**readme for fish data submitted to SeaBASS from EXPORTS and WHOI OTZ projects**

written by Helena McMonagle, 2024-03-21

EXPORTS\_EXPORTSNA\_JC214\_MOCNESS\_fishdata.csv and OTZ\_WHOI\_SG2105\_MOCNESS\_fishdata.csv contain the fish from the R/V Sarmiento's MOCNESS-10 and the R/V James Cook's MOCNESS-1 collected during the North Atlantic NASA EXPORTS WHOI Ocean Twilight Zone Project cruise in May 2021.

A total of 1,553 fish were collected, but we were able to calculate a weight (or were able to use an empirical weight measurement) for 1,428 individuals. The current dataset also excludes fish smaller than 10 mm, which are generally not adult fish and have lesser known migratory behaviors and life histories compared to adults, and thus are not included in the fish carbon flux analysis that this dataset will be used for.

Fish collections at sea were done by Joel Llopiz, Kayla Gardner, Julia Cox, Cristina Garcia-Fernandez and Helena McMonagle on the Sarmiento and led by Deborah Steinberg and Amy Maas on the Cook. Data collection in the lab and with ImageJ was done primarily by Lyndsey Llevebre, Julia Cox and Helena McMonagle.

For more methods information, please refer to the manuscript titled “The contribution of mesopelagic fishes to the biological carbon pump in the Northeast Atlantic Ocean” (McMonagle et al., to be submitted in April 2024)

**Column descriptions are as follows:**

date: the date that each fish was collected from the study site

time: Local time during collection (same as UTC in this case)

day\_night: time of day during collection

lat: Same as lat\_start (needed for SeaBASS column header convention)

lat\_start: Latitude at the start of the tow

lat\_end: Latitude at the end of the tow

lon: Same as lon\_start (needed for SeaBASS column header convention)

lon\_start: Longitude at the start of the tow

lon\_end: Longitude at the end of the tow

depth: Same as depth\_start (needed for SeaBASS column header convention)

depth\_start: depth that net started collection (collections were done upon net recovery, so this is the deepest depth for that net)

depth\_end: depth that net ended collection (collections were done upon net recovery, so this is the shallowest depth for that net)

tow: the number for the entire haul, including all nets on the MOCNESS

net: the depth-stratified net associated with each tow. Note that on the MOCNESS-1 datasheets, net 1 was defined as that which remained open from the surface to 1000 m, whereas on the MOCNESS-10 datasheets, this net was called net zero. Catch from net 1 on the Cook was not saved, so we were not able to compare fish counts between the two ships for this net. This net was not as useful in any case because it is not depth-stratified.

volfilt: volume filtered by each net as determined by a flowmeter connected to the shipboard computer during MOCNESS deployment, in liters

wt: water temperature in degrees C (may need to refer to CTD data for this cruise because MOCNESS sensors did not work on the Sarmiento)

sal: water salinity (though similar issue as above with regard to sensors not recording data in the MOCNESS software on the Sarmiento)

sample: whether gut contents were identified visually (“visual”) or using metabarcoding (abbreviated “mb”)

scientificNameID\_manual: whether individual fish taxa were identified visually (“visual”) or with barcoding (“genetic”)

id: initials of the person who processed these fish in Llopiz lab (JC for Julia Cox, or LSL for Lyndsey Levebvre). All other fish were morphologically IDed by Helena McMonagle from photographs (for smaller Sarmiento fish sorted from zooplankton samples, and for all Cook fish that were either sorted from zooplankton and imaged with Ecotaxa to provide silhouettes, or that were sorted from zooplankton at sea and photographed)

SN: these are the individual fish IDs. These are assigned to the larger Sarmiento fish

that were flash frozen at sea and later processed in Llopiz Lab, or the fish from the Cook used in the analysis. For the smaller of the Sarmiento fish specimens, Julia Cox (Llopiz Lab) assigned each individual fish an ID from the fish she imaged and then measured in ImageJ. These fish were preserved in 90% ethanol prior to being photographed.

scientificName\_manual: definition of taxon according to Worms database.

scientificNameID\_manual\_aphiaID: Worms database information with Aphia ID

data\_provider\_category\_manual: Lowest taxonomic level identified. For the larger Sarmiento fish specimens, a much lower taxonomic level was often recorded from laboratory morphological IDs or genetic barcoding IDs.

length\_fish: Standard length of each fish in millimeters.

weight\_fish: Empirically measureed weight taken by Lyndsey Llevebre for the larger Sarmiento specimens that were flash frozen at sea and then processed individually in Llopiz Lab.

weight\_fish\_1id: weight estimated from a length-weight regression for the lowest taxon possible, where empirical weights were not available

weight\_fish\_2id: the empirical weight measured for that individual where available, or the estimated weight from the length-weight regression where an empirical weight measurement is not available.

weight\_fish\_3id: weight of the gut when full, for the relatively small proportion of Sarmiento fish that had gut content analysis done

weight\_fish\_4id: weight of the gut when empty (after gut contents were removed for ID with either metabarcoding or visual ID with microscope)

associated\_files: This corresponds to the photos taken of the Sarmiento fish.

associated\_file\_types: type of files (all are photos)

FracBiom: This was removed from the SeaBASS data file, because fractions had already been taking into account in the final number of rows here. This is relevant for the MOCNESS-1 samples, which did not always use the same split. The smaller fish from the MOCNESS-10 samples that Julia measured with ImageJ were all from quarter splits, and were multiplied by 4 accordingly. The smaller fish from the MOCNESS-1 that Helena measured with ImageJ using Ecotaxa images (after formalin preservation) were from various different splits, indicated in this column (see Github for the original file prior to formatting edits for uploading to SeaBASS). So, if FracBiom = 4, this fish came from a quarter split and we duplicated this individual 4 times in the dataset to represent its estimated presence in the entire catch. If FracBiom = 2, we duplicated this individual twice.

Associated Github: https://github.com/hmcmonagle/Fish-carbon-flux-N-Atlantic