1. Zooplankton surveys

The Interagency Ecological Program (IEP) is a consortium of State and federal agencies that has been conducting cooperative ecological investigations since the 1970s. The IEP runs over twenty long-term monitoring surveys on biological components of the Upper San Francisco Estuary. Surveys monitor phytoplankton, zooplankton, benthic invertebrates, water quality, and many types of fish. Several fish surveys sample zooplankton concurrently, and information on zooplankton species composition and abundance can be coupled with fish diet studies. The IEP long-term surveys that monitor zooplankton are the Environmental Monitoring Program (EMP; also known as the IEP zooplankton study), 20-mm Survey (20mm), Fall Midwater Trawl (FMWT), Summer Townet Survey (STN), the Yolo Bypass Fisheries Monitoring Survey (not included in this integrated dataset), and the Fish Restoration Program (FRP). An overview of these programs is provided in study\_metadata.csv. Locations of fixed sampling locations for surveys with fixed sampling designs are provided in stations.csv, and coordinates for every sample are provided in environment.csv. Zooplankton surveys sample 3 different size classes of zooplankton, by towing nets with different sized mesh (or in one case a pump that pumps water into a microzooplankton net). Typically, the zooplankton targeted by these different sampling instruments are:

1. Macro (500-505 µm net) for amphipods and mysids
2. Meso (150 - 160 µm net) for copepods and cladocerans
3. Micro (43-50 µm net) for copepods and rotifers.

Every IEP survey that collects zooplankton samples with a mesozooplankton net, which targets adult copepods and cladocerans, because these taxa are believed to comprise the majority of zooplankton in juvenile and adult planktivorous fish diets. Some surveys also sample with micro- or macro-zooplankton nets.

1.1. Environmental Monitoring Program

The Environmental Monitoring Program (EMP) Zooplankton Study (also known as the IEP Zooplankton Study) began in 1972 in order to assess trends in fish food resources ranging from San Pablo Bay to the east Delta, as well as to detect and assess the impacts of recently introduced zooplankton species on native species. The study is mandated by Water Right Decision 1641 for operation of the State Water Project and Central Valley Project (SWRCB 2000). The EMP study is conducted by the California Department of Fish and Wildlife (CDFW), California Department of Water Resources (DWR), and the US Bureau of Reclamation (Reclamation) and currently samples 17 fixed stations and 2 floating entrapment zone (EZ) stations (locations where the bottom salinity is 2 and 6 PPT). There are also 3 additional stations located in Carquinez Strait and San Pablo Bay, which are sampled during periods of high outflow and low salinity. Historically (prior to 1995) the survey sampled at a much larger number of stations.

Zooplankton for EMP are sampled in 3 different size ranges: microzooplankton are sampled using a pump, mesozooplankton are sampled using a modified Clarke-Bumpus net, and macrozooplankton are sampled using a mysid net. All EMP zooplankton are collected monthly at fixed stations year-round in open channels at high slack tide and preserved in 10% formalin dyed with rose bengal. Macrozooplankton and Mesozooplankton are collected using 10-minute oblique tows with a 124 cm long net with 505 µm mesh, and a 73 cm long net with 160 µm mesh, respectively. Prior to 1974, macrozooplankton were sampled with a 930 µm mesh net. Microzooplankton are collected with a Teel marine pump while the intake hose is raised through the water column and pumped into a net with 43 µm mesh. Pump samples collected approximately 1.5 - 1.9 L from 1972 – 2007, and 75 L from 2008 - present, measured by a digital flowmeter connected to the hose.

Microzooplankton samples are passed through a 154 µm mesh sieve nested on top of a 43 µm mesh sieve in the lab, and only the smaller size fraction that passes through the larger sieve and is retained on the smaller sieve is counted. Lengths are recorded for macrozooplankton (but not included in this dataset), and biomass is estimated by length-weight equations for macrozooplankton and by average values for mesozooplankton and microzooplankton (biomass\_mesomicro.csv). Recorded environmental variables for all samples include time, depth, surface and bottom conductivity, surface temperature, Secchi depth, and chlorophyll-a.

More information on EMP and its methods can be found on the EMP zooplankton study website (https://wildlife.ca.gov/Conservation/Delta/Zooplankton-Study) or zooplankton data publication (Barros 2022), or environmental data publication (Battey and Perry 2022).

1.2. 20mm Survey

The 20-mm survey was initiated in 1995 by the California Department of Fish and Wildlife to monitor postlarval-juvenile Delta Smelt (Hypomesus transpacificus) distribution, abundance, and timing throughout their historical spring range in the Delta. The survey is mandated under the Endangered Species Act Biological Opinion for operation of the State and Central Valley water projects (USFWS 2019). 20-mm refers to the length of the fish targeted by the net. Zooplankton samples are collected concurrently with fish samples to monitor Delta Smelt food supply. Between 41 and 55 stations have been sampled each year since the survey began.

Zooplankton are sampled twice per month between March and July at fixed stations in open channels. Mesozooplankton are sampled using 10-minute stepped-oblique tows with a 73 cm long 160 µm mesh modified Clarke-Bumpus net. The net is attached to the top of the 20-mm net frame and a flowmeter is mounted in the mouth. Samples are preserved in 10% formalin. Lengths are not recorded, and biomass is estimated by literature values. Recorded environmental variables include times, tidal stage, depth, surface and bottom conductivity, surface temperature, Secchi depth, and turbidity.

More information on the 20-mm survey and its methods can be found on the 20-mm survey website (California Department of Fish and Wildlife 2021).

1.3. Fall Midwater Trawl

The Fall Midwater Trawl (FMWT) was initiated by the California Department of Fish and Wildlife in 1967 in order to determine the relative abundance and distribution of age-0 Striped Bass (Morone saxatilis), but the data has also been used for other upper estuary pelagic fish species, including Delta Smelt (Hypomesus transpacificus), Longfin Smelt (Spirinchus thaleichthys), American Shad (Alosa sapidissima), Splittail (Pogonichthys macrolepidotus), and Threadfin Shad (Dorosoma petenense). The FMWT is currently mandated by the 2019 Delta Smelt Biological Opinion for the coordinated operation of the Central Valley Project and State Water Project (USFWS 2019). The FMWT samples 122 stations each month from September to December ranging from San Pablo Bay to Stockton, Hood, and The Sacramento Deep Water Ship Channel. FMWT samples both macrozooplankton and mesozooplankton at a subset of these stations since 2011, with some pilot studies in earlier years.

Zooplankton samples are collected along with the fish trawl at fixed stations in open channels using 10-minute oblique tows. Macrozooplankton are sampled using a 124 cm long net with 505 µm mesh, while mesozooplankton is sampled using a 73 cm long modified Clark-Bumpus net with 160 µm mesh. For both zooplankton sizes, samples are preserved in 10% formalin dyed with rose bengal. Lengths are recorded for macrozooplankton but not mesozooplankton, biomass is estimated for both as in EMP. Recorded environmental variables include time, tidal stage, depth, surface and bottom conductivity, surface and bottom temperature, Secchi depth, Microcystis presence, and turbidity.

More information on FMWT and its methods can be found on the FMWT data publication (Burdi et al. 2022).

1.4. Summer Townet Survey

The Summer Townet Survey (STN) was initiated by the California Department of Fish and Wildlife in 1959 in order to determine the relative abundance and distribution of upper estuary pelagic species, namely age-0 Striped Bass (Morone saxatilis). As with the FMWT, the STN is currently mandated by the 2019 Delta Smelt Biological Opinion (USFWS 2019) and began in response to the development of the Central Valley Project pumping plants. The Summer Townet Survey collects mesozooplankton samples from 32 historic stations and 8 supplemental stations ranging from San Pablo Bay to Rio Vista, Stockton, Cache Slough, and the Deep-Water Ship Channel. Zooplankton monitoring began in 2005 with samples collected every 2 weeks between June and August.

STN samples only mesozooplankton during their fish trawl with a net attached to the townet frame. Zooplankton samples are collected during 1 of the fish tows at each fixed station in open channels using 10-minute oblique tows. Mesozooplankton are sampled using a 73 cm long modified Clark-Bumpus net with 160 µm mesh and preserved in 10% formalin dyed with rose bengal. Biomass is estimated and recorded environmental variables include time, tidal stage, depth, surface and bottom conductivity, surface and bottom temperature, Secchi depth, Microcystis presence, and turbidity.

More information on STN and its methods can be found on the STN data publication (Burdi et al. 2022).

1.5. Fish Restoration Program

The Fish Restoration Program (FRP) is devoted to restoring 8,000 acres of tidal habitat in the Delta and Suisun Marsh to provide Delta Smelt habitat and 800 acres of low salinity habitat to benefit Longfin Smelt. These restoration projects are pursuant to requirements in the 2019 Biological Opinions for state and federal water project operations (USFWS 2019). The FRP Monitoring Team monitors fish and their food resources (including zooplankton) within these restored wetlands in order to better understand the benefits of the restored habitats to native fish species. The FRP Monitoring Team surveys zooplankton in shallow waters, generally near tidal marshes or sites that will soon be converted to tidal marsh. The FRP has worked closely with some other IEP surveys to compare zooplankton communities in shallow water to the open-water channel samples collected by the long-term surveys (Contreras et al. 2018).

Zooplankton are sampled annually to monthly between March and December beginning in 2015. Samples are taken from haphazardly selected locations within fixed sites at restored and existing wetlands and adjacent open-water areas across the Delta and Suisun Marsh. Macrozooplankton are collected with 10-minute horizontal surface tows using a 0.4 m x 0.4 m mouth net (500 µm mesh size). Mesozooplankton are collected with 5-minute surface tows using a 14.6 cm diameter net (150 µm mesh size). A flowmeter is attached to the net for both zooplankton size collections. Samples are preserved in 70% ethanol with rose bengal. Lengths are recorded for macrozooplankton but not mesozooplankton, biomass is estimated by literature values for both. Recorded environmental variables include time, tidal stage, surface conductivity, surface temperature, Secchi depth, turbidity, Microcystis, pH, chlorophyll, and dissolved oxygen.

More information on FRP and its methods can be found on the FRP data publication (California Department of Fish and Wildlife et al. 2019).

1.6. Yolo Bypass Fish Monitoring Program

The Yolo Bypass Fish Monitoring Program (YBFMP) was initiated by the California Department of Water Resources in 1998 (zooplankton in 1999) to collect baseline data on lower trophic levels (phytoplankton, zooplankton, and insect drift), juvenile and adult fish, hydrology, and water quality parameters. The baseline data are intended to be used to evaluate the success of restoration projects and more broadly understand the ecological functions of the Yolo Bypass. Zooplankton samples are collected from 2 sites, one in the Yolo Bypass (he base of the Toe Drain) and one in the Sacramento River (Sherwood Harbor). In 1999 and 2000, samples were collected for a couple of months each year, and then from 2001-2010 samples were collected in roughly winter/spring. Since 2011, samples are collected twice monthly during most of the year, and weekly when the bypass is inundated. Samples are collected during ebb tides. In the Yolo Bypass, nets are deployed from the rotary screw trap that is anchored in the middle of the Toe Drain Channel. In the Sacramento River, nets are deployed from a dock when flows are higher than about 2 fps, or from a boat moving 2-3 mph when flows are lower.

YBFMP samples with micro (50 μm) and meso (150 μm) zooplankton nets. Both nets have a 0.50 m diameter mouth and are 2 m long with a polyethylene cod-end jar. The micro net is held just below the surface for 2 minutes and the meso net is held just below the surface for 5 minutes, although tow times may be shorter with debris interference. Samples are washed in deionized water then preserved in 10% formalin with Rose Bengal. Before counting, samples are rinsed through sieves (106 μm for the meso sample and 45 μm for the micro sample) to remove large debris, then samples are transferred to 8% Lugol’s Iodine solution.

More information on YBFMP and its methods can be found in the YBFMP data publication (Interagency Ecological Program et al. 2021).

1.7. Directed Outflow Project Lower Trophic Study

The Directed Outflow Project was initiated by the United States Bureau of Reclamation in 2017 to evaluate the effectiveness of fall outflow actions on Delta Smelt. Field sampling is paired with the United States Fish and Wildlife Service Enhanced Delta Smelt Monitoring. Three random sites are chosen per region (Suisun Bay, Suisun Marsh, the Lower Sacramento River, the Cache Slough Complex, and the Sacramento Deep Water Ship Channel) and time period using a Generalized Random Tessellation Stratified design. Samples were collected once every two weeks in 2017 and weekly thereafter. Sampling is conducted in the fall and, starting in 2019, spring and summer seasons have also been sampled. In 2017, stations were sampled in 3 additional regions: West of the Benicia Bridge, Lower San Joaquin, and Upper Sac River. At each station, sample collection is attempted at both shoal (less than or equal to 10 feet deep) and channel (greater than 10 feet deep) habitat. Channels are sampled at the surface (0.5 m deep) and, if deeper than 20 feet, also at the bottom 1/2 to 1/3 of the water column. In 2017, oblique tows were also conducted to compare with the horizontal tows.

DOP samples with meso (150 μm) and macro (500 μm) zooplankton nets that are each 2 m in length. In 2017, both nets were 50 cm in diameter but starting in 2018 a smaller 20 cm diameter net was used for the mesozooplankton samples. Samples are stored in 10% formalin.

2. Descriptions of common methods

2.1. Mesh sizes

Nets/sieves typically sample zooplankton species whose smallest dimension is larger than the mesh size, but may also capture some organisms smaller than the mesh size (which are under-sampled since some of these smaller plankters are washed through the mesh). Furthermore, organisms significantly larger than the net mesh may be able to avoid the net and thereby evade capture.

Since the meso- and micro-zooplankton data overlap in sampled taxa, we investigated sampling biases of these 2 mesh sizes by comparing taxa counted in both. We used EMP data, filtered to include only stations and dates when both meso- and micro-zooplankton samples were collected.

For each taxon (or life stage) represented in both datasets, we compared the total summed catch per unit effort (individuals per m3 of water sampled; CPUE) from the mesozooplankton net (153 µm mesh, net) and the microzooplankton (43 µm mesh, pump) to assess where each method may be under-sampling. In almost all cases, the two methods had drastically different total CPUEs, with the microzooplankton (pump) sample collecting substantially more individuals (19 out of 23 taxa). The mesozooplankton (net) sample was only better at capturing Cirripedia larvae, Cyclopoida adults, and Oithona similis adults. The two catches were very similar for Cyclopoida juveniles (mesozooplankton/net captured 80% of the catch of microzooplankton/pump). Using this information, we developed a list of taxa and life stages under sampled by each method (excluding only Cyclopoida juveniles since the catch was so close), included as undersampled.csv. These under sampled plankton are retained in the integrated zooplankton dataset but can be flagged and removed using undersampled.csv.

It is important to note that, prior to counting in the lab, the EMP microzooplankton (pump) samples are passed through a 154 µm sieve in lab and the larger size fraction is counted separately (under the assumption that those individuals are better sampled by the mesozooplankton/net sample). Thus, some of the under sampling of larger taxa by the microzooplankton (pump) samples may be an artifact of this lab methodology rather than an effect of the net mesh size. Therefore, these results may not apply to other zooplankton studies.

2.2. Tow duration and tow type

2.2.1. Stepped-Oblique net tow

The most commonly used sampling technique employed by IEP’s long-term monitoring surveys is the stepped-oblique net tow. In this method, the zooplankton net is attached to a metal sled. This sled may be solely used for meso and macro-zooplankton (as in EMP and FMWT), or it may be attached to a larger fish sampling net (as in 20mm and STN). The sled is deployed off the stern or side of a boat using a winch or a-frame with a cable attached to a winch. The cable is spooled out to a standardized length based on the depth of the water. The boat proceeds at slow speeds while a specified amount of cable is slowly drawn in at specified time intervals following a tow schedule. As the cable is drawn in, the sled rises through the water in a stepwise fashion, sampling each strata of the water column.

2.2.2. Horizontal net tow

In a horizontal net tow, the net is held at a constant depth while the boat proceeds forward at slow speeds. FRP, YBFMP, and DOP uses horizontal tows in which the net is held just below the surface of the water. DOP also conducts horizontal tows at the bottom of the water column (in sites deeper than 20 feet).

2.2.3. Stationary sampling

In some stations, FRP samples by holding the zooplankton net in a constant position and allowing the current to flow through the net for a pre-defined period of time (IEP Tidal Wetlands Monitoring Project Work Team 2017). YBFMP also uses this approach within the Yolo Bypass and in the Sacramento River when flows are high enough. This works best when sampling from shore or a stable structure, to attach the net to, and is most commonly used on ebb tides to sample water flowing out of a wetland.

2.2.4. Pump

Pumps are used by EMP for sampling smaller zooplankton (rotifers, copepod nauplii, etc.; Hennessy 2019). Pumps are advantageous for microzooplankton because the filtered volume is easier to measure, and net clogging is easier to monitor. However, larger organisms can escape the narrow mouth of a pump intake (Harris et al. 2000).

2.3. Measurement of environmental variables

2.3.1. Salinity and temperature

Specific conductivity and temperature are measured by all surveys using YSI probes. Surface measurements are taken in the upper 1 m of the water, while EMP, FMWT, STN, 20mm, and DOP also collect conductivity measurements at the bottom of the water column. For this dataset, we have converted conductivity to salinity using the ec2pss function from the wql package (Jassby et al. 2017) for the R statistical programming language (R Core Team 2023). This function converts electrical conductivity to salinity using the Practical Salinity Scale 1978 for salinities between 2 and 42 (Fofonoff and Millard Jr 1983) and the extension of the Practical Salinity Scale (Hill et al. 1986) for salinities below 2.

2.3.2. Turbidity and Secchi depth

All surveys measure Secchi depth, the depth at which a black and white disk is no longer visible from the surface. Secchi depth is recorded from the shaded side of the boat by an observer not wearing sunglasses. Secchi depth is inversely related to turbidity, which is measured by some surveys (20mm, FMWT, STN, FRP, YBFMP, DOP) using YSI or Hach turbidity meters starting in more recent years.

2.4. Target organisms identified

Depending on the goals of the study, some surveys will enumerate different organisms than others, and identify them to a different level of taxonomic resolution. For example, FMWT macrozooplankton samples are only processed for mysids and amphipods. Other invertebrates (insects, isopods, etc.), are not counted. FRP macrozooplankton samples are processed for all macrozooplankton and micronekton, however insects are only identified to the family level, whereas mysids are identified to species. Comparing these 2 data sets requires understanding and accounting for these differences to avoid erroneously believing that FMWT samples had lower diversity than FRP samples.

2.5. Subsampling methods in the lab

Due to the patchy distribution of zooplankton in the water column, most surveys collect relatively large samples and process a randomly selected subsample of the original sample. The accuracy of an abundance estimate based on a sample is directly related to the number of organisms counted, assuming they are randomly distributed with a Poisson distribution (Harris et al. 2000). Therefore, the size of the original sample and proportion of the sample enumerated will determine accuracy of any derived abundance estimates. If one program collects significantly larger samples or enumerates a higher number of individuals in its subsample, comparing abundance estimates between the two surveys could be confounded by their differing accuracies. In addition, differences in subsampling method can impact precision of an estimate (Guelpen et al. 1982). For these surveys, subsampling is conducted with 1-ml pipetted aliquots for micro- and mesozooplankton, and divider trays for macrozooplankton.

2.5.1. Aliquots

Mesozooplankton samples are typically sampled with a micropipette (for specifics, see (Fujimura et al. 2017; Hennessy 2019). The sample is first diluted to achieve a zooplankton concentration of between 200 and 400 organisms ml-1. The sample is then mixed in a beaker and the taxonomist withdraws a 1-ml subsample with a micropipette and places the subsample on a gridded Sedgewick-Rafter glass slide, or a Ward counting wheel. The organisms are then identified under a microscope. Subsamples are processed until a target is reached, but these targets have changed over time. For EMP, FMWT, STN, and 20mm, the target was 200 total organisms from 1972 to 2003, 6% of the total dilution volume from 2004 to 2005, and from 2006 to present organisms were counted until 6% of the dilution volume had been processed and at least 5 and no more than 20 1-ml subsamples were processed (Fujimura et al. 2017; Hennessy 2019). Under current methods and at the target concentration (200-400 organisms per ml), this results in at least 1,000 organisms and a maximum of 8,000 organisms counted per sample. When samples contain debris or detritus, dilution volume is often increased to enable staff to see all the organisms on a slide clearly, which results in lower total organism counts. FRP processes a minimum of five 1-ml subsamples until 400 organisms are counted, or 20 ml total, depending on which occurs first. DOP processes 6 percent of the diluted sample volume with a minimum of five 1 mL subsamples and a maximum of 20 (2017) or 10 (after 2017) subsamples are processed. Between 1,000 organisms to 5,000 organisms are identified and counted per sample. YBFMP counts 300 organisms per sample in microzooplankton samples and 200-250 organisms per sample in mesozooplankton samples.

2.5.2. Divider trays

In the divider-tray method, the macrozooplankton sample is uniformly spread across a plastic tray and a 4-quadrant divider is then dropped on top of the tray. Technicians then enumerate only the invertebrates in the lower right-hand corner (or random position) of the tray. For very heavy samples, this procedure may be repeated so that a 1/16th or a 1/64th fraction of the original sample is enumerated (for specifics, see IEP Tidal Wetlands Monitoring Project Work Team 2017; Hennessy 2019). This technique is simple to conduct; however, it relies on the sample being randomly distributed in the tray. Organisms and detritus may also be stuck under the dividers when they are placed in the tray. From 1972 to 1984, these surveys targeted a minimum count of 220 total organisms before subsampling is completed. From 1984 to present, 400 total organisms were targeted (IEP Tidal Wetlands Monitoring Project Work Team 2017; Hennessy 2019), which gives a precision of +/- 10 % (Harris et al. 2000). For DOP, a minimum of 400 non-gravid mysids, 100 gravid mysids, and 250 amphipods are targeted for total counts per species.

2.6. Calculations

2.6.1. Count per unit effort (CPUE)

CPUE calculations are based on the volume of water sampled. Most IEP surveys estimate volume using a flowmeter in the center of the net mouth (model 2030R, General Oceanics, Inc, Miami Florida). The volume of water sampled is calculated from flowmeter counts, a meter constant, and the net mouth area. For EMP microzooplankton samples, the volume of the water pumped into the net is measured directly using a GPI inline digital flowmeter (Great Plaines Industries, Inc, Sparta, NJ) near the output end of the hose where water enters the net for filtration.

2.6.2. Biomass

Meso- and microzooplankton biomass is most frequently calculated based on average weights derived from literature values. These calculations apply a single value for mg C per individual for all individuals of a given life stage (Culver et al. 1985; Kimmerer et al. 2011). There are no existing biomass values for many species, so related species must be used.

For mysids collected by EMP and FMWT, the first 100 individuals are also measured to the nearest mm. Biomass is than calculated based on length-weight regressions. Length-weight regressions provide a somewhat more accurate estimate of total biomass, however the extent to which a given individual fits the regression will vary based on sex, reproductive state, health, and time of year. Mysid length data and conversion equations are not currently included in the integrated dataset but should be included in future revisions.

We have compiled updated micro- and mesozooplankton biomass conversions from the literature into biomass\_mesomicro.csv. All species and taxonomic groups are not covered, reflecting gaps in the literature, but these conversion values provide a starting point for researchers interested in estimating zooplankton biomass.

3. Data integration methods

Data integration was completed in R version 4.2.2 (R Core Team 2023). All code is available in the R package zooper version v2.5.0 (Bashevkin et al. 2023) and in the R script “Data\_processing.R.” First, we created lookup tables to assist with the data integration. The locations of fixed sampling locations were compiled into stations.csv. Taxonomic classifications were compiled into taxonomy.csv, while the taxonomic resolution of each source dataset and the dates this resolution changed or species were introduced were compiled into taxa\_lists.csv.

Datasets were downloaded from their sources online and reformatted for consistency by converting species codes to scientific names, renaming column names, converting units, and pivoting all datasets to the “long” format (where each row contains just one CPUE value for each taxon and sample). In some datasets, CPUE was reported as 0 in years before the taxa was counted at that taxonomic level. These values should have been “NA” because the abundances of those taxa were unknown before they were counted. To resolve this issue, in years when taxa were not counted, we replaced those incorrect 0s with “NA” values. However, non-native species were left with 0 CPUE before their known introduction year.

The consistent datasets were then bound together by column name. All environmental parameters were not measured by all datasets and those gaps are represented in the combined datasets with “NA” values. To reduce data duplication and file size, the combined dataset was then split into sample-level data (environment.csv; sampling location, date, environmental parameters, etc.) and zooplankton catch data (zooplankton.csv), each retaining the column “SampleID” as a key to rejoin them. The taxonomic resolution of each source dataset is unaltered and thus variable across surveys within the integrated dataset. Information on the taxonomic resolution of source datasets can be found in taxa\_lists.csv.

3.1. Resolving differences in taxonomic resolution: methods used to create zooplankton\_community.csv

The approaches described in this section were not applied to the YBFMP survey due to inconsistencies in taxonomy and life stage differentiation over time and with the other surveys. Differences in taxonomic resolution among studies could result in misleading findings from a community-level analysis of the integrated dataset. For example, EMP and FMWT lump all members of the genus Tortanus together and count them within the category Tortanus spp., while 20mm separates and counts Tortanus discaudatus, Tortanus dextrilobatus, and other Tortanus (Tortanus spp.). A naïve analysis would conclude that Tortanus discaudatus and Tortanus dextrilobatus only appear at the 20mm sampling locations while the lumped Tortanus spp. category is much more prevalent at EMP and FMWT sites. However, these results would be due solely to differences in taxonomic resolution among surveys. To resolve this issue, we developed a method to standardize taxonomic resolution to ensure data are comparable across surveys.

To start, we find all taxa that are not counted in every survey (Tortanus discaudatus and Tortanus dextrilobatus in the example above). Then, we sum these taxa up to a higher taxonomic level that is counted in all surveys (Tortanus spp.) and remove the lower taxa that have been summed (Tortanus discaudatus and Tortanus dextrilobatus) to prevent double counting. Now, all surveys have categories (Tortanus spp.) that represent the same set of taxa (all copepods in the genus Tortanus). Any taxa that are not represented at a higher taxonomic level in all surveys are removed from the dataset. These removed taxa are less-commonly counted taxa such as Annelida, Nematoda, or Insecta. This process considers each life stage of a taxa separately and is applied separately to each size class, so taxa lists for microzooplankton samples are only compared to taxa lists for other microzooplankton samples (and the same for meso- and macrozooplankton). This solution has been applied to the zooplankton\_community.csv table, which also has been merged with the environment.csv, taxonomy.csv, and undersampled.csv tables so that it is an analysis-ready dataset with all taxon- and sample-level information.

3.2. Other considerations and features of the zooper package

The approaches described in this section were not applied to the YBFMP survey due to the amount of inconsistencies in taxonomy and life stage differentiation over time. The taxonomic resolution for many surveys (all except FRP) has changed over time. In some cases, recently introduced taxa were added after their introduction, but in other cases taxa formerly identified e.g., to the genus level were subsequently identified to the species level. Analyses of community change over time must take these changes in taxonomic resolution over time into account to prevent a naïve analysis from discovering increasing diversity over time that is solely attributed to changes in methods. The zooper R package (Bashevkin et al. 2023) can correct for changes in taxonomic resolution over time by reducing the taxonomic resolution of the dataset to its lowest resolution at any point in time. However, this would exclude introduced species from analyses, so the package allows users to input a time-lag for introduced species. If surveys began counting an introduced species within a defined period of years (the time-lag) after its introduction, that species is retained in the time-corrected dataset.

In addition to the incorporation of a fix for changing taxonomic resolution over time, the zooper R package (and its associated interactive point-and-click shiny application: https://deltascience.shinyapps.io/ZoopSynth/) have a number of other options to customize your zooplankton dataset. They allow users to filter the data by date, salinity, temperature, survey, size class, or sampling location. The taxonomic resolution fixes are then applied on the filtered dataset. This ensures the fewest possible alterations to the data are made. Lastly, the R package and shiny application also have an alternative solution for resolving differences in taxonomic resolution among studies. For users interested in querying all available data on certain taxa, the package will return all data on your chosen taxa along with summed categories representing higher taxonomic levels that are comparable across surveys. Unlike the process used to create the zooplankton\_community.csv file described above, this method does not remove lower taxonomic categories that are members of summed groups, so plankton are double counted. Thus, outputs with this option selected should not be used for multivariate or community-level analyses. Users interested in using these advanced options to return a more customizable dataset are encouraged to produce their dataset with the zooper R package or shiny application, instead of using zooplankton\_community.csv.

4. References

Barros, A. E. 2022. Interagency Ecological Program Zooplankton Study. ver 11. Environmental Data Initiative. doi:10.6073/pasta/51bfce9bc26011095d8b99b3a2aee1b8

Bashevkin, S. M., R. Hartman, K. Alstad, and C. Pien. 2023. zooper: an R package to download and integrate zooplankton datasets from the Upper San Francisco Estuary. v2.5.0. Zenodo. doi:10.5281/zenodo.7641064

Battey, M., and S. Perry. 2022. Interagency Ecological Program: Discrete water quality monitoring in the Sacramento-San Joaquin Bay-Delta, collected by the Environmental Monitoring Program, 1975-2021. ver 7. Environmental Data Initiative. doi:10.6073/pasta/3b12b205c1b53290ff555abd809edfcf

Burdi, C. E., S. M. Breining-Aday, and S. B. Slater. 2022. Interagency Ecological Program: Zooplankton and water quality data in the San Francisco Estuary collected by the Summer Townet and Fall Midwater Trawl monitoring programs. Environmental Data Initiative. doi:10.6073/pasta/00db7da0265df448f167c823fb063a9a

California Department of Fish and Wildlife. 2021. CDFW 20-mm Survey meso-zooplankton data. https://filelib.wildlife.ca.gov/Public/Delta%20Smelt.

California Department of Fish and Wildlife, R. Hartman, S. Sherman, D. Contreras, and D. Ellis. 2019. Fish catch, invertebrate catch, and water quality data from the Sacramento-San Joaquin Delta collected by the Fish Restoration Monitoring Program, 2015-2017. Environmental Data Initiative. doi:10.6073/PASTA/86810E72766AD19FCCB1B9DD3955BDF8

Contreras, D., R. Hartman, and S. Sherman. 2018. Fish sampling gear comparisons in littoral and channel habitats. IEP Newsletter 32: 32–47.

Culver, D. A., M. M. Boucherle, D. J. Bean, and J. W. Fletcher. 1985. Biomass of freshwater crustacean zooplankton from length–weight regressions. Canadian Journal of Fisheries and Aquatic Sciences 42: 1380–1390. doi:10.1139/f85-173

Fofonoff, N. P., and R. C. Millard Jr. 1983. Algorithms for the computation of fundamental properties of seawater. UNESCO Technical Papers in Marine Science 44.

Fujimura, B., L. Damon, and T. Tempel. 2017. 20mm survey metadata 1995-2016. California Department of Fish and Wildlife.

Guelpen, V. L., D. F. Markle, and D. J. Duggan. 1982. An evaluation of accuracy, precision, and speed of several zooplankton subsampling techniques. Journal Du Conseil International Pour L’Exploration De La Mer 40: 226–236.

Harris, R. P., P. H. Wiebe, J. Lenz, H. R. Skjoldal, and M. Huntley, eds. 2000. ICES zooplankton methodology manual, Academic Press.

Hennessy, A. 2019. Zooplankton meta data, IEP Bay-Delta Monitoring and Analysis Section, Department of Water Resources, Sacramento, CA.

Hill, K., T. Dauphinee, and D. Woods. 1986. The extension of the Practical Salinity Scale 1978 to low salinities. IEEE Journal of Oceanic Engineering 11: 109–112.

IEP Tidal Wetlands Monitoring Project Work Team. 2017. Tidal wetland monitoring framework for the upper San Francisco Estuary: Standard Operating Procedures.

Interagency Ecological Program, C. L. Pien, J. B. Adams, and N. Kwan. 2021. Interagency Ecological

Program: Zooplankton catch and water quality data from the Sacramento River floodplain and tidal slough, collected by the Yolo Bypass Fish Monitoring Program, 1998-2018. ver 2. Environmental Data Initiative. doi:10.6073/pasta/baad532af96cba1d58d43b89c08ca081

Jassby, A. D., J. E. Cloern, and J. Stachelek. 2017. wql: Exploring Water Quality Monitoring Data.

Kimmerer, W., T. Ignoffo, and L. Sullivan. 2011. Length, weight, carbon, and nitrogen content of common copepods in the San Francisco Estuary. Romberg Tiburon Center, San Francisco State University.

R Core Team. 2023. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing.

Schultz, A. A., and A. J. Kalmbach. 2023. Directed Outflow Project Lower Trophic Study.

Environmental Data Initiative. doi:10.6073/pasta/73b4acc0e1a72cb31d95ba96a4071fbc

SWRCB. 2000. Revised Water Right Decision 1641. In the Matter of Implementation of Water Quality Objectives for the San Francisco Bay/Sacramento-San Joaquin Delta Estuary; A Petition to Change Points of Diversion of the Central Valley Project and the State Water Project in the Southern Delta, and A Petition to Change Places of Use and Purposes of Use of the Central Valley Project, State Water Resources Control Board, California Environmental Protection Agency.

USFWS. 2019. Biological Opinion for the Reinitiation of Consultation of the Coordinated Operations of the Central Valley Project and State Water Project, U.S. Fish and Wildlife Service.