



**NOAA**  
**FISHERIES**

Office of Science and Technology

**Marine Recreational Information Program**

# **Data User Handbook**

Published June 2021

Updated September 2021

# MRIP Data User Handbook

## Table of Contents

<b>1. Introduction</b>	<b>5</b>
1.1 Scope	5
1.2 Additional Resources	5
Recreational Fishing Survey and Data Standards	5
Survey Design and Statistical Methods Manual	6
MRIP Milestones	6
<b>2. An Introduction to MRIP General Survey Data</b>	<b>7</b>
2.1 Data Products	7
General Survey Microdata (SAS, CSV)	7
Frequent Questions: Survey Variables	8
General Survey Estimates (SAS, CSV)	9
Metadata (InPort)	9
Public Fishing Access Site Register (Public Access)	10
For-Hire Vessel Registry (Restricted Access)	10
2.2 Data Availability	10
Data Collection Schedule	10
Table 1: Temporal Survey Coverage	10
Estimation and Publication Schedule	11
2.3 Common Analyses	11
2.4 Explanation of Terms Used in the MRIP Query Tool	11
Spatial Terms	11
Sample Coverage	11
Geographical Area	12
Fishing Area	13
Temporal Terms	13
Waves	13
Summarize By	13
Year Type	14
Fishing Mode Terms	14
Type of Fishing (Mode)	14
Statistical Terms	15
Calibrated Estimates	15
Confidence Interval	16
Percent Standard Error (PSE)	16
Catch Query Terms	17
Species	17
Type of Catch	17
Fish Weight Data	17

Length Distribution	18
Effort Query Terms	19
Directed Trip Species Options	19
Participation Query Terms	19
<b>3. Understanding the Estimates</b>	<b>21</b>
3.1 Statistical Procedures	21
3.2 Non-Sampling Errors	21
Coverage Error	21
Measurement Error	21
Non-response Error	22
3.3 Limitations	22
Revisions	22
Sample Size	22
Time Series	22
Weight Estimates	23
<b>4. Queries</b>	<b>24</b>
4.1 MRIP Query Tool	24
4.2 Types of Queries	24
Catch Data	24
Effort Data	25
Participation Data	25
National Summary Data	25
Data Download	26
4.3 Example Queries	26
Querying Catch Data	25
Graphing Data	26
Exporting Data	26
Finding Targeted Species	27
Finding Uncommon Species	27
4.4 Custom Queries	27
<b>5. Domain Analyses</b>	<b>28</b>
5.1 General Guidance	28
Overview of Template Programs	28
Estimation Methods used by Template Programs	29
Considerations for Domain Analysis	30
Before Producing Estimates: Exploring Sample Sizes and Defining the Domain	30
After Producing Estimates: Additional Data Checks for Users to Consider	31
Table 2: Domain Analyses MRIP Data Generally Can and Cannot Support	31
Figure 1: Sample Size Check Examples	32
5.2 Common Custom Analyses	33
Catch, Effort, and Length Frequencies (General)	33
In SAS	33
In R	37

Custom Geographic Domains	38
Custom Temporal Domains	40
Species Group Domains	43
5.3 Experimental Analyses	44
Multi-Year Averaging	44
Catch Frequency Estimation	44
<b>6. Large Pelagics Survey</b>	<b>46</b>
6.1 Data Products	47
Large Pelagics Survey Microdata and Estimates (SAS, CSV)	47
6.2 Data Availability	47
Data Collection Schedule	47
Estimation and Publication Schedule	47
6.3 Using LPS Data	48

# 1. Introduction

## 1.1 Scope

The [Marine Recreational Information Program](#)'s MRIP Data User Handbook is designed to provide fisheries analysts and stock assessors with detailed information about downloading, exporting, querying, and analyzing NOAA Fisheries' marine recreational fishing data. These data include both **microdata** and **estimates**, and are produced to help scientists and managers assess and maintain sustainable fish stocks.

- **Microdata** are the recreational fishing information gathered through the [Access Point Angler Intercept Survey](#) (APAIS), [Fishing Effort Survey](#) (FES), [For-Hire Survey](#) (FHS), and [Large Pelagics Survey](#) (LPS).
- **Estimates** are the calculated statistical values produced from these microdata.

This handbook is organized into five sections that orient users with informational resources; general survey data; the MRIP Query Tool; and the template R and SAS programs available to support custom domain analyses. A sixth section focuses on the specialized LPS. This handbook does **not** provide guidance on the access or analysis of data collected by surveys other than those listed above. It will be updated as methodologies change, new tools are made available, and customer feedback is incorporated. Please contact [Lauren Dolinger Few](#) with questions about the handbook or suggestions for additional content.

## 1.2 Additional Resources

In accordance with NOAA's [Information Quality Guidelines](#), MRIP works to provide those that are affected by science and management decisions with the information needed to understand the data that informed them and the way these data were produced. This includes providing open access to:

- The data that were used;
- The analytical methods that were applied; and
- The assumptions and statistical procedures that were employed.

By providing open access to MRIP's data products and analytical tools, this handbook supports the program's commitment to transparency. However, it is just one in a suite of resources for data customers, listed here.

### [Recreational Fishing Survey and Data Standards](#)

These standards guide the design, improvement, and quality of the information produced by the recreational fishing surveys administered or funded through MRIP. By providing a single set of guidelines for data collection and estimation, the standards support the program's commitment

to transparency. By promoting data quality, consistency, and comparability across data collection programs, the standards facilitate the shared use of the statistics these programs produce. Ultimately, the standards will further ensure the integrity of our data collection efforts, the quality of our recreational fisheries statistics, and the strength of science-based management decisions.

### [Survey Design and Statistical Methods Manual](#)

This document describes the technical details of the surveys and methods used to produce estimates of total recreational catch. It includes information about current and legacy survey designs, anticipated survey improvements, and the calibration methodologies that have allowed NOAA Fisheries to maintain a long-term time series of marine recreational catch statistics while implementing changes to survey designs.

### [MRIP Milestones](#)

This webpage documents milestones in the program's history, as well as changes to the design and coverage of its recreational fishing surveys.

## 2. An Introduction to MRIP General Survey Data

The Marine Recreational Information Program works with state and regional partners to collect recreational fishing data throughout the year. Public-use data can be downloaded through the links below.

For a complete timeline of additions and updates to data products, visit the [Recreational Fishing Estimate Updates](#) webpage. To receive notifications of updates, [subscribe to the MRIP email service](#).

### 2.1 Data Products

#### General Survey Microdata (SAS, CSV)

General survey microdata, or the recreational fishing information gathered through the [APAIS](#), [FES](#), and [FHS](#), are available in SAS and CSV formats. Each .zip file (labeled either PS\_YYYY\_sas.zip or PS\_YYYY\_csv.zip) contains the following datasets, organized by year (YYYY) and two-month sampling period, or wave (W):

- **CATCH\_YYYYW.sas7bdat:** Catch-level data and variables required for estimation. Contains one record per species per angler trip interview (identified by the variable *id\_code*). Each record contains catch totals by [catch type](#) in numbers, weight (kg), and length (mm). Design variables include:
  - *Strat\_id*: Identifier for the design stratum.
  - *Psu\_id*: Identifier for the primary sampling unit (site-day).
  - *Id\_code*: Identifier for the angler trip.
  - *Wp\_catch*: Post-stratified sampling weight for use in weighted estimation.
- **TRIP\_YYYYW.sas7bdat:** Trip-level data and variables required for estimation. Contains one record per angler trip interview (identified by the variable *id\_code*). Design variables include:
  - *Strat\_id*: Identifier for the design stratum.
  - *Psu\_id*: Identifier for the primary sampling unit (site-day).
  - *Id\_code*: Identifier for the angler trip.
  - *Wp\_int*: Post-stratified sampling weight for use in weighted estimation.

Both the FES, which covers private/rental boat and shore fishing modes, and the FHS, which covers charter and headboat fishing modes, provide estimates for individual collapsed areas of fishing (*area\_x*). For charter and headboat modes, it is possible to have an effort estimate for a given area (inland, ocean  $\leq$  3 miles, ocean  $>$  3 miles) with no corresponding trip data from the APAIS. To account for these situations, charter and/or headboat records are included in the trip dataset. While there are no catch data associated with these records, their inclusion allows data

users to calculate trip totals that match those reported by the [MRIP Query Tool](#) for charter and headboat effort. These records may be identified by:

- *Month*=99
  - *KOD*="xx"
  - Last 4 characters of *psu\_id*="0000"
  - Last 4 characters of *strat\_id*="99xx"
  - *Id\_code*=first 15 digits of appended (year||wave||sub\_reg||st||fl\_reg||mode\_fx||month||kod||"0000"), with *area\_x* appended.
- **SIZE\_YYYYW.sas7bdat:** Fish-level length and weight data and variables required for estimation. Contains one record per fish caught and measured or weighed by the APAIS interviewer. Missing lengths and/or weights are imputed as needed for individual fish records. When *lngh* and *wgt* fields have imputed values, the corresponding *lngh\_imp* and *wgt\_imp* fields have values of 1. Design variables include:
    - *Strat\_id*: Identifier for the design stratum.
    - *Psu\_id*: Identifier for the primary sampling unit (site-day).
    - *Id\_code*: Identifier for the angler trip.
    - *Wp\_size*: Post-stratified sampling weight for use in weighted estimation for the size dataset. Contains an additional adjustment for situations in which only a subset of landed fish have been measured for an angler trip.

More information about variable formats and descriptions can be found in the MRIP\_Survey\_Variables guide on the [Recreational Fishing Data Downloads](#) webpage.

## Frequent Questions: Survey Variables

*In the trip datasets, what is the difference between the prt\_code and leader variables?*

While *prt\_code* and *leader* values will often match, the two variables provide different information. The *prt\_code* is the *id\_code* of the leader of a fishing party; a fishing party includes everyone who fished on the same boat trip. Within a fishing party, there may be multiple groups with their own grouped catches; these groups have distinct *leader* codes. Headboat trips generally have multiple groups with grouped catches (and therefore multiple *leader* codes) within the same fishing party (*prt\_code*). Most private/rental and charter boat trips have only one group within the party, so in these cases, the *prt\_code* value will usually match the *leader* value.

*In the catch datasets, why are there some non-integer catch counts (the variables claim, harvest, landing, release, and tot\_cat)?*

There are two causes of non-integer catch counts: grouped catch and incomplete shore mode trips.



**Grouped Catch:** Type A (*claim*) catch counts will be non-integer values when anglers report grouped catch. This often occurs on private/rental and charter/headboat mode trips, when catch is difficult to separate by individual angler. In standard estimation, Type A grouped catch is weighted differently than Type B1 (*harvest*) or B2 (*release*) catch to avoid the overestimation of catch rates. However, having two different sample weights complicates calculating Type A+B1 (*landing*), so for these public-use datasets, we multiply the *claim* counts (A) by an adjustment for the records with grouped catch. This ultimately reduces the number of steps data users must take to produce their own estimates. As noted in the “catch\_begin2013w2” tab in the MRIP\_Survey\_Variables guide on the [Recreational Fishing Data Downloads](#) webpage, users may determine whether their count data has been adjusted by comparing the value to the variable *claim\_unadj*.

**Incomplete Shore Mode Trips:** Under certain conditions, APAIS interviewers may intercept incomplete shore mode trips. In these cases, anglers are asked to estimate the amount of additional time they will continue to fish. This additional fishing time is used to expand the catch counts recorded during the interview.

*What is the difference between wp\_int, found in both the catch and trip datasets, and wp\_catch, found in the catch datasets?*

*Wp\_int* and *wp\_catch* are the final sample weights for the trip and catch datasets, respectively, which are needed to produce custom estimates. Conceptually, they can be thought of as the number of angler trips being represented by an individual angler trip record. Typically, *wp\_int* and *wp\_catch* will be equal for the same angler trip in the catch and trip datasets, but there can be differences. For instance, they will differ when we have used an annual level estimation method to address small sample size issues. (These cases are listed on the [Recreational Fishing Estimate Updates](#) webpage for 2/21/2017. The variable *alt\_flag* in the survey datasets will equal 1 if this alternative estimation method has been used.) Other differences are due to applications of the APAIS and FES calibrations to the historical time series of catch and effort estimates. (See the [Survey Design and Statistical Methods Manual](#) for small sample size estimation methods and time series calibration methods.)

## [General Survey Estimates \(SAS, CSV\)](#)

Calibrated recreational catch and effort estimates are available in SAS and CSV formats for the Atlantic and Gulf coasts beginning in 1981. More information about variable formats and descriptions can be found in the MRIP\_Estimate\_Variables guide on the [Recreational Fishing Data Downloads](#) webpage.

## [Metadata \(InPort\)](#)

In accordance with NOAA Fisheries' [Data and Information Management Policy Directive](#) (PDF, 5 pages) metadata that describes the “what, where, when, how, and who” of MRIP’s data holdings is available on [InPort](#): a centralized repository of NOAA Fisheries’ data documentation.

## [Public Fishing Access Site Register \(Public Access\)](#)

The Public Fishing Access Site Register is an online database of public saltwater fishing access sites. It includes detailed information about site characteristics, site pressure, and nearby amenities. It is used to generate the sample frame for the APAIS, and is updated regularly with input from anglers and state fisheries staff.

## [For-Hire Vessel Registry \(Restricted Access\)](#)

The For-Hire Vessel Registry is an online database of known for-hire vessels (charter boats and headboats). It includes information such as vessel identifiers, vessel status (i.e., active, inactive, or ineligible) and contact information for vessel representatives. It serves as the sample frame for the FHS, and is updated regularly with input from vessel representatives and state fisheries staff. Please note: “Guest Login” only allows users to view the Public Fishing Access Site Register. To view the For-Hire Vessel Registry, users must log in with a username and password. However, to protect personally identifiable information, NOAA Fisheries only grants access to the For-Hire Vessel Registry to those who require it for their work.

## 2.2 Data Availability

### Data Collection Schedule

To optimize sampling efficiency, different surveys are administered in different regions at different times. In the Northeast, for example, MRIP’s general surveys (i.e., the APAIS, FES, and FHS) are not administered during known periods of low fishing activity (i.e., Wave 1, and in some states, Wave 2).

[Table 1](#) indicates which MRIP surveys are conducted during which wave. Note: The APAIS and FHS do not sample headboats south of Virginia, as these vessels are covered by the Southeast Fisheries Science Center’s [Southeast Region Headboat Survey](#). The APAIS and FHS are not administered in Hawaii, but Hawaii does implement an angler intercept survey alongside the FES. (The [Hawaii Marine Recreational Fishing Survey](#) uses Marine Recreational Fisheries Statistics Survey (MRFSS) methodologies, rather than the APAIS methods implemented on the Atlantic and Gulf coasts.) Data collection efforts are suspended in Puerto Rico as the territory rebuilds following Hurricane Maria (September 2017).

Table 1: Temporal Survey Coverage

Sampling Period	<a href="#">APAIS</a> , <a href="#">FHS</a> and <a href="#">FES</a>
Wave 1 (Jan.-Feb.)	Administered in North Carolina, Florida, Alabama, Mississippi, and Hawaii (FES).
Wave 2 (March-April)	Administered in coastal states from Massachusetts to Mississippi, and in Hawaii (FES). Note: During Wave 2, New Hampshire samples headboats independently of MRIP; the state should be contacted directly for those data.

Wave 3 (May-June)	Administered in coastal states from Maine to Mississippi and in Hawaii (FES).
Wave 4 (July-August)	
Wave 5 (Sept.-Oct.)	
Wave 6 (Nov.-Dec.)	Administered in coastal states from Massachusetts to Mississippi and in Hawaii (FES).

## Estimation and Publication Schedule

Preliminary microdata and estimates from MRIP’s general surveys (i.e., the APAIS, FES, and FHS) are published approximately 45 days after the end of each wave (April 15, June 15, August 15, October 15, December 15, and February 15). Final microdata and estimates from these surveys are published on or around April 15 of the following year.

Note: Once NOAA Fisheries’ [Recreational Fishing Survey and Data Standards](#) are fully implemented, cumulative estimates will be published instead of two-month wave estimates, beginning with the first survey administration of the survey year. These cumulative estimates will be released following the publication schedule described above. Data users who choose to create non-standard wave-level estimates may do so using MRIP’s public-use microdata (see [Section 2.1](#)) and custom domain analysis programs (see [Section 5](#)).

## 2.3 Common Analyses

The [MRIP Query Tool](#) allows users to filter catch and effort data by time series, geographic area, species, mode, and other characteristics. General guidance on how to use the MRIP Query Tool to perform common analyses of NOAA Fisheries’ marine recreational fishing data is described in [Section 4](#) of this handbook; guidance on how to use template R and SAS programs to support custom domain analyses is described in [Section 5](#).

## 2.4 Explanation of Terms Used in the MRIP Query Tool

The terms used in the MRIP Query Tool are defined below and in the online [Recreational Fishing Data Glossary](#). Here, the terms are organized into several sections, beginning with those that are relevant to most queries (spatial terms, temporal terms, fishing mode terms, and statistical terms), and ending with those that are specific to certain query types (catch terms, effort terms, and participation terms).

### Spatial Terms

#### Sample Coverage

Today, NOAA Fisheries conducts its recreational fisheries statistics program on the Atlantic coast from Maine to Florida; the Gulf coast from Florida to Mississippi; and in Hawaii. With the

exception of the National Summary Query, data collected from other state or federal surveys, including those certified and those funded by MRIP, are not included in the MRIP Query Tool<sup>1</sup>.

Generally speaking, estimates from the Atlantic and Gulf coasts may be compared across extended periods of time, because [calibration methodologies](#) have been applied to account for [changes in survey design and sample coverage](#) over the years. When conducting a National Summary Query, estimates from Alaska may be compared across their time series (1996-Present), because the design of the state data collection program has not changed significantly since 1996. Estimates from Texas may be compared across their time series (1983-Present) for the same reason. However, users should exercise caution when using the National Summary Query to make such long-term comparisons for estimates in Louisiana, California, Oregon, Washington, and Hawaii. Caution should also be exercised when comparing for-hire estimates before 2004 (prior to the full implementation of the FHS).

## Geographical Area

Geographical areas are arranged in geographic order by sub-region. These sub-regions include:

- **North Atlantic:** Maine, New Hampshire, Massachusetts, Rhode Island, and Connecticut
- **Mid-Atlantic:** New York, New Jersey, Delaware, Maryland, and Virginia
- **South Atlantic:** North Carolina, South Carolina, Georgia, and East Florida
- **Gulf of Mexico:** West Florida, Alabama, Mississippi, and Louisiana (1981-2013)
- **Pacific:** California, Oregon, and Washington
- **North and Mid-Atlantic**
- **Atlantic Coast**
- **Atlantic and Gulf Coast**
- **Pacific Coast:** Washington, Oregon, and California
- **Alaska**
- **Hawaii**
- **Caribbean:** Puerto Rico (2000-2017)

Estimates are available for an entire sub-region or for states within a sub-region. For example, selecting “North Atlantic” will produce estimates summed across the five states within that sub-region. Selecting “North Atlantic by State” will produce the sub-region’s estimates on a state-by-state basis.

In the MRIP Query Tool, both Florida and California are divided into two regions: East and West Florida, and Southern and Northern California. Selecting “Florida” or “California” will produce estimates for the whole state.

---

<sup>1</sup> The National Summary Query includes data from the following sources, by geographic area: (1) MRIP: ME-MS (1981-Present), LA (1981-2013), PR (2000-2017), HI (2003-Present), CA/OR/WA (1981-2004). (2) [Louisiana Department of Wildlife and Fisheries](#): LA (2014-Present). (3) [Texas Parks and Wildlife Department](#): TX (1981-Present, with design change in 1983). (4) [Pacific Recreational Fisheries Information Network](#): CA/OR/WA (2005-Present). (5) [Alaska Department of Fish and Game](#): AK (1996-Present).

The LPS is conducted on the Atlantic coast from Maine through Virginia. To complement NOAA Fisheries' Highly Migratory Species management, these states are divided into two regions: the Northern Atlantic, which extends from Maine through northern New Jersey, and the Southern Atlantic, which extends from southern New Jersey through Virginia. (New Jersey is divided along the border of Atlantic and Ocean counties.) Due to the limited scale of recreational fishing for large pelagic fish, several states are further combined into two-state groups: Maine and New Hampshire, Connecticut and Rhode Island, and Maryland and Delaware.

HMS Catch Card programs operate in Maryland and North Carolina.

## Fishing Area

- **Inland:** Inshore bodies of saltwater or brackish water (e.g., bays, estuaries, or sounds). Does not include inland freshwater areas.
- **State Territorial Sea:** An area extending three nautical miles from shore in all states and territories except for Puerto Rico and the Gulf coast of Florida, where the seaward state-federal boundary measures three leagues (about 10 miles). Does not include inland areas.
- **State Waters:** Both inland waters and state territorial seas.
- **Federal Exclusive Economic Zone (EEZ):** An area contiguous to all state territorial seas, extending seaward 200 nautical miles from the baseline from which the state territorial sea is measured.
- **Ocean:** Both state territorial seas and the EEZ.

## Temporal Terms

### Waves

Two-month sampling periods.

- Wave 1: January-February
- Wave 2: March-April
- Wave 3: May-June
- Wave 4: July-August
- Wave 5: September-October
- Wave 6: November-December

### Summarize By

- **Annual:** Provides estimates for each year during a select period of time.
- **Wave:** Provides estimates for each two-month sampling period during a select period of time.
- **Cumulative:** Provides cumulative estimates from a select year through a select wave.
- **Single Wave:** Provides estimates for a select wave during a select period of time.

Once NOAA Fisheries' [Recreational Fishing Survey and Data Standards](#) are fully implemented, cumulative estimates will be published instead of two-month wave estimates, beginning with the first survey administration of the survey year. Data users who choose to create non-standard wave-level estimates may do so using MRIP's public-use microdata (see [Section 2.1](#)) and custom domain analysis programs (see [Section 5](#)).

## Year Type

The Catch Data: MRFSS/MRIP Comparisons Query—which compares MRIP estimates to those of its predecessor, the Marine Recreational Fisheries Statistics Survey—provides estimates at the annual level for three year types:

- **Calendar Year:** Jan. 1 – Dec. 31
- **May Fishing Year:** May 1 – April 30
- **July Fishing Year:** July 1 – June 30

Each range begins in the Earliest Year selected in the “From” field. For example, if a user selects From: 2009, To: 2009, and Year Type: July Fishing Year, the Query Tool would return estimates for July 1, 2009, through June 30, 2010.

## Fishing Mode Terms

### Type of Fishing (Mode)

Estimates can be divided into six categories, based on an angler's fishing mode:

- **Shore**
- **Private or Rental Boat**
- **Headboat a.k.a. Party Boat**
- **Charter Boat**
- **Shore and Private/Rental Boat Combined**
- **All For-Hire Modes Combined**, which includes headboat and charter boat.

Vessel size, inspection requirements, fee structure, and angler experience are all factors that may be used to distinguish charter boats from headboats. However, within the context of recreational fishing data collection, anticipated fishing activity is the most important factor that separates the two vessel types. Grouping vessels based on patterns in, and levels of, fishing effort—and sampling the catch of the two groups differently—reduces the potential for bias and improves the precision of catch and effort estimates. Therefore, MRIP defines charter boats and headboats as follows:

- **Charter boats** are vessels that take a group of anglers—usually six or fewer—on a fishing trip with a licensed captain and crew. The anglers hire, or “charter,” the vessel, and pay a fee for the captain's services. Charter boats engage in a full range of fishing techniques, including drift fishing, trolling, and bottom fishing. Charter boat catch is

sampled at public fishing access sites through the APAIS and Large Pelagics Intercept Survey (LPIS).

- **Headboats** are vessels that take multiple individual and/or small groups of anglers on a fishing trip with a licensed captain and crew. Headboats are generally larger than charter boats, and almost always take more than six anglers on a given trip. From Maine through Virginia, headboat catch is sampled at sea through the APAIS. South of Virginia, headboats are sampled by the Southeast Fisheries Science Center's [Southeast Region Headboat Survey](#).

Sampling of the for-hire sector has varied over time.

- From 1981 through 1985, all for-hire boats were sampled as one category, called the **party/charter mode**.
- From 1986 through 2004, all for-hire boats in the North and Mid-Atlantic (from Maine to Virginia) continued to be sampled as one undifferentiated party/charter mode. In the South Atlantic and Gulf of Mexico, the only for-hire vessels MRIP sampled were charter boats. (Party (head) boats were, and continue to be, sampled by the Southeast Fisheries Science Center's Southeast Region Headboat Survey.)
- In 2005, charter boats and party (head) boats began to be sampled by the FHS and stratified APAIS. Separate charter boat and party (head) boat estimates are produced, and undifferentiated party/charter sampling is no longer performed.

LPS and HMS Catch Card estimates are limited to private and charter boat modes. The private boat mode includes vessels with an HMS Angling or Atlantic Tunas General permit. The charter boat mode includes vessels with an HMS Charter/Headboat permit. Vessels targeting large pelagic species without an HMS permit are also sampled through the dockside LPIS.

In North Carolina, shore mode is further divided into **Beach/Bank mode** and **Man-Made mode**.

## Statistical Terms

### Calibrated Estimates

The **BASE** series contains estimates that were available between 1981 and 2017, before the application of two [calibrations](#) to support the modified APAIS and new FES.

- BASE estimates from 1981 to 2003 were derived from the MRFSS.
- BASE estimates from 2004 to 2012 were derived from MRFSS intercept data, but produced using MRIP's weighted estimation methodology.
- BASE estimates from 2013 to 2017 were derived from APAIS data and produced using MRIP's weighted estimation methodology.

The **ACAL** series contains estimates resulting from the application of the **APAIS calibration** to data collected before the 2013 APAIS design change. This calibration applies to estimates between 1981 and 2012, and accounts for the 2013 design change (for all years prior to 2013) and the MRIP weighted estimation methodology (for all years prior to 2004). The ACAL series

was produced, in part, to illustrate the effects of the APAIS and FES calibrations on estimates, and is for comparison purposes only.

The **FCAL** series contains estimates resulting from the full application of both (APAIS and FES) calibrations to data collected between 1981 and 2017. These calibrations account for the APAIS design change and the transition from the Coastal Household Telephone Survey (CHTS) to the FES. The estimates are the estimates that are available in standard catch and effort queries.

More information about the calibration methods used to produce these series can be found on the [MRIP Research](#) webpage and in the [MRIP Survey Design and Statistical Methods Manual](#).

## Confidence Interval

Data users should consider the confidence intervals that surround point estimates before drawing conclusions about these estimates. The dotted lines that appear above and below point estimates in the graph outputs of the MRIP Query Tool illustrate a 95% confidence interval, and encompass the range of values within which we are 95% sure the actual population value is located. In other words, if our surveys were conducted over and over again, 95% of the results would fall within this range. Confidence intervals are calculated as the estimate minus 1.96 times the standard error (the lower limit) and the estimate plus 1.96 times the standard error (the upper limit).

## Percent Standard Error (PSE)

Percent standard error, or PSE, is a measure of precision that expresses the standard error of an estimate as a percentage of that estimate. Precision refers to the variability that results from the dispersion of the sample measurements used to calculate an estimate.

Large PSEs indicate low precision. Small PSEs indicate higher precision, and are therefore more desirable. In accordance with guidelines from the Office of Management and Budget—which state federal agencies should identify their own criteria for determining when errors are too large for an estimate to be publicly released—the MRIP Query Tool presents a warning when a PSE exceeds 30%. Once NOAA Fisheries' [Recreational Fishing Survey and Data Standards](#) are fully implemented, the MRIP Query Tool will not include an estimate when the PSE exceeds 50%.

Generally speaking, precision improves as sample sizes increase. When we group year, state, wave, or mode estimates, for example, our sample size expands and the precision of our estimates improves. Catch estimates for commonly caught species are often more precise than catch estimates for rare-event or pulse fisheries.

Percent standard error is automatically included with all requested information, with the exception of mean lengths and weights for recreational catch in Hawaii.



## Catch Query Terms

### Species

Common species are arranged in alphabetical order in the MRIP Query Tool's "Species" dropdown menu. Less common species are not listed, but can be found by following these steps:

1. Click the button labeled Other Species.
2. Enter the species name in the appropriate field and click the button labeled Search.
3. Select the correct result from the Common Name(s) Returned dropdown menu and click the button labeled Return.
4. This species will now be available in the Species dropdown menu for future queries.

### Type of Catch

- **Harvest Based on Observed Harvest, or Type A:** Fish that are caught and brought back to the dock in a form that can be identified by APAIS interviewers.
- **Harvest Based on Reported Harvest, or Type B1:** Fish that are caught, released dead, used for bait, or filleted, and are identified by individual anglers. This catch type is also called **unobserved harvest**.
- **Released Alive, or Type B2:** Fish that are released alive and identified by individual anglers.
- **Harvest:** Type A + B1.
- **Total Catch:** Type A + Type B1 + Type B2.

LPS catch types include **kept**, or fish retained by the vessel; **alive**, or fish released alive; and **dead**, or dead fish discarded at sea.

### Fish Weight Data

In 2013, MRIP recalculated all estimates of landings by weight (lb and kg) using the same design-based estimation methodology used to recalculate all estimates of catch in numbers of fish. We also developed a new method to handle missing weights. This method uses length-weight modeling (where either length or weight data are missing) and a mix of hot and cold deck imputation (where both length and weight data are missing) to impute, or fill in, missing length and weight values by species at the individual angler trip level.

For individual fish records **where lengths are present but weights are missing**, missing weights are imputed using length-weight modeling of the form  $weight = a \times length^b$ , where where  $a$  and  $b$  are parameters estimated using available length and weight data for the fish species of interest. In most cases, models are fit by species and two-month wave in the current year. Should a current-year model fail to converge, models are fit by species using the most recent decade of data. Should the current-year model and the ten-year model fail to converge, which is very rare, these cases are passed on to hot and cold deck imputation, with both length and weight values set to missing prior to imputation.

For intercepted angler trips with landings **where both length and weight measurements are missing**, paired length and weight observations are imputed from complete cases using hot and cold deck imputation. (Complete cases include records with both length and weight data available, as well as records where we were able to compute a missing length or weight using the length-weight modeling described above.) Up to five rounds of imputation are conducted in an attempt to fill in missing values. These rounds begin with imputation cells that correspond to the most detailed MRIP estimation cells, but are aggregated to higher levels in subsequent rounds to bring in more length-weight data.

- Round 1: Current year, wave, sub-region, state, mode, area fished, species.
- Round 2: Current year, half-year, sub-region, state, mode, species.
- Round 3: Current + most recent prior year, wave, sub-region, state, mode, area fished, species.
- Round 4: Current + most recent prior year, sub-region, state, mode, species.
- Round 5: Current + most recent prior year, sub-region, species.

If fish weights are still missing after each imputation method has been applied, the weight data will remain missing, and the resulting weight estimates should be considered minimums. At this point, it is up to the data user to determine whether to impute and, if so, what imputation approach is most appropriate for these missing data points; we do not make these decisions, because information needs and data sensitivity vary among species.

Missing weights are more common with large fish (which may be too large for a weight measurement to take place) and rare-event species. The existence and/or extent of missing weights for a weight estimates query will be provided in the column labeled “Landings (no.) without Size Information.” This column contains the number of landed (Type A+B1) fish that are not included in the weight estimate column labeled “Harvest (A+B1) Total Weight.” If the “Landings (no.) without Size Information” column contains a value of zero, then all landed fish are included in the weight estimate.

More information about these imputation methods can be found in the [MRIP Survey Design and Statistical Methods Manual](#).

## Length Distribution

Distributions of fork or centerline lengths are available for select species by inch or centimeter group. Inch groups contain fish measuring X.00 to X.99 inches long. For example, inch group nine includes fish between 9.00 and 9.99 inches long.

AP AIS and FES calibrations have been applied to length distribution estimates from Maine through Mississippi (1981-2017) and in Louisiana (1981-2013, since the AP AIS and FES were discontinued in Louisiana in 2014). Estimates from Hawaii (all years) and Puerto Rico (prior to 2014) are based on MRFSS methodology.

“Percent of Fish” and “Landings” (or number-at-length) are provided by length group. Records with missing length group values indicate the number of landed (Type A+B1) fish with no length information. Percent standard error only applies to landings (or number-at-length) estimates.

## Effort Query Terms

### Directed Trip Species Options

- **Primary Target:** Provides an effort estimate for all trips on which an angler identified the select species as their primary target.
- **Secondary Target:** Provides an effort estimate for all trips on which an angler identified the select species as their secondary target.
- **Caught/Available Catch (Type A):** Provides an effort estimate for all trips on which the select species was caught and brought back to the dock in a form that could be identified by an APAIS interviewer.
- **Harvested/Unavailable Catch (Type B1):** Provides an effort estimate for all trips on which the select species was caught and killed but not available for identification by an interviewer.
- **Released/Unavailable Catch (Type B2):** Provides an effort estimate for all trips on which the select species was caught and released alive.

Selecting more than one of these options will provide an effort estimate for all trips on which one or more of your selections was true. For example, selecting Primary Target, Secondary Target, and Released will provide an effort estimate for all trips on which an angler identified the select species as their primary or secondary target on the trip, and/or the select species was caught and released alive.

## Participation Query Terms

Participation estimates are divided into three categories, based on an angler’s area of residence: coastal, non-coastal, and out-of-state.

- **Coastal** designations are based on historical fishing activity, and vary throughout the calendar year. Due to their small geographic areas, all counties in Rhode Island, Connecticut, and Delaware are considered coastal. All Florida counties are considered coastal due to the relatively high rate of fishing throughout the state.
- **Non-coastal** describes anglers from counties outside of a state’s defined coastal zone.
- **Out-of-state**, also known as nonresident anglers, describes anglers from other states or countries who came to the state and fished.

When summing participants across categories and geographic areas:

- All participants within a single state can be added together.
- Coastal and non-coastal anglers across states and sub-regions can be added together.
- Out-of-state anglers should not be added across states or sub-regions. Because an out-of-state angler could have fished in more than one state that is not their state of

residence, the addition of out-of-state anglers across states or sub-regions may result in the double counting of some individuals.

Due to changes in survey coverage and design, it is no longer possible to derive meaningful participation estimates from NOAA Fisheries' marine recreational fishing surveys. Instead, the agency will use the [National Survey of Fishing, Hunting, and Wildlife-Associated Recreation](#) (U.S. Fish and Wildlife Service) to obtain estimates of the number of recreational saltwater anglers. Until these estimates are available at the state level—expected to occur in 2023—2018 participation estimates are the most reliable participation estimates we have.

## 3. Understanding the Estimates

### 3.1 Statistical Procedures

The [Survey Design and Statistical Methods Manual](#) describes the technical details of MRIP's surveys and the methods used to produce estimates of total recreational catch. It includes information about current and legacy survey designs, anticipated survey improvements, and the calibration methodologies that have allowed NOAA Fisheries to maintain a long-term time series of marine recreational catch statistics while implementing changes to survey designs.

### 3.2 Non-Sampling Errors

All surveys are affected by non-sampling errors, which may be random or systematic errors that cause estimates to differ from the true population value. While no real-world data collection program can be error-free, MRIP's methods account for common non-sampling errors, and program staff continue to research how non-sampling errors may be impacting our data and estimates.

#### Coverage Error

Coverage error occurs when members of a target population are omitted, duplicated, or wrongly included in a sample frame. For example, because APAIS interviewers cannot sample private access sites, anglers who fish from private access sites are by necessity omitted from the APAIS sample frame. The use of data from public access sites to calculate recreational catch, therefore, rests on the assumption that anglers who fish from public sites have the same average catch per angler trip as those who fish from private sites.

To reduce the potential for coverage error in the FES and FHS, information gathered through the APAIS is used to calculate adjustment factors that account for off-frame fishing activity (i.e., by residents of non-coastal states or by vessels that aren't listed in the For-Hire Vessel Registry). FES survey weights are also adjusted to ensure respondent data conform to demographic data reported by the U.S. Census Bureau.

#### Measurement Error

Measurement error occurs when interviewers, respondents, or others involved in a survey make mistakes in collecting or reporting survey data. All MRIP surveys use various protocols and strategies to minimize this broad class of non-sampling error. For example, all surveys have been designed to be as short as possible (to minimize errors due to reporting burden), and all questions have been carefully worded to be as clear as possible (to minimize errors due to misunderstanding). All surveys are also designed to mitigate errors due to anglers misremembering trips: the APAIS asks anglers to report their fishing activity immediately following their fishing trip, and the questionnaire includes a calendar and multiple questions about fishing activity over time (i.e., "How many days did you go saltwater fishing in <month> and <month> of <year>? In the past 12 months?") to help respondents accurately recall and

report their fishing trips. As part of the FHS, for-hire captains are interviewed during the week following their fishing trips; a dockside validation survey is used to estimate and adjust for reporting error.

## Non-response Error

Non-response error occurs when respondents are unable or unwilling to respond to a survey, and differ in some key characteristic from those who are able or willing to respond. The APAIS, FES, and FHS assume respondents and nonrespondents respond similarly to key survey questions. This assumption has been evaluated for the FES (see [A Comparison of Recreational Fishing Effort Survey Designs](#)), but has not been tested for the APAIS or FHS. However, it is important to note that response rates to the APAIS and FHS are high, averaging 81% and 73%, respectively. While response rates alone cannot predict bias due to nonresponse error, a higher response rate does decrease the risk of such bias.

## 3.3 Limitations

Data users should be aware of the following data limitations.

### Revisions

Preliminary estimates may be revised before they are published as final, and final estimates may be revised if errors are found. The direction and magnitude of such revisions cannot be predicted. When substantial revisions are made, subscribers to our [email service](#) are notified, and notes are posted to the [MRIP Query Tool](#) and [Recreational Fishing Estimate Updates](#) webpage.

### Sample Size

Small sample sizes may result in imprecise estimates. Catch estimates for rare-event species, for example, are often less precise than catch estimates for commonly caught species. But when we group year, state, wave, or mode estimates together, sample sizes increase and precision improves. For this reason, MRIP estimates are best viewed in aggregate: annually and at the state or regional level.

### Time Series

Generally speaking, estimates from the Atlantic and Gulf coasts may be compared across extended periods of time, because [calibration methodologies](#) have been applied to account for [changes in survey design and sample coverage](#) over the years. When conducting a National Summary Query, estimates from Alaska may be compared across their time series (1996-Present), because the design of the state data collection program has not changed significantly since 1996. Estimates from Texas may be compared across their time series (1983-Present) for the same reason. However, users should exercise caution when using the National Summary Query to make such long-term comparisons for estimates in Louisiana,

California, Oregon, Washington, and Hawaii. Caution should also be exercised when comparing for-hire estimates before 2004 (prior to the full implementation of the FHS).

## Weight Estimates

Use caution with weight data.

In some cases, landed fish may not be represented in weight estimates. This can occur when no fish were observed, or when observed fish were too large for a weight measurement to take place. Furthermore, weight estimates published in the MRIP Query Tool may differ from weight estimates published by the Southeast Fisheries Science Center, which follows a different [weight estimation procedure](#) for South Atlantic and Gulf of Mexico managed species. More information about how weight estimates are produced can be found in the [Fish Weight Data](#) entry of Section 2.4 of this handbook.

## 4. Queries

### 4.1 MRIP Query Tool

The [MRIP Query Tool](#) allows users to filter catch and effort data by time series, geographic area, species, mode, and other characteristics. Query results can be viewed in table or graph form, or downloaded as a CSV file.

Most of the queries pull from MRIP's general survey data (i.e., the recreational fishing information gathered through the APAIS, FES, and FHS). However, the LPS Queries use data collected through the LPS, and the National Summary Query includes data from state programs in Louisiana, Texas, California, Oregon, Washington, and Alaska. More information about using LPS data is available in [Section 6.3](#) of this handbook; more information about the National Summary Query is available in [Section 2.4 \(Sample Coverage\)](#).

### 4.2 Types of Queries

#### Catch Data

- **Snapshot:** Provides a summary of count and weight estimates for all species caught.
- **Time Series:** Provides detailed count and weight estimates for individual species over time, as well as mean lengths and weights for harvested (Type A+B1) fish.
- **Preview of Data Standards:** Provides cumulative estimates at the annual level (i.e., across all available waves for each calendar or fishing year specified) or through the most recent wave for which data are available at the time the query is run. Also provides a preview of the precision threshold, which does not publish statistically unreliable estimates (i.e., estimates with insufficient sample sizes or PSEs above 50%).
- **Calibration Comparisons:** Provides a comparison of estimates available before the 2017 application of two calibrations to support the modified APAIS and new FES, or BASE series; estimates resulting from the application of a calibration to data collected before the 2013 APAIS design change, or ACAL series; and estimates resulting from the full application of both calibrations to data collected between 1981 and 2017, or FCAL series.
- **MRFSS/MRIP Comparisons:** Provides a comparison of original, unweighted MRFSS catch estimates and weighted MRIP catch estimates for Atlantic and Gulf coast states from 2004 through 2011.
- **Length Frequencies:** Provides length distributions for select species by inch or centimeter group.
- **LPS Time Series:** Provides detailed count estimates for individual species over time, as determined through the LPS.



- **LPS Length Frequencies:** Provides length distributions for species covered in the LPS.
- **HMS Catch Cards:** Provides billfish, swordfish, and bluefin tuna count estimates and length distributions reported in Maryland and North Carolina.

## Effort Data

- **Time Series:** Provides detailed angler trip estimates over time.
- **Preview of Data Standards:** Provides cumulative estimates at the annual level (i.e., across all available waves for each calendar or fishing year specified) or through the most recent wave for which data are available at the time the query is run. Also provides a preview of the precision threshold, which does not publish statistically unreliable estimates (i.e., estimates with insufficient sample sizes or PSEs above 50%).
- **Directed Trip:** Provides effort estimates for trips that targeted or caught a specific species. More information about the species options available (Primary Target, Secondary Target, Caught/Available Catch, Harvested/Unavailable Catch, and Released/Unavailable Catch) can be found in the [Directed Trip Species Options](#) entry of Section 2.4 of this handbook.
- **Calibration Comparisons:** Provides a comparison of estimates available before the 2017 application of two calibrations to support the modified APAIS and new FES, or BASE series; estimates resulting from the application of a calibration to data collected before the 2013 APAIS design change, or ACAL series; and estimates resulting from the full application of both calibrations to data collected between 1981 and 2017, or FCAL series.
- **MRFSS/MRIP Comparisons:** Provides a comparison of original, unweighted MRFSS effort estimates and weighted MRIP effort estimates for Atlantic and Gulf coast states from 2004 through 2011.
- **LPS Time Series:** Provides detailed trip estimates of boats targeting species covered in the LPS.

## Participation Data

- **Time Series:** Provides estimates of angler participation over time from 1981 through 2016.

## National Summary Data

- **National Summary:** Provides summaries of catch, effort, and participation data from all coastal states.

## Data Download

- **Data Download:** Provides custom downloads of catch, effort, and participation data as SAS or CSV files.

## 4.3 Example Queries

### Querying Catch Data

*How do I find out how many striped bass were caught in Maryland between 2011 and 2020?*

1. Under the “Catch Data” section of the MRIP Query Tool, select “Time Series” from the dropdown menu. Click the button labeled “Go to Query.”
2. Select 2011 as the EARLIEST YEAR and 2020 as the LATEST YEAR.
3. Select MARYLAND from the GEOGRAPHICAL AREA dropdown menu and STRIPED BASS from the SPECIES dropdown menu. Leave all other categories constant.
4. Click the button labeled “Submit Query.” Whether you decide to output the data as a table, a graph, or a CSV file, your information will be listed by year.

### Graphing Data

*How do I view the results of my effort query as a graph?*

1. Under the “Effort Data” section of the MRIP Query Tool, select “Time Series” from the dropdown menu. Click the button labeled “Go to Query.”
2. Enter your specific search needs (e.g., time period, geographical area, fishing mode). Select GRAPH from the OUTPUT FORM dropdown menu.
3. Click the button labeled “Submit Query.”

### Exporting Data

*How can I obtain red snapper catch data to use in Excel?*

1. Under the “Catch Data” section of the MRIP Query Tool, select “Time Series” from the dropdown menu. Click the button labeled “Go to Query.”
2. Enter your specific search needs (e.g., time period, geographical area, fishing mode). Select RED SNAPPER from the SPECIES dropdown menu and DOWNLOAD CSV FILE from the OUTPUT FORM dropdown menu.
3. Click the button labeled “Submit Query.” Your information will be downloaded to your device as a CSV file that can be used in Excel.

## Finding Targeted Species

*How do I calculate effort estimates for fishing trips targeting summer flounder on which summer flounder was harvested?*

1. Under the “Effort Data” section of the MRIP Query Tool, select “Directed Trip” from the dropdown menu. Click the button labeled “Go to Query.”
2. Enter your specific search needs (e.g., year, geographical area, fishing mode). Select SUMMER FLOUNDER from the SPECIES dropdown menu.
3. Hold down the Shift or CTRL key to select multiple items from the SPECIES OPTIONS dropdown menu. Select PRIMARY TARGET, which provides an effort estimate for all trips on which an angler identified the select species as their primary target; SECONDARY TARGET, which provides an effort estimate for all trips on which an angler identified the select species as their secondary target; HARVESTED - OBSERVED HARVEST (TYPE A), which provides an effort estimate for all trips on which the select species was caught and brought back to the dock in a form that could be identified by an interviewer; and HARVESTED - REPORTED HARVEST (TYPE B1), which provides an effort estimate for all trips on which the select species was caught and killed but not available for identification by an interviewer. (More information about these options can be found in the [Directed Trip Species Options](#) entry of sSection 2.4 of this handbook.)
4. Select the appropriate choice from the OUTPUT FORM dropdown menu. Whether you decide to output the data as a table or a CSV file, your information will be listed by year. Please note: This query may take several minutes to run.

## Finding Uncommon Species

*Why can't I find haddock/tarpon/menhaden in the SPECIES dropdown list?*

Less common species are not available in the SPECIES dropdown menu. When running a Catch Data: Time Series, Effort Data: Directed Trip, and/or National Summary Query, these less common species can be found by following these steps:

1. Click the button labeled “Other Species.”
2. Enter the species name in the appropriate field and click the button labeled “Search”.
3. Select the correct result from the “Common Name(s) Returned” dropdown menu and click the button labeled RETURN.

This species will now be available in the SPECIES dropdown menu for future queries.

## 4.4 Custom Queries

If the information you need is not available through the MRIP Query Tool or the public-use datasets described in [Section 2](#) of this handbook, you may [submit a custom data request](#).

Requests are handled in order of receipt. Please allow four to six weeks for your request to be filled (although in some cases, it may take less time).

## 5. Domain Analyses

In survey statistics, estimates can be produced at the most aggregated level (e.g., using a national dataset to produce a national rate or total) or for smaller sub-groups called **domains** (e.g., using a national dataset to produce state-specific rates or totals). At the most granular level, MRIP's standard domain estimates, available through the [MRIP Query Tool](#), are defined by year, two-month wave, geographic sub-region, state, fishing mode, and area fished (for effort estimates); and year, two-month wave, geographic sub-region, state, fishing mode, area fished, species, and catch type (for catch estimates). The Query Tool can be used to aggregate these base domains to encompass larger sub-groups. However, some users may require estimates for smaller or otherwise different sub-groups than the Query Tool can provide. (For example: Once NOAA Fisheries' [Recreational Fishing Survey and Data Standards](#) are fully implemented, the Query Tool will no longer include wave-level estimates or estimates with PSEs above 50%.) For these reasons, MRIP provides a suite of template analytical programs to support users in producing custom estimates that meet their unique needs<sup>2</sup>.

This section of the handbook provides guidance for producing custom domain estimates. While we have focused on the most common custom domain analyses users may need, we have also included some experimental concepts that advanced users may consider. Note: The template analytical programs are written in SAS and R. We have provided guidance on how to run them, but do recommend users have some proficiency in these languages in order to fully customize their estimates. If users would prefer to use a different software package (e.g., SPSS, Stata), we recommend they familiarize themselves with the [SAS procedure](#), [R package](#), and [statistical methodologies](#) we use in order to effectively replicate the estimation process in their preferred language.

As additional template analytical programs are made available, this section of the handbook will be updated to include the appropriate guidance.

### 5.1 General Guidance

#### Overview of Template Programs

Three sets of programs are currently available to support custom domain analyses of catch, effort, and length data. These template programs can be downloaded from the [Recreational Fishing Data Downloads](#) webpage, and are meant to be used with the public-use datasets from 2004 and beyond. Note: These programs are not meant to be used with the LPS data. More information about the datasets mentioned below can be found in [Section 2.1](#) of this handbook.

- **Catch Estimation for Custom Domains (SAS, R)** is a template program that estimates catch totals using MRIP public-use datasets. It is meant to define custom domains with

---

<sup>2</sup> Note: Custom, user-produced estimates should not be considered official MRIP products. MRIP's standard domain estimates undergo rigorous review prior to publication, and ongoing public review thereafter. While this handbook includes detailed guidance for producing custom estimates, custom analyses, custom estimates, and their use are ultimately the responsibility of the user.

the information in the TRIP\_YYYYW dataset. Catch is estimated within these domains by merging trip information with the CATCH\_YYYYW dataset.

- **Directed Angler Trips (Effort) for Custom Domains (SAS, R)** is a template program that estimates directed trips using MRIP public-use datasets. It is meant to define custom domains with the information in the TRIP\_YYYYW dataset and to estimate total angler trips within these domains. Catch information can be used to define domains by merging the CATCH\_YYYYW datasets onto the TRIP\_YYYYW datasets.
- **Length Frequencies for Single Species within Custom Domains (SAS)** is a template program that estimates length frequencies (i.e., catch totals within length bins) using MRIP public-use datasets. It is meant to calculate length frequencies (in numbers of fish and as proportions) with the information in the SIZE\_YYYYW datasets. Custom estimation domains can be defined by merging the TRIP\_YYYYW datasets onto the SIZE\_YYYYW datasets.

The SAS programs use the [SAS surveymeans procedure](#). The R programs use the [R survey package](#). Both sets of programs provide domain analysis, including point and precision estimates.

## Estimation Methods used by Template Programs

Weighted estimation methods (catch, effort, and length frequencies) are covered in detail in the [Survey Design and Statistical Methods Manual](#). While the manual includes a number of sample weighting steps for each estimate, we've already followed these steps in producing the public-use datasets, which include final sample weights (captured in the variables *wp\_int*, *wp\_catch* and *wp\_size*) to simplify the estimation process for data users. The effort and length frequency estimation processes with the public-use datasets follow MRIP's standard estimation methodology: these estimates are produced as weighted totals for the domains of interest (with length frequency estimates as weighted totals, partitioned into length bins). Catch estimation with the public-use datasets, however, is a simplified process compared to MRIP's standard methodology.

MRIP's standard catch estimation method involves calculating weighted catch rates, and then multiplying these catch rates by weighted total effort to obtain total catch; when using the public-use datasets, catch estimation is simplified using an algebraic rearrangement of the standard calculation. In this simplified method, the sample weights of the catch dataset (*wp\_catch*) are rescaled using the total effort estimates, and total catch is calculated as a weighted sum. With this method, catch estimates can be produced from the public-use datasets in a single step, which streamlines the process for data users as fewer steps and variables are required for estimation.

It is important to note that this simplified method has a limitation when it comes to variance estimation. If users try to replicate domains that are covered by the MRIP Query Tool using the template catch program (which we initially recommend to new users familiarizing themselves

with the programs), they will notice their variance and PSE estimates will not exactly match those produced by the standard method. This mismatch occurs because the effort point estimates are directly incorporated into the sample weights, but the corresponding effort variance cannot be incorporated in a similar manner for estimating total catch variance, so the template program uses a different variance estimation methodology. Based on customer feedback, easier access to estimate production has been deemed worth the trade-off of slightly different estimates of precision, but MRIP statisticians are actively exploring approaches to resolve this limitation.

## Considerations for Domain Analysis

There are two very general steps to domain analysis: (1) Exploring the data in order to inform the domain definition, and (2) producing estimates and reviewing them for any obvious anomalies that may warrant revisiting the domain definition. The key to this process is ensuring enough data are available in the custom domain to produce a robust estimate.

### Before Producing Estimates: Exploring Sample Sizes and Defining the Domain

While users have specific needs that are likely driving their domain definition, it is important to ensure that each level of their domain of interest has at least some observations. There is no threshold for a sample size that is guaranteed to produce a robust domain estimate, but zero data in a given level of a domain should serve as a warning to data users that there is not enough sample in that domain for analysis: because MRIP samples are weighted, each sampled unit represents itself along with all other units in the domain level of the sample frame (e.g., days and sites that have been sampled within a wave are weighted to represent themselves *and* all the other days and sites within the same wave). For this reason, we highly recommend users do not try to produce day- or site-level estimates: MRIP is not designed to support such fine-scale estimates, and if this guidance is not followed, users will almost certainly end up with extreme observations that will result in unreliable estimates. In general, the more data that are in each level of the user's domain definition, the more precise their domain estimates will be, so users should attempt to use the highest level of aggregation possible that can meet their needs. In [Table 2](#), we've included some of the most common examples of domain analyses that MRIP data can and cannot support.

To check the sample sizes of their domain levels, users should filter the data by their chosen domain, and produce frequency distributions of the observations in each domain level. This preliminary data exploration is not built into the template programs, so if users are unfamiliar with how to go about this process, they should consult [SAS](#) or [R](#) documentation on filtering datasets, creating new identification variables based on the content of the other columns, and producing frequency tables/plots. Alternatively, users can complete the sample size check using Microsoft Excel. In SAS, once users filter the data and assign their desired domain levels, the [Surveymeans procedure](#) has a keyword (*nobs*) that can be used to produce the number of observations in each domain level. In R, users can filter their data to the desired level, define their domain as a factor variable, then use the *summary* function to view the number of observations in each of those domain levels.

If, in looking at the frequency distribution, users notice a lot of empty domain levels, or otherwise very few sampled units in each domain level (see [Figure 1](#) for examples), they should consider redefining their domain at an aggregated level. All custom domains will have a geographic and temporal component, so if one component does not have enough data, different combinations of aggregated temporal levels with more granular geographic levels, or vice-versa, may be examined in order to improve the sample size of the domain.

We recommend users complete this check prior to producing estimates because it will help them define a more robust domain from the start of the process.

### After Producing Estimates: Additional Data Checks for Users to Consider

After producing estimates (see [Section 5.2](#) of this handbook for guidance), users should review them for possible issues, such as unusual values or highly imprecise estimates. If issues are found, users should investigate possible drivers. The three most common drivers of anomalous estimates are: (1) Extreme sample weights relative to other sample weights in that domain, (2) extreme observations relative to other observations in that domain, or (3) data that are highly clustered (e.g., all the sample available for a particular domain is limited to one week, or one or two sites, within that domain). These three drivers can work on their own, or in combination with each other.

Users can examine these drivers by producing frequency distributions of the sample weights (using the weight variables *wp\_catch* and *wp\_int*), the per-observation catch counts (using the variables *claim*, *harvest*, *landing*, *release*, or *tot\_cat*) or length measurements (using the variables *lngh*, *tot\_len*, *tot\_len\_A*, or *tot\_len\_b1*), and the sample itself (using the variable *stratid*) in their domain of interest. If any of these factors are causing issues in their estimates, users may want to revisit the sample size check within their domain level that they completed prior to running estimates; it could be that there were some data in all levels of their domain, but there still was not enough sample to produce a robust estimate. In this case, users should again consider redefining their domain at a more aggregated level to help smooth their estimates.

**Table 2: Domain Analyses MRIP Data Generally Can and Cannot Support**

Note: The domains below are NOT an exhaustive list. For any analysis listed in the “Can Do” column, users should still make sure to check sample sizes in their domain prior to producing estimates.

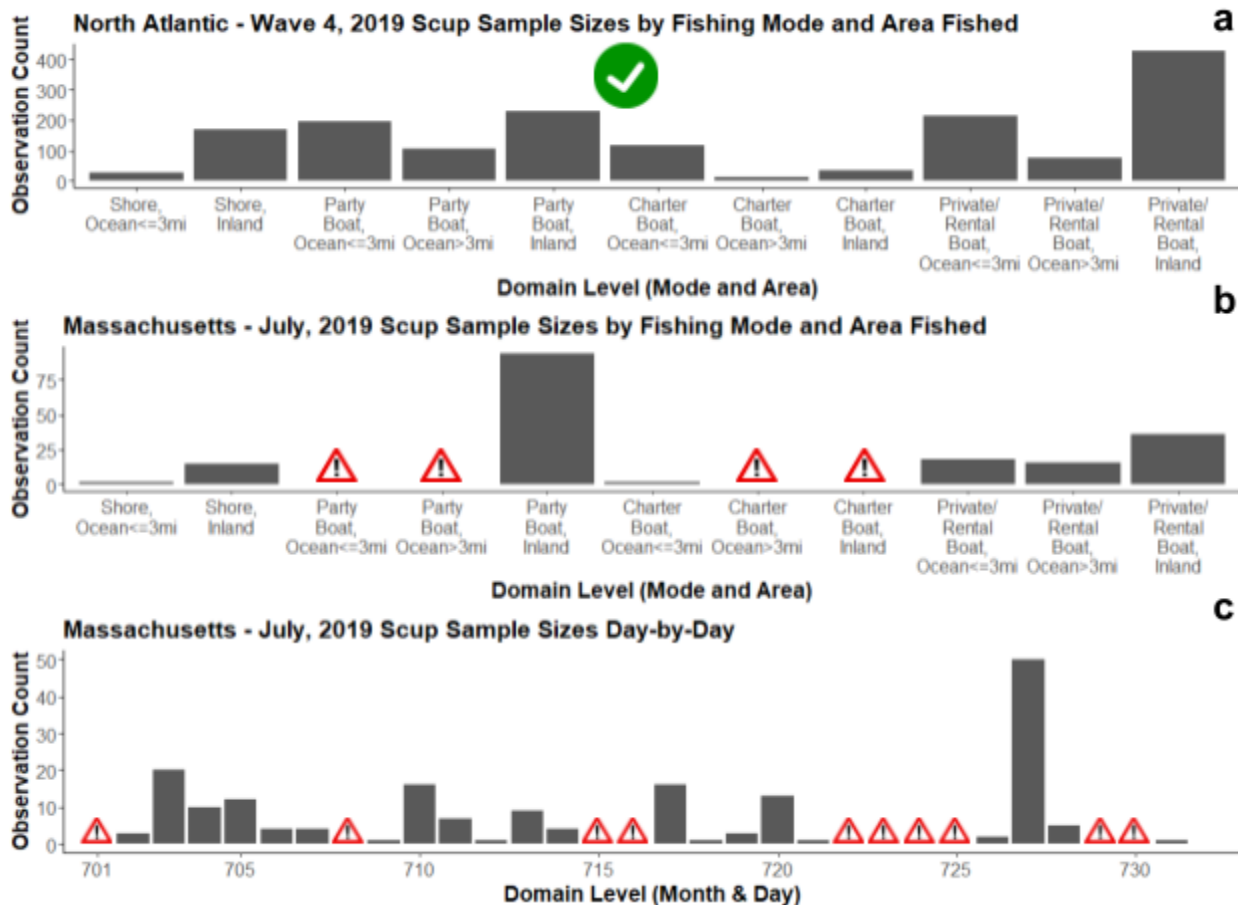
	<b>Example Analyses Users Can Do (Depending on Available Data)</b>	<b>Example Analyses Users Should Avoid (Most Common Pitfalls)</b>
<b>Geographic Domains</b>	<ul style="list-style-type: none"> <li>• County group estimates (e.g., dividing a state into northern and southern halves)</li> <li>• Regional estimates (e.g., north and south of Cape Hatteras; east and west of the Florida Keys)</li> </ul>	<ul style="list-style-type: none"> <li>• Site-level estimates</li> <li>• County-level estimates (possible in select states, but generally not recommended)</li> </ul>
<b>Temporal Domains</b>	<ul style="list-style-type: none"> <li>• Seasonal estimates</li> <li>• Fishing year (instead of calendar year)</li> </ul>	<ul style="list-style-type: none"> <li>• Daily estimates</li> <li>• Weekly estimates</li> </ul>



	estimates • Monthly estimates	
--	----------------------------------	--

Figure 1: Sample Size Check Examples

Example sample size checks for (a) fishing mode and area fished domains for scup in Wave 4 (July and August) 2019 in the North Atlantic, (b) fishing mode and area fished domains for scup in July 2019 in Massachusetts, and (c) daily domains for scup in July 2019 in Massachusetts. Panel a shows a standard MRIP wave domain definition at the regional level. Since all domain levels have at least some sample, users could proceed to produce estimates with this domain definition. The domain definition for panel b is similar to panel a, but at the month and state level; notice that a number of these domain levels are empty, so even though this domain definition falls in the “Can Do” column of [Table 2](#), this particular definition should be aggregated (e.g., by fishing mode or area fished rather than including both variables in the domain definition; or, by keeping fishing mode and area fished as domain levels, but aggregating to the regional level rather than the state level). Panel c demonstrates the lack of data available for daily estimates, providing a visual explanation of why this definition falls in the “Avoid” column of [Table 2](#): even at an aggregated level (i.e., excluding fishing mode and area fished in the domain definition), there are many empty domain levels, with most non-empty levels having few observations.



## 5.2 Common Custom Analyses

### Catch, Effort, and Length Frequencies (General)

The general steps for completing various analyses with the template programs are the same:

- 1. Download all of the data files needed for your analysis. Specify their location.**
- 2. Specify the datasets needed for analysis in the template program.**
  - a. The programs will merge all datasets specified, e.g., multiple year/wave datasets, and/or the catch/trip/size datasets.
- 3. Define your domain.**
  - a. Custom domains will, at a minimum, need to include a geographic component (e.g., sub-region, state) and a temporal component (e.g., year, wave). For catch estimates, users may also want to define their domains using variables that specify trip characteristics, or particular species, of interest. For effort estimates, the domains should be further defined as the effort targeting a specific species if directed trips are of interest. The MRIP\_Survey\_Variables guide on the [Recreational Fishing Data Downloads](#) webpage lists all the variables, and the distinct values they can take on, that can be used to define domains in the template programs.
- 4. Run the program to obtain point and precision estimates.**
  - a. For the length frequency program: Users may also want to customize length bins (e.g., creating two-inch or two-cm bins instead of the standard one-inch and one-cm bins) before running the program.

### In SAS

All three SAS programs are written using a macro (*compile\_data*) to compile and merge the necessary datasets, and the SAS *surveymeans* procedure to produce estimates. At the top of the program, users will need to specify where they have saved the SAS datasets they downloaded from the MRIP website in the *libname* statement, as well as the starting/ending year and starting/ending wave macrovariable parameters (*yr\_st*, *yr\_nd*, *wv\_st*, *wv\_end*) which the program uses to read in and compile the dataset(s) users want to analyze. (The length frequencies program has an additional macrovariable *my\_common*, where users need to enter the common name for their species of interest.) Users will also need to define their domains with the variable *my\_dom\_id* in each program. Depending on what the user wants to do, they may need to add additional code to specify more detailed domains of interest and “other” domains to exclude from the analysis (e.g., adding a data statement specifying that only one state will be used for catch estimates; adding a statement specifying a target species). Note: Annotations are present throughout the programs with simple examples and guidance on how to run the program and specify domains. Boxes 1-3 below provide simple example SAS code for using the template programs to produce catch estimates ([Box 1](#)), directed trip estimates ([Box 2](#)), and length frequency estimates ([Box 3](#)).

### Box 1: Example Catch Estimates in SAS

Catch estimates using MRIP's standard domain definition for Wave 3, 2019, in North Carolina (i.e., by fishing mode and area fished for all species and all catch types).

```
/*Step 1: Download the 2019 SAS TRIP_YYYYW and CATCH_YYYYW datasets: ps\_2019\_sas.zip. Extract to "c:\my_mrip_data" (or another location of the user's choosing). Specify the location of the SAS datasets in the libname statement of the domain_catch_totals program (Line 46 of the program)*/
    libname mrip "c:\my_mrip_data";
/*Step 2: Define the macrovariable parameters yr_st, yr_nd, wv_st, wv_end (lines 53-56 of the program)*/
    %let yr_st=2019;
    %let yr_nd=2019;
    %let wv_st=3;
    %let wv_nd=3;
/*Step 3: Add the following lines to the program defining your domain (above where the surveymeans procedure is called at the end)*/
    data my_nc_catch;
        set mycatch;
        if st=37;
        my_dom_id = compress(year||wave||st||sub_reg||mode_fx||area_x);
    run;
/*Step 4: Edit the surveymeans procedure (line 171 of the program) to run on the my_nc_catch dataset you've created in the previous step.*/
    proc surveymeans data=my_nc_catch sum cvsum varsum missing;
        strata  var_id;
        cluster psu_id;
        weight wp_catch;
        domain my_dom_id*common;
        var    tot_cat landing claim harvest release wgt_ab1;
        ods output domain=my_domain_totals;
    run;
/*Step 5: Run the full program, which will output the domain estimates to the SAS dataset "my_domain_totals" in the work library.*/
```

### Box 2: Example Effort - Directed Trips Estimates in SAS

Annual effort estimates for 2019 in North Carolina targeting Spanish mackerel (by fishing mode and area fished).

```
/*Step 1: If you haven't already, download the 2019 SAS TRIP_YYYYW and CATCH_YYYYW datasets: ps\_2019\_sas.zip. Extract to "c:\my_mrip_data" (or another location of the user's choosing). Specify the location of the SAS datasets in the libname statement of the domain_directed_trips program (Line 57 of the program)*/
    libname mrip "c:\my_mrip_data";
/*Step 2: Define the macrovariable parameters yr_st, yr_nd, wv_st, wv_end (lines 64-67 of the template effort program)*/
    %let yr_st=2019;
    %let yr_nd=2019;
    %let wv_st=1;
```

```
%let wv_nd=6;
```

/\*Step 3: Add the following lines to the program defining your domain (above where the surveymeans procedure is called at the end)\*/

/\*3a: These statements subset NC out of the full mytrip dataset (created by the compile\_data macro in the program) and then create a new variable, *dom\_id*, which equals 1 if the target species was Spanish Mackerel, and 2 if otherwise\*/

```
data my_nc_trip_smackerel;  
  set mytrip;  
  if st=37;  
  dom_id=2;  
  if prim1_common="SPANISH MACKEREL" then dom_id=1;  
run;
```

/\*3b: The sql procedure statement creates a new variable, *rs\_flag* that serves as a "check" that the *dom\_id* values were assigned appropriately. The data step then reassigns *dom\_id* based on the check, and defines the final domain code *my\_dom\_id* using standard estimation fields + the newly generated *dom\_id* variable\*/

```
proc sql;  
  create table my_nc_trip_smackerel as  
    select *,min(dom_id) as rs_flag  
    from my_nc_trip_smackerel  
    group by strat_id,psu_id,leader;  
quit;  
data my_nc_trip_smackerel;  
  set my_nc_trip_smackerel;  
  if rs_flag=1 and dom_id=2 then dom_id=1;  
  my_dom_id = compress(year||st||sub_reg||mode_fx||area_x||dom_id);  
run;
```

/\*3c: The next two statements sort and remove extra records from the above *my\_nc\_trip\_smackerel* dataset that were created by merging the catch data onto the trip dataset. The method assumes that the domain of interest has the smallest numeric code, so if you don't use the example *dom\_id* codes 1 and 2 above, be sure to code the "other" domain as a larger number than the code for the domain of interest\*/

```
proc sort data=my_nc_trip_smackerel;  
  by strat_id psu_id id_code my_dom_id;  
run;  
proc sort data=my_nc_trip_smackerel nodupkey;  
  by strat_id psu_id id_code;  
run;
```

/\*Step 4: Edit the surveymeans procedure (line 249 of the program) to run on the *my\_nc\_trip\_smackerel* dataset you've created in the previous step.\*/

```
proc surveymeans data=my_nc_trip_smackerel sum cvsum varsum missing;  
  strata strat_id;  
  cluster psu_id;  
  weight wp_int;  
  domain my_dom_id;  
  var dtrip;
```

```

ods output domain=my_domain_trip_totals;

run;
/*Step 5: Run the full program, which will output the domain estimates to the SAS dataset
"my_domain_trip_totals" in the work library.*/

```

### Box 3: Example Length Frequency Estimates in SAS

Standard length frequency estimates for Wave 3, 2019, in North Carolina by fishing mode and area fished for striped bass.

```

/*Step 1: If you haven't already, download the 2019 SAS SIZE_YYYYW and TRIP_YYYYW datasets:
ps\_2019\_sas.zip. Extract to "c:\my_mrip_data" (or another location of the user's choosing). Specify the
location of the SAS datasets in the libname statement of the domain_length_freqs program (Line 48 of
the program)*/
libname mrip "c:\my_mrip_data";
/*Step 2: Define the macrovariable parameters yr_st, yr_nd, wv_st, wv_end, and my_common (lines 54-59
of the program)*/
%let yr_st=2019;
%let yr_nd=2019;
%let wv_st=3;
%let wv_nd=3;

%my_common="STRIPED BASS";
/*Step 3: Add the following lines to the program defining your domain (above where the surveymeans
procedure is called at the end)*/
data my_nc_size;
    set mysize;
    if st=37;
    my_dom_id = compress(year||wave||st||sub_reg||mode_fx||area_x);

run;
/*Step 4: Edit the surveymeans procedure (line 155 of the template length frequencies program) to run
on the my_nc_size dataset you've created in the previous step.*/
proc surveymeans data=my_nc_size missing sum varsum cvsum mean var cv;
    strata strat_id;
    cluster psu_id;
    weight wp_size;
    domain my_dom_id*common_dom;
    class l_in_bin; *edit to "l_cm_bin" if desire cm bins instead of inch bins;
    var l_in_bin; *edit to "l_cm_bin" if desire cm bins instead of inch bins;
    ods output domain=my_domain_l_freqs;

run;
/*Step 5: Now run the full program, which will output the domain estimates to the SAS dataset
"my_domain_l_freqs" in the work library.*/

```

## In R

The R programs define a function (“MRIP.catch” in the catch program, and “MRIP.dirtrips” in the effort program) that automatically compiles data and completes the domain analysis when called. Users need to first run the program, which will store the function in their global environment. They can then call and run the function by defining the minimum required arguments: *intdir*, which is the file path to the main directory where the user has saved the CSV datasets downloaded from the website (note: when you download each year of data, you need to organize them in subdirectories named *intyyyy*, e.g., int2019 for all the 2019 data, within this main directory), *st* (state), *styr* (starting year), and *endyr* (ending year). For MRIP.catch, *common* is an optional argument which specifies the species common name (running the function without *common* defined will return estimates for all species). For MRIP.dirtrips, *common* is a required argument. For both functions, the *dom* argument is an optional argument where users can specify their domain. Domains should be defined as a series of lists that specify wave, area fished, and/or fishing mode groupings. If *dom* is not defined, the function will return estimates for each level of wave, area fished, and fishing mode. Boxes 4 and 5 below provide simple example code for producing catch estimates ([Box 4](#)), and directed trip estimates ([Box 5](#)).

Both programs are accompanied by an R documentation (.RD) file with additional guidance, which users should consult prior to running a program. Do not forget to install and load the R survey package prior to calling MRIP.catch or MRIP.dirtrips; both functions rely on functions that are defined in the R survey package, so you will get an error unless the package has been loaded.

The MRIP.catch and MRIP.dirtrip functions were developed by Gary Nelson with the Massachusetts Division of Marine Fisheries. They were written for users to be able to produce estimates using very little code, which makes it very easy to quickly produce some of the most common custom domain estimates. However, there is a caveat to this design: There is slightly less flexibility in customizing domains than in the template SAS programs. Certain custom analyses may require editing the code for the functions themselves, for which users will need a high level of proficiency in R (e.g., to produce custom effort estimates rather than directed trip estimates, since *common* is a required argument for the MRIP.dirtrips function; estimates for a species group rather than individual species, again, since *common* is separate from the *dom* argument which is where you can create groupings; custom geographic domain estimates that span multiple states, since state, *st*, is a required argument for both R functions). If users have additional questions about the R programs, they should contact Gary Nelson ([gary.nelson@state.ma.us](mailto:gary.nelson@state.ma.us)).

### **Box 4: Example Catch Estimates in R**

Catch estimates using MRIP’s standard domain definition for Wave 3, 2019, in North Carolina (i.e., by fishing mode and area fished for all species and all catch types)

## Step 1: Download the 2019 CSV datasets: [ps\\_2019\\_csv.zip](#). Extract to “c:\my\_mrip\_data\int2019” (or another location of the user’s choosing, just make sure to put all the data in that int2019 sub-directory so the function will work).

##Step 2: Load the R survey package. If you have not installed the package yet, make sure to do that first with the statement `install.packages('survey')`.

```
library(survey)
```

##Step 3: Run the MRIP.catch R script to save the function to your global environment.

## Step 4: Define the domain and call the MRIP.catch function, filling in the needed arguments. Assign the output to an object (in this example, we’ve saved to “estimates\_nc\_w3”) so that when the function is done running, you will have an output dataset that contains the domain estimates.

```
dom1 <- list(wave=list(c(3)))  
estimates_nc_w3 <- MRIP.catch(intdir = "c:\my_mrip_data", st = 37,  
                              styr = 2019, endyr = 2019, dom = dom1)
```

### **Box 5: Example Effort - Directed Trips Estimates in R**

Annual effort estimates for 2019 in North Carolina targeting Spanish mackerel (by fishing mode and area fished).

## Step 1: If you haven’t already, download the 2019 CSV datasets: [ps\\_2019\\_csv.zip](#). Extract to “c:\my\_mrip\_data\int2019” (or another location of the user’s choosing, just make sure to put all the data in that int2019 sub-directory so the function will work).

##Step 2: Load the R survey package if you have not already during your current R session. If you have not installed the package yet, make sure to do that first with the statement `install.packages('survey')`.

```
library(survey)
```

##Step 3: Run the MRIP.dirtrips R script to save the function to your global environment (if you have already done this, skip this step).

## Step 4: Define the domain and call the MRIP.dirtrips function, filling in the needed arguments. Assign the output to an object (in this example, we’ve saved to “dirtrips\_smackerel”) so that when the function is done running, you will have an output dataset that contains the domain estimates.

```
dom1 <- list(wave=list(c(1,2,3,4,5,6)))  
dirtrips_smackerel <- MRIP.dirtrips(intdir = 'c:\my_mrip_data',  
                                      common = 'SPANISH MACKEREL', st = 37,  
                                      styr = 2019, endyr = 2019, trips = 1,  
                                      dom = dom1)
```

## **Custom Geographic Domains**

Custom geographic domains can be defined in many ways: users may need to split a state into county groups, define multi-state units, or a combination of both (e.g., grouping the northern half

of North Carolina with other states north through New York; grouping the southern half of North Carolina down through the east coast of Florida; grouping parts of Massachusetts with New Hampshire and Maine for a specific Gulf of Maine stock). In the event that users are only interested in multi-state units but otherwise require standard MRIP domain definitions, users can either use a template program or use the MRIP Query Tool to sum the estimates of their states of interest.

No matter how a custom geographic domain is defined, the analysis steps will conceptually be the same: users should group the data into the desired geographic domains, and then include that grouping in their domain definition. If using SAS, users will need to add a data step to create a new variable that will flag which data are in the domain of interest or not, and then edit the domain definition (*my\_dom\_id*) to include this new variable. Users will then need to edit the *surveymeans* procedure that can be found at the bottom of each template program so that it will run on the dataset generated by the user's new data step (see [Box 6](#)). If using R, users will have to define their desired geographic groupings in the *dom* argument (see [Box 7](#)). Note that in R, users can aggregate geographic groupings at the sub-state level but cannot currently produce multi-state aggregated estimates without editing the *MRIP.catch* function and/or the required dataset, since *st* is a required argument.

#### **Box 6: Example Catch Estimates with Custom Geographic Domains in SAS**

Catch estimates for Wave 3, 2019, in North Carolina (counties grouped into North NC and South NC) by fishing mode and area fished for all species and all catch types

```

/*Step 1: If you haven't already, download the 2019 SAS TRIP_YYYYW and CATCH_YYYYW datasets:
ps\_2019\_sas.zip. Extract to "c:\my_mrip_data" (or another location of the user's choosing). Specify the
location of the SAS datasets in the libname statement of the domain_catch_totals program (Line 46 of
the program)*/
    libname mrip "c:\my_mrip_data";
/*Step 2: Define the macrovariable parameters yr_st, yr_nd, wv_st, wv_end (lines 53-56 of the program)*/
    %let yr_st=2019;
    %let yr_nd=2019;
    %let wv_st=3;
    %let wv_nd=3;
/*Step 3: Add the following lines to the program defining your domain (above where the surveymeans
procedure is called at the end). This splits the state of North Carolina into two groups of counties: North
NC and South NC*/
    data my_nc_catch;
        set mycatch;
        if st=37;
        dom_id=0;
        if st=37 and cnty in (15 29 41 53 55 139 143 177 187) then dom_id=1;      *North NC;
        if st=37 and cnty in (13 19 31 49 95 129 133 137 141 147) then dom_id=2; *South NC;
        my_dom_id = compress(year||wavel|st|sub_reg||mode_fx||area_x||dom_id);
    run;
/*Step 4: Edit the surveymeans procedure (line 171 of the program) to run on the my_nc_catch dataset
you've created in the previous step.*/

```



```

proc surveymeans data=my_nc_catch sum cvsum varsum missing;
  strata  var_id;
  cluster psu_id;
  weight  wp_catch;
  domain  my_dom_id*common;
  var     tot_cat landing claim harvest release wgt_ab1;
  ods output domain=my_domain_totals;
run;

```

/\*Step 5: Run the full program, which will output the domain estimates to the SAS dataset "my\_domain\_totals" in the work library.\*/

### Box 7: Example Catch Estimates with Custom Geographic Domain in R

Catch estimates for Wave 3, 2019, in North Carolina (counties grouped into North NC and South NC) by fishing mode and area fished for all species and all catch types

## Step 1: Download the 2019 CSV datasets: [ps\\_2019\\_csv.zip](#). Extract to "c:\my\_mrip\_data\int2019" (or another location of the user's choosing, just make sure to put all the data in that int2019 sub-directory so the function will work).

##Step 2: Load the R survey package if you have not already during your current R session. If you have not installed the package yet, make sure to do that first with the statement `install.packages("survey")`.

```
library(survey)
```

##Step 3: Run the MRIP.catch R script to save the function to your global environment (if you have already done this, skip this step).

## Step 4: Define the domain and call the MRIP.catch function, filling in the needed arguments. Assign the output to an object (in this example, we've saved to "nc\_cnty\_catch\_w3") so that when the function is done running, you will have an output dataset that contains the domain estimates. Note that the *cnty* list in the domain definition (*dom1*) is split using two combine functions (`c()`), which creates the two county groups for north NC and South NC.

```

dom1 <- list(wave = list(c(3)), cnty = list(c(15, 29, 41, 53, 55, 139, 143, 177, 187),
                                         c(13, 19, 31, 49, 95, 129, 133, 137, 141, 147)))
nc_cnty_catch_w3 <- MRIP.catch(intdir = 'c:\my_mrip_data', st = 37, styr = 2019, endyr = 2019,
                              dom = dom1)

```

## Custom Temporal Domains

Custom domains do not have to be defined by two-month waves and calendar years, and can generally be divided however the user chooses using the temporal information contained in the variable *id\_code* (e.g., to produce estimates by month, season, or for fisheries that cross calendar years). This variable is a unique 16 digit identifier for each APAIS sampling assignment, and is defined as the assignment number (1 digit), the interviewer code (4 digit), the date (in YYYYMMDD format), and the interview number (3 digit).

In SAS, users will have to parse the year/month/day information out of *id\_code*, and then write a conditional statement that creates a new variable that identifies the years/months/days they want to include in their domain (see [Box 8](#)). This variable can then be added to their domain definition.

In R, if users only require cross-year domains but otherwise standard wave estimates, these can be generated using the R functions by defining the *styr* and *endyr* arguments as their starting and ending years of interest, and including *wave* as part of the *dom* definition. However, defining sub-wave domains can be more complicated because the steps of reading in and filtering the required data are all hard-coded as part of the MRIP.catch function. Also, state (*st*) is a required argument to run the function, so estimates can only be produced state-by-state, and cannot be easily aggregated without editing the function itself. If state-level estimates are sufficient, the easiest way to generate custom temporal domains in R that does not require editing the functions themselves, is to read the CSV wave datasets that need to be filtered into R, filter them, and write the edited datasets back out to the original CSV file location with the same names. Then, users can run the MRIP.catch function, which will read in the filtered file(s) to produce estimates (see [Box 9](#)). This isn't an ideal method, because users will need to remember that those wave datasets have some data deleted and, if they later need the full wave datasets again, they will need to re-download them. However, this is currently the easiest option for custom temporal domain analysis using the R template.

#### **Box 8: Example Catch Estimates for Seasonal Domains in SAS**

2019 catch estimates for black sea bass in Connecticut during the federal season (May 15 through December 31)

```
/*Step 1: If you haven't already, download the 2019 SAS TRIP_YYYYW and CATCH_YYYYW datasets:  
ps\_2019\_sas.zip. Extract to "c:\my_mrip_data" (or another location of the user's choosing). Specify the  
location of the SAS datasets in the SAS libname statement of the domain_catch_totals program (Line 46  
of the program)*/
```

```
libname mrip "c:\my_mrip_data";
```

```
/*Step 2: Define the macrovariable parameters yr_st, yr_nd, wv_st, wv_end (lines 53-56 of the template  
program). In this case, we'll need to include waves 3-6 data in order to subset wave 3 to produce the  
seasonal estimate*/
```

```
%let yr_st=2019;
```

```
%let yr_nd=2019;
```

```
%let wv_st=3;
```

```
%let wv_nd=6;
```

```
/*Step 3: Add the following lines to the program defining your domain (above where the surveymeans  
procedure is called at the end). This data step parses out the YYYYMMDD field from the id_code, and  
then uses a conditional statement to flag everything after May 15, 2019 as part of the domain of interest.  
Note that in this example, we don't really need to parse out the year information since the required data  
are all within the same calendar year, but the year field will be needed if users are trying to do any  
cross-calendar year analyses*/
```

```
data my_seasonal_bsb_catch;
```

```
set mycatch;
```

```
if common="BLACK SEA BASS";
```

```

if st=09;
year_month_day = substr(id_code, 6, 8);
if year_month_day >= 20190515 then dom_id=1;
else dom_id=0;
my_dom_id = compress(year||st||sub_reg||mode_fx||area_x||dom_id);
run;

```

/\*Step 4: Edit the surveymeans procedure (line 171 of the program) to run on the my\_seasonal\_bsb\_catch dataset you've created in the previous step.\*/

```

proc surveymeans data=my_seasonal_bsb_catch sum cvsum varsum missing;
strata var_id;
cluster psu_id;
weight wp_catch;
domain my_dom_id*common;
var tot_cat landing claim harvest release wgt_ab1;
ods output domain=my_domain_totals;
run;

```

/\*Step 5: Now run the full program, which will output the domain estimates to the SAS dataset "my\_domain\_totals" in the work library.\*/

### Box 9: Example Catch Estimates for Seasonal Domains in R

2019 catch estimates for black sea bass in Connecticut during the federal season (May 15 through December 31)

##Step 1: Read the catch\_20193 dataset into R,, filter it to May 15 and beyond, and write it back out to the original file. This step parses out the MMDD field from *id\_code*, and then uses the dplyr filter function (if the user doesn't have the dplyr package already and wants to use the template code below, they should install it first) to exclude data prior to May 15, 2019. It then removes the newly created variable *mmdd*, so that the data can be written back out to the original file with the same number of columns, so that the MRIP.catch function can read it properly.

```

library(dplyr)
catch_20193 <- read.csv('c:\my_mrip_data\int2019\catch_20193.csv')
catch_20193$mmdd <- substr(catch_20193$ID_CODE, 10,13)
catch_20193$mmdd <- as.numeric(catch_20193$mmdd)
catch_20193_seasonal <- filter(catch_20193, mmdd>=515)
catch_20193_seasonal <- subset(catch_20193_seasonal,select=-c(mmdd))
write.csv(catch_20193_seasonal, 'c:\my_mrip_data\int2019\catch_20193.csv', row.names=T)

```

##Step 2: Load the R survey package if you have not already during your current R session. If you have not installed the package yet, make sure to do that first with the statement `install.packages("survey")`.

```
library(survey)
```

##Step 3: Run the MRIP.catch R script to save the function to your global environment (if you have already done this, skip this step).

##Step 4: Define the domain and call the MRIP.catch function, filling in the needed arguments. Assign the output to an object (in this example, we've saved to "seasonal\_dom") so that when the function is done running, you will have an output dataset that contains the domain estimates.

```
dom1 <- list(wave=list(c(3,4,5,6)))
```

```
seasonal_dom <- MRIP.catch(intdir='c:\my_mrip_data', common="BLACK SEA BASS",
st=09, styr=2019, endyr=2019, dom=dom1)
```

## Species Group Domains

Users may want to define species groups either for species complexes (e.g., produce catch estimates for a snapper/grouper complex in the South Atlantic), or for a sensitivity analysis (e.g., determine the effect of uncertainty that species identification has on an estimate, since the MRIP data includes some general taxonomic categories for fish that couldn't be identified down to the species level).

To create species groups in SAS, users should use a data step with conditional statements to generate a new species field using either the species common name (*common*) or the species code (*sp\_code*). There are various ways to accomplish this; one way would be to code the new field as an indicator variable (e.g., 1 for the species of interest, while all other species get a 0), and add it to their domain definition (see [Box 10](#) for an example).

Since *common* is a required argument to run both the MRIP.catch and MRIP.dirtrip functions in R, there is no way to currently group species using the R template program without editing the function or the underlying datasets.

### Box 10: Example Catch Estimates with Species Group Domains in SAS

Catch for Wave 4, 2019, in Florida for a five-species group of snappers and groupers

```
/*Step 1: If you haven't already, download the 2019 SAS SAS SIZE_YYYYW and TRIP_YYYYW datasets:
ps\_2019\_sas.zip. Extract to "c:\my_mrip_data" (or another location of the user's choosing). Specify the
location of the SAS datasets in the SAS libname statement of the catch estimation program (Line 46 of
the template catch program)*/
```

```
libname mrip "c:\my_mrip_data";
```

```
/*Step 2: Define the macrovariable parameters yr_st, yr_nd, wv_st, wv_end (lines 53-56 of the template
catch program)*/
```

```
%let yr_st=2019;
```

```
%let yr_nd=2019;
```

```
%let wv_st=4;
```

```
%let wv_nd=4;
```

```
/*Step 3: Add the following lines to the program defining your domain (above where the surveymeans
procedure is called at the end). This is a conditional statement that filters FL data, and creates a new
variable dom_id that groups the five specified snapper and grouper species*/
```

```
data my_snapper_grouper_catch;
```

```
set mycatch;
```

```
if st = 12;
```

```
if common = "RED GROUPER" | common = "YELLOWTAIL SNAPPER" | common = "RED
SNAPPER" | common = "VERMILLION SNAPPER" | common = "NASSAU
GROUPER" then dom_id = 1;
```

```
else dom_id = 0;
```

```

        my_dom_id = compress(year||wave||st||sub_reg||mode_fx||area_x||dom_id);
run;
/*Step 4: Edit the surveymeans procedure (line 171 of the template catch program) to run on the
my_snapper_grouper_catch dataset.*/
proc surveymeans data=my_snapper_grouper_catch sum cvsum varsum missing;
    strata    var_id;
    cluster   psu_id;
    weight    wp_catch;
    domain    my_dom_id;
    var       tot_cat landing claim harvest release wgt_ab1;
    ods output domain=my_domain_totals;
run;
/*Step 5: Now run the full program, which will output the domain estimates to the SAS dataset
"my_domain_totals" in the work library.*/

```

## 5.3 Experimental Analyses

MRIP periodically develops experimental estimation tools for users that, once made publicly available, will be added to this document with accompanying guidance. Please note that the methods below currently require more study, and it is not yet an official NOAA Fisheries recommendation to use them.

### Multi-Year Averaging

While most species are encountered frequently enough by the APAIS to produce robust and reliable estimates, others are not. These “rare-event” species have highly variable estimates, or estimates that are missing in some years. These types of estimates can be difficult to use in stock assessments and management, so MRIP has been evaluating different methods for improving estimates for these species. MRIP statistical consultants have specifically evaluated two different multi-year averaging methods: moving averages, where the surrounding years before and after the target year are used in estimation; and time series predictions, where the years leading up to the target year are used in estimation. Preliminary findings suggest that these methods can improve the precision of individual estimates and stabilize time series (i.e., produce estimates for the target year even when there is no data in that year for their species of interest, and minimize extreme year-to-year variability of the estimates). In theory, estimates produced using moving averages would be best suited for improving the overall time series for a rarely encountered species, while time series predictions would be better suited for use in management, such as setting catch limits for the upcoming year. For more information on these methodologies and a 21-species example case study, see [Opsomer et al 2020](#).

### Catch Frequency Estimation

Species-specific catch frequencies, or catch per unit effort estimates, are often used as indicators of stock abundance. MRIP data can theoretically be used to produce counts of angler

trips by the number of landed fish per trip (e.g., in the defined domain, an estimated 1000 angler trips landed one summer flounder, 500 angler trips landed two summer flounder, 200 angler trips landed three summer flounder, and so on), as well as proportions of total angler trips that landed a certain number of fish per trip (e.g., 10% of the total angler trips in the domain landed one summer flounder, 5% of the total angler trips landed two summer flounder, 2% of the total angler trips landed three summer flounder, and so on). A SAS template program for catch frequencies is currently under development, and will be made available to users in the near future. Users will ultimately be able to customize domains in the same manner as the other template programs.

## 6. Large Pelagics Survey

The [LPS](#) consists of three components: the Large Pelagics Intercept Survey (LPIS), Large Pelagics Telephone Survey (LPTS), and Large Pelagics Biological Survey (LPBS).

- The **LPIS** collects catch information from vessel trips on which anglers targeted large pelagic species using rod and reel or handline gear. The target population is stratified by year, month, state, and fishing mode (private or charter). Sampling units are offshore LPS fishing trips returning to public boat sites.
- The **LPTS** collects effort information from a list of vessels with the NOAA Fisheries permits that are required to fish for Atlantic tunas, billfishes, swordfish, and sharks using rod and reel or handline gear. These data are used to estimate the total number of boat trips on which anglers targeted large pelagic species using rod and reel or handline gear. The target population is stratified by year, sample reporting period, state, and fishing mode (private or charter). Sampling units are individual vessels with an HMS permit.
- The **LPBS** collects length and weight data and biological samples (e.g., otoliths, white muscle tissue, first dorsal spines, and gonads) for Atlantic bluefin tuna and other large pelagic species. Unlike the LPIS or LPTS, the resulting data are not incorporated into NOAA Fisheries' recreational catch estimates. Instead, the data and samples are used by the [Southeast Fisheries Science Center](#) to assess the bluefin tuna stock's age structure, population genetics, growth rates, and reproduction rates.

Fishing modes within the LPS consist of private and charter boat.

- **Private** mode includes vessels with the HMS Angling Category permit or the Atlantic Tunas General Category permit.
- **Charter** mode includes vessels with the HMS Charter/Headboat Category permit.
- Vessels targeting large pelagic species without an HMS permit may be intercepted by the LPIS. The fishing mode of these vessels (private vs charter) is determined by their responses to LPIS questions. The data collected through these intercepts are used in the effort estimation process to account for off-frame fishing trips.

The LPS is conducted on the Atlantic coast from Maine through Virginia. To complement NOAA Fisheries' Highly Migratory Species management, these states are divided into two regions: the Northern Atlantic, which extends from Maine through northern New Jersey, and the Southern Atlantic, which extends from southern New Jersey through Virginia. (New Jersey is divided along the border of Atlantic and Ocean counties.) Due to the limited scale of recreational fishing for large pelagic fish, several states are further combined into two-state groups: Maine and New Hampshire, Connecticut and Rhode Island, and Maryland and Delaware.

## 6.1 Data Products

Each LPS .zip file (labeled either LPS\_YYYY\_sas.zip or LPS\_YYYY\_csv.zip) from 2011 to present contains **monthly** (June-October) LPIS and LPTS datasets, which include monthly catch, main, and size files; a monthly LPTS file; and a single All Estimates file (labeled all\_ests\_YYYY). Each LPS .zip file from 2002-2010 contains **annual** LPIS and LPTS datasets, which include annual catch, main, and size files; an annual LPTS file; and a single All Estimates file.

### Large Pelagics Survey Microdata and Estimates (SAS, CSV)

The recreational fishing information gathered through the LPS and the catch estimates produced from these microdata are available in SAS and CSV formats and include:

- **CATCH\_YYYYMM.sas7bdat or CATCH\_YYYYMM.csv:** Catch-level LPIS data and variables required for estimation. Each record contains catch totals by species caught for major catch disposition types (kept, released alive, or discarded dead) in numbers.
- **MAIN\_YYYYMM.sas7bdat or MAIN\_YYYYMM.csv:** Boat-level LPIS data and variables required for estimation. Contains one record per respondent interviewed.
- **SIZE\_YYYYMM.sas7bdat or SIZE\_YYYYMM.csv:** Fish-level length data and other variables. Contains one record per fish that is measured by the interviewer during the intercept survey.
- **LPTS\_YYYYMMM:** Private mode phone interview data. The variable LPTRIPS indicates the number of LPS trips reported. Contains one record for each trip profiled. (Note: The LPTS\_2002 dataset is not available for download. It will be provided at a later date.)
- **All\_ests\_YYYY.sas7bdat or All\_ests\_YYYY.csv:** Catch and effort estimates by month, mode, area, and species.

More information about variable formats and descriptions can be found in the LPIS\_Survey\_Variables (.xls), LPTS\_Survey\_Variables (.xls), and LPS\_Estimates\_Variables (.xls) guides located within the [LPS Data Downloads](#) folder.

## 6.2 Data Availability

### Data Collection Schedule

The LPS is administered from June through October in coastal states from Maine through Virginia.

### Estimation and Publication Schedule

Preliminary LPS estimates are published approximately two months after the end of the month of data collection (e.g., June estimates are published in early September, July estimates are



published in early October, and so on). Final estimates, as well as updates to the data products described in [Section 6.1](#) of this handbook, are published in April of the following year.

Note: Once NOAA Fisheries' [Recreational Fishing Survey and Data Standards](#) are fully implemented, cumulative estimates will be published instead of monthly estimates, beginning with the first survey administration of the survey year. The switch to cumulative LPS estimates is expected to occur following the certification of the improved LPS design.

## 6.3 Using LPS Data

Because trips targeting large pelagic and highly migratory species are specialized, they often fall outside of the APAIS, FES, and FHS sample frames. The LPS was designed to fill this gap in coverage, and is conducted during the months where the majority of large pelagic and HMS fishing activity occur. As such, LPS estimates for tuna, sharks, and billfish are generally more precise than MRIP's general survey estimates for these species. For this reason, we recommend using LPS Time Series queries to access catch and effort estimates for those species in the geographic areas and during the months covered by the LPS (Maine through Virginia; June through October). For catch estimates outside of these areas and months, the more general Time Series queries can be used, although estimates should be viewed with caution as PSEs increase beyond 30%.

Guidance for using LPS data to conduct custom domain analyses will be provided at a later date. At present, the LPS data collection and estimation methods are undergoing a significant redesign. (The primary issues with the current LPS include deviations from formal probability sampling in the LPIS, and the use of unweighted estimation methods that don't account for the survey's complex sampling design.) The redesign is scheduled for completion by 2024. Once the survey's statistical issues have been addressed, template analytical programs and accompanying guidance will be added to this handbook.