPUE data and scale fishing effort for unsampled ports

The current iteration of the expansion algorithm is referred to as the base expansion and is compared with alternative pooling algorithm and representative port formulations.

Pooling algorithm

The pooling algorithm is required to ensure that CPUE estimates are available for each estimation domain. Some estimation domains naturally reflect low levels of fishing activity and consequently tend to lack interviews. Without interviews, CPUE cannot be estimated and with only a single interview, the variance of CPUE cannot be estimated. Interview pooling is used when fewer than three interviews are available for an estimation domain. This slightly higher limit is taken to reduce the influence of outlier interviews on expansion results.

The pooling algorithm can be viewed as a series of assumptions about which estimation domains are believed to have similar interview information. The pooling algorithm within the base expansion is complex, so three simpler pooling algorithms are compared to the current algorithm:

1. No interview pooling: only those interviews collected for an estimation domain are used, and no expanded results are produced for domains with no interviews available
2. Cross-year pooling only: interviews can be borrowed from the same estimation domain from prior years, starting with the most recent year
3. Cross-day and cross-year pooling only: interviews can be borrowed from the estimation domain differing in type of day, and subsequently from the estimation domain for either type of day from prior years, starting with the most recent year

The four pooling algorithms produce very similar total catch estimates for every surveyed year (Figure 17). Even in the years where the algorithms differ most (e.g., 1996), the estimates from each algorithm are still within a standard deviation of each other and the difference among algorithms is small relative to the interannual difference within an algorithm. Thus, there is no significant difference in total catch among pooling algorithms for any given year and the four algorithms all indicate similar trends over time. Since the estimation domains that contribute most to the total catch tend to have sufficient interviews (i.e., at least 3) and consequently do not

require interview pooling, it is expected that the choice of pooling algorithm should have minimal effect on the total catch, as suggested by these results.

Though there is little difference in the total catch across pooling algorithms, it is likely that species-level catch differs more. Particular estimation domains that require interview pooling could contribute little to the total catch but greatly to the catch for some species. Figure 18 (a) illustrates the relationship between average species-level catch and the degree to which that estimate differs among pooling algorithms. On the horizontal axis, species are presented in order of decreasing catch averaged across the four algorithms between 2017 and 2019. Since this species-level catch does not decrease linearly, the solid black line shows the cumulative percent of the total catch represented by the top species. Dashed lines delineate the number of species required for 95 and 99 percent of the total catch. Thus, the top 19 and top 61 species are required for 95 and 99 percent of the total catch, respectively. Figure 18 (b) reduces the x-axis scale to focus on these top species.

The orange and red lines depict the cumulative percent of the top species with catch that differs by more than one and more than two standard deviations, respectively, between the base expansion and one of the three alternate pooling algorithms. Overall, 34% of the 281 species differ by at least one standard deviation between the base expansion and an alternate pooling algorithm and 28% differ by at least two standard deviations (Figure 18 (a)). However, this percentage is much smaller when only the top species are considered. Of those species that contribute 99% of the total catch, only 8% and 5% differ by at least one and two standard deviations, respectively (Figure 18 (b)). Overall the lesser-caught species have a higher probability of exhibiting significant differences among pooling algorithms. Figure 19 confirms this for the top three species, each of which is remarkably consistent among the four pooling algorithms.

These results make sense given that lesser caught species will tend to occur in estimation domains with less catch and fewer interviews. When interview pooling is required for these estimation domains, the choice of algorithm could significantly affect species-level catch, to the extent that some species may only have non-zero catch if an interview containing them is selected by the pooling algorithm. Historically, only those species with greater data availability have been assessed. As shown in Figure 18 (b), the choice of pooling algorithm should only have a significant impact on the vast minority (less than 10%) of these more frequently caught species. Figure A. 4 and Figure A. 5 provide species-level catch from each of the four pooling algorithms for species in the bottomfish management unit complex (BMUS; WPRFMC 2020b). Fortunately, estimates among algorithms are not significantly different for all of the thirteen BMUS species.

Representative ports

Since fishing activity occurs at several ports surveyed for relative fishing effort but where catch interviews are not conducted, assumptions are necessary to estimate total effort and CPUE for these ports. The expansion algorithm selects a subset of the sampled ports to be representative of fishing activity at the unsampled ports. Interviews from the representative ports provide proxy CPUE data used for the unsampled ports, and fishing activity at the sampled ports is used to scale trailer counts at the unsampled ports to total effort. The base expansion uses Agat and Merizo as representative ports, and three alternate selections are compared to this current algorithm:

1. Agana and Agat are the representative ports
2. Agana and Merizo are the representative ports
3. All three sampled ports are the representative ports

The four representative port selections give very similar total catch estimates (Figure 20). Here, we only analyze the years from 1995 to 2019 since Agat was not sampled and different representative ports were used prior to 1995. As with the pooling algorithms, estimates from the four port selections are always within one standard deviation of each other in a given year and follow temporal trends in total catch very closely. Since unsampled ports contribute minimally to the total catch and effort after 2000 (Figure 12 and Figure 13), it is not surprising that the choice of representative ports has little effect on the estimate.

Even at the species level, there is little difference in the species-level catch across the representative port selections. Figure 21 (a) illustrates the relationship between average species-level catch and the degree to which that estimate differs among port selections. Similar to the results from the pooling sensitivity scenarios, only the top 19 and 59 species are required to account for 95 and 99 percent of the total catch, respectively. Only 15 rarely caught species have differences between 1 and 2 standard deviations from the base scenario, and none of the species that contribute 99% of the catch exhibit differences greater than one standard deviation (Figure 21 (b)). Total catch for the top three species was similar among port selection scenarios (Figure 22), indicating that representative port selection has little effect on the catch estimate for common species.

Very few species are sensitive to the selection of representative ports, and these differences are not as severe as those found in the interview pooling algorithm sensitivity analysis. This may be due to the fact that only species caught at sampled ports can be present in the estimated catch from unsampled ports, whereas interview pooling can introduce a species not otherwise caught in the domain of interest. Combined with the observation that fishing activity is lower at unsampled ports (recently about 7% of the total trips and catch) than sampled ports, this suggests that the potential effects of the selection of representative ports is small compared to the effects of the choice of pooling algorithm. Figure A. 6 and Figure A. 7 provide species-level catch from each of the four port selections for top bottomfish species that could be assessed. Similar to the results of the pooling algorithm sensitivity analyses, estimates differ minimally among algorithms for all these species.

Discussion

Interview pooling

Historically, the interview pooling algorithm has been a point of contention among survey supervisors, expansion developers, and data users. Many decisions led to the current algorithm, and the complexity has caused concern over how well-justified these decisions are and how they influence catch estimates. While the sensitivity analyses evaluate the effects of these decisions, several decisions deserve elaboration here.

The first step in each interview pooling algorithm is to aggregate interviews from the alternate day type. At first, this may seem counterintuitive since sampling is stratified by day type, presumably due to fishing differences between weekdays and weekends. However, the data provide some support for this step in the algorithm. Though the CPUE for non-charter trolling appears to be higher during weekdays than weekends in Guam (Figure A. 3), interview pooling typically is not needed for trolling since interviews are abundant. On the other hand, there is no apparent difference in CPUE between day types for bottomfishing (Figure A. 3), making interview pooling across day types less controversial. With decades of data available for each fishing survey, further analyses could indicate which domains are most justified to aggregate across.

The final step of each interview pooling algorithm is to aggregate interviews from previous years. Progressively earlier years are included until at least three interviews are available, or all survey years have been exhausted. This raises the question of whether future years should be considered and at what point they should be included. Presumably, catch rates from one year in the future will be more similar to rates in a given year than rates from early years of the survey that are more distant in time. In the past, the interview pooling algorithm had considered future years, but this aspect was removed for the base expansion so that expansion results for a given year will remain stable over time. Given that catch limits are set in consideration of historical catch, it is worrisome for managers when catch estimates for a specific year vary over time. The need to borrow information from more recent years could be reduced if further steps could be justified to borrow data from within the year. However, analyses would need to be conducted to indicate which is less likely to bias catch rate estimates.

When the interview pooling algorithm is unable to gather three interviews, the last resort is to borrow a catch rate estimate for the fishing method of interest from a reference table. However, this reference table only contains the proxy values needed to compute the CPUE of aggregated catch and provides no basis to estimate the species composition. While this does not affect the total catch estimate, it does introduce problems in the species-level catch because it is not clear how to distribute the proxy-estimated catch among species. Further investigations could be made to create an alternative reference table that includes both aggregated catch and catch for individual species, or otherwise allow the species-level catch to be estimated for domains lacking pooled interviews.

Sensitivity analyses

The complexity of the expansion algorithms makes it tedious to analyze the sensitivity of results to all aspects of the expansion formulations. Two important aspects of the Guam boat-based

expansion algorithm were considered, and both the choice of interview pooling algorithm and selection of representative ports were found to have minimal impact on total catch estimates. The species-level catch was insensitive for most species, but the sensitivity increased for species with high catch by uncommon fishing methods. Species-level catch was more sensitive to the choice of interview pooling algorithm than the selection of representative ports. Though only two aspects of the expansion algorithm were analyzed, this does help identify some important considerations for data managers.

Greater attention should be paid when disseminating species-level catch estimates. While the estimation domains that contribute most to the total catch tend to have sufficient data and are straightforward to expand, species-level catch estimates are much more susceptible to data limitations. Consequently and as is evident from the sensitivity analyses, the specific algorithmic decisions that are made to handle data limitations will be more impactful at the species level. This is not to say that species-level catch estimates are unreliable. Indeed, the two major algorithmic decisions analyzed had minimal impact on the most-caught trolling and bottomfishing species, which tend to be of greater interest to data users. Rather, the number of interviews and estimation domains the species is present in should be considered when providing species-level data. If either is limited, it is more likely that the estimate may be sensitive to algorithmic decisions. In such cases, the variance estimates of aggregated catch may underestimate the uncertainty in the species-level catch, and sensitivity analyses specific to the data limitations of the species may be warranted.

At the other end of the data production pathway, managers should consider the appropriate level of simplicity when choosing the expansion algorithm. Sensitivity results indicate that some of the finer-scale elements of the expansion algorithm have little impact on the broader-scale data products that are actually used, such as species-level catch estimates that have been summed across domains. Though these fine-scale elements may be appealing for handling specific situations, they also complicate the ultimate presentation of the data products. Thus, there is a tradeoff between the complexity of the expansion algorithm and its ease of conveyance to data users and stakeholders. This is true both at the level of detail within a specific expansion algorithm, as well as across the expansion algorithms from the territories. There is some value in having a largely consistent algorithm across the territories, but there are inevitable differences across the corresponding fisheries. As a potential rule of thumb, perhaps only those differences that potentially affect highly-caught species or multiple estimation domains should be considered for specific exceptions in the expansion algorithm.

Future research

The survey data and catch estimates from the Guam boat-based survey have been reviewed extensively in this document. Similar work may be needed for surveys in American Samoa and the CNMI, even though the expansion process is broadly similar among the three territories. The sensitivity analyses for catch by species have been made for trolling and bottomfishing methods. Additional sensitivity analyses for other species of interest could be added based on the need from stock assessments and/or fisheries management for the territories. Future studies should address the required data/information (i.e., base sampling weight, and sampling weight adjustments due to interviews missed or refused) so that the catch and effort expansion can also

be made using available survey packages (Lumley 2010) or survey procedures (e.g., Ma et al. 2018).

The territorial bottomfish complexes are one of the most important federally-managed fisheries for the PIR. The annual catch estimates of aggregated shallow and deep bottomfish species in Guam generally have reasonable precision. However, the catch estimates for highly targeted bottomfish, such as Onaga, are imprecise, with a coefficient of variation around 50%. The boat-based survey design and implementation would need to be revisited if catch estimates for individual bottomfish species are needed for future stock assessments and management. For instance, the survey could be strategically designed to encounter more bottomfishing trips if this fishery has the highest priority. In addition to improving fishing surveys, a mandatory trip reporting program has been recommended recently (Turner et al. 2019) for the bottomfish fishery in the Pacific island territories. Currently, an electronic platform has been developed for fishers to report their bottomfishing trips. This application has the potential to capture the commercial sector when mandatory reporting requirements and enforcement measures are in place. Such a mandatory reporting program can be integrated with voluntary fishing surveys to better characterize the combined catch from commercial and non-commercial fishery sectors.

References

Allen S, Bartram P. 2008. Guam as a fishing community. Pacific Islands Fish. Sci. Cent., Natl. Mar. Fish. Serv.,NOAA, Honolulu, HI 96822-2396. Pacific Islands Fish. Sci. Cent. Admin. Rep. H-08-01, 61 p.

Amesbury SS, Sherwood TS, Davis GW. 1991. Monitoring a tropical island reef fishery. American Fisheries Society Symposium 12: 84-87.

Breidt FJ, Lessor V, Opsomer JD. 2012. Consultant's report: Preliminary review of Hawaii Marine Recreational Fishing Survey.

Craig P, Ponwith B, Aitaoto F, Hamm D. 1993. The Commercial, Subsistence, and Recreational Fisheries of American Samoa. Marine Fisheries Review 55(2): 109-116.

Crawley MJ. 2007. The R Book. John Wiley & Sons, Ltd.

Dalzell P, Adams T, Polunin N. 1996. Coastal fisheries in the South Pacific. Oceanography and Marine Biology Annual Review 33: 395-531.

Goodman LA. 1960. On the exact variance of products. Journal of the American Statistical Association 55: 708-713.

Hamm DC. 1993. The Western Pacific Fishery Information Network: A Fisheries Information System. Marine Fisheries Review 55(2): 102-108.

Hensely RA, Sherwood TS. 1993. An Overview of Guam's Inshore Fisheries. Marine Fisheries Review 55(2): 129-138.

Hospital J, Beavers C. 2014. Economic and Social Characteristics of Small Boat Fishing in the Commonwealth of the Northern Mariana Islands. Pacific Islands Fish. Sci. Cent., Natl. Mar. Fish. Serv., NOAA, Honolulu, HI 96818-5007. Pacific Islands Fish. Sci. Cent. Admin. Rep. H-14-02, 58 p.+ Appendices.

Jasper W, Matthews T, Gutierrez J, Flores T, Tibbatts B, Martin N, Bass J, Wusstig S, Franquez R, Manibusan F, Ducusin J, Regis A, Lowe MK, Quach M. 2016. DAWR Creel Survey Methodology. Division of Aquatic & Wildlife Resources (DAWR), Guam Department of Agriculture. Tech. Rept. 1, 110 p.

Langseth B, Syslo J, Yau A, Cavalho F. 2019. Stock Assessment of the Bottomfish Management Unit Species of Guam, the Commonwealth of the Northern Mariana Islands, and American Samoa, 2019. NOAA Tech. Memo. NMFS-PIFSC-86, 177 p.

Lumley T. 2010. Complex Surveys: A Guide to Analysis Using R. Hoboken, N.J.: John Wiley.

Ma H, Ogawa T, Sminkey T, Breidt F., Lesser V, Opsomer J, Foster J, Van Voorhees D. 2018. Pilot surveys to improve monitoring of marine recreational fisheries in Hawaii. Fisheries Research 204: 197-208.

Myers RF. 1993. Guam's small-boat-based fisheries. Marine Fisheries Review 55 (2): 117-128.

Nadon MO. 2019. Stock Assessment of Guam Coral Reef Fish, 2019. NOAA Tech. Memo. NMFS-PIFSC-82, 107 p.

NMFS. 2021. Fisheries of the United States, 2019. U.S. Department of Commerce, NOAA Current Fishery Statistics No. 2019.

Oram, R., T. Flores Jr., B. Tibbatts, J. Gutierrez, J. P. Gesner, S. Wusstig, A. Regis, M. Quach, D. Hamm, and P. Tao. Edited by K. Lowe and K. Brousseau. 2014. Guam Shore-based Creel Survey. Pacific Islands Fish. Sci. Cent. , Natl. Mar. Fish. Serv., NOAA, Honolulu, HI 96818-5007. Pacific Islands Fish. Sci. Cent. Admin. Rep. H-14-, var. p.

Turner S, Suter J, Ryznar B. 2019. Pacific Insular Fisheries Monitoring & Assessment Planning Summit (PIFMAPS) Data Summit.

Wolter KM. 2010. Introduction to variance estimation, 2nd Edition. Springer.

WPRFMC. 2020a. Annual Stock Assessment and Fishery Evaluation Report Pacific Island Pelagic Fishery Ecosystem Plan 2019. Remington T, Fitchett M, Ishizaki A, DeMello J. (Eds.) Western Pacific Regional Fishery Management Council. Honolulu, Hawaii 96813 USA. 372 pp. + Appendices.

WPRFMC. 2020b. Annual Stock Assessment and Fishery Evaluation Report for the Mariana Archipelago Fishery Ecosystem Plan 2019. Remington T, Sabater M, Ishizaki A. (Eds.) Western Pacific Regional Fishery Management Council. Honolulu, Hawaii 96813 USA. 204 pp. + Appendices.

WPRFMC. 2020c. Annual Stock Assessment and Fishery Evaluation Report for the American Samoa Archipelago Fishery Ecosystem Plan 2019. Remington T, Sabater M, Ishizaki A. (Eds.) Western Pacific Regional Fishery Management Council. Honolulu, Hawaii 96813 USA. 141 pp. + Appendices.

Tables

Table 1. Effort and catch rate survey implementation details for the American Samoa, the CNMI, and Guam shore-based surveys. WD = weekday and WE = weekend.

	American Samoa	CNMI	Guam
Effort Frequency	Minimum 40 survey shifts/month (36 WD and 4 WE shifts), split across regions	32 shifts/quarter (evenly split across shifts, sequences, and WD/WE)	8 shifts/month (2 shifts on 2 WD and 2 WE)
Effort Times	06:30-13:00, 10:30-19:00, 16:30-24:00	00:01-06:00, 06:01-12:00, 12:01-18:00, 18:01-24:00	06:30 start and 19:00 start, both until finished
Effort Order	Alternating start point with the two ends of the route	Alternating start point with the two ends of the route	Random start point and alternating direction
Catch Rate Frequency	With effort, alternating effort and catch runs	With effort, alternating effort and catch runs	8/month (2 shifts on 2 WD and 2 WE, split across 3 regions)
Catch Rate Times	With effort, alternating effort and catch runs	With effort, alternating effort and catch runs	06:30 – 12:00, 19:00 – 24:00
Number of Regions	3 (west, central, east)	1 (western lagoon)	3 for catch (roughly Tumon and Agana Bays, west, and combined south and east), combined into one region for effort
Aerial Survey	No	No	2 shifts/month (1 WD and 1 WE)

Table 2. Symbols used in equations 1-11 for shore-based expansion.

Symbol	Description
n_c	Number of calendar days in a year by day type
n_s	Number of effort-survey sample days for a day type
$\sum g$	Sum of gear counts in an estimation domain from the effort survey (aggregated across sample days)
$\bar{g}$	Average number of gears per survey run in a domain
ghr	Expanded gear hours in a domain
$var(\bar{g})$	Variance of average number of gears per survey run
var(ghr)	Variance of expanded gear hours
$\sum w$	Sum of total catch weight from catch interviews in a domain
$\sum ghr$	Sum of gear hours fished from catch interviews
cpue_ghr	Catch per gear hour
w_tot	Expanded catch weight
c_sp	Catch weight of a species from all catch interviews in a domain
$\sum c\_sp$	Catch weight of all species from all catch interviews in a domain
w_tot_sp	Expanded catch weight for a species
p_1	Adjustment for temporal under-coverage of time intervals in American Samoa
p_2	Ratio of fishing effort in unsampled areas to effort in sampled areas

Table 3. Effort and catch rate survey implementation details for the American Samoa, the CNMI, and Guam boat-based surveys.

	American Samoa	CNMI	Guam
Effort Frequency	Minimum 28 survey shifts/month (24 WD and 4 WE shifts)	6 days/month (3 WD and 3 WE)	8 days/month (4 WD + 4 WE, split across 3 ports)
Effort Times	06:00, 09:00, 12:00, 15:00, 18:00, 21:00	10:00, 14:00, 20:00, 22:00	05:00/05:30/06:00 – 11:00/12:00 (depending on port), 16:00 – 24:00
Catch Rate Frequency	Concurrent with effort	Concurrent with effort, but only covering one port	Concurrent with effort
Catch Rate Times	05:00-13:30, 13:00-21:30, 21:00-05:30	10:00-18:00, 18:00-02:00	Concurrent with effort
Number of Ports	4 (Pago Pago, Fagatogo, Utulei, Faga'alu)	3 for catch (Smiling Cove, Sugar Dock, Fishing Base), plus unsampled ports for effort	3 for catch (Agaña, Agat, Merizo), plus unsampled ports for effort

Table 4. Symbols used in equations 12-22 for boat-based expansion.

Symbol	Description
n_c	Number of calendar days in a year for a day type
n_s	Number of sample days at a port for a day type
n_i	Number of catch interviews in a domain
w_{ij}	Total catch weight in a catch interview (j) of a domain (i)
n_t	Number of fishing trips with a known fishing method in a domain, aggregated over sample days
cpue_t	Catch per trip, for a specific fishing method
$var(w_i)$	Variance of catch weight among interviews in a domain (i)
var(cpue_t)	Variance for catch per unit effort
n_t$_{adj}$	Adjusted number of fishing trips with a known fishing method
$\sum n\_t$	Number of fishing trips with known fishing methods in the boat log data, aggregated over fishing methods
n_t_mukn	Number of fishing trips with unknown fishing method
n_t_fkn	Number of boat trips with known fishing status (either fished and did not fish)
n_t_fukn	Number of boat trips with unknown fishing status (i.e., status unknown whether fishing occurred or not)
a_2	Adjustment factor for unknown fishing method
a_1	Adjustment factor for unknown fishing status
p_1	Adjustment factor for temporal under-coverage
p_2	Adjustment for spatial under-coverage
$\bar{t}$	Adjusted number of fishing trips for a specific fishing method per sample day
t_tot	Expanded number of fishing trips
$var(\bar{t})$	Variance of average number of fishing trips per sample day
var(t_tot)	Variance of expanded fishing trips

w_tot	Expanded catch weight

Table 5. (a) Survey values used to compute the average number of fishing trips per sample day, and (b) the intermediate and final values used for these computations. Bolded survey value entries require computation.

(a)		Survey Values						
Estimation domain		n_s	n_t	n_t_mukn	sum(n_t)	n_t_fkn	n_t_fukn	p_1
Agana	WD	24	25	0	360	407	7	0.85
	WE	24	85	2	624	707	9	0.85
Agat	WD	11	8	0	102	116	2	0.85
	WE	12	29	0	131	155	2	0.85
Merizo	WD	12	24	0	46	57	3	0.85
	WE	11	19	0	48	83	3	0.85
Unsampled	WD	**12**	**32**	0	0	0	0	0.85
	WE	**12**	**48**	0	0	0	0	0.85

(b)		Computed Values				
Estimation domain		a_1	a_2	p_2	n_t_adj	$\bar{t}$
Agana	WD	1.02	1.000	1.00	29.92	1.25
	WE	1.01	1.003	1.00	101.60	4.23
Agat	WD	1.02	1.000	1.00	9.57	0.87
	WE	1.01	1.000	1.00	34.56	2.88
Merizo	WD	1.05	1.000	1.00	29.72	2.48
	WE	1.04	1.000	1.00	23.16	2.11
Unsampled	WD	1.00	1.000	0.23	8.82	0.74
	WE	1.00	1.000	0.23	13.24	1.10

Table 6. The survey values used to compute the average CPUE, and the intermediate and final values for these computations. Bolded survey value entries require computation.

Estimation domain		Survey Value n_i	Computed Values		
			n_i_pooled	w_i	cpue_t
Agana	WD	12	12	127.45	10.62
	WE	46	46	488.56	10.62
Agat	WD	3	3	33.66	11.22
	WE	21	21	241.70	11.51
Merizo	WD	2	15	89.20	5.95
	WE	13	13	65.92	5.07
Unsampled	WD	**5**	5	56.94	11.39
	WE	**34**	34	307.62	9.05

Table 7. (a) The survey value and previous computed values used to compute the expanded values, and (b) the intermediate and final values for these computations.

(a)		Survey Value	Previous Computed Values		
Estimation domain		n_c	$\bar{t}$	cpue_t	w_i
Agana	WD	250	1.25	10.62	127.45
	WE	116	4.23	10.62	488.56
Agat	WD	250	0.87	11.22	33.66
	WE	116	2.88	11.51	241.70
Merizo	WD	250	2.48	5.95	89.20
	WE	116	2.11	5.07	65.92
Unsampled	WD	250	0.74	11.39	56.94
	WE	116	1.10	9.05	307.62

(b)		Computed Values			
Estimation domain		t_tot	w_tot	w_i_species	w_tot_species
Agana	WD	311.64	3309.90	0.00	0.00
	WE	491.05	5215.43	76.45	816.11
Agat	WD	217.59	2441.38	0.00	0.00
	WE	334.06	3844.87	125.75	2000.38
Merizo	WD	619.20	3682.15	0.00	0.00
	WE	244.24	1238.49	0.00	0.00
Unsampled	WD	183.85	2093.73	0.00	0.00
	WE	127.96	1157.76	125.75	473.27

Figures
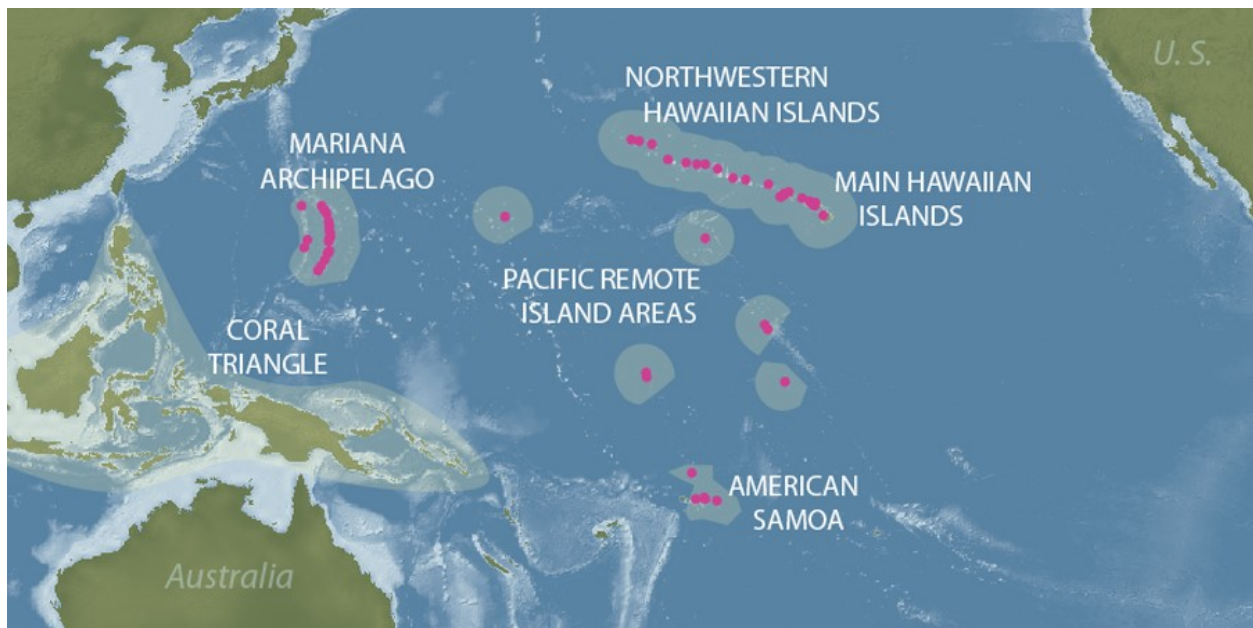
Figure 1. Map of the United States-associated areas in the Pacific Ocean, including the territories of Guam and the CNMI within the Mariana archipelago, as well as American Samoa. Islands, reefs, and shoals are represented by the pink dots.

Figure 2. Sampled ports (Agana Boat Basin, Agat Marina, and Merizo Pier) for the Guam boat-based fishing survey. On a sample day, a single sampled port is covered for catch interviews and fishing trip counts. Separate trailer roving surveys collect effort data for all unsampled ports.
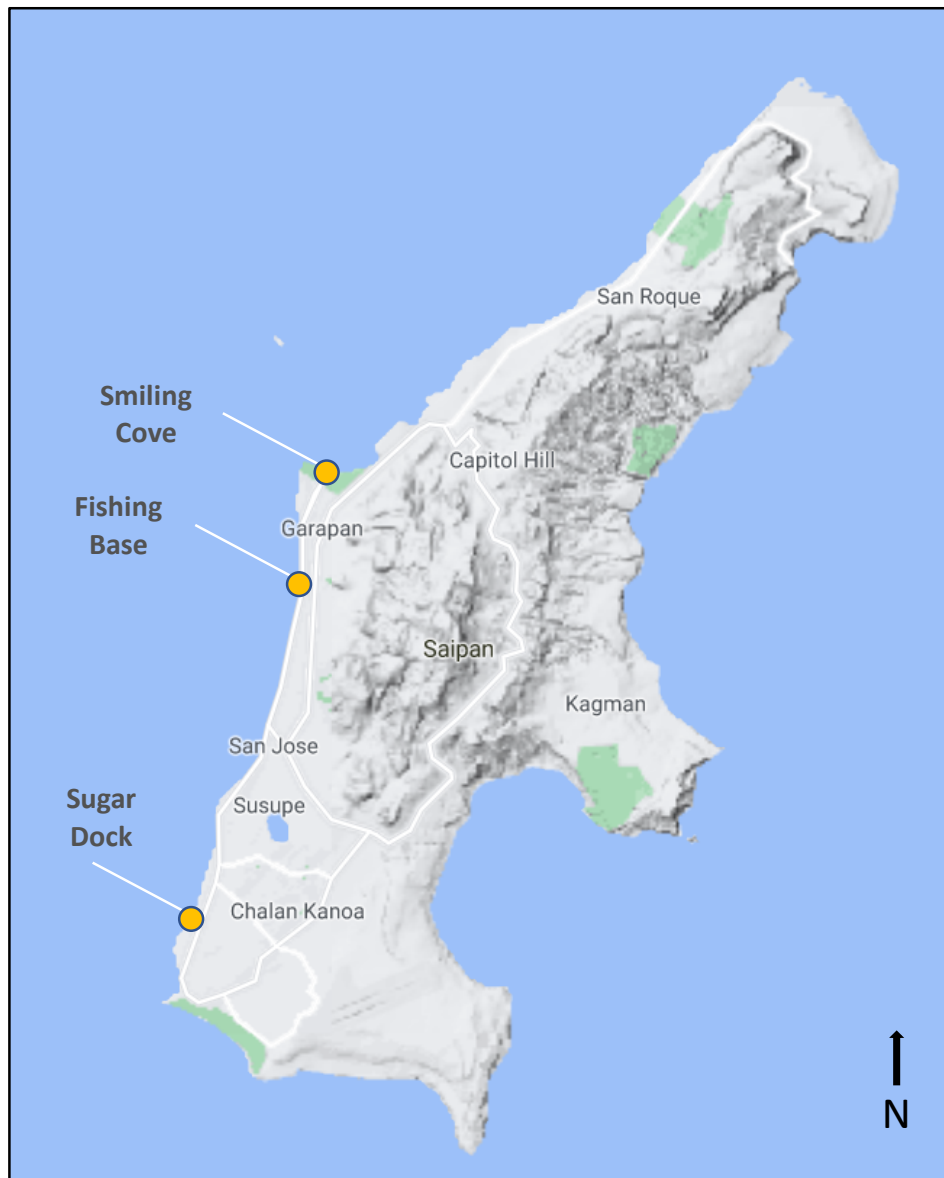
Figure 3. Sampled ports (Smiling Cove, Fishing Base, and Sugar Dock) for the CNMI boat-based fishing survey on Saipan. On a sample day, all boat trailers are counted at three sampled ports and several unsampled ports on a sample day while catch interviews are conducted at a single sampled port.
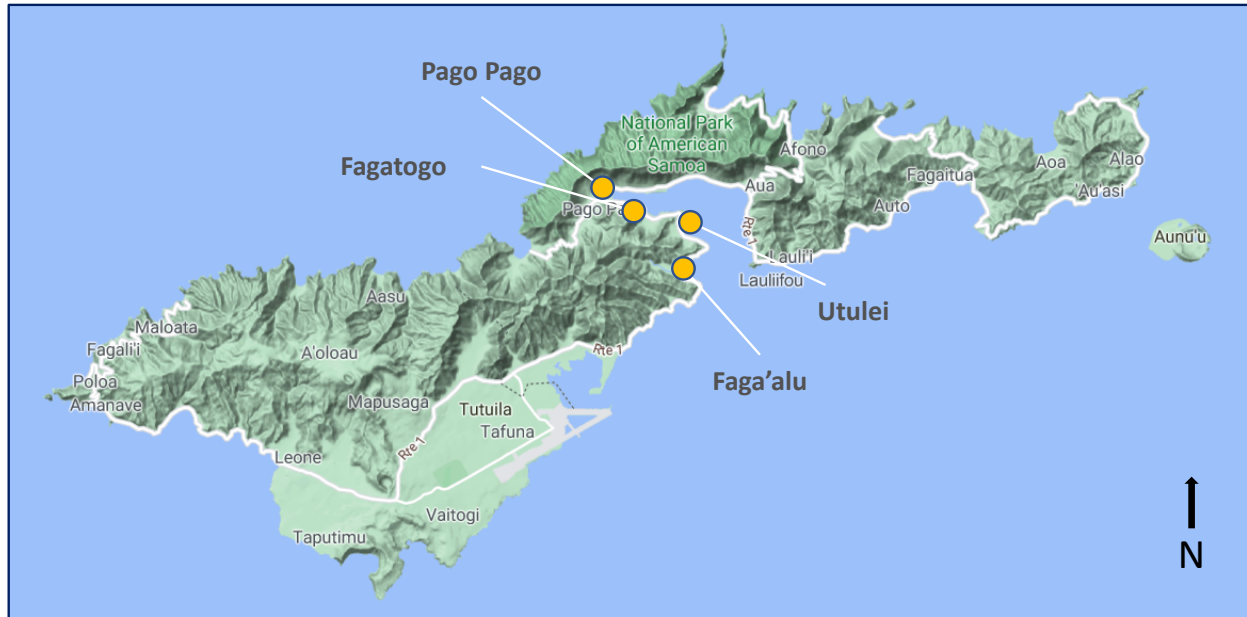
Figure 4. Sampled ports (Pago Pago, Fagatogo, Utulei, and Faga'alu) for the American Samoa boat-based fishing survey on Tutuila. On a sample day, all four sampled ports are covered for catch interviews and fishing trip counts.

guam BB.R

guam BB interview pooling.R

year, pool_f

run_expansion

pool_interviews

year, method,
type of day, port,
charter status

pooled
interviews

year, pool_f

total catch,
species-level catch

pooled
interviews

year, method,
type of day,
charter status

year, method, type of day,
port, charter status

df_expansion

year, pool_f, method

df_method_expansion

calc_df

method-level total
catch and species-
level catch

domain-level
total catch

guam BB expansion.R

Figure 5. Calls among functions in the three expansion R files. Functions are bolded. When fewer than three interviews are available for a domain, the function "pool_interviews" is called by functions "calc_df" and "df_method_expansion" when computing total catch and species-specific catch, respectively.

46
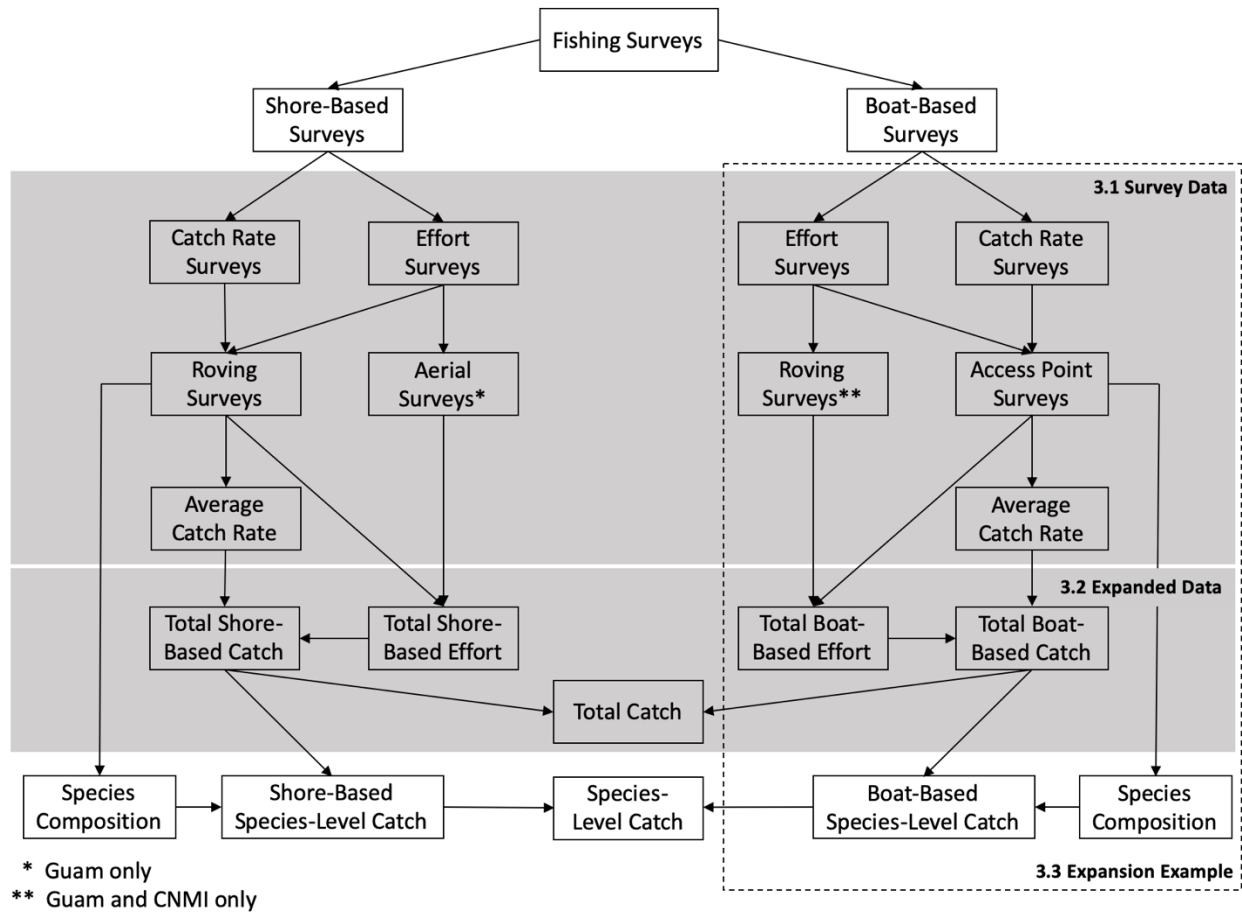
Figure 6. Fishing survey components and their connections with the expansion process. Components relevant to the survey data, expanded data, and expansion example sections are indicated as such.
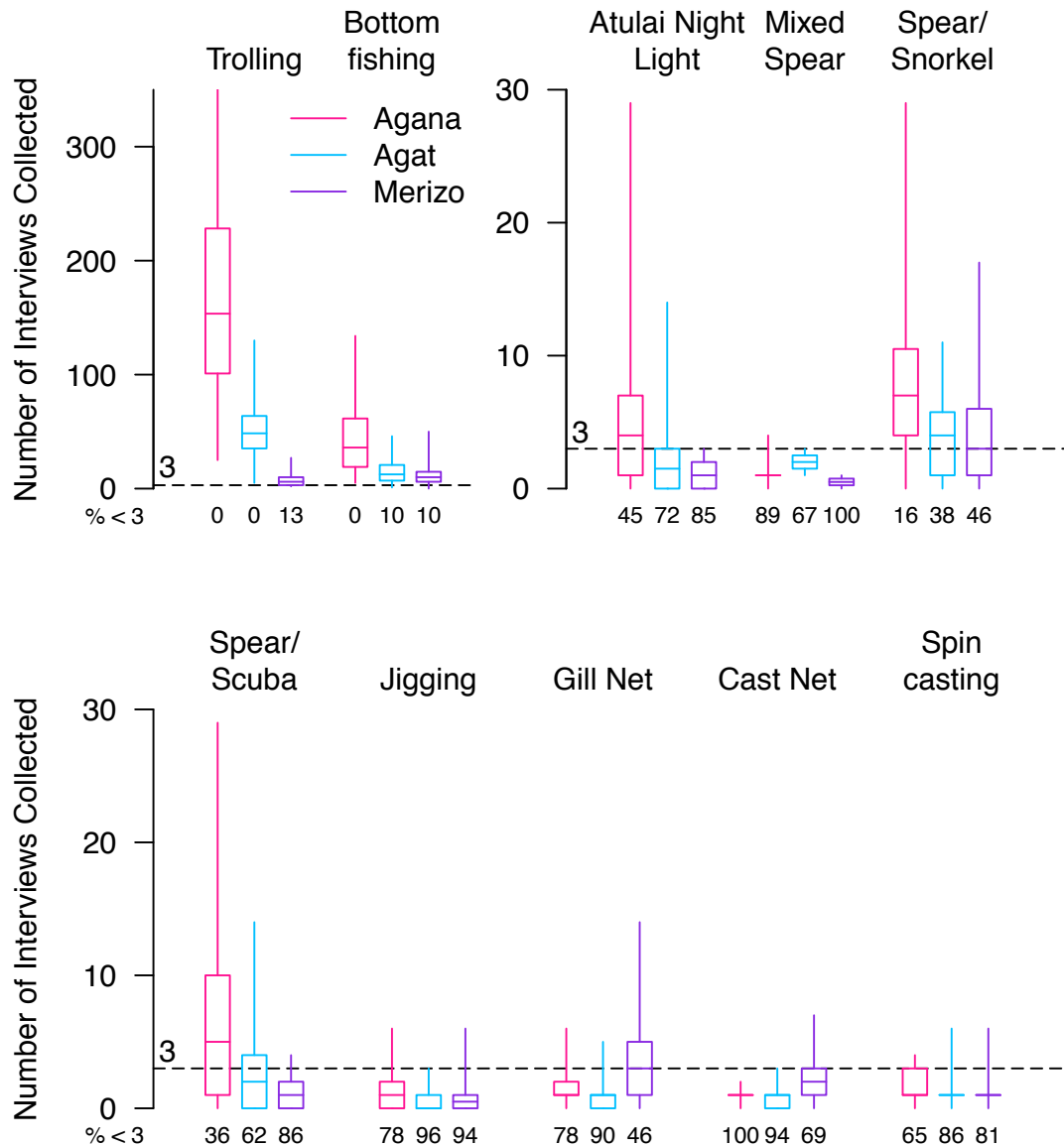
Figure 7. Boxplots for number of non-charter interviews in individual estimation domains at Agana, Agat, and Merizo. Boxes indicate the 25, 50 and 75th percentiles of interview counts across estimation domains, and lines extend from the box to the minimum and maximum counts. The percentage of domains with fewer than 3 interviews (% < 3) out of all domains with any effort or catch data is shown below each boxplot.
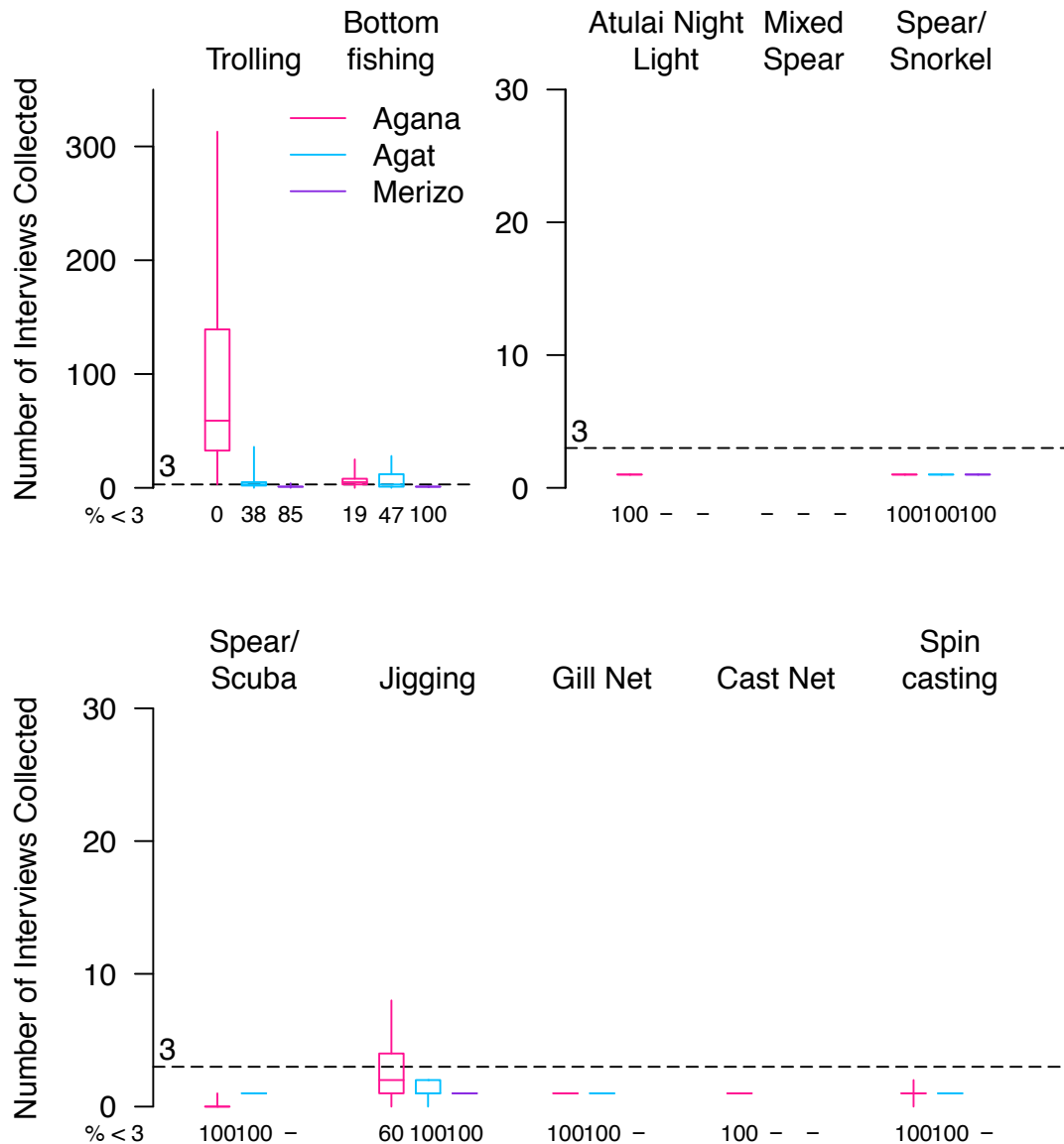
Figure 8. Boxplots for number of charter interviews in individual estimation domains at Agana, Agat, and Merizo. Boxes indicate the 25, 50 and 75th percentiles of interview counts across estimation domains, and lines extend from the box to the minimum and maximum counts. The percentage of domains with fewer than 3 interviews (% < 3) out of all domains with any effort or catch data is shown below each boxplot.
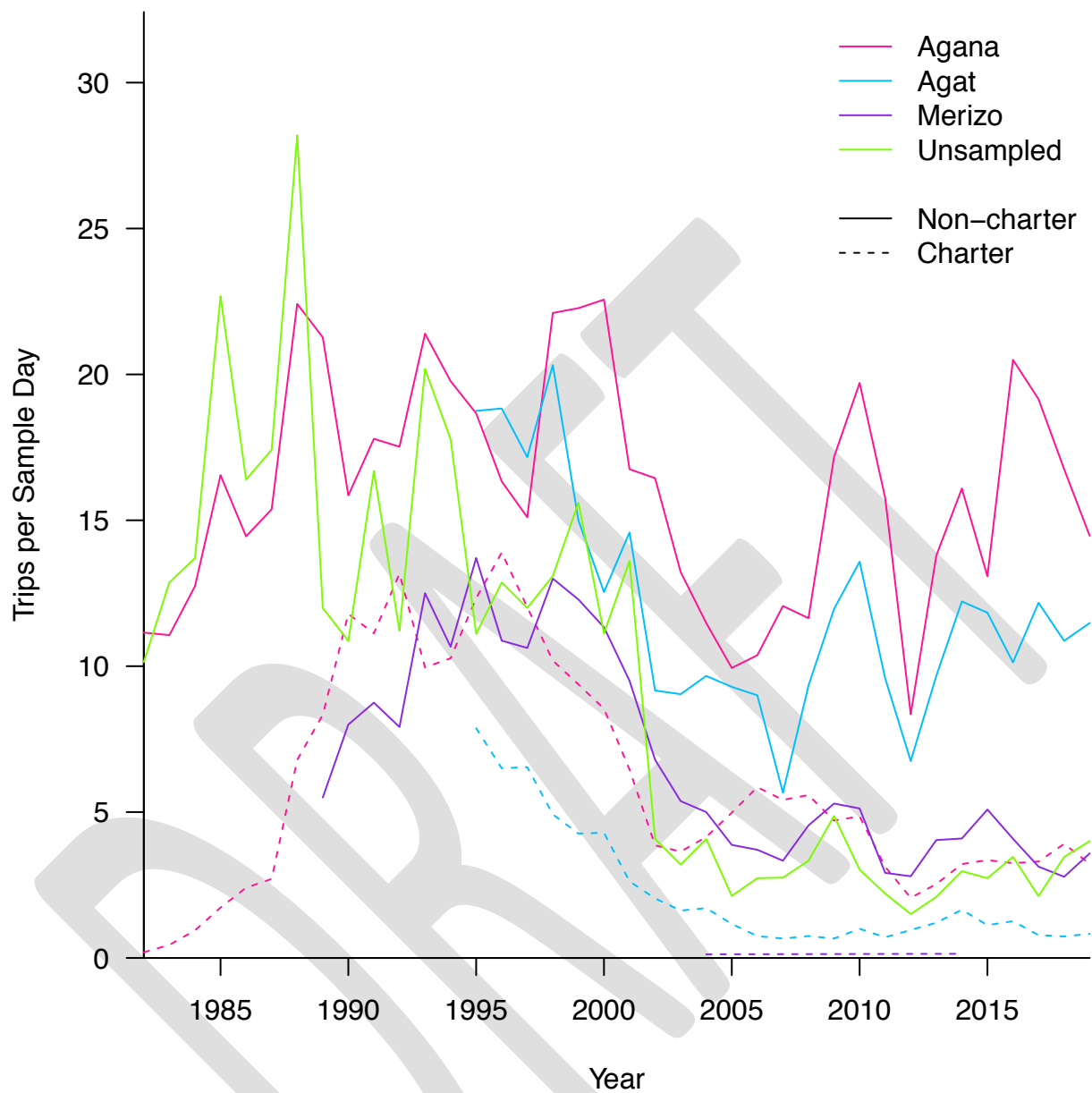
Figure 9. Number of boat trips per sample day at Agana, Agat, and Merizo. Trips per sample day at unsampled ports represents the trip counts at representative ports, scaled by the ratio of trailer counts of unsampled to representative ports. The only representative port in years prior to 1989 is Agana and all other ports are unsampled ports. During 1989-1994, representative ports are Agana and Merizo, and all other ports are unsampled ports. For years in 1995 and afterwards, representative ports are Agat and Merizo and all ports other than three presently sampled ports are unsampled ports.
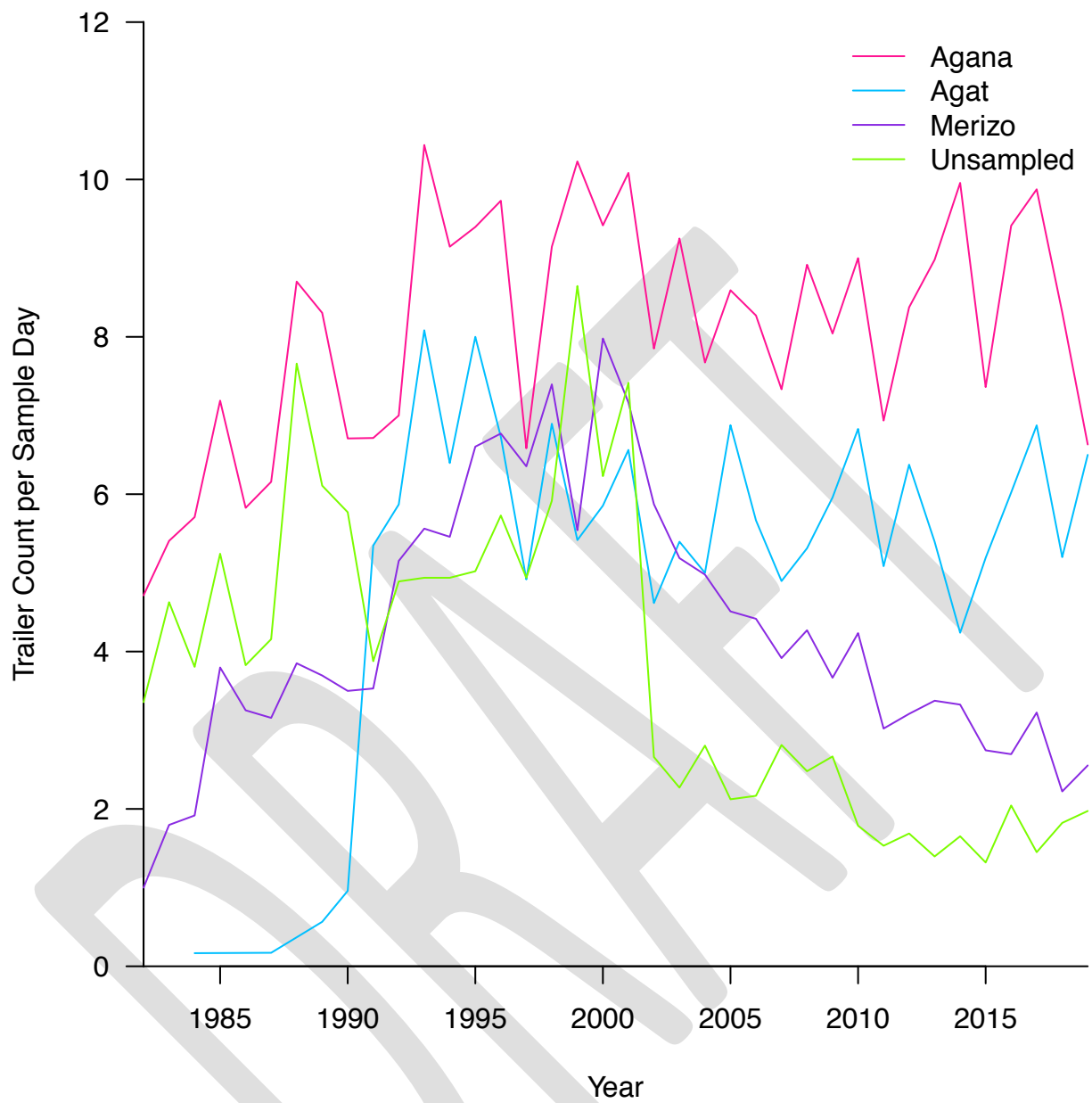
Figure 10. Number of boat trailers at Agana, Agat, Merizo, and unsampled ports. In this figure, "Unsampled" is for ports other than the three presently sampled ports.
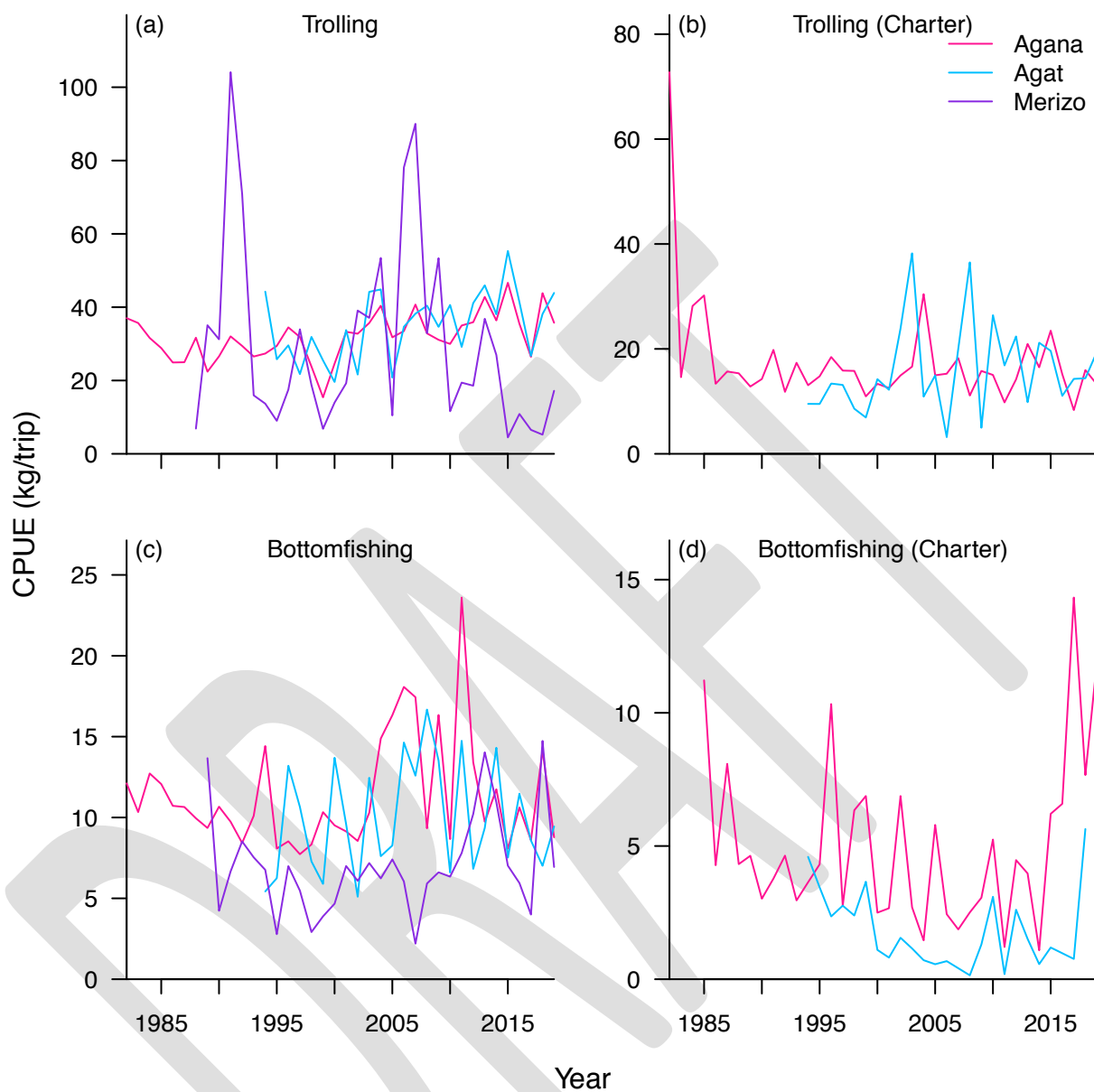
Figure 11. CPUE for trolling and bottomfishing (charter and non-charter) at individual ports based on the raw data from onsite interviews. Interviews from weekdays and weekends are aggregated together to compute the CPUE.
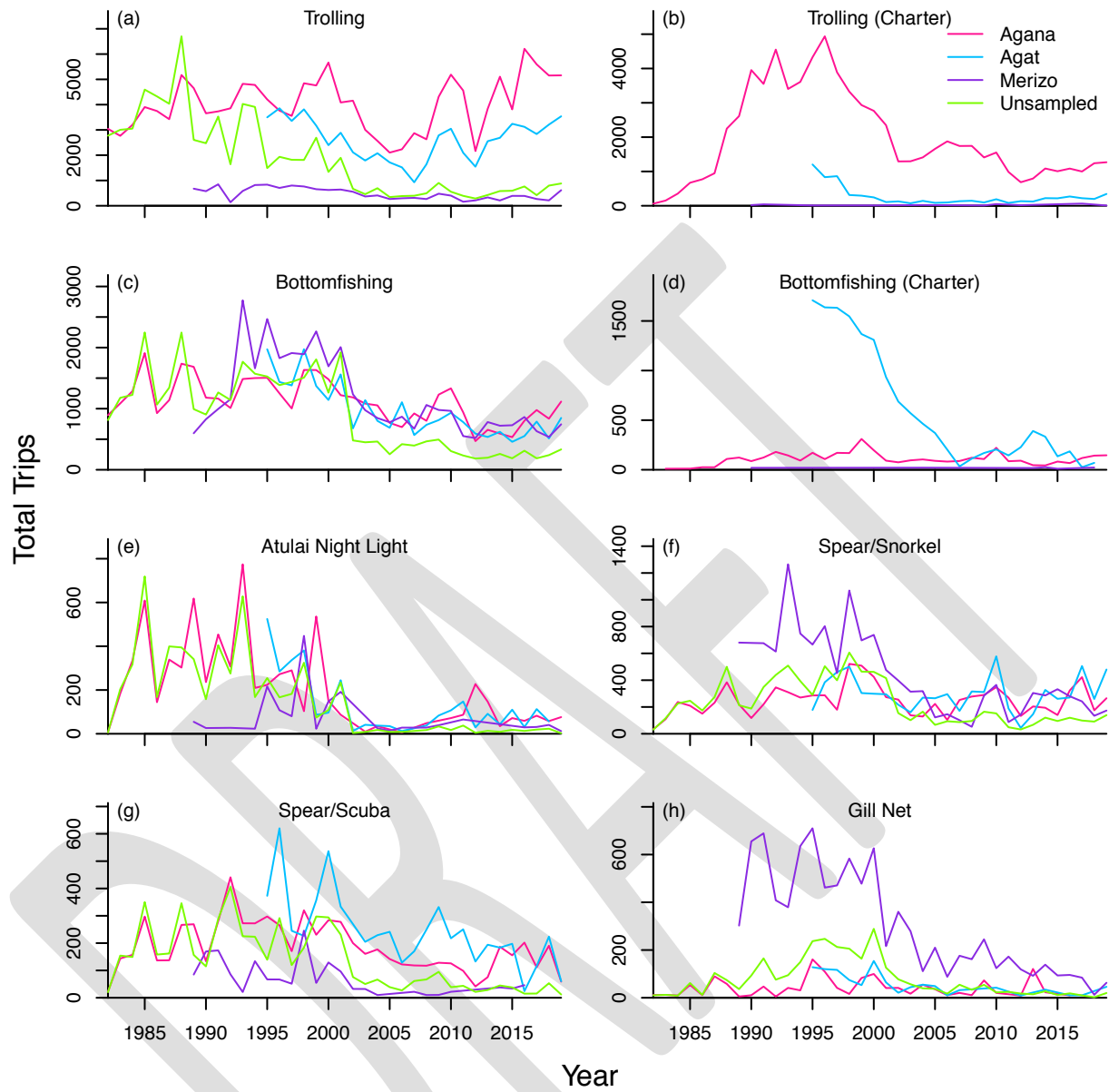
Figure 12. Total trips (charter and non-charter) by different fishing methods from 1982 to 2019. Unsampled ports in years prior to 1989 include all ports except for Agana. During 1989-1994, unsampled ports are the ports other than Agana and Merizo. For years in 1995 and afterwards, unsampled ports are the ports other than three presently sampled ports.
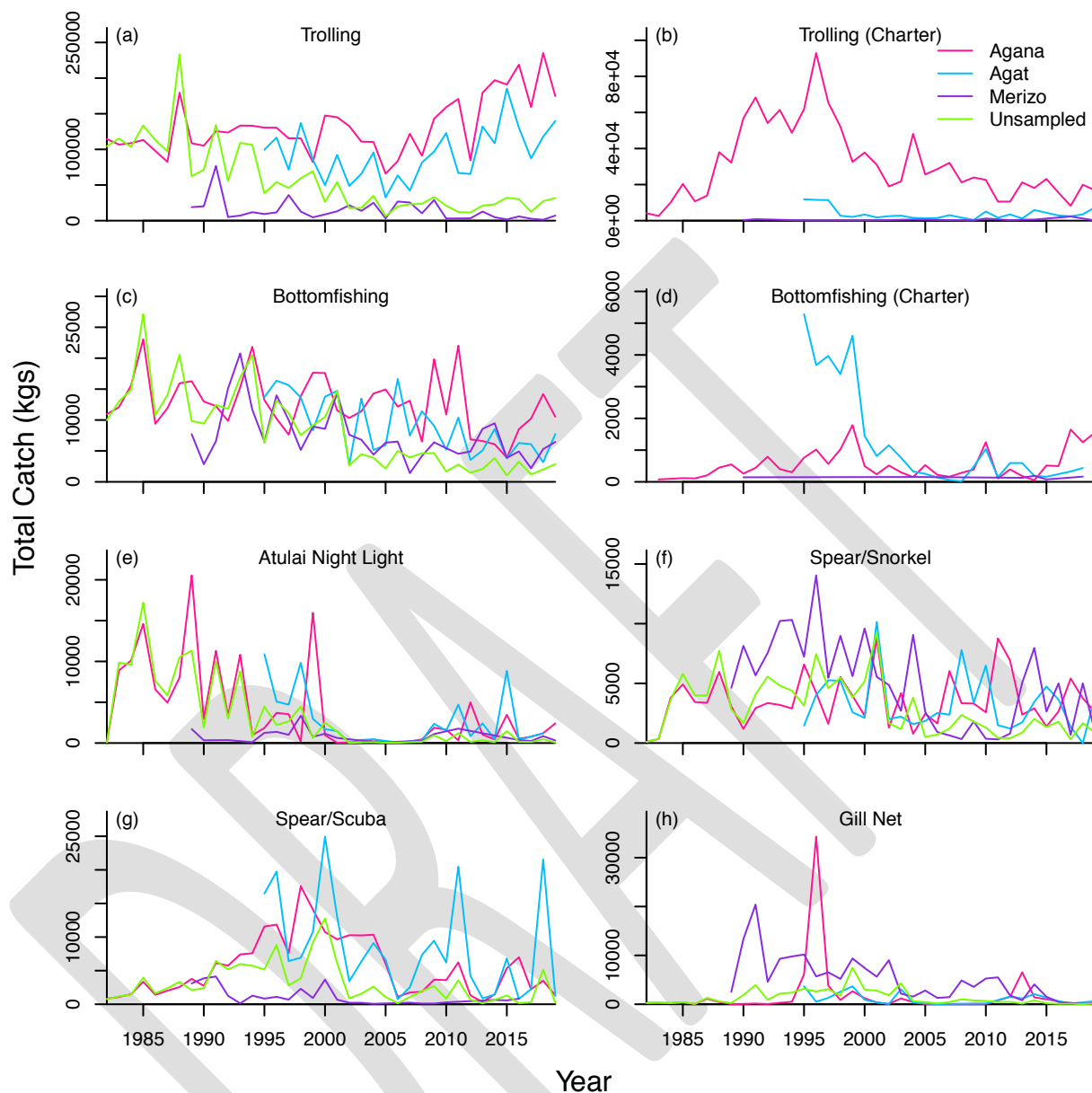
Figure 13. Total catch (charter and non-charter fishing) by different fishing methods from 1982 to 2019. Unsampled ports in years prior to 1989 include all ports except for Agana. During 1989-1994, unsampled ports are the ports other than Agana and Merizo. For years in 1995 and afterwards, unsampled ports are the ports other than three presently sampled ports.
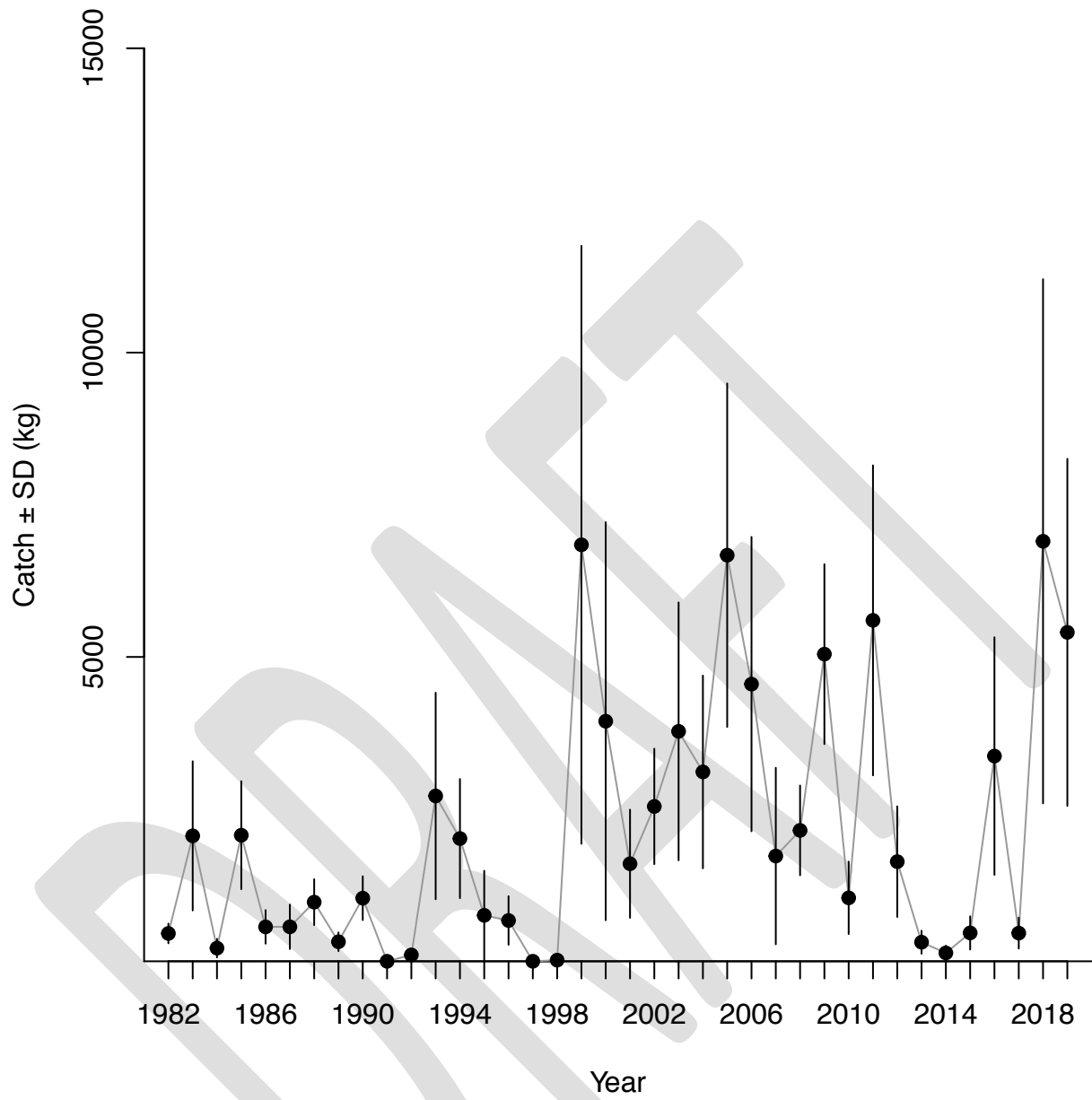
Figure 14. Onaga (*Etelis coruscans*) catch estimates from 1982 to 2019.
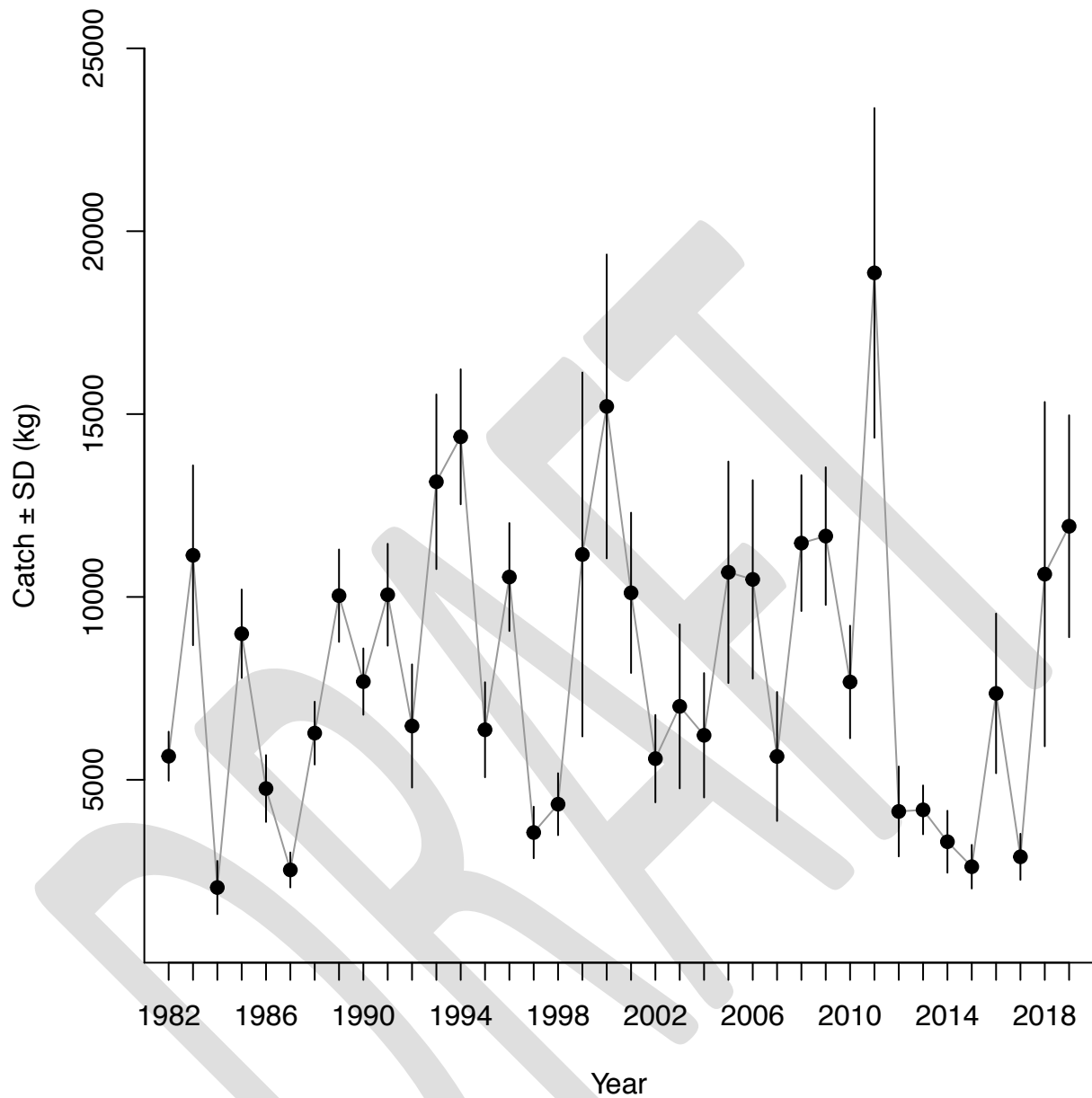
Figure 15. Deep bottomfish catch estimates from 1982 to 2019. The deep-water bottomfish include silver jaw jobfish (*Aphareus rutilans*), squirrelfish snapper (*Etelis carbunculus*), onaga (*E. coruscans*), yelloweye opakapaka (*Pristipomoides flavipinnis*), von Siebold's snapper (*P. sieboldii*), Brighams snapper (*P. zonatus*), yellowtail snapper (*P. auricilla*), and pink snapper (*P. filamentosus*).
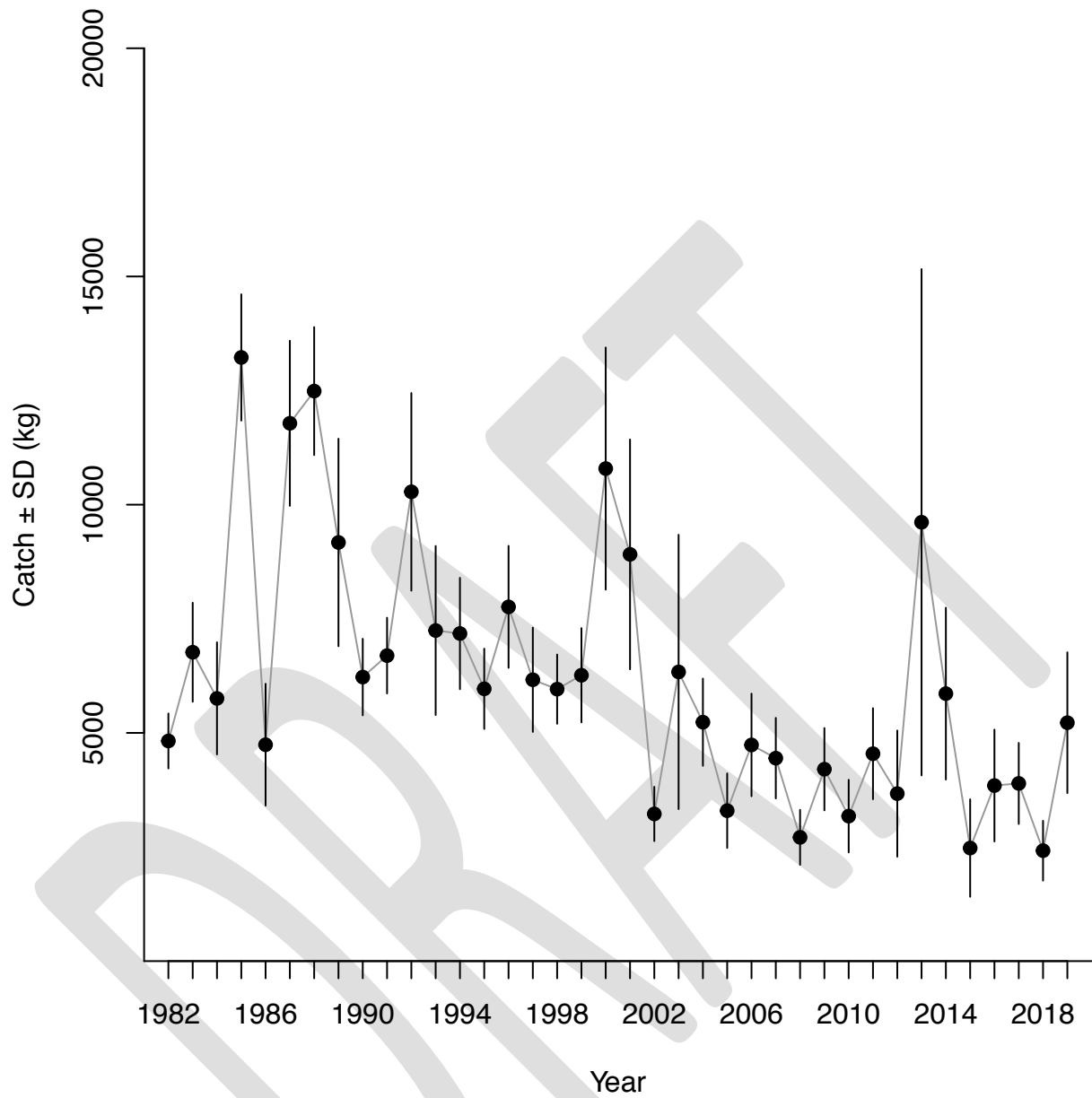
Figure 16. Shallow bottomfish catch estimates from 1982 to 2019. The shallow-water bottomfish include giant trevally (*Caranx ignobilis*), black trevally (*Caranx lugubris*), redgill emperor (*Lethrinus ruberioperculatus*), blueline snapper (*Lutjanus kasmira*), and lunartail grouper (*Variola louti*).
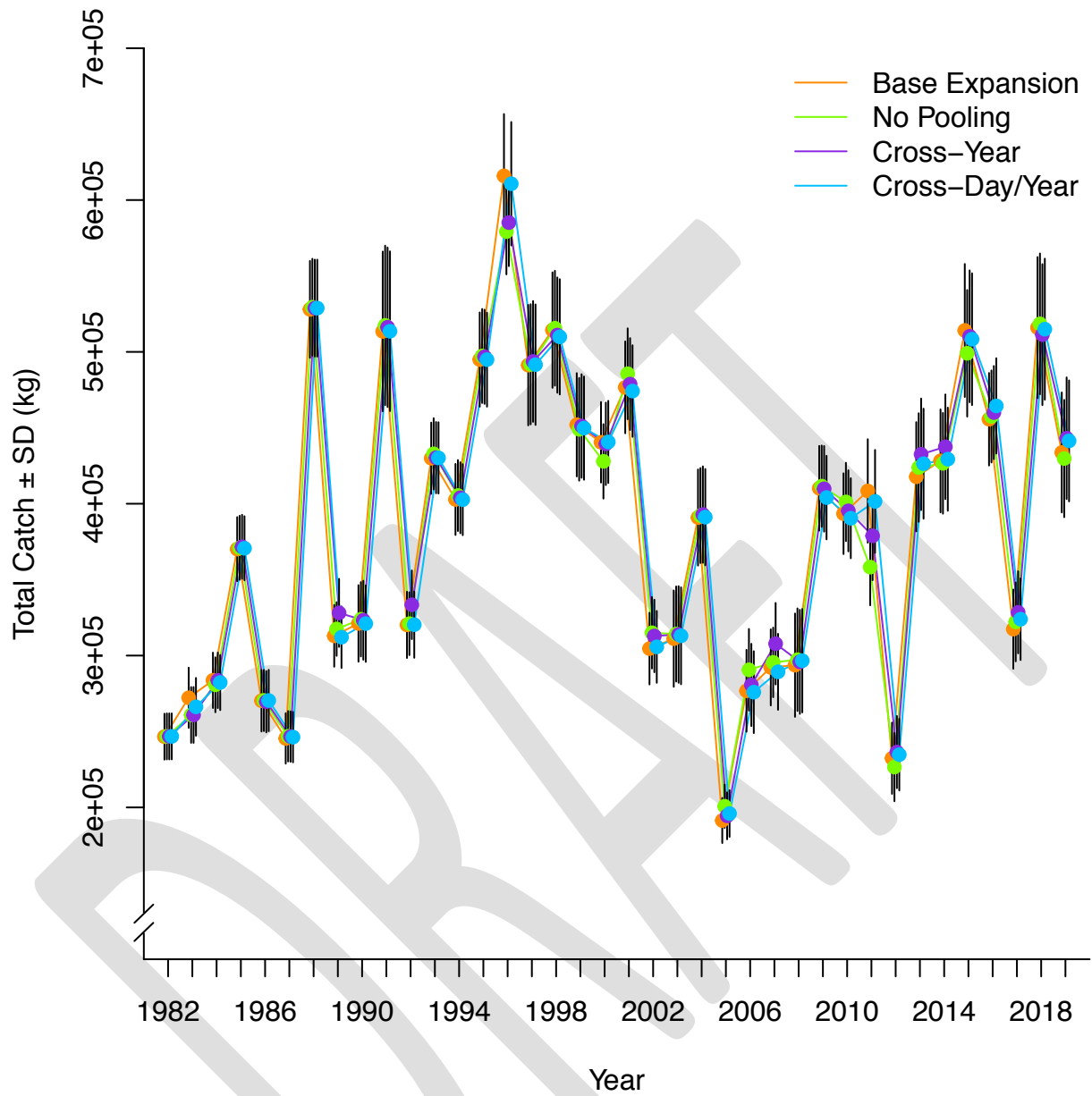
Figure 17. Total catch from 1982 to 2019 computed using the base expansion (full pooling algorithm) and three alternative interview pooling algorithms (No Pooling, Cross-Year, and Cross-Day/Year). In the No Pooling scenario, only existing interviews in a domain are used without borrowing from other domains.
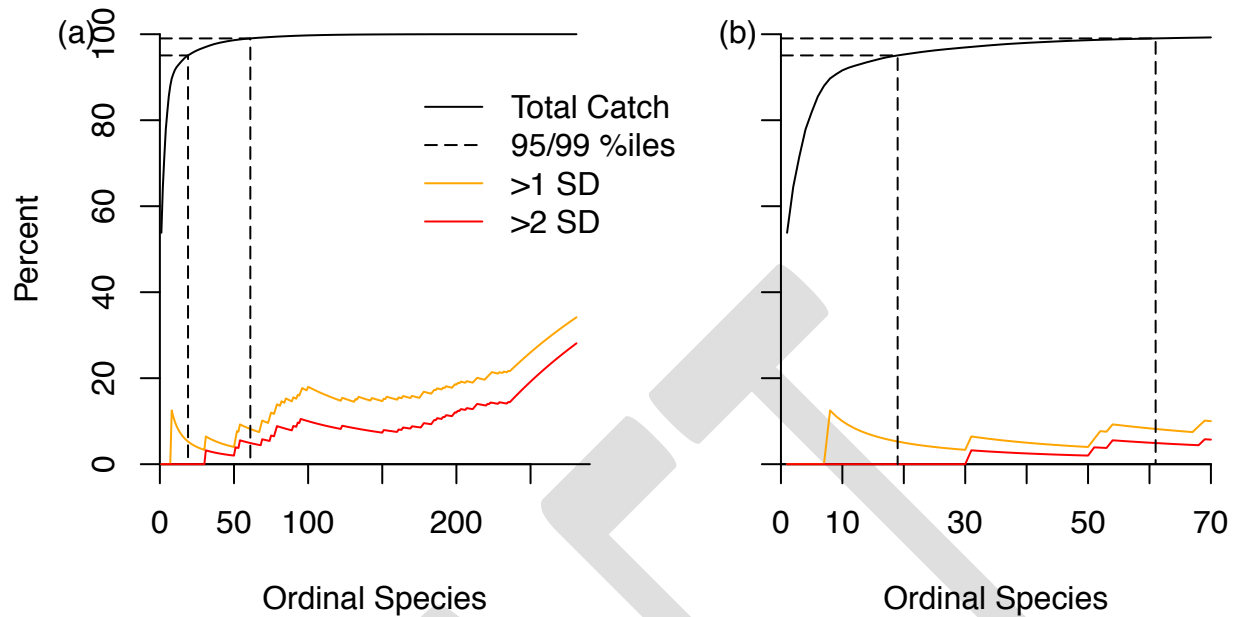
Figure 18. The percent of species for which species-level catch using an alternate interview pooling algorithm differed by more than one and more than two standard deviations from the base expansion estimate, averaged over 2017 to 2019. Species are ordered by decreasing catch along the x-axis, and the cumulative percent of total catch (black line) and percent of cumulative species differing by more than one and two standard deviations (orange and red lines) when using alternate pooling algorithms are displayed. Dashed lines indicate the number of species required to represent 95% and 99% of the total catch. (a) shows all species and (b) reduces the x-axis to magnify those species that represent approximately 99% of the total catch.
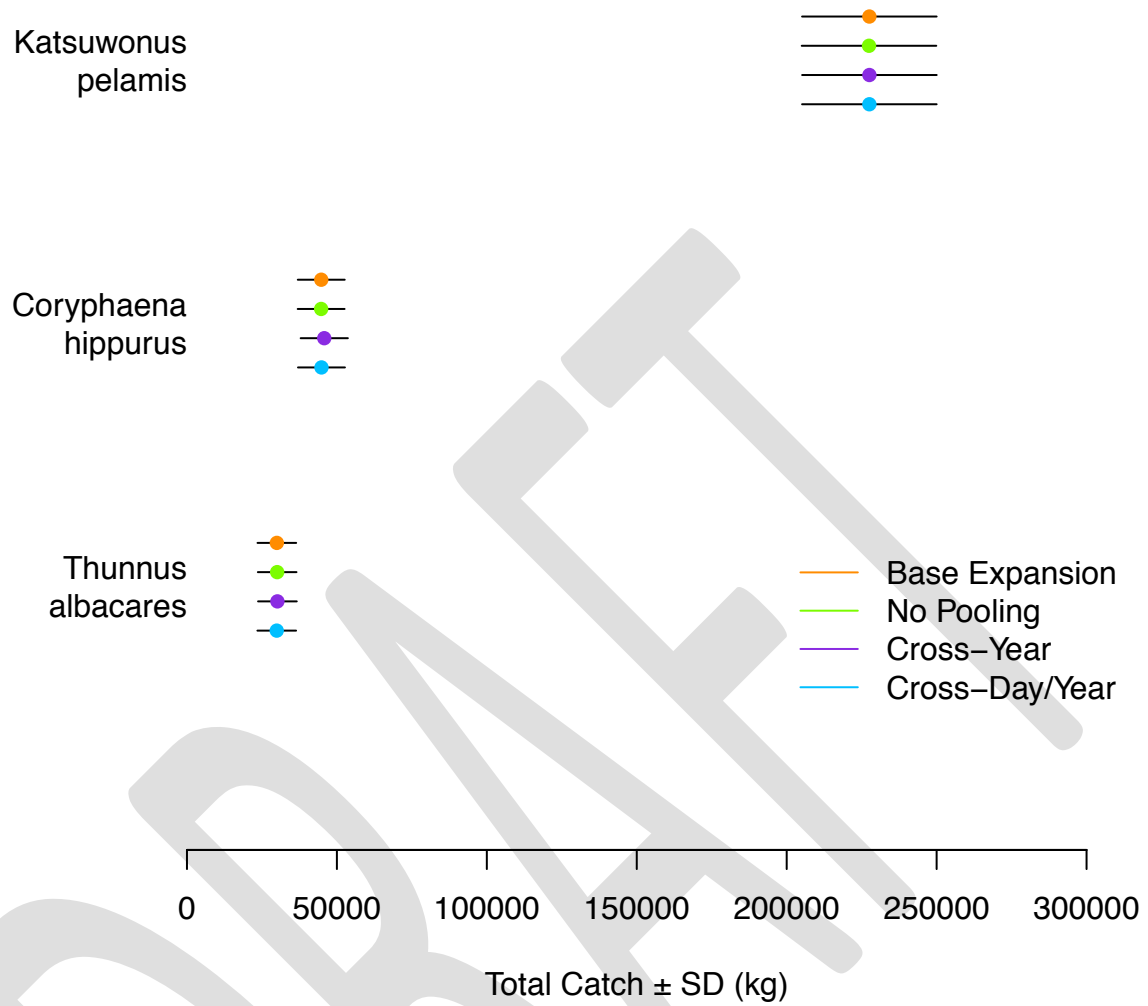
Figure 19. Species-level catch for the three most caught species, averaged over 2017 to 2019, computed using the base expansion and three alternative interview pooling algorithms.
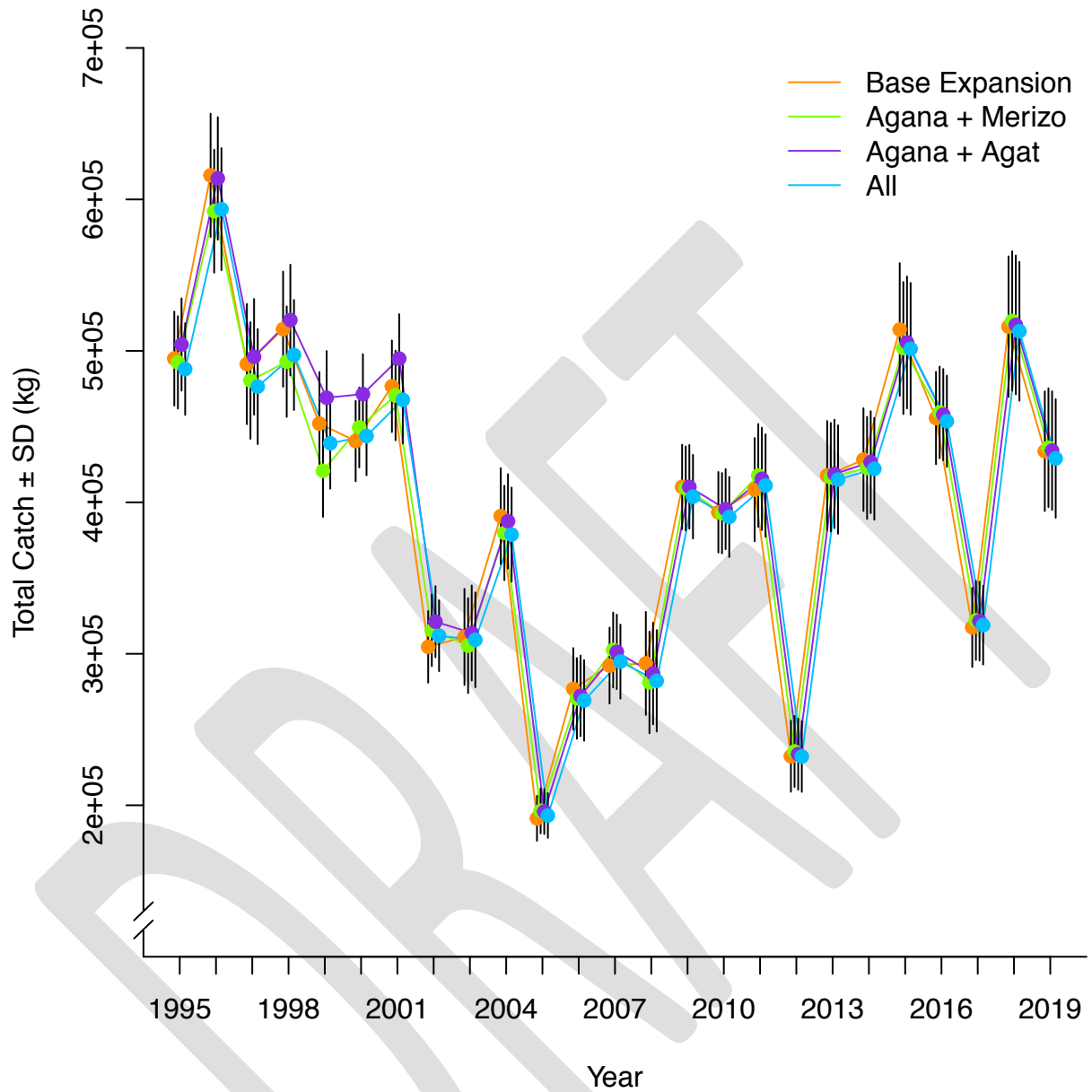
Figure 20. Total catch from 1995 to 2019 computed using the base expansion (Agat + Merizo) and three alternative representative port selections (Agana + Merizo, Agana + Agat, and All). In the Base Expansion scenario, Agat and Merizo are used as representative ports. In the All scenario, Agana, Agat, and Merizo are all used as representative ports.
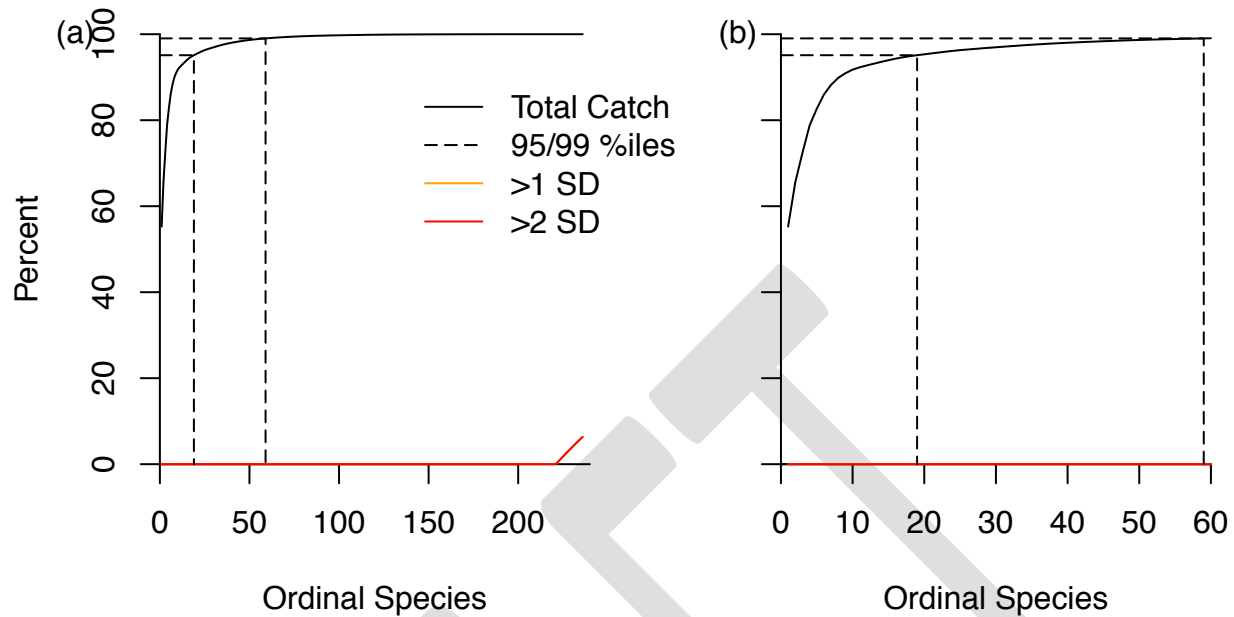
Figure 21. The percent of species for which species-level catch using an alternate representative port selection differed by more than one and more than two standard deviations from the base expansion estimate, averaged over 2017 to 2019. Species are ordered by decreasing catch along the x-axis, and the cumulative percent of total catch (black line) and percent of cumulative species differing by more than one and two standard deviations (orange and red lines) when using alternate representative port selections are displayed. Dashed lines indicate the number of species required to represent 95% and 99% of the total catch. (a) shows all species and (b) reduces the x-axis to magnify those species that represent approximately 99% of the total catch.
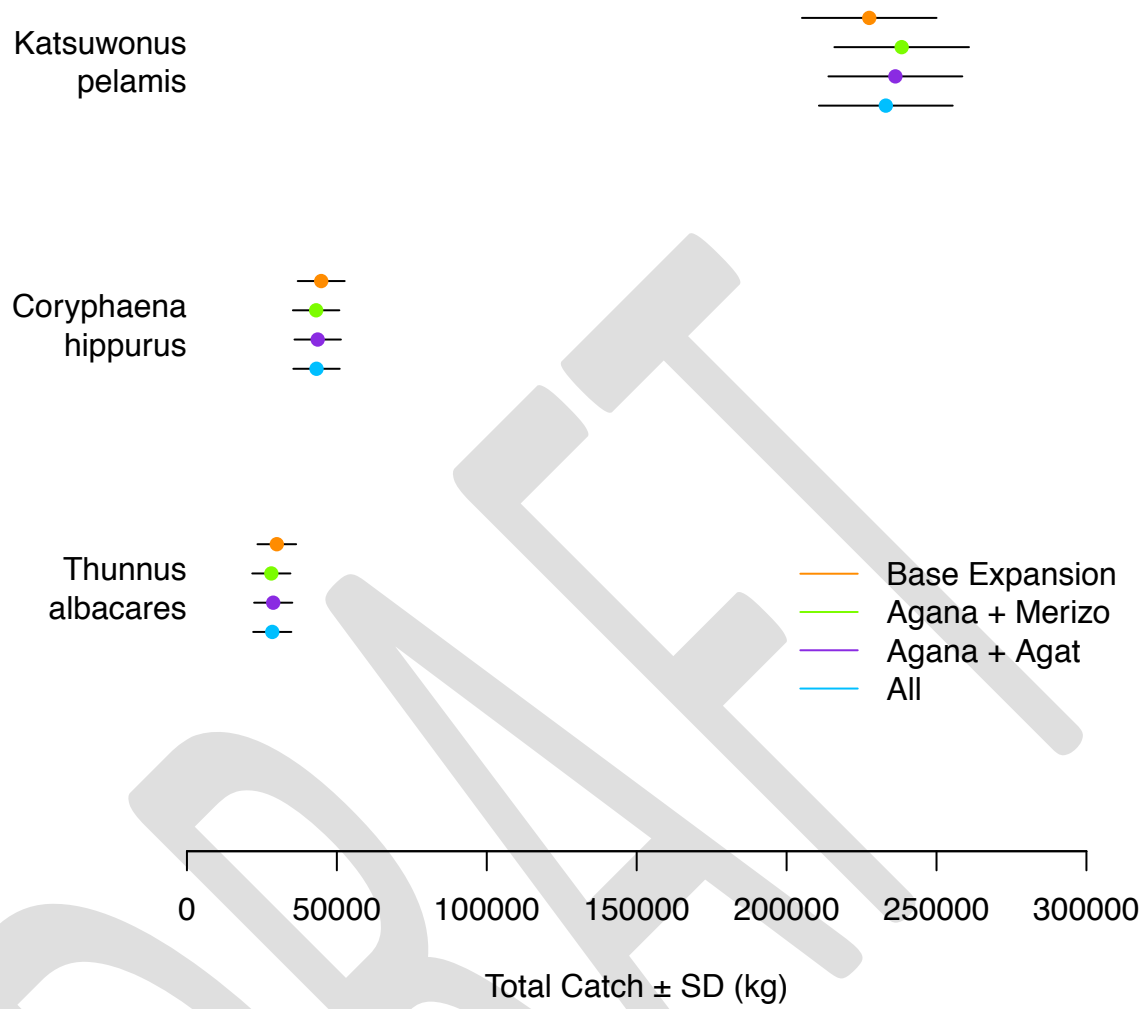
Figure 22. Species-level catch for the three most caught species, averaged over 2017 to 2019, computed using the base expansion and three alternative representative port selections.

Appendices

Appendix A. Supplementary tables and figures for the Guam boat-based expansion.

Table A. 1. Number of non-charter interviews at Agana. WD = weekday and WE = weekend. The last row (% <3) is the percentage of years with fewer than three interviews among years with non-zero interviews. This is slightly different from the percentages in Figure 7, which represent the percent of domains with positive effort and fewer than three interviews.

Method Year	Trolling		Bottomfishing		Atulai light		Mix spear		Spear/snorkel		Spear/scuba		Jigging		Gillnet		Castnet		Spincasting	
	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE
1982	115	240	30	67		1		1		5		1								3
1983	90	223	30	54	2	1			1	3	6	7								3
1984	131	216	43	81	4	1			4	16	5	7			1				1	2
1985	187	265	81	134	16	9	1		8	11	21	15			1	2			1	4
1986	82	227	11	88	5	2	1	1	4	10	2	6				1			1	1
1987	77	267	20	91	7	20	1	1	10	10	5	6				5			1	
1988	128	358	33	127	4	8			16	17	7	18			1	3		1		
1989	111	308	35	125	13	29		3	5	15	8	13			1	1	1		1	1
1990	92	264	33	97	4	13	1	1	2	10	5	9								
1991	74	277	22	90	7	17		1	4	10	13	12			1	2	2			4
1992	93	264	23	82	4	13			7	19	17	29		2		1				3
1993	132	308	36	75	14	11			8	17	12	16		1		1				1
1994	123	306	42	91	4	11			11	14	16	9			1		1			1
1995	101	258	39	88	6	5			8	15	12	18			5	3	2	1		
1996	95	234	38	50	4	2			8	12	10	10		1	2	2				1
1997	101	232	23	56	7	12	1	2	4	12	9	8			2	2	1			4
1998	126	349	36	104	1	4		1	17	29	14	18				3	1	1		
1999	109	326	31	84	5	8			14	15	16	6	1	5	3	2	1			3
2000	154	334	23	70		5			6	13	13	4	3	1	2	6	2			3
2001	116	197	25	55	2	1			4	3	9	5			1	3			1	
2002	85	153	20	51			1	1	2	5	3	1			1					4
2003	60	91	8	35		1			3	4	4	3		1						
2004	74	115	18	46	1	1			1	4	2	1	1		3					
2005	43	117	7	63		1			1	11	4	1			2	1				
2006	25	171	9	47								7	3							
2007	70	130	9	43					4	7			1				1			
2008	45	150	8	38	1				6	10		2	1	4						
2009	124	217	19	61	1	2			2	8		1		2				1		1
2010	168	251	15	85	1	5			5	8				1	1	1				1
2011	144	184	12	38	1	4			4	5				1			1			1
2012	72	87	6	14	5	11			3	2	1	2	3	1		2			3	4
2013	128	154	5	32		9			1	8		1			6	5				
2014	154	144	13	27		3		1	5	4		1			1					
2015	97	179	9	19	2			4		6	2			2						
2016	203	298	12	46		4			1	10	7	11		2			2			
2017	181	238	15	48	1	6			7	12	1	3	6	4						
2018	191	219	13	33		2			6	6	6	4	1	4						
2019	155	204	12	50		1			2	8	4		1	5						
(% < 3)	0	0	0	0	38	36	100	82	26	3	20	27	70	64	76	68	100	100	88	47

Table A. 2. Number of non-charter interviews at Agat. WD = weekday and WE = weekend. The last row (% <3) is the percentage of years with fewer than three interviews among years with non-zero interviews. This is slightly different from the percentages in Figure 7, which represent the percent of domains with positive effort and fewer than three interviews.

Method	Trolling		Bottomfishing		Atulai light		Mix spear		Spear/snorkel		Spear/scuba		Jigging		Gillnet		Castnet		Spincasting	
Year	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE
1994	3	17	1	4						1	1		1							
1995	49	116	11	36	4	14			1	10	11	6				2	1	1		3
1996	58	130	12	42	2	4		1	10	7	13	10			1	4		2	1	6
1997	44	111	8	38	1	7		2	5	11	4	5		1	1	5		3		1
1998	57	127	18	46		10			7	9	4			1	1	1	1		1	1
1999	74	115	19	33		2			5	5	14	4	1		2			1		4
2000	38	56	14	19		2			3	8	14	6	1		5	1				
2001	33	53	16	14	1	2			4		10	1			1				1	1
2002	23	44	4	22		1			6	4	2	4				2	2			1
2003	19	42	12	18	1				1	2	2	4			1				1	
2004	20	47	4	14						6	5	2				1		1		1
2005	11	54	2	23						5	5	2				1		1		1
2006	9	35	10	14						1		8		2		1		1		1
2007	5	26	1	13						1		1		1						
2008	11	46	5	19		2			1	9	3	1	1	1						
2009	36	88	8	23	1				1	8		2					1			
2010	39	64	3	25	2	3			6	4						2				
2011	21	63	7	20		5			1	4	3	1								
2012	11	48	4	11					1	1								1		
2013	47	58	2	10		3				1	1	1								1
2014	25	82	2	22		1	3		1	5	1				1					
2015	61	86	2	12	3	2	1	3	4	4	1									
2016	63	47	3	21	1					4	3	1	1							
2017	43	82	7	23	1	4		1	5	3						3		1		
2018	55	75	7	8	2				2	2	4	2				1				
2019	59	89	8	21						5		2						1		1
(% < 3)	0	0	23	0	82	47	50	75	45	24	45	59	100	83	88	83	100	88	100	70

Table A. 3. Number of non-charter interviews at Merizo. WD = weekday and WE = weekend. The last row (% <3) is the percentage of years with fewer than three interviews among years with non-zero interviews. This is slightly different from the percentages in Figure 7, which represent the percent of domains with positive effort and fewer than three interviews.

Method	Trolling		Bottomfishing		Atulai light		Mix spear		Spear/snorkel		Spear/scuba		Gillnet		Castnet		Spincasting	
Year	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE	WD	WE
1989	2	7	4	5	1				2	2	1		2	2				
1990	2	11	3	10		1			3	3		3	2	6		1		
1991	8	6	8	11					3	3	1		5	3	3	1		
1992	2	8	6	11					1	3		2	1	5	4			
1993	7	8	21	10					10	3		1	3	4	3	2	1	1
1994	10	11	14	17		1			7	4	2	3	6	5		3	1	
1995	6	27	25	42	2	3			13	9	2	3	14	11	6	1		3
1996	11	27	20	50		2			16	9	1	4	9	7	4	4	1	1
1997	12	18	18	27		1			6	3	1		10	7	2	6	1	1
1998	11	17	15	38	3	3			17	17	3	1	11	11	1	7		1
1999	5	9	15	36		1			8	3		1	3	4	1	2		6
2000	9	9	8	18		3			7	10	2	2	10	13	2	2		4
2001	11	17	14	48					2	6	2	2	4	4	1	3	1	5
2002	7	10	12	24					4	8		1	3	3	1		1	1
2003	6	6	6	13					4	8			2	3	1	2		
2004	5	7	2	13			2		3	2		1	1		1	5		1
2005	3	5	5	6								1	4	2	2	3		
2006	3	7	8	7	1				2	1					2			4
2007	2	4	1	5								1	2	2				
2008	3	7	5	11									2	6	1			
2009	4	6	10	10					6	3			4	3				
2010	3	6	6	9					2	2	1		1	2				
2011	2	4	6	6	2	1						1		1	1			
2012	2	10	1	7					1				1	1				
2013	3	3	2									2			5			1
2014	3	3	4	14				1	1	1	1		1	4				
2015	2	8	3	15				1	1	2			1	3				
2016	4	6	2	13			1			7		2	1	4		1		
2017	3	4	4	8					3	1			1	2		1		1
2018	2	11	7	10								3			1	2		1
2019	7	12	7	12	1	1			3	2								
(% <3)	26	0	16	0	83	70	100	100	33	39	92	67	50	31	71	59	100	64

Department of Agriculture, Guam
Division of Aquatic and Wildlife Resources
Fisheries Section

Page ___ of ___

BOAT-BASED SURVEY BOAT LOG

Interviewer   Start Time   End Time _____

Date _____   1 WD/2 WE

Port _____

AM:	/	/
PM:	/	/

Log No.	Int. No.	Depart Time	Return Time	Boat No. or Name	Fish? (Y/N/U)	Charter (Y/N/U)	Type of Activity	Vehicle License No.	Remarks
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									
28									
29									
30									
31									
32									
33									
34									
35									

Revised (02/20/2019)

Figure A. 1. Guam DAWR boat-based fishing survey boat log form.

Department of Agriculture, Guam
Division of Aquatic and Wildlife Resources
Fisheries Section

Page _____ of _____

BOAT-BASED CREEL INTERVIEW FORM

☐ Opportunistic

Date _____ 1 WD/2 WE

Port _____

Boat # _____ Charter (y/n/u) _____ Berthed (y/n/u) _____ Towing Vehicle's License # _____

Interview # _____

Interviewer _____

Interview time _____

Method	# of Gears	Hours Fished	Area Fished
1. Trolling	_____	_____	_____
2. Bottom (s, d, m)	_____	_____	_____
3. Atulai night jigging	_____	_____	_____
4. Snorkel spearfishing	_____	_____	_____
5. Scuba spearfishing	_____	_____	_____
6. Other	_____	_____	_____
7. Other	_____	_____	_____

No. of people on board _____

No. of guests (charter only) _____

Weather _____ Cloud cover _____

Wind direction _____ Speed _____

Tropical storm/Typhoon condition _____

Warnings: Small craft (y/n/u) _____

High surf (y/n/u) _____

Species/Code	Length (mm)	Wt. (kg)	Length (mm)	Wt. (kg)	Length (mm)	Wt. (kg)	Total No. Act.	Total No. Est.	Total Weight Act.	Total Weight Est.

Bycatch: Did you release or throw back any fish? ( ) No   ( ) Yes (if yes, list below)

Species/Code	Method	Bycatch type Alive (1)	Bycatch type Dead (2)	Length (mm)	Wt. (kg)	Length (mm)	Wt. (kg)	Total No. Act.	Total No. Est.	Total Weight Act.	Total Weight Est.

Disposition of fish landed:

Method	% Not Sold	% Sold	Buyer

Remarks:

Shark interaction: ( ) No   ( ) Yes

Trip Cost Information: Refused   ( ) No   ( ) Yes

Gallons of fuel used	gallons
Fuel price per gallon	$
Cost of ice used	$
Cost of bait/chum	$
Cost of gear lost	S
Engine type	2 stroke 4 stroke Diesel
GPS onboard	yes no
Used fish finder	yes no
Used electric/hydraulic gear	yes no

Revised (02/20/2019)

Figure A. 2. Guam DAWR boat-based fishing survey interview form.
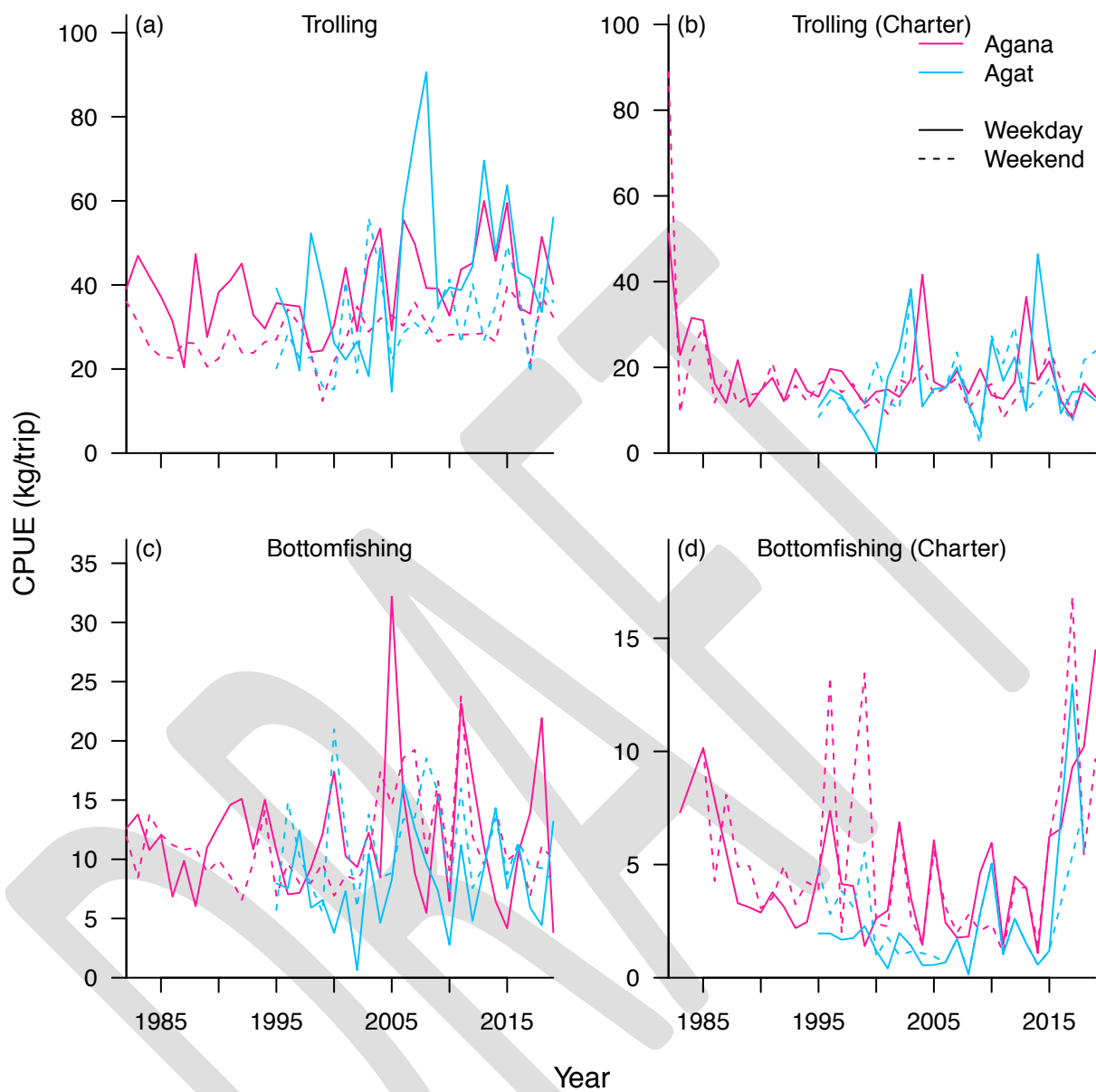
Figure A. 3. CPUE (charter and non-charter) of trolling and bottomfishing between weekdays and weekends at Agana and Agat, based on data after expansion (i.e. after interview pooling when/if it is needed).
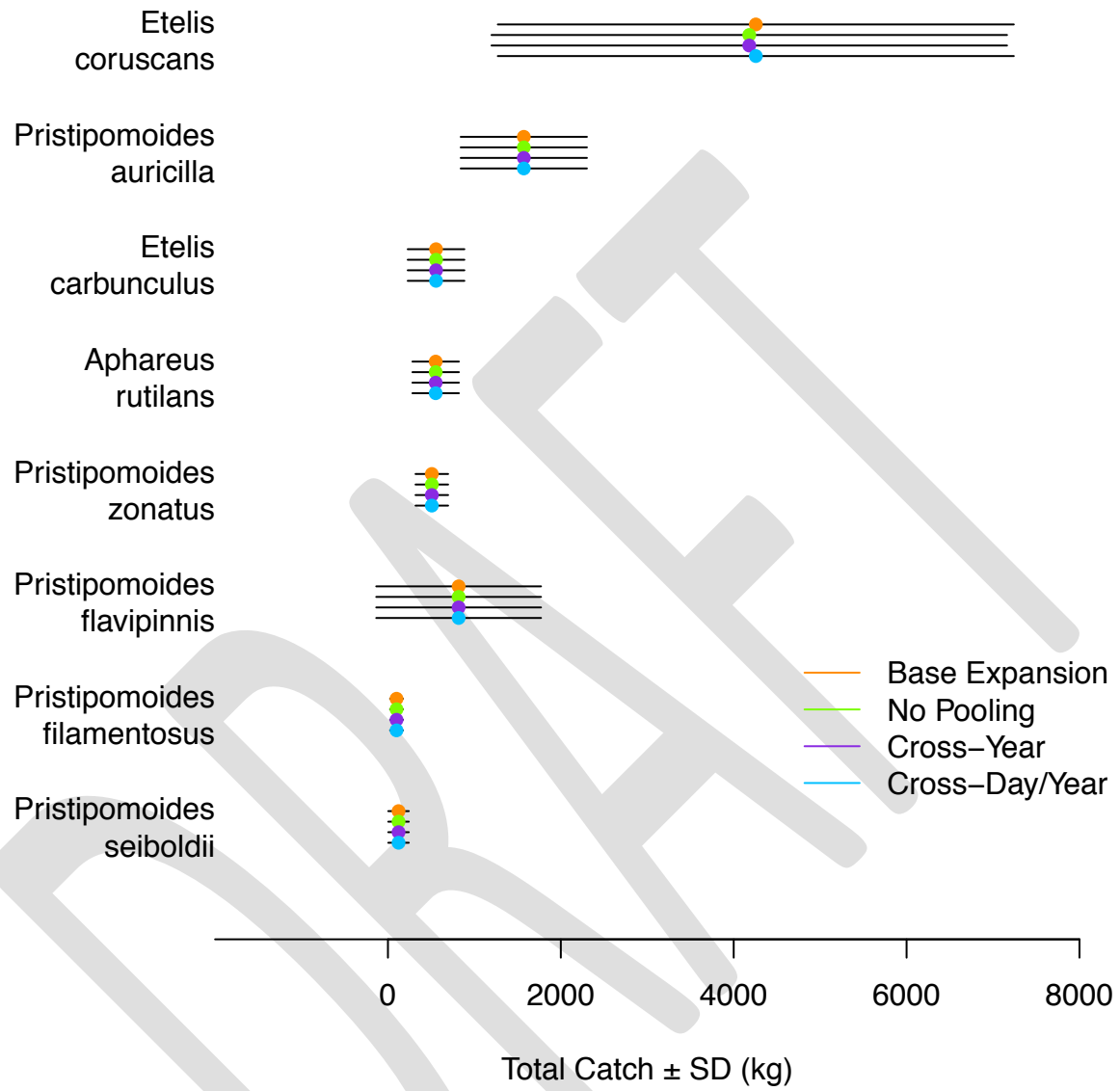
Figure A. 4. Species-level catch for the most caught deep bottomfish species, averaged over 2017 to 2019, computed using the base expansion and three alternative interview pooling algorithms (No Pooling, Cross-Year, and Cross-Day/Year).
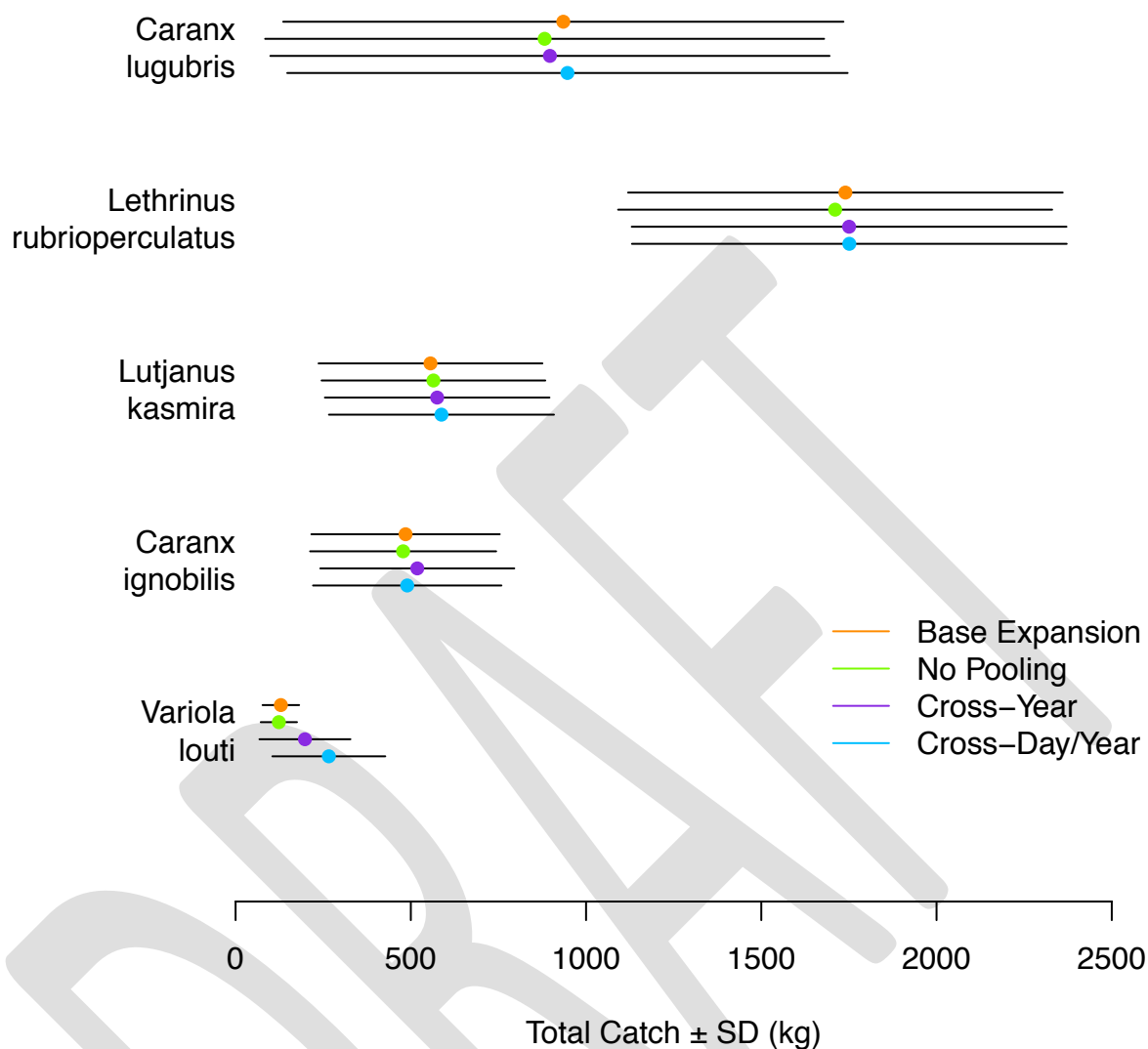
Figure A. 5. Species-level catch for the most caught shallow bottomfish species, averaged over 2017 to 2019, computed using the base expansion and three alternate interview pooling algorithms (No Pooling, Cross-Year, and Cross-Day/Year).
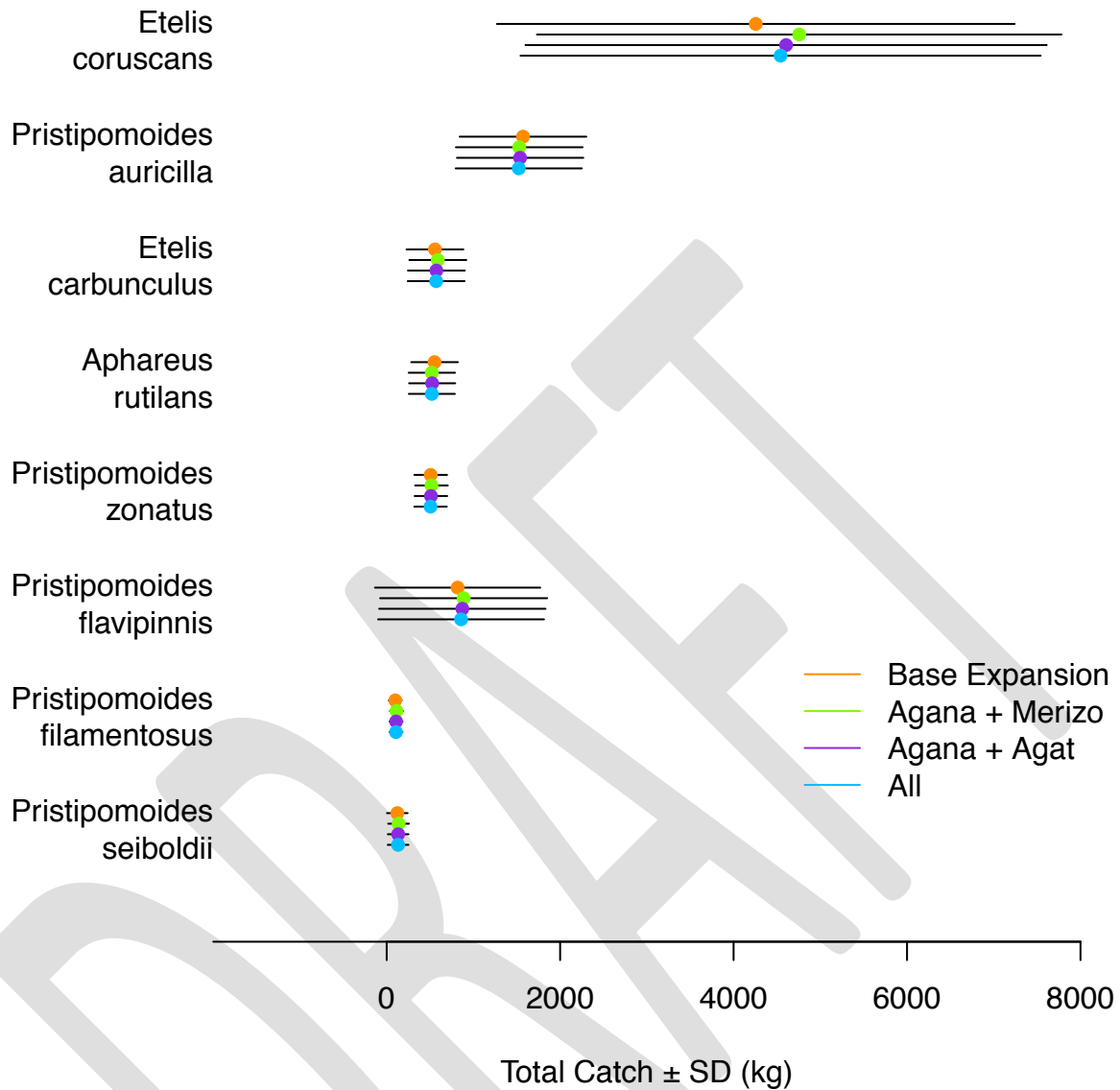
Figure A. 6. Species-level catch for the most caught deep bottomfish species, averaged over 2017 to 2019, computed using the base expansion and three alternate representative port selections (Agana + Merizo, Agana + Agat, and All).
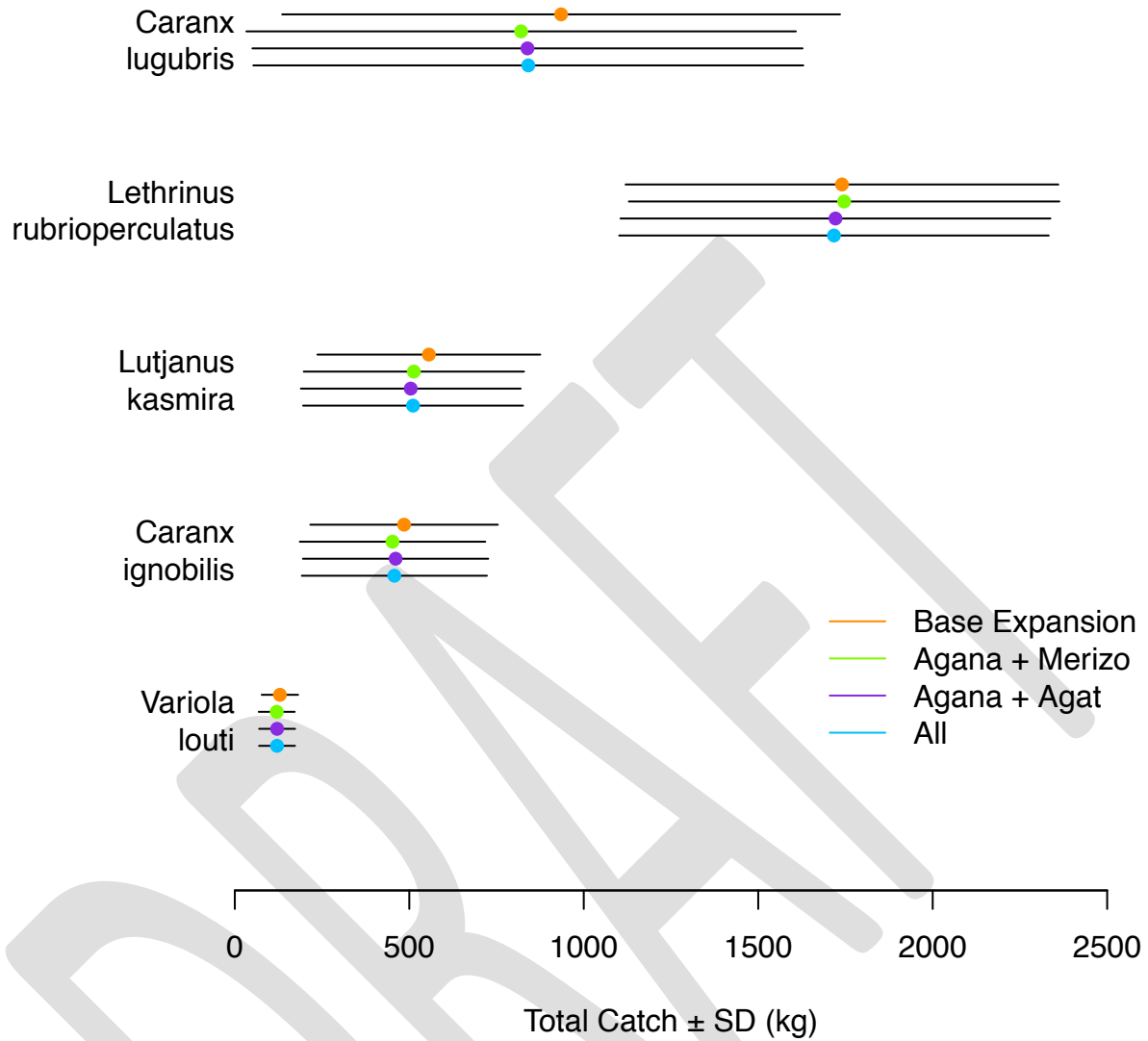
Figure A. 7. Species-level catch for the most caught shallow bottomfish species, averaged over 2017 to 2019, computed using the base expansion and three alternate representative port selections (Agana + Merizo, Agana + Agat, and All).

Appendix B. Supplemental notes on the available resources, scheduling process, survey coverage, on-the-ground implementation, and data expansion for the Guam, CNMI, and American Samoa shore-based surveys.

Guam shore-based survey

Summary: Effort and catch rate surveys are conducted separately. Effort surveys are stratified by type of day, with two effort shifts that both encompass all three catch rate survey regions on each survey day. The starting location and direction of travel for the effort survey are previously determined. Catch rate surveys are stratified by type of day and region, again with two shifts on each survey day. However, an effort and catch rate survey cannot be scheduled on the same day, and no more than one of either survey can be scheduled on any day. Expansion domains are then specified by the type of day, time of day (morning [06:00-18:00] and night [18:00-02:00], characterized by the 06:30 and 19:00 shifts, respectively), and fishing method. Expansion domains for the hook and line fishing method are further divided by region.

Available resources
- 3-4 staff whose primary task is to conduct shore-based and boat-based surveys, though not full-time
- Another ~6 higher-level staff who occasionally conduct surveys
- One staff is assigned per shift
- 2 regularly available vehicles, shared with shore-based surveys

Scheduling process (monthly)
1. Schedule 4 effort survey days. Randomly select 2 weekdays and 2 weekend days without replacement.
   a. Select a random start location from the list of effort survey areas for each survey day.
   b. Go through the survey days from the first day to the last, alternating the direction between clockwise and counterclockwise, ignoring the type of day.
   c. On each survey day two shifts are conducted, one starting at 06:30 and another starting at 19:00. Shifts cover all three survey regions in a period of about 4-5 hours.
2. Schedule 1 catch rate survey day for region 1. The type of day will be the opposite of the type for the region 1 catch rate survey day during the previous month. Randomly select a day of this type that does not already have a shore-based survey scheduled.
   a. Two shifts are conducted on this survey day, one from 06:30-12:00 and another from 19:00-24:00.
3. Schedule 1 catch rate survey day for region 2. This type of day will be the opposite of the type for the region 2 catch rate survey day during the previous month, which is also the opposite of the type for the region 1 catch rate survey day during this month. Randomly select a day of this type that does not already have a shore-based survey scheduled.
   a. Two shifts are conducted on this survey day, one from 06:30-12:00 and another from 19:00-24:00.
4. Schedule 2 catch rate survey days for region 3. Randomly select 1 weekday and 1 weekend day that do not already have a shore-based survey scheduled.

a. Two shifts are conducted on this survey day, one from 06:30-12:00 and another from 19:00-24:00.

Survey coverage
- Weekdays
  - 2 weekday effort survey days per month (with 2 shifts each survey day and covering all three regions) and 2 weekday interview survey days per month (with 2 shifts each survey day and always one survey day for region 3, with the other survey day for either region 1 or 2), though effort and interview surveys cannot be conducted on the same day
  - Approximately 20 weekdays per month (represented as the number of weekdays per year, n_c, in EQ 1 and EQ 2), depending on the number of holidays
- Weekends/holidays
  - Assume that fishing effort and CPUE are similar across Saturdays, Sundays, and holidays, but only able to sample on Saturdays and Sundays
  - 2 weekend effort survey days per month (with 2 shifts each survey day and covering all three regions) and 2 weekend interview survey days per month (with 2 shifts each survey day and always one survey day for region 3, with the other survey day for either region 1 or 2), though effort and interview surveys cannot be conducted on the same day
  - Approximately 10 weekend/holidays per month (represented as the number of weekend days per year, n_c, in EQ 1 and EQ 2), depending on the number of holidays; only 8-9 weekend days are actually available for sampling

On-the-ground implementation
- Effort
  1. Drive to the designated starting point within the survey route.
  2. Begin driving in the designated direction along the survey route.
  3. Record the fishing method and number of gears for any shore-based fishing activity observed.
  4. The survey is complete when back at the starting point.
- Catch rate
  1. Drive back and forth within the region, speaking with any fishers in the process of (but have also been active for at least 30 minutes) or that have just finished fishing to collect data on their effort and catch. Emphasize interviews for fishers engaging in less frequently observed fishing methods.

Data expansion basics
- An expansion domain is specified by the:
  - Type of day (weekday or weekend)
  - Time of day (12 day hours represented by the 06:30 shifts and 8 night hours represented by the 19:00 shifts)
  - Region, but only for the hook and line fishing method
  - Fishing method

CNMI shore-based survey

Summary: Effort and catch rate surveys are conducted on alternating two-hour runs during the same six-hour shift. Surveys are stratified by type of day and shift, within logistical restraints that effectively prevent back-to-back surveys. For each survey the end of the route to begin at and the type of first run (effort or catch rate) are previously determined. Expansion domains are then specified by the two stratifying variables (type of day and shift, which is later summarized to day and night) as well as the fishing method.

Available resources
- 3-4 staff whose primary task is to conduct shore-based and boat-based surveys, though not full-time
- Two staff are assigned for each shift (one to drive and one to record)
- 1 regularly available vehicle

Scheduling process (quarterly)
1. Schedule 16 weekend surveys. Iterate through the combinations 4 shifts (00-06, 06-12, 12-18, 18-24) and 2 run orders (IPI and PIP, where I = interview = catch rate survey and P = participation = effort survey). For each combination select two random days on which to schedule the resulting survey, with the following restrictions according to the shift:
   a. 00-06 shift: the survey cannot occur the day after a boat-based survey or a 18-24 shore-based survey, or on the same day as any other survey except for a 12-18 or 18-24 shore-based survey
   b. 06-12 shift: the survey cannot occur on the same day as any other survey except for a 18-24 shore-based survey
   c. 12-18 shift: the survey cannot occur on the same day as any other survey except for a 00-06 shore-based survey
   d. 18-24: the survey cannot occur the day before a 00-06 shore-based survey, or on the same day as any other survey except for a 00-06 or 06-12 shore-based survey
2. Schedule 16 weekday surveys. Iterate through the same 8 combinations of shift and run orders and for each combination select two random days on which to schedule the resulting survey. The previous shift-specific restrictions still apply and, more restrictively, the surveys cannot occur on the same day as any other survey.
3. Assign the starting point for each survey. Go through the surveys from the first day to the last, alternating the starting point between north and south, ignoring the type of day.

Survey coverage
- Weekdays
  o 16 weekday shifts per quarter; up to two shifts can be scheduled on the same day, so fewer than 16 distinct weekdays may have surveys – also note that these weekend surveys are being scheduled in conjunction with 9 boat-based weekday surveys per quarter and shore-based and boat-based surveys cannot occur on the same day
  o Approximately 60 weekdays per quarter (represented as the number of weekdays per year during effort expansion)

- Weekends/holidays
  - Assume that fishing effort and CPUE are similar across Saturdays, Sundays, and holidays, but only able to sample on Saturdays and Sundays
  - 16 weekend shifts per quarter; up to two shifts can be scheduled on the same day, so fewer than 16 distinct weekdays may have surveys – also note that these weekend surveys are being scheduled in conjunction with 9 boat-based weekend surveys per quarter and shore-based and boat-based surveys cannot occur on the same day
  - Approximately 30 weekend/holidays per quarter (represented as the number of weekend days per year during effort expansion), depending on the number of holidays; only approximately 26 weekend days are actually available for sampling

On-the-ground implementation
1. Drive to the designated starting point (either the north or south end of the region).
2. Begin the first run (either interview or participation, as dictated by the order). Each run is scheduled for two hours.
   a. Interview: Drive along the designated survey route and speak with any fishers in the process of or that have just finished fishing to collect data on their effort and catch. Emphasize interviews for fishers engaging in less frequently observed fishing methods.
   b. Participation: Drive along the designated survey route and make visual observations of shore-based fishing activity occurring, recording the fishing method and number of gears.
3. Conduct the second run, driving back toward the original starting point. This run is of the opposite type of the first run (i.e. if the first run was for effort, the second run is for catch rate, and vice versa).
4. Conduct the third run, beginning from the original starting point. This run is of the same type as the first run.

Data expansion basics
- An expansion domain is specified by the:
  - Type of day (weekday or weekend)
  - Time of day (day = 06-18 or night = 18-06); effort data is retained at the shift level, but CPUE data is aggregated into day and night averages
  - Fishing method
  - Note that there is only a single survey region encompassing the western lagoon

American Samoa shore-based survey

Summary: Effort and catch rate surveys are conducted on alternating one-hour runs during the same six-hour survey period. Surveys are stratified by type of day, shift, and region, within logistical restraints that prevent survey teams from working two shifts without a break in between and further restraints that prevent two survey teams from being in the same region at the same time. For each survey the type of first run (effort or catch rate) is randomized. Expansion domains are then specified by the three stratifying variables (type of day, region, and shift, which is later summarized to day [06:00-18:00] and night [18:00-06:00]) as well as the fishing method.

Available resources
  - 4 staff who conduct shore-based surveys full time, with 4 additional staff who conduct boat-based surveys full time
  - Two staff are assigned for each shift (one to drive and one to record)
  - 2 regularly available vehicles

Scheduling process (monthly)
  1. Schedule 2 weekend survey days independently for each of the 2 survey teams. Weekend surveys only occur on Saturdays. Randomly pick two Saturdays for each team, and for each survey team-day randomly assign a shift (06:30-13:00, 10:30-19:00, 16:30-24:00; these shifts have changed regularly over time and represent the worker shifts, with only 6 hours actually dedicated to surveys and the rest used for transportation and a meal break) and region (west, central, east) within the restriction that both teams cannot be in the same region during identical or overlapping shifts.
  2. For each weekend survey day, randomly select a weekday off during the corresponding week for that survey team.
  3. Schedule the weekday surveys for each of the 2 survey teams. For each weekday, assign each survey team that is not off on that day a random shift and region within the restrictions for weekend surveys and furthermore that survey teams need at least 7 hours off between shifts.
  4. For each survey, randomly assign the type of first run (effort or catch rate).

Survey coverage
  - Weekdays
    o 2 shifts (one for each team and randomly assigned across 3 regions) on each weekday, minus 4 shifts per month (two for each team) that are moved to Saturdays; these 4 shifts may be on 2-4 weekdays
    o Approximately 20 weekdays per month (represented as the number of weekdays per year, n_c, in EQ 5 and EQ 6),
  - Weekends/holidays
    o Assume that fishing effort and CPUE are similar on Saturdays and holidays, but only able to sample on Saturdays
    o Assume that no fishing occurs on Sundays since it is a day for church
    o 4 Saturday shifts (two for each team and randomly assigned across 3 regions) each month; these shifts may be on 2-4 Saturdays

      o Approximately 5-6 Saturdays/holidays per month (represented as the number of weekend days per year, n_c, in EQ 5 and EQ 6), depending on the number of holidays; only 4-5 Saturdays are actually available for sampling

On-the-ground implementation
1. Drive to the closer end of the survey region.
2. Begin the first run (either effort or catch rate, as selected during scheduling). Each run is scheduled for one hour.
    a. Catch rate: Drive along the designated survey route and speak with any fishers in the process of or that have just finished fishing to collect data on their effort and catch. Emphasize spearfishing interviews.
    b. Effort: Drive along the designated survey route and make visual observations of shore-based fishing occurring, recording the fishing method and number of gears.
3. Continue driving between the two ends of the survey region, alternating catch rate and effort runs until six runs have been completed.

Data expansion basics
- An expansion domain is specified by the:
    o Type of day (weekday or weekend)
    o Time of day (day = 06-18 or night = 18-24)
    o Region
    o Fishing method

Appendix C. Supplemental notes on the available resources, scheduling process, survey coverage, on-the-ground implementation, and data expansion for the Guam, CNMI, and American Samoa boat-based surveys.

Guam boat-based survey

Summary: Three main ports are surveyed for effort and catch rate on separate days, and additional trailer count surveys cover all three main ports and lesser boat launch areas to estimate island-wide effort. Port surveys are stratified by type of day and port, with two shifts on each day. Trailer count surveys are stratified by type of day and have two shifts on each day. Multiple port surveys cannot occur on the same day, but a port survey can occur on the same day as a trailer count survey. Expansion domains are then specified by type of day, port, charter status, and fishing method.

Available resources
- 3-4 staff whose primary task is to conduct boat-based and shore-based surveys, though not full-time
- Another ~6 higher-level staff who occasionally conduct surveys
- One staff is assigned per shift
- 2 regularly available vehicles, shared with shore-based surveys

Scheduling process (monthly)
1. Schedule 4 trailer count survey days (conducted together with the shore-based effort survey). Randomly select 2 weekdays and 2 weekend days without replacement.
   a. Select a random start location from the list of shore-based effort survey areas for each survey day.
   b. Go through the survey days from the first day to the last, alternating the direction between clockwise and counterclockwise, ignoring the type of day.
   c. On each survey day two shifts are conducted, one starting at 06:30 and another starting at 19:00. Shifts cover all three shore-based survey regions in a period of about 4-5 hours.
2. Schedule 4 Agana Boat Basin survey days. Randomly select 2 weekdays and 2 weekend days without replacement.
   a. Two shifts are conducted on these survey days, one from 05:00-12:00 and another from 16:00-24:00.
3. Schedule 2 Agat Marina survey days. Randomly select 1 weekday and 1 weekend day that do not already have a port survey scheduled.
   a. Two shifts are conducted on these survey days, one from 05:30-12:00 and another from 16:00-24:00.
4. Schedule 2 Merizo Pier survey days. Randomly select 1 weekday and 1 weekend day that do not already have a port survey scheduled.
   a. Two shifts are conducted on these survey days, one from 06:00-12:00 and another from 16:00-24:00.

Survey coverage
- Weekdays

- o 2 weekday trailer count survey days per month (with 2 shifts each survey day) and 4 weekday port survey days per month (with 2 shifts each survey day and split 2:1:1 across three ports)
  - o Trailer count surveys can be scheduled on the same day as port surveys, but only a single port survey can be scheduled on each day
  - o Approximately 20 weekdays per month (represented as the number of weekdays per year, n_c, in EQ 19 and EQ 21), depending on the number of holidays
- Weekends/holidays
  - o Assume that fishing effort and CPUE are similar across Saturdays, Sundays, and holidays, but only able to sample on Saturdays and Sundays
  - o 2 weekend trailer count survey days per month (with 2 shifts each survey day) and 4 weekend port survey days per month (with 2 shifts each survey day and split 2:1:1 across three ports)
  - o Trailer count surveys can be scheduled on the same day as port surveys, but only a single port survey can be scheduled on each day
  - o Approximately 10 weekend/holidays per month (represented as the number of weekend days per year, n_c, in EQ 19 and EQ 21), depending on the number of holidays; only 8-9 weekend days are actually available for sampling

On-the-ground implementation
- Port survey
  1. While at the port, maintain a boat log of all trips departing or returning, including their fishing methods. Also attempt to conduct interviews for all returning trips to collect information on effort and catch.
- Trailer count
  1. Drive to the designated starting point within the survey route.
  2. Begin driving in the designated direction along the survey route.
  3. Record the location of any fishing trailers observed.
  4. The survey is complete when back at the starting point

Data expansion basics
- An expansion domain is specified by the:
  - o Type of day (weekday or weekend)
  - o Port
  - o Charter status (charter or non-charter)
  - o Fishing method

CNMI boat-based survey

Summary:
Three main ports are surveyed for fishing method composition and catch rate on separate days, while the number of fishing trailers at all ports and boat ramps is recorded each survey day. Trailer counts are used to estimate overall fishing effort at each location and the fishing method composition at main ports is used to partition this overall effort across each fishing method. Surveys are stratified by type of day and port, with two shifts on each day. No more than one weekday and one weekend survey can occur each week. Expansion domains are then specified by type of day, port, charter status, and fishing method.

Available resources
- 3-4 staff whose primary task is to conduct shore-based and boat-based surveys, though not full-time
- Two staff are assigned for each shift (to make interview collection easier and so one can leave the port to conduct trailer count runs)
- 1 regularly available vehicle

Scheduling process (monthly)
1. Randomly select 3 weekend port survey days that are from different weeks.
2. Randomly select 3 weekday port survey days that are from different weeks and do not have a boat-based survey the day before or after (i.e., a Monday cannot be selected if the Sunday before has a survey and a Friday cannot be selected if the Saturday after has a survey).
3. Go through the port survey days from the first day to the last, ignoring type of day, and assign ports by rotating through Smiling Cove, Sugar Dock, and Fishing Base. Two shifts are conducted on each port survey day, one from 10:00-18:00 and another from 18:00-02:00.

Survey coverage
- Weekdays
  o 3 weekend survey days per month with two shifts on each survey day, split across 3 ports; these surveys must occur in different weeks
  o Mondays and Fridays will be selected less frequently than other weekdays
  o Approximately 20 weekdays per month (represented as the number of weekdays per year, n_c, in EQ 19 and EQ 21), depending on the number of holidays; these will occur over 4-6 weeks
- Weekends/holidays
  o Assume that fishing effort and CPUE are similar across Saturdays, Sundays, and holidays, but only able to sample on Saturdays and Sundays
  o 3 weekday survey days per month with two shifts on each survey day, split across 3 ports; these surveys must occur in different weeks
  o Approximately 10 weekend/holidays per month (represented as the number of weekend days per year, n_c, in EQ 19 and EQ 21), depending on the number of holidays; only 8-9 weekend days are actually available for sampling and these will occur over 4-6 weeks

On-the-ground implementation
- Port survey
    1. While at the port, maintain a boat log of all trips departing or returning, including their fishing methods. Also attempt to conduct interviews for all returning trips to collect information on effort and catch.
    2. At predetermined times during each shift (10:00 and 14:00 for the first shift and 20:00 and 22:00 for the second shift) surveyors drive the western lagoon route (as for the shore-based survey) and record the number of fishing trailers at all ports and boat ramps.

Data expansion basics
- An expansion domain is specified by the:
    o Type of day (weekday or weekend)
    o Port
    o Charter status (head boat charter, 6-pack charter, or non-charter)
    o Fishing method

American Samoa boat-based survey

Summary: Four main ports are surveyed for effort and catch rate together on each survey day. Surveys are stratified by type of day. Two shifts (05:00-13:30, 13:00-21:30) occur on most days and a graveyard shift (21:00-05:30) occurs on weekdays at a low frequency. Expansion domains are then specified by type of day, charter status, and fishing method.

Available resources
- 4 staff who conduct boat-based surveys full time, with 4 additional staff who conduct shore-based surveys full time
- Two staff are assigned for each shift (to make interview collection easier)
- 2 regularly available vehicles

Scheduling process (monthly)
1. Schedule the first market invoice collection day. If the 15th is a weekday, select it, and otherwise select the first weekday after the 15th. Both survey teams are scheduled on this day.
2. Randomly select three more weekdays after the first market invoice collection day as follow-up market invoice collection days. Both survey teams are scheduled on these days.
3. For the first survey team, schedule every other Saturday (with the status of the first Saturday determined by the status of the last Saturday of the previous month) with a randomly selected non-graveyard shift (05:00-13:30 or 13:00-21:30). In those weeks where a Saturday shift is scheduled, randomly select a weekday where both it and the next day are unscheduled (thus, a Friday cannot be selected in this way). Schedule a graveyard shift (21:00-05:30) on that day and a day off on the following day. Finally, randomly select a non-graveyard shift for each unscheduled weekday.
4. For the second survey team, repeat the same process as with the first survey team except that each randomly selected shift cannot be the same as the first team's shift on that day, if the first team is also scheduled. The two teams may or may not be scheduled to work on the same Saturdays.

Survey coverage
- Weekdays
  - Each team surveys one shift on all but 6-7 weekdays (for 4 invoice collection days and 2-3 Saturday surveys) each month
  - The invoice collection days will be the same for both teams and go completely unsampled, while the weekdays missed for Saturday surveys will not be the same for both teams
  - 2-3 graveyard shifts are scheduled per team each month, while all other shifts are one of the two non-graveyard shifts
  - Approximately 20 weekdays per month (represented as the number of weekdays per year, n_c, in EQ 19 and EQ 21)
- Weekends/holidays
  - Assume that fishing effort and CPUE are similar on Saturdays and holidays, but only able to sample on Saturdays
  - Assume that no fishing occurs on Sundays since it is a day for church

- o Each team surveys one shift on every other Saturday, giving 2-3 Saturday surveys each month
- o The Saturday survey days will either be the same or completely opposite for the two teams
- o No graveyard shifts are conducted on Saturdays
- o Approximately 5-6 Saturdays/holidays per month (represented as the number of weekend days per year, n_c, in EQ 19 and EQ 21), depending on the number of holidays; only 4-5 Saturdays are actually available for sampling

On-the-ground implementation
- Port survey
  1. From the office, watch out for and be alert for notice of any boats departing from or returning to one of the four ports.
  2. When a boat is departing or returning, record it on the boat log, including the fishing method. Also attempt to conduct interviews for all returning trips to collect information on effort and catch.

Data expansion basics
- An expansion domain is specified by the:
  - o Type of day (weekday or weekend)
  - o Charter status (charter or non-charter)
  - o Fishing method

```r
1    # Note: This file contains the two user inputs (year and pool_f) and the main executable function (run_expansion).
2    #       All reading of input data from the MySQL database is performed in this code, and passed on to code in the
3    #       file "guam BB expansion.R" to perform the expansion calculations.
4
5    #library(RMySQL)  *** Use csv files for input rather than MySQL tables**** Use modified interview pooling program!!!
6    library(dplyr)
7
8    setwd("/Users/Toby/Documents/Work/20200415 CIE Expansion Review/Sensitivity Section/Guam BBS/")
9    source("guam BB expansion.R")
10   source("guam BB interview pooling.R")
11
12   #source("C:/Users/Hongguang.Ma/Documents/TeleWork/CIE_Territory Surveys/Demo_withCSVinput/Guam BBS/guam BB
     expansion_FY20csv.R")
13   #source("C:/Users/Hongguang.Ma/Documents/TeleWork/CIE_Territory Surveys/Demo_withCSVinput/Guam BBS/guam BB interview
     pooling FY20csv.R")
14   # vectors to define the indices for levels of each survey stratum
15   days_refer = NULL # type of day
16   charters_refer = NULL # charter status
17   ports_refer = NULL # survey port
18   methods_refer = NULL # fishing method
19
20   # Input: vector of strings or single string for type of day ("WD" or "WE")
21   # Output: the type of day indices for the input string(s) (1 = "WD", 2 = "WE")
22   get_days = function(type_of_days) {
23     match(type_of_days, c("WD", "WE"))
24   }
25
26   # Input: a single string or boolean for charter status
27   # Output: the charter indices for the input string/boolean (1 = charter, 2 = non-charter)
28   get_charter = function(charter) {
29     if(charter == TRUE || charter == "T" || charter == "TRUE") {
30       return(1)
31     }
32     else if(charter == FALSE || charter == "F" || charter == "FALSE") {
33       return(2)
34     }
35     else {
36       warning(paste("Charter ", charter, " not valid"))
37       return(-1)
38     }
39   }
40
41   # Input: vector of strings/boolean or single string/boolean for charter status
42   # Output: the charter indices for the input string(s)/boolean(s) (1 = charter, 2 = non-charter)
43   get_charters = function(charters) {
44     unlist(lapply(charters, get_charter))
45   }
46
47   # Input: vector of integers for port keys
48   # Output: the port indices for the input integer(s)
49   get_ports = function(port_fks) {
50     match(port_fks, ports_refer)
51   }
52
53   # Input: vector of integers for method keys
54   # Output: the method indices for the input integer(s)
55   get_methods = function(method_fks) {
56     match(method_fks, methods_refer)
57   }
58
59   # Input: x = data frame where each row represents the index of a value, along with the value itself (as
60   #            returned by the summarise method in library dplyr)
61   #        l = list of reference vectors whose lengths give the dimensions of the output array (first vector
62   #            length = first dimension of the output, etc.)
63   # Output: array whose dimensions match the dimensions of l and whose entries represent data for the corresponding
64   #         indices from x. Non-specified entries are assumed to be zero. For example, if inputs are:
65   #
66   #         x = 1 1 7      l = list(c(1,2), c(1,7,9))
67   #             1 2 5             ^       ^
68   #             2 2 1             |       |
69   #             2 3 2             |       second dimension is of length length(c(1,7,9)) = 3
70   #             ^ ^ ^             first dimension is of length length(c(1,2)) = 2
71   #             | | |
72   #             | | values for each index
73   #             | second dimension indices
74   #             first dimension indices
75   #
76   #         Then the output will be:
77   #
```

```r
78  #             res = [7 5 0   <-- 2 x 3 array with values as specified by the last column of x, and other entries zero
79  #                   0 1 2]
80  df_to_array = function(x, l) {
81    # initialize dimensions of result matrix based on reference vectors given
82    res = array(0, sapply(l, function(f) length(f)))
83
84    indices_df = data.matrix(x[, -length(x)]) # get rid of the last column, which holds the counts
85    indices = t(apply(indices_df, 1, function(f) as.vector(f))) # indices corresponding to each count
86    values = x[[length(x)]] # get only the last column, which holds the counts
87
88    res[indices] = values
89
90    return(res)
91  }
92
93  # Input:
94  #   year = year of the expansion
95  #   pool_f = whether to pool interviews when fewer than 3 are available for a stratum
96  # Output: two data frames containing all expansion and species composition fields for each expansion stratum
97  run_expansion = function(year, pool_f, species) {
98    sample_days = read.csv("CIE_sample_days_allyears.csv", stringsAsFactors = F)
99    sample_days_current_year = filter(sample_days, YEAR == year)
100
101   # Ma used MySQL package to read MySQL tables into R and use write.csv to create the csv files for use in this program
102
103   # bl_head = read.csv("G_BBS_BLHD.csv", stringsAsFactors = F)
104   # bl_detail = read.csv("G_BBS_BLDT.csv", stringsAsFactors = F)
105   # bl = inner_join(bl_head, bl_detail, by = c("BLHD_PK" = "BLHD_FK"))
106
107   # Ma joined the two tables with MySQL in R and output the result as "CIE_bl_allyears.csv"
108    bl = read.csv("CIE_bl_allyears.csv", stringsAsFactors = F)
109    bl_current_year = filter(bl, substr(SAMPLE_DATE, 1, 4) == year)
110
111    p1 = read.csv("CIE_p1_allyears.csv", stringsAsFactors = F)
112    p1_current_year = filter(p1, YEAR == year)
113
114    days = read.csv("CIE_days_allyears.csv", stringsAsFactors = F)
115    days_current_year = filter(days, YEAR == year)
116
117    interviews_raw = read.csv("CIE_interviews_raw.csv", stringsAsFactors = F)
118    interviews_raw = mutate(interviews_raw, YEAR = strtoi(substr(SAMPLE_DATE, 1, 4)))
119    catch_raw = read.csv("CIE_catch_raw.csv", stringsAsFactors = F)
120    # 6/23/20 added by TM
121    if(!is.na(species)) {
122      catch_raw = filter(catch_raw, SPECIES_FK %in% species)
123      for(i in 1:nrow(interviews_raw)) {
124        interview_key = interviews_raw[i, "INTERVIEW_PK"]
125        interviews_raw[i, "TOT_EST_KGS"] = sum(filter(catch_raw, INTERVIEW_FK == interview_key)$EST_KGS, na.rm = T)
126      }
127    }
128    interviews_current_year = filter(interviews_raw, YEAR == year)
129    catch_current_year = filter(catch_raw, substr(INTERVIEW_FK, 1, 4) == year)
130
131    iwc = read.csv("CIE_iwc_allyears.csv", stringsAsFactors = F)
132    iwc_current_year = filter(iwc, substr(SAMPLE_DATE, 1, 4) == year)
133
134    reference_raw = read.csv("CIE_reference_raw.csv", stringsAsFactors = F)
135
136    # vector whose length is the number of types of day to use when type of day is a dimension of interest
137    days_refer <<- c(1, 2)
138    # vector whose length is the number of charter statuses to use when charter status is a dimension of interest
139    charters_refer <<- c(1, 2)
140    # vector whose length is the number of ports to use when port is a dimension of interest
141    # the length is obtained by finding all of the different port keys referenced in data
142    ports_refer <<- sort(unique(c(sample_days_current_year$PORT_FK, interviews_current_year$PORT_FK,
143  bl_current_year$PORT_FK, p1_current_year$PORT_FK)), decreasing = FALSE)
143    # vector whose length is the number of methods to use when method is a dimension of interest
144    # the length is obtained by finding all of the different method keys referenced in data
145    methods_refer <<- sort(unique(c(interviews_current_year$METHOD_FK, bl_current_year$METHOD_FK,
146  p1_current_year$METHOD_FK)), decreasing = FALSE)
146
147    # remove duplicate columns
148    bl_current_year = bl_current_year[!duplicated(names(bl_current_year))]
149    # create columns for the strata indices to make later manipulations easier
150    sample_days_current_year = mutate(sample_days_current_year, port = get_ports(PORT_FK), day = get_days(TYPE_OF_DAY))
151    bl_current_year = mutate(bl_current_year, port = get_ports(PORT_FK), day = get_days(TYPE_OF_DAY), method =
152  get_methods(METHOD_FK), charter = get_charters(CHARTER_F))
152    p1_current_year = mutate(p1_current_year, port = get_ports(PORT_FK), method = get_methods(METHOD_FK))
153
```

```r
154      #***Total days***
155      # [TYPE_OF_DAY]
156      days = c(sum(days_current_year$NUM_WD), sum(days_current_year$NUM_WE))
157
158      #***Sample days***
159      # [PORT_FK, TYPE_OF_DAY]
160      sample_days = df_to_array(summarise(group_by(sample_days_current_year, port, day), count = n()), list(ports_refer,
       days_refer))
161
162      #***Boat log***
163      # [PORT_FK, TYPE_OF_DAY, METHOD_FK, CHARTER_F]
164      # number of trips with known fishing method
165      bl = array(0, c(length(ports_refer), length(days_refer), length(methods_refer), length(charters_refer)))
166      # sum of squared number of trips with known fishing method, aggregated by sample date
167      bl2 = array(0, c(length(ports_refer), length(days_refer), length(methods_refer), length(charters_refer)))
168      # [PORT_FK, TYPE_OF_DAY, CHARTER_F]
169      # number of trips with unknown fishing method
170      bl_unknown_method = array(0, c(length(ports_refer), length(days_refer), length(charters_refer)))
171      # number of trips with known fishing status
172      bl_known_fished = array(0, c(length(ports_refer), length(days_refer), length(charters_refer)))
173      # number of trips with unknown fishing status
174      bl_unknown_fished = array(0, c(length(ports_refer), length(days_refer), length(charters_refer)))
175
176      # boat log entries with known fishing method (these entries are used in the main calculations)
177      subsample_bl_days = filter(bl_current_year, METHOD_FK != 0, METHOD_FK != 99, FISHED == "Y")
178      # boat log entries with unknown fishing method
179      subsample_bl_unknown_method_days = filter(bl_current_year, METHOD_FK == 0 | METHOD_FK == 99, FISHED == "Y")
180      # boat log entries with known fishing status
181      subsample_bl_known_fished_days = filter(bl_current_year, !((FISHED != "Y" & FISHED != "N") | is.na(FISHED)))
182      # boat log entries with unknown fishing status
183      subsample_bl_unknown_fished_days = filter(bl_current_year, (FISHED != "Y" & FISHED != "N") | is.na(FISHED))
184
185      if(nrow(subsample_bl_days) > 0) {
186        bl = df_to_array(summarise(group_by(subsample_bl_days, port, day, method, charter), trips = n()), list(ports_refer,
       days_refer, methods_refer, charters_refer))
187        bl2_temp = summarise(group_by(subsample_bl_days, port, day, method, charter, SAMPLE_DATE), square = (n()) ^ 2) #
       squares of the number of trips on each sample day
188        bl2_temp = aggregate(bl2_temp$square, by = list(port = bl2_temp$port, day = bl2_temp$day, method = bl2_temp$method,
       charter = bl2_temp$charter), FUN = sum) # sum over sample days
189        bl2 = df_to_array(bl2_temp, list(ports_refer, days_refer, methods_refer, charters_refer))
190      }
191
192      if(nrow(subsample_bl_unknown_method_days) > 0) {
193        bl_unknown_method = df_to_array(summarise(group_by(subsample_bl_unknown_method_days, port, day, charter), trips =
       n()), list(ports_refer, days_refer, charters_refer))
194      }
195
196      if(nrow(subsample_bl_known_fished_days) > 0) {
197        bl_known_fished = df_to_array(summarise(group_by(subsample_bl_known_fished_days, port, day, charter), trips = n()),
       list(ports_refer, days_refer, charters_refer))
198      }
199
200      if(nrow(subsample_bl_unknown_fished_days) > 0) {
201        bl_unknown_fished = df_to_array(summarise(group_by(subsample_bl_unknown_fished_days, port, day, charter), trips =
       n()), list(ports_refer, days_refer, charters_refer))
202      }
203
204      #***Temporal adjustment factor p1***
205      # Note: Separate adjustment factors are given for each quarter, so to compute an annual adjustment factor these
       quarterly factors are averaged
206      # [PORT_FK, METHOD_FK]
207      # sum of charter P1 values
208      charter_p1 = df_to_array(summarise(group_by(p1_current_year, port, method), p1 = sum(CHARTER_P1)), list(ports_refer,
       methods_refer))
209      # sum of non-charter P1 values
210      non_charter_p1 = df_to_array(summarise(group_by(p1_current_year, port, method), p1 = sum(NON_CHARTER_P1)),
       list(ports_refer, methods_refer))
211      # number of charter/non-charter P1 values
212      count = df_to_array(summarise(group_by(p1_current_year, port, method), count = n()), list(ports_refer, methods_refer))
213      # calculate average charter P1 values
214      charter_p1 = charter_p1 / count
215      # calculate average non-charter P1 values
216      non_charter_p1 = non_charter_p1 / count
217      # [PORT_FK, METHOD_FK, CHARTER_F]
218      # average P1 values
219      p1 = array(c(charter_p1, non_charter_p1), dim = c(length(ports_refer), length(methods_refer), 2))
220
221      #***Spatial adjustment factor p2***
222      iwc_representative = 0
```

```r
223    iwc_unsampled = 0
224    if(year < 1989) { # Agana is the only sampled port and reference port
225      iwc_representative = sum((filter(iwc_current_year, PORT_FK == 1))$NUM_VESSEL)
226      iwc_unsampled = sum((filter(iwc_current_year, PORT_FK != 1))$NUM_VESSEL)
227      # The following line computes iwc_unsampled the same way as the VFP code, which is incorrect
228 #     iwc_unsampled = sum((filter(iwc_current_year, PORT_FK > 3))$NUM_VESSEL)
229    }
230    else if(year < 1995) { # Agana and Merizo are the sampled ports and reference ports
231      iwc_representative = sum((filter(iwc_current_year, PORT_FK == 1 | PORT_FK == 3))$NUM_VESSEL)
232      iwc_unsampled = sum((filter(iwc_current_year, PORT_FK != 1 & PORT_FK != 3))$NUM_VESSEL)
233      # The following line computed iwc_unsampled the same way as the VFP code, which is incorrect
234 #     iwc_unsampled = sum((filter(iwc_current_year, PORT_FK > 3))$NUM_VESSEL)
235    }
236    else { # Agana, Agat, and Merizo are the sampled ports and Agat and Merizo are the reference ports
237      iwc_representative = sum((filter(iwc_current_year, PORT_FK == 2 | PORT_FK == 3))$NUM_VESSEL)
238      iwc_unsampled = sum((filter(iwc_current_year, PORT_FK > 3))$NUM_VESSEL)
239    }
240    p2 = ifelse(iwc_representative > 0, iwc_unsampled / iwc_representative, 0)
241
242    res = df_expansion(year, pool_f, species, bl, bl2, bl_unknown_method, bl_unknown_fished, bl_known_fished, p1, p2,
    sample_days, days, interviews_current_year, interviews_raw, catch_current_year, catch_raw, methods_refer, ports_refer,
    reference_raw)
243
244    return(res)
245 }
246
247 # Whether to pool interviews when fewer than 3 are available for a stratum
248 pool_f = T
249 species = NA # NA to expand for all species, or a vector of SPECIES_PK to only include specific species
250
251 start_year = 1982
252 end_year = 2019
253
254 expansion = data.frame()
255 species_composition = data.frame()
256 for(year in start_year:end_year) {
257    res = run_expansion(year, pool_f, species)
258    expansion = rbind(expansion, res[[1]])
259    species_composition = rbind(species_composition, res[[2]])
260 }
```

Appendix E. R source code for "guam BB expansion.R," a support file that provides functions for computing expanded values.

```r
1   # Note: This file contains all supporting functions to compute expansions. It receives input data from code in the
2   #       file "guam BB.R" and relies on code in the file "guam BB interview pooling.R" to perform
3   #       interview pooling if the option is selected by the user and strata with few interviews exist.
4
5   # Input:
6   #   year = year of the expansion
7   #   pool_f = whether to pool interviews when fewer than 3 are available for a stratum
8   #   species = a vector of species or all species (when spcies = NA)
9   #   bl = counts of boat log entries with known fishing method for each stratum ([PORT_FK, TYPE_OF_DAY, METHOD_FK,
        CHARTER_F])
10  #   bl2 = squared counts of boat log entries with known fishing method for each stratum ([PORT_FK, TYPE_OF_DAY,
        METHOD_FK, CHARTER_F])
11  #   bl_unknown_method = counts of boat log entries with unknown fishing method for each stratum ([PORT_FK, TYPE_OF_DAY,
        CHARTER_F])
12  #   bl_unknown_fished = counts of boat log entries with unknown fishing status for each stratum ([PORT_FK, TYPE_OF_DAY,
        CHARTER_F])
13  #   bl_known_fished = counts of boat log entries with known fishing status for each stratum ([PORT_FK, TYPE_OF_DAY,
        CHARTER_F])
14  #   p1 = adjustment factors to account for missed vessels outside of shift times for each stratum ([PORT_FK, METHOD_FK,
        CHARTER_F])
15  #   p2 = adjustment factor to account for vessels outside of the sampled ports
16  #   sample_days = the number of sample days for each stratum ([PORT_FK, TYPE_OF_DAY])
17  #   days = the number of calendar days for each stratum ([TYPE_OF_DAY])
18  #   interviews = interviews for the expansion period
19  #   interviews_raw = all interviews for all years, for use when pooling
20  #   catch = interview catch data for the expansion period
21  #   catch_raw = interview catch data for all years
22  #   methods_refer = vector to convert method keys to method indices
23  #   ports_refer = vector to convert port keys to port indices
24  #   reference = reference values for use when insufficient interviews are available for a stratum
25  # Output: two data frames containing all expansion and species composition fields for each expansion stratum
26  df_expansion = function(year, pool_f, species, bl, bl2, bl_unknown_method, bl_unknown_fished, bl_known_fished, p1, p2,
      sample_days, days, interviews, interviews_raw, catch, catch_raw, methods_refer, ports_refer, reference) {
27    expansion = data.frame()
28    species_composition = data.frame()
29
30    for(m in 1:dim(bl)[3]) { # go through methods
31      res = df_method_expansion(year, pool_f, species, bl, bl2, bl_unknown_method, bl_unknown_fished, bl_known_fished, p1,
        p2, sample_days, days, interviews, interviews_raw, catch, catch_raw, m, methods_refer[m], ports_refer, filter(reference,
        METHOD_FK == methods_refer[m]))
32
33      expansion = rbind(expansion, res[[1]])
34      species_composition = rbind(species_composition, res[[2]])
35    }
36
37    return(list(expansion, species_composition))
38  }
39
40  # Input: a single integer for port key
41  # Output: the survey zone key corresponding to the input port key
42  get_survey_zone = function(port_fk) {
43    if(port_fk == 1) {
44      # Agana Boat Basin
45      22 # SURVEY_ZONE_PK for Agana Boat Basin-A1994
46    }
47    else if(port_fk == 2) {
48      # Agat Marina
49      23 # SURVEY_ZONE_PK for Agat Marina-A1994
50    }
51    else if(port_fk == 3) {
52      # Merizo Pier
53      21 # SURVEY_ZONE_PK for Merizo Pier-A1994
54    }
55    else {
56      warning(paste("Port ", port_fk, " not valid"))
57      -1
58    }
59  }
60
61  # Input: a vector of values
62  # Output: the variance of the average of those values
63  calc_var_average = function(x) {
64    x = x[is.finite(x)]
65    return(var(x)/length(x))
66  }
67
68  # Input: two vectors of values, of the same length
69  # Output: the variance of the quotient of two means (mean of first vector divided by mean
70  #         of second vector) = variance of the quotient of two sums - see EQ7&EQ8 of the tech memo
```

```r
calc_var_quotient = function(x, y) {
  good_indices = is.finite(x) && is.finite(y)
  x = x[good_indices]
  y = y[good_indices]
  mean1 = mean(x)
  mean2 = mean(y)
  var1 = var(x)/length(x)
  var2 = var(y)/length(y)
  cov12 = cov(x, y)/sqrt(length(x)*length(y))

  return((mean1 / mean2) ^ 2 * (var1 / (mean1) ^ 2 - 2 * cov12 / (mean1 * mean2) + var2 / (mean2) ^ 2))
}
#the function calc_var_quotient is modified in May 2019 by Ma for variance of {mean(x)/mean(y)}

# Input: varibles x & y and their variances var_x & var_y
# Output: the variance of the product of the two variables x & y - see EQ 10 of the tech memo
calc_var_product = function(x, var_x, y, var_y) {
  return(var_x * y ^ 2 + var_y * x ^ 2 - var_x * var_y)
}

# Input:
#   year = year of the expansion
#   pool_f = whether to pool interviews when fewer than 3 are available for a stratum
#   species = a vector of species or all species (when specie = NA)
#   bl = counts of boat log entries with known fishing method for each stratum ([PORT_FK, TYPE_OF_DAY, METHOD_FK, CHARTER_F])
#   bl2 = squared counts of boat log entries with known fishing method for each stratum ([PORT_FK, TYPE_OF_DAY, METHOD_FK, CHARTER_F])
#   bl_unknown_method = counts of boat log entries with unknown fishing method for each stratum ([PORT_FK, TYPE_OF_DAY, CHARTER_F])
#   bl_unknown_fished = counts of boat log entries with unknown fishing status for each stratum ([PORT_FK, TYPE_OF_DAY, CHARTER_F])
#   bl_known_fished = counts of boat log entries with known fishing status for each stratum ([PORT_FK, TYPE_OF_DAY, CHARTER_F])
#   p1 = adjustment factors to account for missed vessels outside of shift times for each stratum ([PORT_FK, METHOD_FK, CHARTER_F])
#   p2 = adjustment factor to account for vessels outside of the sampled ports
#   sample_days = the number of sample days for each stratum ([PORT_FK, TYPE_OF_DAY])
#   days = the number of calendar days for each stratum ([TYPE_OF_DAY])
#   interviews = interviews for this stratum
#   interviews_raw = all interviews for all years, for use when pooling
#   method_index = vector index of the current method
#   method_key = database key of the current method
#   ports_refer = vector to convert port keys to port indices
#   reference = reference values for the current method
#   port_index = vector index of the current port
#   type_of_day_index = vector index of the current type of day
#   charter_index = vector index of the current charter status
#   other_ports = boolean for whether this expansion is for the other (unsampled) ports (survey_zone = 24)
# Output: a data frame with a single row containing all expansion fields for the current stratum
calc_df = function(year, pool_f, species, bl, bl2, bl_unknown_method, bl_unknown_fished, bl_known_fished, p1, p2, sample_days, days, interviews, interviews_raw, method_index, method_key, ports_refer, reference, port_index, type_of_day_index, charter_index, other_ports) {
  type_of_day = ifelse(type_of_day_index == 1, "WD", "WE")
  process_type = "R" # default process type to "R" (reference) unless sufficient interviews are found to be available
  fishery_type = ifelse(charter_index == 1, "C", "N")
  port_fk = ifelse(other_ports, NA, ports_refer[port_index])
  survey_zone_fk = ifelse(other_ports, 24, get_survey_zone(port_fk))
  method_fk = method_key
  sys_create_time = Sys.time()
  sys_modify_time = Sys.time()
  hash = NA


  period_type = "Y"
  quarter = 1

  exp_pk = paste0(period_type,
                  year,
                  sprintf("%02d", quarter),
                  sprintf("%04d", method_fk),
                  sprintf("%03d", survey_zone_fk),
                  type_of_day,
                  fishery_type,
                  process_type)

  num_calendar_days = days[type_of_day_index]
  num_sample_days = sample_days[port_index, type_of_day_index]
  num_kn_method_trip = ifelse(other_ports, 0, sum(bl[port_index, type_of_day_index, , charter_index]))
```

```
142    num_unkn_method_trip = ifelse(other_ports, 0, bl_unknown_method[port_index, type_of_day_index, charter_index])
143    num_kn_fished_trip = ifelse(other_ports, 0, bl_known_fished[port_index, type_of_day_index, charter_index])
144    num_unkn_fished_trip = ifelse(other_ports, 0, bl_unknown_fished[port_index, type_of_day_index, charter_index])
145    num_trip = ifelse(other_ports, 0, bl[port_index, type_of_day_index, method_index, charter_index])
146    num_trip2 = ifelse(other_ports, 0, bl2[port_index, type_of_day_index, method_index, charter_index])
147
148    # Initialize the other fields to NA since, depending on the inputs, not all may be calculable
149    num_interview = NA
150    num_interview_pooled = NA
151    num_days_fished = NA
152    sum_fisher = NA
153    sum_fisher2 = NA
154    avg_fisher = NA
155    var_avg_fisher = NA
156    sum_hour = NA
157    sum_hour2 = NA
158    avg_hour = NA
159    var_avg_hour = NA
160    sum_fisher_hour = NA
161    sum_fisher_hour2 = NA
162    avg_fisher_hour = NA
163    var_avg_fisher_hour = NA
164    sum_gear = NA
165    sum_gear2 = NA
166    avg_gear = NA
167    var_avg_gear = NA
168    sum_tot_kgs = NA
169    sum_tot_kgs2 = NA
170    sum_est_trip = NA
171    sum_est_trip2 = NA
172    avg_est_trip = NA
173    var_est_trip = NA
174    std_est_trip = NA
175    cpue_trip = NA
176    var_cpue_trip = NA
177    std_cpue_trip = NA
178    cpue_hour = NA
179    var_cpue_hour = NA
180    std_cpue_hour = NA
181    cpue_gear = NA
182    var_cpue_gear = NA
183    std_cpue_gear = NA
184    cpue_fisher_hour = NA
185    var_cpue_fisher_hour = NA
186    std_cpue_fisher_hour = NA
187    exp_trip = NA
188    var_exp_trip = NA
189    std_exp_trip = NA
190    exp_hour = NA
191    var_exp_hour = NA
192    std_exp_hour = NA
193    exp_fisher = NA
194    var_exp_fisher = NA
195    std_exp_fisher = NA
196    exp_fisher_hour = NA
197    var_exp_fisher_hour = NA
198    std_exp_fisher_hour = NA
199    exp_gear = NA
200    var_exp_gear = NA
201    std_exp_gear = NA
202    exp_tot_kgs = NA
203    var_exp_tot_kgs = NA
204    std_exp_tot_kgs = NA
205    flag_pooled = NA
206
207    # If we can calculate an A1 value, use it. Otherwise, just use A1 = 1, which will have no effect as an adjustment
       factor
208    a1 = ifelse(num_kn_fished_trip == 0, 1, (num_kn_fished_trip + num_unkn_fished_trip) / num_kn_fished_trip)
209    # If we can calculate an A2 value, use it. Otherwise, just use A2 = 1, which will have no effect as an adjustment
       factor
210    a2 = ifelse(num_kn_method_trip == 0, 1, (num_kn_method_trip + num_unkn_method_trip) / num_kn_method_trip)
211    # If a valid P1 value is available, use it. Otherwise, just use P1 = 1, which will have no effect as an adjustment
       factor
212    p1_val = 0
213    if(!is.nan(p1[port_index, method_index, charter_index]) && p1[port_index, method_index, charter_index] > 0) {
214      p1_val = p1[port_index, method_index, charter_index]
215    }
216    else {
217      if(fishery_type == "C") {
```

```r
218          p1_val = 1
219        }
220        else if(!is.na(port_fk)){
221          if(port_fk == 1) {
222            p1_val = 0.98
223          }
224          else if(port_fk == 2) {
225            p1_val = 0.95
226          }
227          else if(port_fk == 3) {
228            p1_val = 0.8
229          }
230          else {
231            p1_val = 1
232          }
233        }
234        else {
235          p1_val = 1
236        }
237    }

239    num_interview = nrow(interviews)
240    num_interview_no_pool = num_interview

242    if(num_interview >= 3) {
243      # If sufficient interviews are available

245      process_type = "G"
246    }
247    else if(num_interview < 3 && pool_f) {
248      # If insufficient interviews are available and pooling is to be used

250      # Call the pooling function and get the pooled interviews back
251     # interviews_pooled = pool_interviews(interviews_raw, year, method_key, type_of_day_index, ports_refer[port_index],
    charter_index, FALSE)[[1]]
252      interviews_pooled = pool_interviews(interviews_raw, year, method_key, type_of_day_index, port_fk, charter_index,
    other_ports)[[1]]
253      flag_pooled = pool_interviews(interviews_raw, year, method_key, type_of_day_index, port_fk, charter_index,
    other_ports)[[2]]
254      # Ma changed the last argument (above) from FALSE to other_ports and ports_refer[port_index] to port_fk
255      if(is.data.frame(interviews_pooled) && nrow(interviews_pooled) >= 3) {
256        # If pooling was able to generate at least 3 interviews
257        process_type = "G"
258        interviews = interviews_pooled
259        num_interview_pooled = nrow(interviews)
260        print("port_fk&index"); print(c(port_fk, ports_refer[port_index])) # use port_fk, not ports_refer[port_index], to
    track ports because other_ports were also assgined a port_index ranging from 1 to 2
261        print("Pooled interview # for EXP"); print(num_interview_pooled)
262        num_interview = num_interview_pooled # use pooled interviews for all calculations
263      }
264    }

266    if(process_type == "R" && nrow(reference) != 0) {
267      # If we are using reference values and these reference values are available
268      # the reference table could be updated
269      num_interview = reference$NUM_INT
270      num_days_fished = 1
271      sum_fisher = reference$SUM_FISHER
272      sum_fisher2 = sum_fisher ^ 2
273      avg_fisher = sum_fisher / num_interview
274      sum_hour = reference$SUM_HOUR
275      sum_hour2 = sum_hour ^ 2
276      avg_hour = sum_hour / num_interview
277      sum_fisher_hour = reference$SUM_FISHER_HOUR
278      sum_fisher_hour2 = sum_fisher_hour ^ 2
279      avg_fisher_hour = sum_fisher_hour / num_interview
280      sum_gear = 1
281      sum_gear2 = sum_gear ^ 2
282      avg_gear = sum_gear / num_interview
283      sum_tot_kgs = reference$SUM_KGS
284      sum_tot_kgs2 = sum_tot_kgs ^ 2
285      sum_est_trip = num_trip * a1 * a2 / p1_val
286      if(other_ports && fishery_type != "C") {
287        sum_est_trip = bl[port_index, type_of_day_index, method_index, charter_index] * a1 * a2 / p1_val * p2
288      }
289      sum_est_trip2 = num_trip2 * (a1 * a2 / p1_val) ^ 2
290      if(other_ports && fishery_type != "C") {
291        sum_est_trip2 = bl2[port_index, type_of_day_index, method_index, charter_index] * (a1 * a2 / p1_val * p2) ^ 2
292      }
```

```
293        avg_est_trip = sum_est_trip / num_sample_days
294        if(num_sample_days > 1) {
295          var_est_trip = (sum_est_trip2 - sum_est_trip ^ 2 / num_sample_days) / (num_sample_days * (num_sample_days - 1))
296          std_est_trip = 100 * var_est_trip ^ 0.5 / avg_est_trip
297        }
298        cpue_trip = sum_tot_kgs / num_interview
299        cpue_hour = sum_tot_kgs / sum_hour
300        cpue_gear = sum_tot_kgs / sum_gear
301        cpue_fisher_hour = sum_tot_kgs / sum_fisher_hour
302        exp_trip = avg_est_trip * num_calendar_days
303        if(num_sample_days > 1) {
304          var_exp_trip = var_est_trip * num_calendar_days ^ 2
305          std_exp_trip = 100 * var_exp_trip ^ 0.5 / exp_trip # This is 100* coefficient of variation (CV)
306        }
307        exp_hour = avg_hour * exp_trip
308        exp_fisher = avg_fisher * exp_trip
309        exp_fisher_hour = avg_fisher_hour * exp_trip
310        exp_gear = avg_gear * exp_trip
311        exp_tot_kgs = cpue_trip * exp_trip
312      }
313      else if(process_type == "G") {
314        # If we are using the available interviews for the stratum
315
316        num_days_fished = sum(interviews$NUM_DAYS_FISHED)
317        sum_fisher = sum(interviews$NUM_FISHER)
318        sum_fisher2 = sum(interviews$NUM_FISHER ^ 2)
319        avg_fisher = sum_fisher / num_interview
320        sum_hour = sum(interviews$HOURS_FISHED)
321        sum_hour2 = sum(interviews$HOURS_FISHED ^ 2)
322        avg_hour = sum_hour / num_interview
323        sum_fisher_hour = sum(interviews$NUM_FISHER * interviews$HOURS_FISHED)
324        sum_fisher_hour2 = sum((interviews$NUM_FISHER * interviews$HOURS_FISHED) ^ 2)
325        avg_fisher_hour = sum_fisher_hour / num_interview
326        sum_gear = sum(interviews$NUM_GEAR)
327        sum_gear2 = sum(interviews$NUM_GEAR ^ 2)
328        avg_gear = sum_gear / num_interview
329        sum_tot_kgs = sum(interviews$TOT_EST_KGS)
330        sum_tot_kgs2 = sum(interviews$TOT_EST_KGS ^ 2)
331        sum_est_trip = num_trip * a1 * a2 / p1_val
332        if(other_ports && fishery_type != "C") {
333          sum_est_trip = bl[port_index, type_of_day_index, method_index, charter_index] * a1 * a2 / p1_val * p2
334        }
335        sum_est_trip2 = num_trip2 * (a1 * a2 / p1_val) ^ 2
336        if(other_ports && fishery_type != "C") {
337          sum_est_trip2 = bl2[port_index, type_of_day_index, method_index, charter_index] * (a1 * a2 / p1_val * p2) ^ 2
338        }
339        avg_est_trip = sum_est_trip / num_sample_days
340        if(num_sample_days > 1) {
341          var_est_trip = (sum_est_trip2 - sum_est_trip ^ 2 / num_sample_days) / (num_sample_days * (num_sample_days - 1))
342          std_est_trip = 100 * var_est_trip ^ 0.5 / avg_est_trip
343        }
344        cpue_trip = sum_tot_kgs / num_interview
345        cpue_hour = sum_tot_kgs / sum_hour
346        cpue_gear = sum_tot_kgs / sum_gear
347        cpue_fisher_hour = sum_tot_kgs / sum_fisher_hour
348        exp_trip = avg_est_trip * num_calendar_days
349        if(num_sample_days > 1) {
350          var_exp_trip = var_est_trip * num_calendar_days ^ 2
351          std_exp_trip = 100 * var_exp_trip ^ 0.5 / exp_trip
352        }
353        exp_hour = avg_hour * exp_trip
354        exp_fisher = avg_fisher * exp_trip
355        exp_fisher_hour = avg_fisher_hour * exp_trip
356        exp_gear = avg_gear * exp_trip
357        exp_tot_kgs = cpue_trip * exp_trip
358
359        var_avg_fisher = calc_var_average(interviews$NUM_FISHER)
360        var_avg_hour = calc_var_average(interviews$HOURS_FISHED)
361        var_avg_fisher_hour = calc_var_average(interviews$NUM_FISHER * interviews$HOURS_FISHED)
362        var_avg_gear = calc_var_average(interviews$NUM_GEAR)
363        var_cpue_trip = calc_var_average(interviews$TOT_EST_KGS)
364        std_cpue_trip = 100 * var_cpue_trip ^ 0.5 / cpue_trip # "std" = 100*CV in the block here
365        var_cpue_hour = calc_var_quotient(interviews$TOT_EST_KGS, interviews$HOURS_FISHED)
366        std_cpue_hour = 100 * var_cpue_hour ^ 0.5 / cpue_hour
367        var_cpue_gear = calc_var_quotient(interviews$TOT_EST_KGS, interviews$NUM_GEAR)
368        std_cpue_gear = 100 * var_cpue_gear ^ 0.5 / cpue_gear
369        var_cpue_fisher_hour = calc_var_quotient(interviews$TOT_EST_KGS, interviews$NUM_FISHER * interviews$HOURS_FISHED)
370        std_cpue_fisher_hour = 100 * var_cpue_fisher_hour ^ 0.5 / cpue_fisher_hour
371        var_exp_hour = calc_var_product(exp_trip, var_exp_trip, avg_hour, var_avg_hour)
```

```
372        std_exp_hour = 100 * var_exp_hour ^ 0.5 / exp_hour
373        var_exp_fisher = calc_var_product(exp_trip, var_exp_trip, avg_fisher, var_avg_fisher)
374        std_exp_fisher = 100 * var_exp_fisher ^ 0.5 / exp_fisher
375        var_exp_fisher_hour = calc_var_product(exp_trip, var_exp_trip, avg_fisher_hour, var_avg_fisher_hour)
376        std_exp_fisher_hour = 100 * var_exp_fisher_hour ^ 0.5 / exp_fisher_hour
377        var_exp_gear = calc_var_product(exp_trip, var_exp_trip, avg_gear, var_avg_gear)
378        std_exp_gear = 100 * var_exp_gear ^ 0.5 / exp_gear
379        var_exp_tot_kgs = calc_var_product(exp_trip, var_exp_trip, cpue_trip, var_cpue_trip)
380        std_exp_tot_kgs = 100 * var_exp_tot_kgs ^ 0.5 / exp_tot_kgs
381
382        if(!is.na(num_interview_pooled)) {
383          # If we pooled interviews, num_interview currently represents the pooled number of interviews since
384          # that number was needed in the above computations
385
386          num_interview = num_interview_no_pool # restore num_interview to the original number of interviews without pooling
387        }
388      }
389
390      # Create the data frame to return
391      df = data.frame(EXP_PK = exp_pk,
392                      PERIOD_TYPE = period_type,
393                      YEAR = year,
394                      QUARTER = quarter,
395                      TYPE_OF_DAY = type_of_day,
396                      PROCESS_TYPE = process_type,
397                      FISHERY_TYPE = fishery_type,
398                      PORT_FK = port_fk,
399                      SURVEY_ZONE_FK = survey_zone_fk,
400                      METHOD_FK = method_fk,
401                      NUM_CALENDAR_DAYS = num_calendar_days,
402                      NUM_SAMPLE_DAYS = num_sample_days,
403                      NUM_INTERVIEW = num_interview,
404                      NUM_INTERVIEW_POOLED = num_interview_pooled,
405                      NUM_DAYS_FISHED = num_days_fished,
406                      NUM_TRIP = num_trip,
407                      NUM_KN_METHOD_TRIP = num_kn_method_trip,
408                      NUM_UNKN_METHOD_TRIP = num_unkn_method_trip,
409                      NUM_KN_FISHED_TRIP = num_kn_fished_trip,
410                      NUM_UNKN_FISHED_TRIP = num_unkn_fished_trip,
411                      SUM_EST_TRIP = sum_est_trip,
412                      SUM_EST_TRIP2 = sum_est_trip2,
413                      SUM_HOUR = sum_hour,
414                      SUM_HOUR2 = sum_hour2,
415                      SUM_FISHER = sum_fisher,
416                      SUM_FISHER2 = sum_fisher2,
417                      SUM_FISHER_HOUR = sum_fisher_hour,
418                      SUM_FISHER_HOUR2 = sum_fisher_hour2,
419                      SUM_GEAR = sum_gear,
420                      SUM_GEAR2 = sum_gear2,
421                      SUM_TOT_KGS = sum_tot_kgs,
422                      SUM_TOT_KGS2 = sum_tot_kgs2,
423                      AVG_FISHER = avg_fisher,
424                      AVG_HOUR = avg_hour,
425                      AVG_FISHER_HOUR = avg_fisher_hour,
426                      AVG_GEAR = avg_gear,
427                      AVG_EST_TRIP = avg_est_trip,
428                      EXP_TRIP = exp_trip,
429                      EXP_HOUR = exp_hour,
430                      EXP_FISHER = exp_fisher,
431                      EXP_FISHER_HOUR = exp_fisher_hour,
432                      EXP_GEAR = exp_gear,
433                      EXP_TOT_KGS = exp_tot_kgs,
434                      CPUE_TRIP = cpue_trip,
435                      CPUE_HOUR = cpue_hour,
436                      CPUE_FISHER_HOUR = cpue_fisher_hour,
437                      CPUE_GEAR = cpue_gear,
438                      VAR_AVG_FISHER = var_avg_fisher,
439                      VAR_AVG_HOUR = var_avg_hour,
440                      VAR_AVG_FISHER_HOUR = var_avg_fisher_hour,
441                      VAR_AVG_GEAR = var_avg_gear,
442                      VAR_EST_TRIP = var_est_trip,
443                      VAR_EXP_TRIP = var_exp_trip,
444                      VAR_EXP_HOUR = var_exp_hour,
445                      VAR_EXP_FISHER = var_exp_fisher,
446                      VAR_EXP_FISHER_HOUR = var_exp_fisher_hour,
447                      VAR_EXP_GEAR = var_exp_gear,
448                      VAR_EXP_TOT_KGS = var_exp_tot_kgs,
449                      VAR_CPUE_TRIP = var_cpue_trip,
450                      VAR_CPUE_HOUR = var_cpue_hour,
```

```
451                     VAR_CPUE_GEAR = var_cpue_gear,
452                     VAR_CPUE_FISHER_HOUR = var_cpue_fisher_hour,
453                     STD_EST_TRIP = std_est_trip,
454                     STD_EXP_TRIP = std_exp_trip,
455                     STD_EXP_HOUR = std_exp_hour,
456                     STD_EXP_FISHER = std_exp_fisher,
457                     STD_EXP_FISHER_HOUR = std_exp_fisher_hour,
458                     STD_EXP_GEAR = std_exp_gear,
459                     STD_EXP_TOT_KGS = std_exp_tot_kgs,
460                     STD_CPUE_TRIP = std_cpue_trip,
461                     STD_CPUE_HOUR = std_cpue_hour,
462                     STD_CPUE_GEAR = std_cpue_gear,
463                     STD_CPUE_FISHER_HOUR = std_cpue_fisher_hour,
464                     SYS_CREATE_TIME = sys_create_time,
465                     SYS_MODIFY_TIME = sys_modify_time,
466                     FLAG_POOLED = flag_pooled,
467                     HASH = hash)
468
469     return(df)
470 }
471
472 # Input:
473 #   year = year of the expansion
474 #   pool_f = whether to pool interviews when fewer than 3 are available for a stratum
475 #   species = a vector of species or all species (when species = NA)
476 #   bl = counts of boat log entries with known fishing method for each stratum ([PORT_FK, TYPE_OF_DAY, METHOD_FK,
     CHARTER_F])
477 #   bl2 = squared counts of boat log entries with known fishing method for each stratum ([PORT_FK, TYPE_OF_DAY,
     METHOD_FK, CHARTER_F])
478 #   bl_unknown_method = counts of boat log entries with unknown fishing method for each stratum ([PORT_FK, TYPE_OF_DAY,
     CHARTER_F])
479 #   bl_unknown_fished = counts of boat log entries with unknown fishing status for each stratum ([PORT_FK, TYPE_OF_DAY,
     CHARTER_F])
480 #   bl_known_fished = counts of boat log entries with known fishing status for each stratum ([PORT_FK, TYPE_OF_DAY,
     CHARTER_F])
481 #   p1 = adjustment factors to account for missed vessels outside of shift times for each stratum ([PORT_FK, METHOD_FK,
     CHARTER_F])
482 #   p2 = adjustment factor to account for vessels outside of the sampled ports
483 #   sample_days = the number of sample days for each stratum ([PORT_FK, TYPE_OF_DAY])
484 #   days = the number of calendar days for each stratum ([TYPE_OF_DAY])
485 #   interviews = interviews for the expansion period
486 #   interviews_raw = all interviews for all years, for use when pooling
487 #   catch = interview catch data for the expansion period
488 #   catch_raw = interview catch data for all years
489 #   method_index = vector index of the current method
490 #   method_key = database key of the current method
491 #   ports_refer = vector to convert port keys to port indices
492 #   reference = reference values for the current method
493 # Output: two data frames with rows containing all expansion and species composition fields for each stratum of the
     current method
494 df_method_expansion = function(year, pool_f, species, bl, bl2, bl_unknown_method, bl_unknown_fished, bl_known_fished,
     p1, p2, sample_days, days, interviews, interviews_raw, catch, catch_raw, method_index, method_key, ports_refer,
     reference) {
495     stratum_interviews_m = filter(interviews, METHOD_FK == method_key) # restrict interviews to the current method
496
497     expansion = data.frame()
498     species_composition = data.frame()
499
500         dimensions = dim(bl_unknown_method)
501
502         sampled_ports = c()
503         if(year < 1989) {
504           sampled_ports = c(1)
505         }
506         else if(year < 1995) {
507           sampled_ports = c(1, 3)
508         }
509         else {
510           sampled_ports = c(1, 2, 3)
511         }
512
513         for(i in 1:dimensions[1]) { # Go through the ports
514                 if(ports_refer[i] %in% sampled_ports) {
515                     # If the current port is one of the sampled ports
516
517                     stratum_interviews_mp = filter(stratum_interviews_m, PORT_FK == ports_refer[i]) # restrict interviews
     to the current method and port
518         for(j in 1:dimensions[2]) { # Go through the types of day
```

```r
519                              stratum_interviews_mpd = filter(stratum_interviews_mp, TYPE_OF_DAY == ifelse(j == 1, "WD",
     "WE")) # restrict interviews to the current method, port, and type of day
520                                for(k in 1:dimensions[3]) { # Go through the charter statuses
521                                  if(sum(bl[i, j, , k]) > 0 && sample_days[i, j] > 0) { # If data has been
     collected for the current stratum
522                                    stratum_interviews_mpdc = filter(stratum_interviews_mpd, CHARTER_F == ifelse(k
     == 1, T, F)) #!! For CHARTER_F, change "T" and "F" to T and F with csv input files
523                                    num_interview = nrow(stratum_interviews_mpdc)
524
525                                    df_g = calc_df(year, pool_f, species, bl, bl2, bl_unknown_method,
     bl_unknown_fished, bl_known_fished, p1, p2, sample_days, days, stratum_interviews_mpdc, interviews_raw, method_index,
     method_key, ports_refer, reference, i, j, k, F)
526                                    expansion = rbind(expansion, df_g)
527
528
529                                    #added block in May 2019 for pooling intervews for SPC
530                                    if(num_interview < 3 && pool_f) { # num_interview is restored to the number of
     interview_nopool at the end of calc_df
531                                      # If insufficient interviews are available and pooling is to be used
532
533                                      # Call the pooling function and get the pooled interviews back
534                                      interviews_pooled = pool_interviews(interviews_raw, year, method_key, j,
     ports_refer[i], k, FALSE)[[1]] # j and k are type of day index and charter index
535                                      if(is.data.frame(interviews_pooled) && nrow(interviews_pooled) >= 1) { # keep any
     pooling results for SPC and changed >=3 to >= 1
536                                        # If pooling was able to generate at least 3 interviews
537                                        # process_type = "G"
538                                        stratum_interviews_mpdc = interviews_pooled
539                                        print("SPCpoolingPort&Interview#"); print(ports_refer[i]);
     print(nrow(stratum_interviews_mpdc))# print # of rows to make sure pooling works
540                                        num_interview = nrow(stratum_interviews_mpdc)
541                                      }
542                                    }
543
544
545
546                                    catch_df = data.frame() # species catch information
547                                    if(num_interview > 0) {
548                                      for(l in 1:num_interview) {
549                                        catch_interview = filter(catch_raw, INTERVIEW_FK ==
     stratum_interviews_mpdc[l,]$INTERVIEW_PK) # all species catch data from all matched interviews
550
551                                        catch_df = rbind(catch_df, catch_interview)
552                                      }
553                                    }
554
555                                    if(nrow(catch_df) > 0) {
556                                      # if there was any catch for the current stratum interviews
557
558                                      # 6/23/20 added by TM
559                                      species_df = data.frame()
560                                      species_list = unique(catch_df$SPECIES_FK)
561                                      if(is.na(species)) {
562                                        num_interview = nrow(stratum_interviews_mpdc)
563                                        sum_hour = sum(stratum_interviews_mpdc$HOURS_FISHED)
564                                        sum_gear = sum(stratum_interviews_mpdc$NUM_GEAR)
565                                        sum_fisher_hour = sum(stratum_interviews_mpdc$NUM_FISHER *
     stratum_interviews_mpdc$HOURS_FISHED)
566
567                                        for(l in 1:length(species_list)) {
568                                          specie = species_list[l]
569
570                                          all_catch_specie = rep.int(0, nrow(stratum_interviews_mpdc))
571                                          for(m in 1:nrow(stratum_interviews_mpdc)) {
572                                            catch_entries = filter(catch_df, SPECIES_FK == specie & INTERVIEW_FK
     == stratum_interviews_mpdc[m, "INTERVIEW_PK"])
573                                            if(nrow(catch_entries) > 0) {
574                                              all_catch_specie[m] = sum(catch_entries$EST_KGS)
575                                            }
576                                          }
577
578                                          sum_tot_kgs = sum(all_catch_specie)
579                                          cpue_trip = sum_tot_kgs / num_interview
580                                          cpue_hour = sum_tot_kgs / sum_hour
581                                          cpue_gear = sum_tot_kgs / sum_gear
582                                          cpue_fisher_hour = sum_tot_kgs / sum_fisher_hour
583                                          var_cpue_trip = calc_var_average(all_catch_specie)
584                                          std_cpue_trip = 100 * var_cpue_trip ^ 0.5 / cpue_trip # std = 100*CV
```

```r
585                                                    var_cpue_hour = calc_var_quotient(all_catch_specie,
       stratum_interviews_mpdc$HOURS_FISHED)
586                                                    std_cpue_hour = 100 * var_cpue_hour ^ 0.5 / cpue_hour
587                                                    var_cpue_gear = calc_var_quotient(all_catch_specie,
       stratum_interviews_mpdc$NUM_GEAR)
588                                                    std_cpue_gear = 100 * var_cpue_gear ^ 0.5 / cpue_gear
589                                                    var_cpue_fisher_hour = calc_var_quotient(all_catch_specie,
       stratum_interviews_mpdc$NUM_FISHER * stratum_interviews_mpdc$HOURS_FISHED)
590                                                    std_cpue_fisher_hour = 100 * var_cpue_fisher_hour ^ 0.5 /
       cpue_fisher_hour
591                                                    # added on July 2, 2020
592                                                    var_kgs_caught = calc_var_product(df_g$EXP_TRIP, df_g$VAR_EXP_TRIP,
       cpue_trip, var_cpue_trip)
593                                                    species_df = rbind(species_df, data.frame(SPECIES_FK = specie,
594                                                                                               CPUE_TRIP = cpue_trip,
595                                                                                               CPUE_HOUR = cpue_hour,
596                                                                                               CPUE_GEAR = cpue_gear,
597                                                                                               CPUE_FISHER_HOUR =
       cpue_fisher_hour,
598                                                                                               VAR_CPUE_TRIP = var_cpue_trip,
599                                                                                               VAR_KGS_CAUGHT =
       var_kgs_caught,
600                                                                                               STD_CPUE_TRIP = std_cpue_trip,
601                                                                                               VAR_CPUE_HOUR = var_cpue_hour,
602                                                                                               STD_CPUE_HOUR = std_cpue_hour,
603                                                                                               VAR_CPUE_GEAR = var_cpue_gear,
604                                                                                               STD_CPUE_HOUR = std_cpue_hour,
605                                                                                               VAR_CPUE_FISHER_HOUR =
       var_cpue_fisher_hour,
606                                                                                               STD_CPUE_FISHER_HOUR =
       std_cpue_fisher_hour))
607                                                }
608                                            }
609
610                                        # group catch by species and calculate summary fields
611                                        catch_by_species = summarise(group_by(catch_df, SPECIES_FK), num_interview =
       n(), kgs_caught = sum(EST_KGS), num_kept = sum(NUM_KEPT), price_lb = mean(PRICE_LB))
612
613                                        species_composition_stratum = data.frame(SPC_PK =
       paste0(substring(df_g$EXP_PK, 1, 14),
614                                                                                            sprintf("%05d",
       catch_by_species$SPECIES_FK),
615
       substring(df_g$EXP_PK, 15, 18)),
616                                                                                 EXP_FK = df_g$EXP_PK,
617                                                                                 FLAG_POOLED = df_g$FLAG_POOLED,
618                                                                                 SPECIES_FK =
       catch_by_species$SPECIES_FK,
619                                                                                 NUM_INTERVIEW =
       catch_by_species$num_interview,
620                                                                                 NUM_INTERVIEW_POOLED = num_interview,
621                                                                                 KGS_CAUGHT =
       catch_by_species$kgs_caught / sum(catch_by_species$kgs_caught) * df_g$EXP_TOT_KGS,
622 #                                                                                VAR_LBS_CAUGHT = NA,
623 #                                                                                STD_LBS_CAUGHT = NA,
624 #                                                                                LBS_SOLD = NA,
625                                                                                 NUM_KEPT = catch_by_species$num_kept,
626                                                                                 PRICE_LB = catch_by_species$price_lb,
627                                                                                 SYS_CREATE_TIME = Sys.time(),
628                                                                                 SYS_MODIFY_TIME = Sys.time(),
629                                                                                 HASH = NA)
630
631                                        # 6/23/20 added by TM
632                                        if(is.na(species)) {
633                                            species_composition_stratum = inner_join(species_composition_stratum,
       species_df, by = c("SPECIES_FK" = "SPECIES_FK"))
634                                        }
635
636                                        species_composition = rbind(species_composition, species_composition_stratum)
637                                    }
638                                }
639                            }
640                        }
641                    }
642                }
643
644            for(j in 1:dimensions[2]) { # Go through the types of day
645                # These 'rep' values will hold the combined data for the representative ports
646                bl_rep = bl
```

```r
647                bl2_rep = bl2
648                bl_unknown_method_rep = bl_unknown_method
649                bl_unknown_fished_rep = bl_unknown_fished
650                bl_known_fished_rep = bl_known_fished
651                p1_rep = p1
652                sample_days_rep = sample_days
653                port1_index = which(ports_refer == 1) # array index for port 1 (Agana)
654                port2_index = which(ports_refer == 2) # array index for port 2 (Agat)
655                port3_index = which(ports_refer == 3) # array index for port 3 (Merizo)
656
657                port_index = 0 # array index to use information from in the expansion
658                stratum_interviews_mpd = NA
659
660                if(year < 1989) { # Agana is the only sampled port and reference port
661                  port_index = port1_index
662                  stratum_interviews_mpd = filter(stratum_interviews_m, PORT_FK == 1, TYPE_OF_DAY == ifelse(j == 1, "WD",
       "WE")) # restrict interviews to the current method, port, and type of day
663                }
664                else if(year < 1995) { # Agana and Merizo are the sampled ports and reference ports
665                  # Combine the data for PORT_FK = 1 and PORT_FK = 3 into PORT_FK = 1, and make entries for PORT_FK = 3 zero.
666                  # This allows use to use the same method as with the single-port expansions, by simply passing it port key 1
       as
667                  # the port parameter.
668                  bl_rep[port1_index, , ,] = bl[port1_index, , ,] + bl[port3_index, , ,]
669                  bl_rep[port3_index, , ,] = 0
670                  bl2_rep[port1_index, , ,] = bl2[port1_index, , ,] + bl2[port3_index, , ,]
671                  bl2_rep[port3_index, , ,] = 0
672                  bl_unknown_method_rep[port1_index, ,] = bl_unknown_method[port1_index, ,] + bl_unknown_method[port3_index,
       ,]
673                  bl_unknown_method_rep[port3_index, ,] = 0
674                  bl_unknown_fished_rep[port1_index, ,] = bl_unknown_fished[port1_index, ,] + bl_unknown_fished[port3_index,
       ,]
675                  bl_unknown_fished_rep[port3_index, ,] = 0
676                  bl_known_fished_rep[port1_index, ,] = bl_known_fished[port1_index, ,] + bl_known_fished[port3_index, ,]
677                  bl_known_fished_rep[port3_index, ,] = 0
678                  p1_rep[port1_index, ,] = (p1[port1_index, ,] + p1[port3_index, ,]) / 2
679                  p1_rep[port3_index, ,] = 0
680                  sample_days_rep[port1_index,] = max(sample_days[port1_index,], sample_days[port3_index,])
681                  sample_days_rep[port3_index,] = 0
682
683                  port_index = port1_index
684                  stratum_interviews_mpd = filter(stratum_interviews_m, PORT_FK == 1 | PORT_FK == 3, TYPE_OF_DAY == ifelse(j
       == 1, "WD", "WE")) # restrict interviews to the current method, port, and type of day
685                }
686                else { # Agana, Agat, and Merizo are the sampled ports and Agat and Merizo are the reference ports
687                  # Combine the data for PORT_FK = 2 and PORT_FK = 3 into PORT_FK = 2, and make entries for PORT_FK = 3 zero.
688                  # This allows use to use the same method as with the single-port expansions, by simply passing it port key 2
       as
689                  # the port parameter.
690                  bl_rep[port2_index, , ,] = bl[port2_index, , ,] + bl[port3_index, , ,]
691                  bl_rep[port3_index, , ,] = 0
692                  bl2_rep[port2_index, , ,] = bl2[port2_index, , ,] + bl2[port3_index, , ,]
693                  bl2_rep[port3_index, , ,] = 0
694                  bl_unknown_method_rep[port2_index, ,] = bl_unknown_method[port2_index, ,] + bl_unknown_method[port3_index, ,]
695                  bl_unknown_method_rep[port3_index, ,] = 0
696                  bl_unknown_fished_rep[port2_index, ,] = bl_unknown_fished[port2_index, ,] + bl_unknown_fished[port3_index, ,]
697                  bl_unknown_fished_rep[port3_index, ,] = 0
698                  bl_known_fished_rep[port2_index, ,] = bl_known_fished[port2_index, ,] + bl_known_fished[port3_index, ,]
699                  bl_known_fished_rep[port3_index, ,] = 0
700                  p1_rep[port2_index, ,] = (p1[port2_index, ,] + p1[port3_index, ,]) / 2
701                  p1_rep[port3_index, ,] = 0
702                  sample_days_rep[port2_index,] = max(sample_days[port2_index,], sample_days[port3_index,])
703                  sample_days_rep[port3_index,] = 0
704
705                  port_index = port2_index
706                  stratum_interviews_mpd = filter(stratum_interviews_m, PORT_FK == 2 | PORT_FK == 3, TYPE_OF_DAY == ifelse(j ==
       1, "WD", "WE")) # restrict interviews to the current method, port, and type of day
707                }
708
709                for(k in 1:dimensions[3]) { # Go through the charter statuses
710                  if(sum(bl_rep[port_index, j, , k]) > 0 && sample_days[port_index, j] > 0) { # If data has been collected for
       the current stratum, changed "bl" to "bl_rep"
711                    stratum_interviews_mpdc = filter(stratum_interviews_mpd, CHARTER_F == ifelse(k == 1, T, F)) # !! For csv
       file input change "T"&"F" to T & F for CHARTER_F
712                    num_interview = nrow(stratum_interviews_mpdc)
713
714              df_other_ports_g = calc_df(year, pool_f, species, bl_rep, bl2_rep, bl_unknown_method_rep, bl_unknown_fished_rep,
       bl_known_fished_rep, p1_rep, p2, sample_days_rep, days, stratum_interviews_mpdc, interviews_raw, method_index,
       method_key, ports_refer, reference, port_index, j, k, T)
```

```
715            expansion = rbind(expansion, df_other_ports_g)
716
717         # pooling interviews for SPC, check with Toby about assinging NA to port!
718         if(num_interview < 3 && pool_f) {
719           # If insufficient interviews are available and pooling is to be used
720
721           # Call the pooling function and get the pooled interviews back
722           interviews_pooled = pool_interviews(interviews_raw, year, method_key, j, NA, k, TRUE)[[1]] # port = NA and
unsampled_ports = TRUE, j&k are indices
723           if(is.data.frame(interviews_pooled) && nrow(interviews_pooled) >= 1) { # keep pooling results for SPC, changed
>=3 to >=1
724             # If pooling was able to generate at least 3 interviews
725             # process_type = "G"
726             stratum_interviews_mpdc = interviews_pooled
727             print("OtherPortSPCpoolingInterview#"); print(nrow(stratum_interviews_mpdc))
728             # num_interview_pooled = nrow(interviews)
729             num_interview = nrow(stratum_interviews_mpdc) # use pooled interviews for all calculations
730           }
731         }
732
733            catch_df = data.frame() # species catch information
734            if(num_interview > 0) {
735              for(l in 1:num_interview) {
736                catch_interview = filter(catch_raw, INTERVIEW_FK == stratum_interviews_mpdc[l,]$INTERVIEW_PK) # all
species catch data for the current interview
737
738                catch_df = rbind(catch_df, catch_interview)
739              }
740            }
741
742            if(nrow(catch_df) > 0) {
743              # if there was any catch for the current stratum interviews
744
745              # 6/23/20 added by TM
746              species_df = data.frame()
747              species_list = unique(catch_df$SPECIES_FK)
748              if(is.na(species)) {
749                num_interview = nrow(stratum_interviews_mpdc)
750                sum_hour = sum(stratum_interviews_mpdc$HOURS_FISHED)
751                sum_gear = sum(stratum_interviews_mpdc$NUM_GEAR)
752                sum_fisher_hour = sum(stratum_interviews_mpdc$NUM_FISHER * stratum_interviews_mpdc$HOURS_FISHED)
753
754                for(l in 1:length(species_list)) {
755                  specie = species_list[l]
756
757                  all_catch_specie = rep.int(0, nrow(stratum_interviews_mpdc))
758                  for(m in 1:nrow(stratum_interviews_mpdc)) {
759                    catch_entries = filter(catch_df, SPECIES_FK == specie & INTERVIEW_FK == stratum_interviews_mpdc[m,
"INTERVIEW_PK"])
760                    if(nrow(catch_entries) > 0) {
761                      all_catch_specie[m] = sum(catch_entries$EST_KGS)
762                    }
763                  }
764
765                  sum_tot_kgs = sum(all_catch_specie)
766                  cpue_trip = sum_tot_kgs / num_interview
767                  cpue_hour = sum_tot_kgs / sum_hour
768                  cpue_gear = sum_tot_kgs / sum_gear
769                  cpue_fisher_hour = sum_tot_kgs / sum_fisher_hour
770                  var_cpue_trip = calc_var_average(all_catch_specie)
771                  std_cpue_trip = 100 * var_cpue_trip ^ 0.5 / cpue_trip # std = 100*CV
772                  var_cpue_hour = calc_var_quotient(all_catch_specie, stratum_interviews_mpdc$HOURS_FISHED)
773                  std_cpue_hour = 100 * var_cpue_hour ^ 0.5 / cpue_hour
774                  var_cpue_gear = calc_var_quotient(all_catch_specie, stratum_interviews_mpdc$NUM_GEAR)
775                  std_cpue_gear = 100 * var_cpue_gear ^ 0.5 / cpue_gear
776                  var_cpue_fisher_hour = calc_var_quotient(all_catch_specie, stratum_interviews_mpdc$NUM_FISHER *
stratum_interviews_mpdc$HOURS_FISHED)
777                  std_cpue_fisher_hour = 100 * var_cpue_fisher_hour ^ 0.5 / cpue_fisher_hour
778                  var_kgs_caught = calc_var_product(df_other_ports_g$EXP_TRIP, df_other_ports_g$VAR_EXP_TRIP,
cpue_trip, var_cpue_trip)
779                  species_df = rbind(species_df, data.frame(SPECIES_FK = specie,
780                                                            CPUE_TRIP = cpue_trip,
781                                                            CPUE_HOUR = cpue_hour,
782                                                            CPUE_GEAR = cpue_gear,
783                                                            CPUE_FISHER_HOUR = cpue_fisher_hour,
784                                                            VAR_CPUE_TRIP = var_cpue_trip,
785                                                            VAR_KGS_CAUGHT = var_kgs_caught,
786                                                            STD_CPUE_TRIP = std_cpue_trip,
787                                                            VAR_CPUE_HOUR = var_cpue_hour,
```

```
788                                                        STD_CPUE_HOUR = std_cpue_hour,
789                                                        VAR_CPUE_GEAR = var_cpue_gear,
790                                                        STD_CPUE_HOUR = std_cpue_hour,
791                                                        VAR_CPUE_FISHER_HOUR = var_cpue_fisher_hour,
792                                                        STD_CPUE_FISHER_HOUR = std_cpue_fisher_hour))
793                    }
794                }
795
796          # group catch by species and calculate summary fields
797       catch_by_species = summarise(group_by(catch_df, SPECIES_FK), num_interview = n(), kgs_caught =
    sum(EST_KGS), num_kept = sum(NUM_KEPT), price_lb = mean(PRICE_LB))
798
799       species_composition_stratum_other = data.frame(SPC_PK = paste0(substring(df_other_ports_g$EXP_PK, 1, 14),
800                                                      sprintf("%05d",
    catch_by_species$SPECIES_FK),
801                                                      substring(df_other_ports_g$EXP_PK, 15,
    18)),
802                                                  EXP_FK = df_other_ports_g$EXP_PK,
803                                                  FLAG_POOLED = df_other_ports_g$FLAG_POOLED,
804                                                  SPECIES_FK = catch_by_species$SPECIES_FK,
805                                                  NUM_INTERVIEW = catch_by_species$num_interview,
806                                                  NUM_INTERVIEW_POOLED = num_interview,
807                                                  KGS_CAUGHT = catch_by_species$kgs_caught /
    sum(catch_by_species$kgs_caught) * df_other_ports_g$EXP_TOT_KGS,
808    #                                              VAR_LBS_CAUGHT = NA,
809    #                                              STD_LBS_CAUGHT = NA,
810    #                                              LBS_SOLD = NA,
811                                                  NUM_KEPT = catch_by_species$num_kept,
812                                                  PRICE_LB = catch_by_species$price_lb,
813                                                  SYS_CREATE_TIME = Sys.time(),
814                                                  SYS_MODIFY_TIME = Sys.time(),
815                                                  HASH = NA)
816
817          # 6/23/20 added by TM
818          if(is.na(species)) {
819            species_composition_stratum_other = inner_join(species_composition_stratum_other, species_df, by =
    c("SPECIES_FK" = "SPECIES_FK"))
820          }
821
822          species_composition = rbind(species_composition, species_composition_stratum_other)
823        }
824      }
825    }
826  }
827
828    return(list(expansion, species_composition))
829 }
```

Appendix F. R source code for "guam BB interview pooling.R," a support file that provides a function for aggregating interview data.

```r
1  # Note: This file contains a single function to conduct interview pooling. It is called by code in the file "guam
2  #       BB expansion.R".
3
4  # Input:
5  #    - interviews: all interviews for all years
6  #    - year: year of the expansion
7  #    - method: fishing method of the current stratum expansion
8  #    - type_of_day: type of day of the current stratum expansion
9  #    - port: port of the current stratum expansion
10 #    - charter: charter status of the current stratum expansion
11 #    - unsampled_ports: boolean representing whether the current pooling is for unsampled ports
12 # Output:
13 #    - data frame containing the pooled interviews for the current stratum; if pooling could not produce at least 3
   interviews, NA is returned
14
15 pool_interviews = function(interviews, year, method, type_of_day, port, charter, unsampled_ports) {
16         type_of_day_str = ""
17         if(type_of_day == 1) {
18                 type_of_day_str = "WD"
19         }
20         else {
21                 type_of_day_str = "WE"
22         }
23
24         charter_str = ""
25         if(charter == 1) {
26                 charter_str = TRUE # from "T" to True for csv file
27         }
28         else {
29                 charter_str = FALSE # from "F" to FALSE for csv file
30         }
31
32         interviews_current_year = interviews[interviews$YEAR == year,]
33         interviews_no_pool = interviews_current_year[(interviews_current_year$METHOD_FK == method) &
   (interviews_current_year$TYPE_OF_DAY == type_of_day_str) & (interviews_current_year$CHARTER_F == charter_str),]
34         if(unsampled_ports) {
35                 interviews_no_pool = interviews_no_pool[interviews_no_pool$PORT_FK == 2 | interviews_no_pool$PORT_FK ==
   3,]
36         }
37         else {
38                 interviews_no_pool = interviews_no_pool[interviews_no_pool$PORT_FK == port,]
39         }
40
41         if(nrow(interviews_no_pool) >= 3) {
42                 return(list(interviews_no_pool, "None"))
43         }
44
45         # D: Pool data from other type of day
46         pool_d = interviews_current_year[(interviews_current_year$METHOD_FK == method) &
   (interviews_current_year$CHARTER_F == charter_str),]
47         if(unsampled_ports) {
48                 pool_d = pool_d[pool_d$PORT_FK == 2 | pool_d$PORT_FK == 3,]
49         }
50         else {
51                 pool_d = pool_d[pool_d$PORT_FK == port,]
52         }
53
54         if(nrow(pool_d) >= 3) {
55                 return(list(pool_d, "D"))
56         }
57
58         if(method == 4){
59                 # S: For method 4, pool methods 5 and 6
60                 pool_s = interviews_current_year[(interviews_current_year$METHOD_FK >= 4) &
   (interviews_current_year$METHOD_FK <= 6) & (interviews_current_year$CHARTER_F == charter_str),]
61                 if(unsampled_ports) {
62                         pool_s = pool_s[pool_s$PORT_FK == 2 | pool_s$PORT_FK == 3,]
63                 }
64                 else {
65                         pool_s = pool_s[pool_s$PORT_FK == port,]
66                 }
67
68                 if(nrow(pool_s) >= 3) {
69                         return(list(pool_s, "S"))
70                 }
71         }
72
73         if(unsampled_ports || (port == 1 || port == 2)) {
74                 # P: For Agana Boat Basin (port 1), pool Agat Marina (port 2), and vice versa
```

```r
                pool_p = interviews_current_year[interviews_current_year$CHARTER_F == charter_str,]
                if(unsampled_ports) {
                        pool_p = pool_p[(pool_p$PORT_FK >= 1) & (pool_p$PORT_FK <= 3),]
                }
                else {
                        pool_p = pool_p[(pool_p$PORT_FK >= 1) & (pool_p$PORT_FK <= 2),]
                }

                if(method == 4) {
                        pool_p = pool_p[(pool_p$METHOD_FK >= 4) & (pool_p$METHOD_FK <= 6),]
                }
                else {
                        pool_p = pool_p[pool_p$METHOD_FK == method,]
                }

                if(nrow(pool_p) >= 3) {
                        return(list(pool_p, "P"))
                }
        }

        if((unsampled_ports || port == 3) && (method == 1 || method == 3)) {
                # Q: For trolling (method 1) and atulai night light (method 3) at Merizo Pier (port 3), pool Agana
                # Boat Basin (port 1) and Agat Marina (port 2)
                pool_q = interviews_current_year[(interviews_current_year$PORT_FK >= 1) &
(interviews_current_year$PORT_FK <= 3) & (interviews_current_year$METHOD_FK == method) &
(interviews_current_year$CHARTER_F == charter_str),]

                if(nrow(pool_q) >= 3) {
                        return(list(pool_q, "Q"))
                }
        }

        # apply D, then S and P or Q, if applicable
        temp = interviews[interviews$CHARTER_F == charter_str,]

        if(method == 4) {
                temp = temp[(temp$METHOD_FK >= 4) & (temp$METHOD_FK <= 6),]
        }
        else {
                temp = temp[temp$METHOD_FK == method,]
        }

        if(unsampled_ports) {
                temp = temp[temp$PORT_FK >= 1 & temp$PORT_FK <= 3,]
        }
        else if(port == 1 || port == 2) {
                temp = temp[(temp$PORT_FK >= 1) & (temp$PORT_FK <= 2),]
        }
        else if(port == 3) {
                if(method != 1 && method != 3) {
                        temp = temp[temp$PORT_FK == port,]
                }
        }

        years = sort(unique(temp$YEAR), decreasing = T)
        years = years[years <= year]

        if(length(years) >= 1) { # years to pool from
          for(y in 1:length(years)) {
            i = temp[temp$YEAR %in% years[1:y],]

            if(nrow(i) >= 3) {
              return(list(i, paste0("-", year - years[y])))
            }
            else if(y == length(years)) { # pooled all previous years and still not enough interviews
              return(list(i, "All Years"))
            }
          }
        }
        else { # no past/present years to pool from
          temp = temp[temp$YEAR %in% years,]
          return(list(temp, "No Past/Present Interviews"))
        }
}
```