discaRd steps for CAMS

Ben Galuardi

Discard Rate Current Method Summary (J. Michael Lanning summary)

1. Rates determine by observer reported values (gear, area, etc)
2. Incomplete observed trips have missing 'hauls' prorated by observed information from that trip
3. Trips with observer get reported/calculated observed discards of that specific trip
4. Unobserved trips get discards from the rate calculated from 1)
5. QM is only interested in the summary total of discards for each trip, not subtrips. Often the interested number is a summary of trips, ie. the herring total of bycatch for an area/season or a sector's season's total of GB Cod.
6. Other others are driven by regs. Here I would place transition rates and EM methods.

Tables created and steps to date

Rates determine by observer reported values (gear, area, etc)
Observed discards

`make_obdbs_table_cams_v2.sql`

created:

- `apsd.bg_obdbs_cams_mock2018`
- `apsd.bg_obdbs_cams_mock2019`
- `apsd.bg_obdbs_cams_mock2020`

Prorated discards

Incomplete observed trips have missing 'hauls' prorated by observed information from that trip

Prorate observed discards on unobserved hauls within a subtrip. This is done by applying a ratio of kept all on the entire trip to kept all on the unobserved hauls only

$$d_{total} = d_{observedhauls} * (1 + KALL_{unobservedhauls} / KALL_{subtrip})$$

`make_obdbs_prorate.sql`

created:

- `apsd.obs_cams_prorate`
  - this table was made using `apsd.bg_obdbs_cams_mock2018` and `apsd.bg_obdbs_cams_mock2019`

Use prorated observed discard values

Trips with observer get reported/calculated observed discards of that specific trip

Match observed hauls to subtrips

`explore_link3_mesh_match.sql`

This step matches on `AREA`, `GEAR` and `MESHGROUP` (sm, lg, xlg). This is a hard match and will go awry if there is a mismatch in the data.

Use prorated observed discard values (cont.)

tables created:

- `apsd.bg_cams_catch_mock`
  - follows the steps layed out for mid-Atlantic discard estimation. `Gear`, `mesh`, `region`, `half of year` CASE statements should be replaced at some point with table driven code.
  - Utilizes the current apportionment table: `apsd.cams_apport_20201230`
- `apsd.bg_obs_cams_tmp1`
  - links to `dmis.d_match_obs_link` and `apsd.bg_cams_catch_mock`; **multiple** subtrips only.
- `apsd.bg_obs_cams_tmp2` is used in the `squid example` and include all trips.

R functions

- `get_obs_disc_vals`
- `make_assumed_rate`
- `make_bdat_focal`
- `run_discard`

Running it

- refresh Oracle tables?
- import to R
- define species/year (time period?!?)
- run `run_discard`

output

```
> dest_strata %>% slice(grep('Otter Trawl_sm*', dest_strata$STRATA))
                STRATA    N   n orate       drate     KALL disc_est   CV
1 Otter Trawl_sm_N_1 1307 122  0.09 0.007531621 21959570   165391 0.41
2 Otter Trawl_sm_N_2 1943 212  0.11 0.004653792 29085106   135356 0.67
3 Otter Trawl_sm_S_1 2023 373  0.18 0.054646780 25624532  1400298 0.23
4 Otter Trawl_sm_S_2 1930 432  0.22 0.005260938 33174274   174528 0.44
```

```
$res
# A tibble: 111,024 x 24
   VTRSERNO DISCARD DMIS_TRIP_ID  YEAR MONTH REGION HALFOFYEAR GEARNM GEARCODE GEARTYPE NEGEAR MESHGROUP CAREA
   <chr>      <dbl> <chr>        <dbl> <dbl> <chr>        <dbl> <chr>  <chr>    <chr>    <chr>  <chr>     <chr>
 1 10258111       0 410416_1905~  2019     5 N                1 DREDG~ DRS      Scallop~ 132    NA        526
 2 10761479       0 410571_1905~  2019     5 S                1 DREDG~ DTC      Scallop~ 132    NA        622
 3 10765215       1 330672_1908~  2019     8 N                2 OTTER~ OTF      Otter T~ 050    lg        511
 4 11177077       0 250547_1902~  2019     2 N                1 OTTER~ OTF      Otter T~ 050    lg        513
 5 11177654       0 330167_1912~  2019    12 S                2 DREDG~ DRS      Scallop~ 132    NA        622
 6 11179863       0 410019_1905~  2019     5 N                1 DREDG~ DRS      Scallop~ 132    NA        526
 7 11198443       1 410195_1906~  2019     6 N                1 DREDG~ DTC      Scallop~ 132    NA        526
 8 11498925       0 320422_1901~  2019     1 S                1 DREDG~ DRS      Scallop~ 132    NA        616
 9 11498942       0 320422_1908~  2019     8 S                2 DREDG~ DTC      Scallop~ 132    NA        615
10 11540799       0 310915_1905~  2019     5 S                1 DREDG~ DRS      Scallop~ 132    NA        622
# ... with 111,014 more rows, and 11 more variables: LIVE_POUNDS <dbl>, DOCID <chr>, STRATA <chr>, SEADAYS <dbl>,
#   DISC_RATE <dbl>, CV <dbl>, STRATA_ASSUMED <chr>, ARATE_IDX <int>, ARATE <dbl>, CRATE <dbl>, EST_DISCARD <dbl>
```

compare to ACL summary

	NESPP3	SPPNM	`discaRd CAMS`	`ACL Discard`
	<chr>	<chr>	<dbl>	<dbl>
1	011	ANGLER	239114	NA
2	012	ANGLER	88454364.	NA
3	051	BUTTERFISH	2638784.	3025979
4	125	SAND-DAB FLOUNDER	2492716.	NA
5	212	ATLANTIC MACKEREL	607012.	490416
6	801	LONGFIN SQUID	1077751.	870063
7	802	ILLEX SQUID	2349929.	2715331

TO DO

- utilize support tables: `STRATA` especially..
- add `SECTOR` for multispecies (see above)
- utilize discard mortality table (species/gear/mesh)
- decide on time periods
- decide on criteria for assumed (fallback) rates: how simplified must this be?
- implement transitions (if using fixed time period)
- refine exact operational process
- ?? what else