Body Size Data Processing

NEON collects macroinvertebrate data via fixed-area samplers (e.g., Surber samples) and measures insect body lengths to the nearest mm. While the samplers vary, all mesh sizes are the same (243 um). Fish are collected using 3-pass removal electrofishing within stream reaches that vary in area across study sites. For each collection, the first 50 fish per taxon are measured for total length in mm (Monahan et al. 2020). Thus, the fish data consist of two subsets of data, one for counts of the total number of fish per pass per taxon per collection event, and another with the length of the first 50 fish caught per taxon per reach per collection event (but not per pass). To convert both types of samples into a single format containing abundance (per square meter) of each dry mass, whether macroinvertebrate or fish, multiple data steps were required as described below.

*Macroinvertebrate*

We obtained 165,316 macroinvertebrate length measurements, ranging from 1 to 86 mm. However, as shown below, the samplers appeared to undercount insects less than ~3 mm (Figure X). Insects of 3 mm or less tend to have head widths that are less than 250 um (Stoffels et al. 2003), suggesting the possibility that insects of this size or smaller could pass through the mesh. We assumed that this might generate undercounts for insects as seen in Figure X and removed lengths less than 3mm from the analysis.

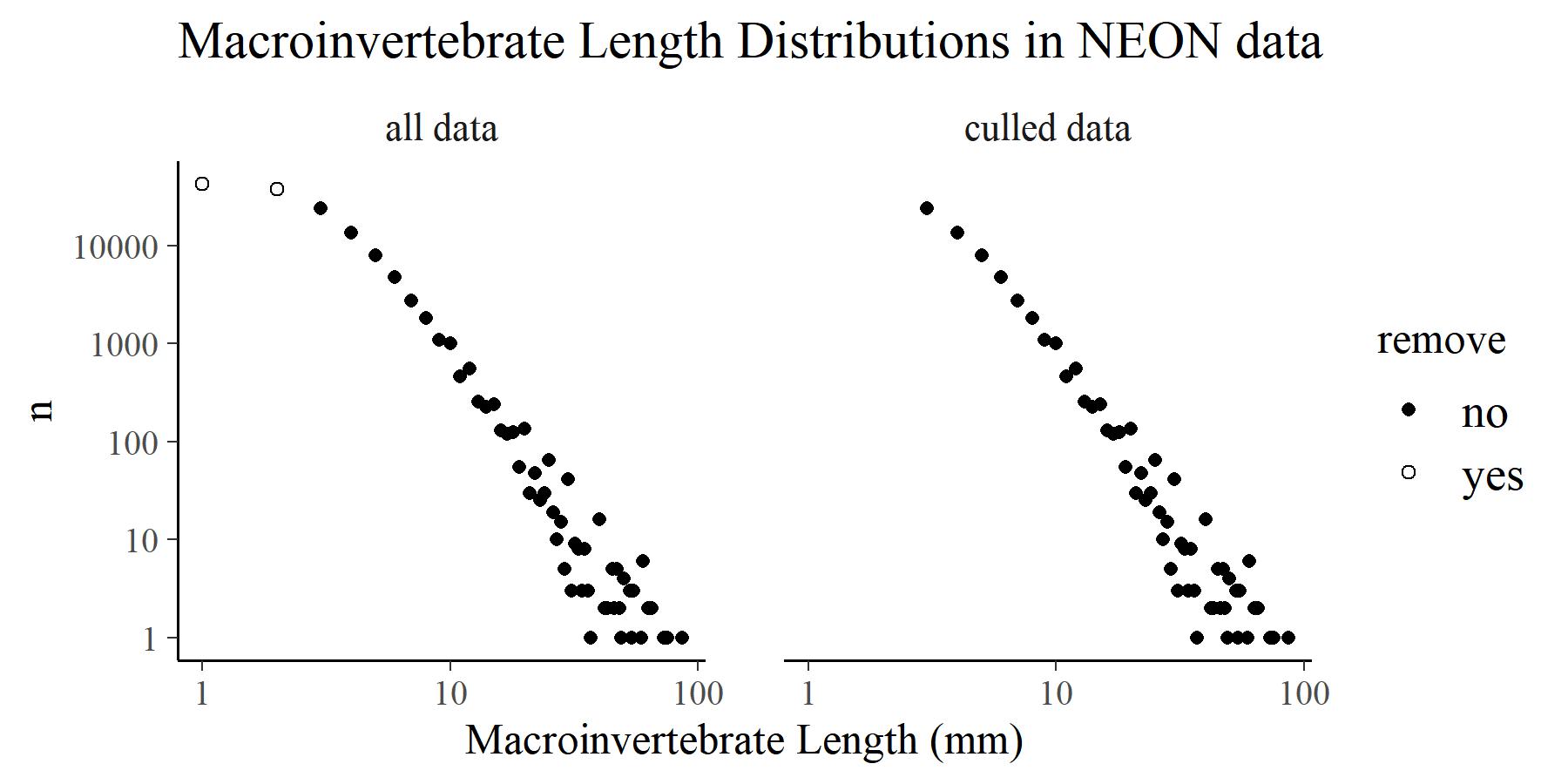


Figure X. The total number (n) of macroinvertebrates in each length class. Data are plotted for all measures in the neon macroinvertebrate dataset totaling 141,246 length measures.

*Fish*

We obtained 64,940 measurements of individual fish masses (mg wet mass) from the *fsh\_perFish* table in data product DP1.20107.001 (NEON 2022). For each collection event (i.e., collection date, reach, and site), *fsh\_perFish* contains up to 50 individual mass measures per species. Wet mass was converted to dry mass estimates (mg) by assuming dry\_mass = wet\_mass\*0.2.

NEON also collects total length data on each fish. We chose to use dry mass measures because fish lengths appeared to show bias against smaller fish (Figure 1b). The same was true in the dry mass estimates, but the impact appeared to be less drastic (Figure 1a).

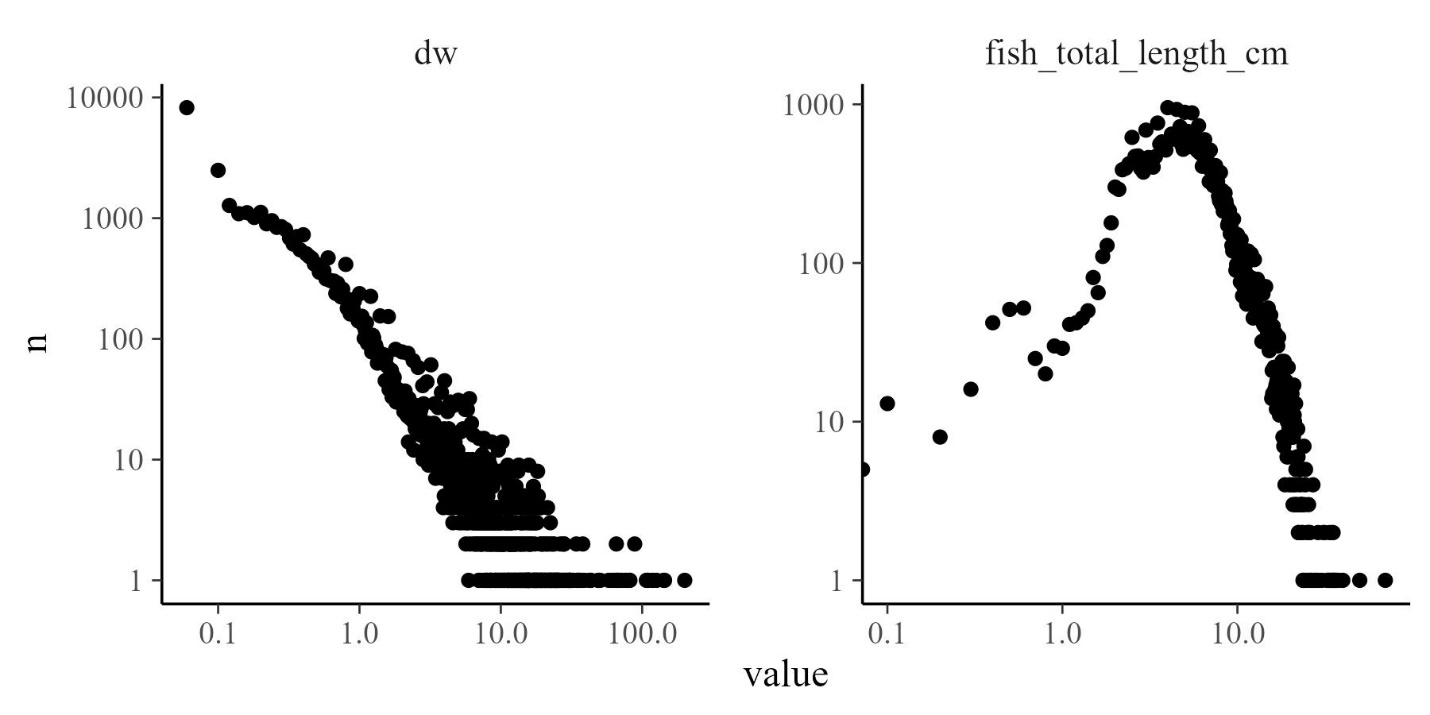


Figure X. Distributions of ~70,000 individual fish lengths. After converting lengths to diameters, we excluded fish that were less than 2 times the width of the mesh diameters to limit bias against small fish. Rugs on the x-axis show the distribution of data in each dataset.

After measuring the first 50 or so fish, NEON then bulk counts the rest and reports those counts per pass in the table *fsh\_bulkCount*. For example, if 100 Rainbow Trout were collected on the first pass, then *fsh\_bulkCount* would indicate 50 fish with the other 50 indicated by the total count in *fsh\_perFish*. Adding these totals gives the total number of fish per reach per pass per taxon. However, if a pass collects zero fish, it is not directly entered as zero in the data (unfortunately!). Instead, it is noted as a true zero in a third file: *fsh\_perPass*. We used that file to identify true zeros. Around 5% of the samples had missing data that could not be identified as a true zero. We removed those samples.

After wrangling the data, we estimated the total population size (number/m2) of fish in each collection using a multinomial Poisson depletion model (Royle et al. 2004). We specified the model in R using the *ubms* package (Kellner et al. 2021). The response variable was the number of fish caught per pass and the predictor variable was the collection id (site\_date\_reach). The model resulted in a population estimate for each collection. We then multiplied that population estimate by the relative abundance of each fish species, resulting in an estimated density (number/m2) of each fish species in each collection. Finally, we merged those estimates with the length measurements and resampled with replacement the body sizes of each species in each collection. To ensure a large enough sample size, we converted each density estimate to number/10,000 m2. For each collection, we then summed up the total number of individual body sizes, along with their density estimates.

*Combining fish and macroinvertebrates*

We combined the fish and macroinvertebrate body size datasets and then tallied the number of body sizes, along with their density. Fish and macroinvertebrates were collected on different dates, with macroinvertebrates collected three times per year and fish collected twice. Therefore, to combine fish and macroinvertebrate samples, we limited the data to only collections that occurred within 30 days of each other. For example, if macroinvertebrates were collected on June 10 and fish collected on June 20, those samples were treated as one. If more than one sample was in this window (e.g., another fish collection on June 21), we included on the most recent collection. The resulting data set is a list of 18,050 unique body sizes ranging nine orders of magnitude (0.003 to 200,000 mg) and their densities for 151 collections across 23 sites over 5 years.