Data Extraction Notes

General overview

1. The code imports data from the Ecosystem Monitoring and Assessment (EMA) program at Auke Bay lab. The EMA data are from the northern Bering Sea surveys and BASIS surveys not housed in the EcoDAAT database. These data were not sorted in Poland and were sorted under a variety of different persons and protocols. I spend time in the code converting these rather messy data to match the EcoDAAT structure. There is double counting of taxa of the same stage from the same station in these data, which I account for in the code below.
2. EcoDAAT data have been collected by EcoFOCI since the 1990s, but have significant spatial and temporal variability. Keep this in mind. We also had protocol changes in how the samples were processed over time, these are represented by different SPECIMEN\_FORM codes in the data set.
   1. Up to 2011, we only identified specific taxa
   2. In 2012, Jeff Napp changed the protocol to identify all taxa in the samples, but made some decisions about which taxa come from which nets that were different from what came prior, so particular taxa must be selected from particular nets to build a consistent data set over time
   3. Dave Kimmel introduced a new protocol in 2019 to correct the data moving forward.
   4. Gear description with net type, diameter, and mesh size are found in the Gear folder.

Overview of chunks in the extraction code **BPSynthesis-Zoop-Data.Rmd**

* This Rmd file compiles all of the raw zooplankton data from various sources within NOAA, filters for the NBS and Chukchi, filters for the correct gears, provides an alternate coarse taxa list for ease of analysis, and selects the taxa from the correct nets and mesh sizes

1. Load packages
2. Import EMA data
   1. Import EMA data (BASIS and north Bering Sea cruises not housed in EcoDAAT database. The raw data are in the Raw-Data folder.
   2. These data were collected by the EMA program at ABL and all methods questions should be directed to EMA staff.
3. Convert EMA fields into EcoDAAT fields
   1. This code takes the EMA data and converts the information to match the format of the EcoDAAT data
4. Import EcoDAAT data
   1. This code connects to the EcoDAAT Oracle database and imports all available zooplankton data
   2. One needs a username and password to connect to the EcoDAAT database, so this will not work for those without access.
5. Do some data tidying for the EcoDAAT data set and combine the EcoDAAT and EMA datasets.
   1. Correct errors in the EcoDAAT mesh codes
   2. Check to see if duplicate cruises are in the EMA database and the EcoDAAT database and remove them.
6. Quick map of the raw data
   1. This quick map shows all of the data coverage prior to filtering
7. Data filtering for BP Synthesis Project
   1. Spatial Filter
      1. Filter for all data north of 60°N
         1. Take a look at a map of the filtered data
      2. Filter for all data west of 150°W
         1. Take a look at a map of the filtered data
   2. Temporal Filter
      1. Remove stand alone year of 1996
   3. Gear filter
      1. Remove the following gears: CALVET (53 um mesh), V60BON, and 80BON for low sample sizes. Remove SLED as is biased toward bottom sample
   4. Remove FAILED and QUESTIONABLE Hauls
   5. Write the raw file for the BP Synthesis project: **NBS\_Zoop\_Raw.csv**
8. Bring in the Coarse Taxa List to aid in lumping and filter out some taxa
   1. This brings in a reference file **TaxaList\_Coarse.csv** that adds a column called TAXA\_COARSE and a NOTE column. It serves to lump rarer taxa into combined groups and removes rare taxa that are not required for analysis. The reference files are in the folder Taxa-Lists. This folder also contains a file **Taxa\_List\_Coarse\_Summary.csv** that shows which nets each taxa will be selected from in the subsequent code.
9. The next chunks cycle through each Coarse Taxa and selects taxa and stages from nets listed in the **Taxa\_List\_Coarse\_Summary.csv**. This is necessary to prevent double counting of taxa in the EMA dataset and ensure consistency across protocol changes.
10. Biomass conversions
    1. Bring in the annotated biomass data set **Biomass-Annotated.xlsx**
       1. This has all of the annotated individual wet, dry, and carbon masses as well as growth rate estimates from measurements and literature
       2. Wet weights are estimated from Seward Line data of Russ Hopcroft
       3. Dry weights are converted from wet weights, or come from literature
       4. Carbon masses are converted from dry weights
    2. Do some tidying
    3. BIOMASS\_WW\_MG\_M3
       1. Multiply individual wet weight (mg) by abundance (individual m-3) to yield biomass (mg wet weight m-3) as both mean and SD
    4. BIOMASS DW\_MG\_M3
       1. Multiply individual dry weight (mg) by abundance (individual m-3) to yield biomass (mg wet weight m-3) as both mean and SD
    5. BIOMASS C\_MG\_M3
       1. Multiply individual carbon weight (mg) by abundance (individual m-3) to yield biomass (mg wet weight m-3) as both mean and SD
    6. PRODUCTION\_MG\_C\_D
       1. Multiply abundance (individual m-3) by individual carbon weight (mg) and growth rate (h-1)
       2. Multiply by 24 to convert hours to days
11. Write final data set
    1. **NBS\_Zoop\_Process\_Final.csv,** in Processed Data folder.