

---

## CA MRFSS/CRFS and OR MRFSS Data Formatting

Data made available on 9/29/2009, via digital download from:

Wade Van Buskirk ([wade@psmfc.org](mailto:wade@psmfc.org))

RecFIN/PSMFC

205 SE Spokane St., Suite 100

Portland, OR 97202-6413

Phone: (503) 595-3100

Fax: (503) 595-3232

<http://www.recfin.org/>

### Modification/Formatting

Location information for MRFSS/CRFS data is stored as start and stop coordinates made on commercial passenger fishing vessel trips. Van Buskirk was not able to easily summarize this data by fishing block, so the data was requested in a raw format for conversion and summary.

The raw data consisted of three regions (Southern California, Northern California, and Oregon), each divided into three tables (Locations, Catch Counts, and Boat Information).

### Locations:

- 43,904 Total locations in original raw data.
- Locations were stored in four coordinate formats: DDMM.MM, Site Code, DDMMSS, and Loran.
  - Coordinate type was designated by a GFORMAT field (1-4)
    - Only 10 total records were associated with Site Codes, and 45 records associated with Loran coordinates. These were excluded.
  - Longitude values are missing the left-most "1" (survey forms have the "1" pre-printed and this is not added back to the coordinates during data-entry).
    - The "1" was added back to all longitude coordinates.
  - Some records had starting coordinates (3964 records), but no ending coordinates. Others had ending coordinates, but no starting coordinates (67).
    - Locations with only starting coordinates are valid according to the MRFSS Interviewer Manual.
    - Locations with only ending coordinates were considered invalid based on personal communication with Van Buskirk.
    - Some records had noticeably invalid values for a lat or long in a coordinate pair. These were not used as point locations since they seemed to indicate a transect that was entered incorrectly.
- Locations of type GFORMAT 1 (DDMM.MM) and 3 (DDMMSS) were extracted into their own tables. New fields were added so that the latitude and longitudes could be formatted with the correct number of digits.

- DD.DD coordinates were then calculated for all extracted records so that the XY events could be generated in ArcMap.
- Points were created for those coordinates with only single values.
- Lines were generated for records with pairs of coordinates (a *Hawth's Tools* table tool was used to create the lines from the coordinate set).
  - Some transects were obviously incorrect (extremely long and/or on land). Van Buskirk suggested calculating the mph and excluding any transects over 20 mph.
    - Length of transects in miles were calculated and total minutes were calculated using recorded start and end times.
    - Mph was calculated by dividing the distance by the total minutes and multiplying that value by 60.
    - **Transects that were greater than 10 miles or had speeds greater than 20 mph were excluded.** (Van Buskirk stated that the trawls shouldn't be more than a few miles. We chose 10 miles as the cut-off because we are assigning fishing block IDs to the locations based on the center-points of the transects. A 10 mile restriction limits the maximum number of fishing blocks that a single transect could cross to two.)
- Points were created from good transects by using the "Feature to Point" tool in ArcMap. Each point was placed at the midpoint of the transect lines.
- All points were intersected with the CA Fishing blocks and the OR/WA Fishing blocks to find the block ID of the block that each location fell into. 38,804 final point locations fell inside the fishing blocks.

#### Table Modifications:

- Species codes in the Catch tables were not in a consistent format.
  - After bringing this up to Wade Van Buskirk, he confirmed that they were a mixture of either 10-digit numeric, ODFW, RecFIN, or Alpha5 codes.
  - Fields were added to Catch tables to separate the code types.
  - Sub-catch tables were separated by catch code type and joined to a master species table provided by Van Buskirk, then merged back together.
    - 30 SoCal Catch records had missing or invalid species codes.
    - 44 NorCal Catch records had missing or invalid species codes.
    - 76 OR Catch records had missing or invalid species codes
- Boat, Location, and Fish Kept tables were joined together based on RecFIN's ASSN number. Only records that had matching ASSN's among all three tables were kept.
  - Oregon: the final joined table contained 40,014 records (Fish Kept originally had 44,992 records). 88.94% of the records were retained.
  - NorCal: the final joined table contained 78,549 records (Fish Kept originally had 79,598 records). 98.68% of the records were retained.
  - SoCal: the final joined table contained 200,328 records (Fish Kept originally had 208,104 records). 96.26% of the records were retained.

- The joined, flat-file tables for each region were summarized by Year, Month, Fishing Block, and Species to uphold RecFIN confidentiality requirements.
  - Tables were summarized using an “Aggregate Features/Records” tool from *XTools Pro for ArcGIS desktop*.
  - Catch per unit effort (CPUE) was calculated pre-summary by dividing the kept fish (KEPT) by observed anglers (OBSANG).
    - avg\_CPUE and sum\_CPUE were included in the summary.
  - Stops were calculated by dividing the sum\_CPUE by the avg\_CPUE
  - KEPT values were summed.
    - Null values represent zero fish of a particular species caught on a specific stop on a trip.
  - MAXDEPTH and MINDEPTH (in feet) were averaged. Unfortunately, not all values are present, so some mindepth values are greater than maxdepth values, or the two are identical because that’s how it was reported.
  - Oregon, SoCal, and NorCal RETD, RETDALIV, and RETDDEAD values were at least 90% null/invalid, so those fields were dropped (Van Buskirk indicated that this sort of information has only been recorded more recently).
  - The following fields were not included as part of the summary or final data set: ASSN, ELAT, ELON, ETEMP, FTYPE, GFORMAT, LOCNUM, PINNIPED, PLBAT, PLFISH, PLGEAR, PLTIME, PRMOVE, SITENAME, SLAT, SLON, STEMP, STIME, WAVE, ALPHA5, COUNTER, DROP, LOCNUM, SPNUM, BOATNAME, BOATNUM, CNTY, INTSITE, INTVEUER, LANDING, NUMLOCS, NUMSP, ST TRPTYP, AREA, CAPTAIN, SG\_CODE, SUPER, GROUP, ORDER, FAMILY, GENUS, HART, MLEE, ESCH, LOVE, A\_FL, B\_FL, A\_TL, B\_TL, A\_FT, B\_FT, NB\_ST, NB\_CNTY, SB\_ST, SB\_CNTY, N3, N2, REGION, CSG, CSG\_NAME, CG\_NAME, CG, NAME, NODC7, NODC8, TSN, CDFGSP, GP\_CODE, RECFINSP, P1, P2, FMP\_CODE, ODFWSP, N\_FL, SP\_PSBS, SP\_WABDS, SP\_PACFIN, and F48.

Field Definitions

**YEAR:** Year of catch.

**MONTH:** Month of catch.

**COMMON\_NAME:** AFS Common Name.

**SCI\_NAME:** AFS Scientific Name.

**SP\_CODE:** MRFSS Species Code.

**KEPT:** Number of fish kept.

**BLOCK10\_ID:** Fishing block identifier code.

**AVG\_MAXDEPTH:** Mean of maximum fishing depth values in feet.

**AVG\_MINDEPTH:** Mean of minimum fishing depth values in feet.

**avg\_CPUE:** Mean of catch per unit effort values (CPUE was calculated by dividing the number of fish kept by the number of observed anglers prior to summary).

**sum\_CPUE:** Sum of catch per unit effort values (CPUE was calculated by dividing the number of fish kept by the number of observed anglers prior to summary).

**Stops:** Summarized number of stops made by vessels.