

JUNE 10 2022

Constructed a dataset that represents the fish species caught by the household

Input files:

- Catch & Consumption_Q7.xlsx
- Catch & Consumption_Q7sub.xlsx

Output file:

- ccm_q7.csv

Notes:

- Current version of the file contains only species that were caught by household (if a species isn't listed, it was not caught)
- Current file contains weight of the species (catch_iweight) and its use breakdown (soldfresh_iweight, atefresh_iweight, process_iweight, lost_iweight, other_iweight). In cases where there were multiple species per "species group", we divided the weights evenly. This was 26% of groups. These cases are indicated by the "duplicate" variable.
- Current file contains binary variables indicating whether a given species was used in a given way (e.g. sold, eaten, lost, processed, etc)
- The weights reported in the use breakdowns do not always sum to the total weight calculated. If/when we use the use breakdown *weights* we need to look into this more.
 - Ideas for further investigation: Recalculate breakdowns based on percentage allocations; check to see if merge with species level data is amplifying discrepancies
- Current file contains binary variables indicating whether a given species was used in a given way (e.g. sold, eaten, lost, processed, etc)

Constructed a dataset that represents the fish species collected during biomonitoring activities

Input files:

- Biological Monitoring_FormBSub.xlsx
- Biological Monitoring_BasicInfo.xlsx

Output file:

- biom_formb.csv

Notes:

- This file contains the list of species that was caught during biomonitoring in each CFR and the weight of each species caught at a given time point. All replicates are summed so that the data are at the CFR-date-species level.
- All OAA have been dropped from this file.
- There are no group codes given in the biomonitoring data. There are several cases of fish species that appear in biomonitoring data, but not in catch data, and seem to have no group code in any of the files. To distinguish these from any other potentially missing group codes that might appear in other parts of the data construction process, I assign them group code 0.

- Any analysis at the group level (which I don't think we are planning) should drop these as they share a group code, but they don't necessarily have characteristics in common.
- If needed, these species could be classified into existing groups by someone familiar with these fish.

Set up time keys to combine catch and biomonitoring datasets

Because the catch and biomonitoring data were collected on different schedules, they do not merge on the same month-year time points. The alignment of data collection is located in [timekey_mapping.xlsx](#) which contains the three time keys that I created.

- timekey1: aligns all catch data with the biomonitoring data point that precedes it. The distance in time from biomonitoring to catch data ranges from 0 months to 2 months (e.g. timekey1 = 5 includes data from November and January.)
- timekey2: aligns all catch data with the biomonitoring data point that precedes it within 2 months (e.g. timekey2 = 5 only contains data from November. timekey2 = 2 contains data from February and March). The distance in time from biomonitoring to catch data ranges from 0-1 month. This is more conservative. Catch data that does not have biomonitoring data from 2 months prior or less is dropped (~15% of catch observations)
- timekey3: uses only cases where the year-month align perfectly. This is the most conservative and involves dropping roughly half of catch observations.

ISSUE: Catch appears in the data to be more diverse than CFRs

Using combined biomonitoring and catch data, there is a clear pattern where more species are caught by fishers in the CFR than are present in the CFR biomonitoring. Troubleshooting steps taken:

- If I merge the files using timekey1 (least conservative), up to 67% of all cases (CFR-time combinations) have more fish species caught than present in CFR biomonitoring.
- If I merge with the more conservative timekey3, this reduces to 50%.
- I also spot-checked several cases (CFR-time) that had very low numbers of species in the raw data to confirm that no mistake was made in the data construction that deflated the CFR biomonitoring species numbers.

Possible explanations:

1. The universe in which fishers fish is larger than the CFR zone of influence.
2. Fishers are more skilled at catching different fish species than the people doing the CFR biomonitoring.
3. Some fish species likely spend minimal or no time in the CFR due to habitat preferences/needs, so would be unlikely to appear in biomonitoring but potentially likely to be caught elsewhere.

Possible solutions:

- Use the data in Q5 of the catch survey (shown below) to understand how much fishers are fishing within the CFR zone of influence. Not all categories in Q5 are clearly in or out of the CFR ZOI, so some assumptions would have to be made. Further assumptions would have to be made to connect to the Q7 group level data. I haven't had time to think about this carefully, but at first glance it seems like it might be tricky.

Q5: Estimate the catch of the Fish and OAA from different habitats

[illegible]