MRIP Estimate Datasets

NOAA Fisheries developed survey and data standards in 2020 to guide the design, improvement, and quality of data produced by the agency's recreational fishing surveys. The standards serve as a single set of guidelines on critical aspects of recreational fisheries data collection and estimation. The shared use of a single list of survey requirements and guidelines helps promote data collection and distribution consistency across the national network of recreational fishing surveys.

NOAA Fisheries switched from producing estimates for catch and effort in 2-month increments, or waves, to producing cumulative estimates to meet our commitment to providing quality data in line with our recreational fishing survey and data standards. Rather than seeing an estimate produced in 2-month wave increments, data users will now see a cumulative estimate that will improve in precision throughout the year. This change increases our sample sizes, thereby producing more reliable estimates of recreational catch that improve in precision throughout the year.

Catch and effort estimate datasets available on the Recreational Fishing Data Downloads page now provide cumulative estimates. These datasets contain estimates at multiple levels of aggregation in order to increase ease of use. Catch and effort estimates are provided at the following levels in each of the available estimate datasets:

- Level 1: year, sub-region, state, mode of fishing, area of fishing, species (for catch estimates only)
- Level 2: year, sub-region, state, mode of fishing, species (for catch estimates only)
- Level 3: year, sub-region, state, species (for catch estimates only)
- Level 4: year, sub-region, species (for catch estimates only)

MRIP Survey Datasets

Each PS_YYYYW.zip file contains three files per year/wave:

- TRIP_YYYYW
- CATCH_YYYYW
- SIZE_YYYYW

See below for descriptions of these files and additional survey datasets.

Imputation is a statistical technique in which missing data points are filled in with estimated values based on known information. In rare cases (due to no or limited survey responses, incomplete responses, natural disasters, temporary suspension of survey sampling, or other unforeseen issues), MRIP survey data are imputed to fill in gaps in data. These imputed data— also known as proxy or replacement data— are from previous years and match the time (wave or month(s)), place, and fishing mode combinations that would have been sampled had the given survey continued uninterrupted. When used, imputed data are combined with observed data to produce catch estimates through our standard estimation methodology.

Imputed data can be identified using the IMP_REC variable in the TRIP_YYYYW, CATCH_YYYYW, and SIZE_YYYYW files. In cases where data were imputed, the original ID_CODE was maintained. The original year the data was collected can be found in characters 6-9 of the ID_CODE. Additionally, there are variables in the catch time series query and general survey catch estimate downloads that show the weighted percentage of catch rate information that can be attributed to imputed catch data.
Note: *The weighted percentage does* ***NOT*** *represent the percent difference between the estimate with and without imputed data incorporated. For more information on weighting catch rate data, visit our estimation methods webpage.*

—
Most notably, MRIP imputed data following disruptions due to COVID. In 2020, there were waves with no APAIS field sampling, and disruptions to data collection continued into 2022 due to state field sampling staffing shortages. NOAA Fisheries was able to fill gaps in 2020-2022 catch data with data collected in 2018 and 2019. Imputed data were combined with any available observed data to produce catch estimates using our standard estimation methodology.
Note: *In this instance, the use of imputed catch data had minimal impact on the agency's effort estimates, as the Fishing Effort Survey (mail) and for-hire effort telephone survey continued largely uninterrupted. Overall, COVID led to relatively minimal impacts on recreational fishing data trends and patterns. To ensure imputed data were not over-represented against observed data, the original sample weights for the 2018 and 2019 catch records were down-weighted.*

TRIP_YYYYW.sas7bdat

Trip level data (analogous to MRFSS i1 dataset) and variables required for use in estimation. Contains one record per angler-trip interview (identified by id_code).

Survey design variables include:

- Strat_id identifies the design stratum
- Psu_id identifies the primary sampling unit (site-day)
- Id_code identifies the angler-trip
- Wp_int is the post-stratified sampling weight to use in weighted estimation

The For-Hire Effort survey provides Charter boat/Headboat effort estimates for individual collapsed areas of fishing (area_x). It is possible to have a Charter or Headboat effort estimate for a given area (e.g., inland, ocean < 3 miles, ocean > 3 miles) with no corresponding trip data from the APAIS (dockside intercept survey). Charter and/or Headboat records are added to the trip dataset to account for these situations. While there are no catch data associated with these records, including them allows the data user to calculate trip totals that match those reported in the MRIP effort web queries 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.

CATCH_YYYYW.sas7bdat

Catch level data and variables required for use in estimation. Contains one record per species for every angler-trip interview. Each record contains catch totals by major catch types (A, B1, B2) in numbers, weight (kg), and length (mm).

Survey design variables include:

- Strat_id identifies the design stratum
- Psu_id identifies the primary sampling unit (site-day)
- Id_code identifies the angler-trip
- Wp_catch is the post-stratified sampling weight to use in weighted catch estimation

SIZE_YYYYW.sas7bdat

Fish level length and weight data and variables required for use in estimation. Contains one record per fish caught and measured or weighed by interviewer. Missing lengths and/or weights are imputed as needed for individual fish records. The lngth_imp and wgt_imp fields have values of '1' when the corresponding lngth and wgt fields have imputed values.

Survey design variables include:

- Strat_id identifies the design stratum
- Psu_id identifies the primary sampling unit (site-day)
- Id_code identifies the angler-trip
- Wp_size is the post-stratified sampling weight to use in weighted estimation for the size dataset only; it contains an additional adjustment for situations when only a subset of landed fish are measured for an angler-trip

FES_HH_YYYYW

Household-level data collected from completed FES surveys. Contains one record for each household (identified by the variable HH_id). Each record includes the number of household residents, household demographic information and the total number of shore and private boat fishing trips taken by household residents during the reference wave.

Design variables include:

- Stratum_ID: Identifier for the design stratum,
- HH_ID: Unique household identifier
- Final_wt: Final survey weight used in weighted estimation.

FES_Person_YYYYW

FES_PERSON_YYYYW: Person-level data collected from households that responded to the FES. Contains one record for each person (identified by the variable PERSON_id) residing in a responding household. Each record includes demographic information and the total number of shore and private boat fishing trips taken by the household resident during the reference wave.

Design variables include:

- Stratum_ID: Identifier for the design stratum,
- HH_ID: Unique household identifier
- Person_ID: Person identifier unique within each household
- Final_wt: Final survey weight used in weighted estimation.

MRIP Template Programs

** The estimation approach used in the MRIP template programs is different from our standard estimation process because the analysis weights ("wp_" weights) incorporate the effort estimates directly allowing for a streamlined version of estimation. The streamlined estimation approach will return the correct point estimates, but the variance estimates may differ from those of the official estimates.

- **domain_catch_totals.sas**
  MRIP catch estimation for custom domains.

  This is a template program for estimating catch totals using the MRIP public-use datasets.

  The program is setup to use information in the trip_yyyyw dataset to define custom domains. The catches are estimated within the domains by merging the trip information onto the catch_yyyyw datasets. See program for additional information.

  For more information on total catch estimation for public use datasets, please see section 6.2.3 Total Catch Estimation for Public Use Datasets in the MRIP Survey Design and Estimation Methods for Recreational Fisheries Statistics document.

- **domain_directed_trip.sas**
  MRIP directed angler-trips (effort) for custom domains.

  This is a template program for estimating directed trips using the MRIP public-use datasets.

  The program is setup to use trip_yyyyw datasets to define custom domains and estimate total angler-trips within domains. Catch information can be used in defining the domains by merging the catch_yyyyw datasets onto the trip_yyyyw datasets. See program for additional information.

- **domain_length_freqs.sas**
  MRIP length frequencies for single species within custom domains.

This is a template program for estimating length frequencies using the MRIP public-use datasets.

The program is setup to calculate length frequencies (in numbers of fish and as proportions) using the size_yyyyw datasets. Custom estimation domains may be defined by merging trip information from the trip_yyyyw datasets onto the size datasets. See program for additional information.

- **domain_FES.sas**
  FES effort estimation for custom domains.

  This is a template program for estimating resident angler effort totals or distributions using the MRIP public-use datasets.

  The program is setup to use information in the FES_yyyyw dataset to define custom domains. See program for additional information.

Domain: a defined sub-level, geographic or otherwise, of the stratified estimation design. Typically, these are sub-state geographic divisions based on county groups, or even sampled site groups if the domain cannot be defined by county boarders.