

# Package: zooper (via r-universe)

November 25, 2024

**Title** Download and Integrate Zooplankton Datasets from the Upper San Francisco Estuary

**Version** 2.5.0.9000

**Date** 2019-12-13

**Description** This package downloads and integrates zooplankton datasets from the Sacramento San Joaquin Delta. Datasets are manipulated into a consistent format and bound together, then differences in taxonomic resolution among datasets are resolved using one of two methods, depending on whether the user wishes to analyze community or taxa-specific trends. Ancillary environmental data are retained in the final dataset. Users can also filter the dataset by a number of parameters.

**Depends** R (>= 3.5.0)

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**LazyDataCompression** xz

**Imports** lubridate, readxl, dplyr (>= 1.1.0), dplyr (>= 1.0.4), tidyr, readr, purrr, tibble, stringr, rlang, magrittr, data.table, tidyselect (>= 1.0.0), stats, utils, wq, rvest

**URL** <https://github.com/InteragencyEcologicalProgram/zooper>

**BugReports** <https://github.com/InteragencyEcologicalProgram/zooper/issues>

**RoxygenNote** 7.3.2

**Suggests** spelling, testthat (>= 2.1.0), covr

**Language** en-US

**Config/testthat/edition** 3

**Config/pak/sysreqs** libicu-dev libxml2-dev libssl-dev libx11-dev

**Repository** <https://sbashevkin.r-universe.dev>

**RemoteUrl** <https://github.com/InteragencyEcologicalProgram/zooper>

**RemoteRef** HEAD

**RemoteSha** df83c6182f35e2d3a7dd03c9f1ca5eb0d99c59fa

## Contents

biomass_macro . . . . .	2
biomass_mesomicro . . . . .	3
Commontaxer . . . . .	4
completeTaxaList . . . . .	5
crosswalk . . . . .	5
LCD_Taxa . . . . .	7
SourceTaxaKeyer . . . . .	8
startDates . . . . .	9
stations . . . . .	10
stationsEMPEZ . . . . .	10
Taxnamefinder . . . . .	11
Uncountedyears . . . . .	12
undersampled . . . . .	13
Zoopbiomass . . . . .	14
zoopComb . . . . .	14
Zoopdownloader . . . . .	16
zoopEnvComb . . . . .	18
zooper . . . . .	19
Zoopsynther . . . . .	21
zoop_urls . . . . .	25
<b>Index</b>	<b>27</b>

---

biomass_macro	<i>Macro zooplankton length-weight equations</i>
---------------	--

---

### Description

Length-weight equations for macro zooplankton to be used for biomass conversions. The equations relate length in mm to dry mass in milligrams. Dry mass can be converted to carbon mass by assuming 40 Uye, S. 1982. Length-weight relationships of important zooplankton from the Inland Sea of Japan. Journal of the Oceanographical Society of Japan 38:149–158.

### Usage

biomass\_macro

### Format

a tibble with 23 rows and 9 columns.

**Taxname** Current scientific name.

**Level** Taxonomic level of the taxa.

**Preservative** Preservative used to store sample before individuals were measured to develop the equations.

**Weight\_type** The type of weight measurement.

**N** The number of individuals used in developing the equation.

**Min\_length** Minimum length (mm) of individuals used in developing the equation.

**Max\_length** Maximum length (mm) of individuals used in developing the equation.

**a** Coefficient a in the equation  $\text{Weight (mg)} = a * \text{Length (mm)} ^ b$ .

**b** Coefficient b in the equation  $\text{Weight (mg)} = a * \text{Length (mm)} ^ b$ .

#### See Also

[biomass\\_mesomicro](#)

---

biomass_mesomicro	<i>Meso and Micro zooplankton average biomass values</i>
-------------------	--

---

#### Description

Average carbon biomass values for meso and micro zooplankton to be used for biomass conversions

#### Usage

biomass\_mesomicro

#### Format

a tibble with 44 rows and 4 columns.

**Taxname** Current scientific name.

**Level** Taxonomic level of the taxa.

**Lifestage** Plankton lifestage.

**Carbon\_mass\_micrograms** Average carbon mass of an individual in micrograms.

#### See Also

[biomass\\_macro](#)

---

Commontaxer

*Detect common taxonomic names across all source datasets*

---

### Description

Calculates taxa by life stage combos present in all source datasets

### Usage

```
Commontaxer(Source_taxa_key, Taxa_level, Size_class)
```

### Arguments

Source_taxa_key	A dataframe with columns named Source, Lifestage, SizeClass, and the value provided to the parameter Taxa_level. This dataframe should list all Taxa_level by Lifestage combinations present for each source dataset. You can provide it with the output of <a href="#">SourceTaxaKeyer</a> .
Taxa_level	Taxonomic level you would like to perform this calculation for. E.g., if you wish to determine all Genus x lifestage combinations present in all datasets, provide Taxa_level = "Genus". The value provided here must be the name of a column in the dataset provided to Source_taxa_key.
Size_class	The size class(es) you would like this function to consider. You should generally only supply 1 size class.

### Details

This function is designed to work on just one size class. To apply to multiple size classes, use [map](#) or [apply](#) functions to apply across size classes.

### Value

A tibble with a column for Taxa\_level and another for Lifestage representing all combinations of these values present in all source datasets.

### Author(s)

Sam Bashevkin

### See Also

[Zoopsynther](#), [crosswalk](#), [SourceTaxaKeyer](#)

**Examples**

```
## Not run:
library(rlang)
library(purrr)
SourceTaxaKey <- SourceTaxaKeyer(zoopComb, crosswalk)
Size_classes <- set_names(c("Micro", "Meso", "Macro"))
Commontax <- map(Size_classes, ~ Commontaxer(SourceTaxaKey, "Taxname", .))

## End(Not run)
```

---

completeTaxaList	<i>All taxonomic names</i>
------------------	----------------------------

---

**Description**

A complete list of all valid taxonomic names included in the full dataset. Used to limit choices for filtering by taxa.

**Usage**

```
completeTaxaList
```

**Format**

a character vector of length 454.

**See Also**

[Taxnamefinder](#), [Zoopsynther](#), [zooper](#)

---

crosswalk	<i>Taxonomic crosswalk among datasets</i>
-----------	---

---

**Description**

A crosswalk table relating the taxonomic code used by each dataset to current scientific names, life stages, and taxonomic hierarchies.

**Usage**

```
crosswalk
```

**Format**

a tibble with 404 rows and 34 variables

**EMP\_Micro** Taxonomic codes used in the Environmental Monitoring Program microzooplankton (43  $\mu\text{m}$ ) mesh dataset

**EMP\_Meso** Taxonomic codes used in the Environmental Monitoring Program mesozooplankton (160  $\mu\text{m}$ ) mesh dataset

**EMP\_Macro** Taxonomic codes used in the Environmental Monitoring Program macrozooplankton (505  $\mu\text{m}$  mesh) dataset

**STN\_Meso** Taxonomic codes used in the Townet Survey mesozooplankton (160  $\mu\text{m}$  mesh) dataset

**STN\_Macro** Taxonomic codes used in the Townet Survey macrozooplankton (505  $\mu\text{m}$  mesh) dataset

**FMWT\_Meso** Taxonomic codes used in the Fall Midwater Trawl mesozooplankton (160  $\mu\text{m}$  mesh) dataset

**FMWT\_Macro** Taxonomic codes used in the Fall Midwater Trawl macrozooplankton (505  $\mu\text{m}$  mesh) dataset

**twentyymm\_Meso** Taxonomic codes used in the 20mm Survey mesozooplankton (160  $\mu\text{m}$  mesh) dataset

**FRP\_Meso** Taxonomic codes used in the Fish Restoration Program mesozooplankton (150  $\mu\text{m}$  mesh) dataset

**FRP\_Macro** Taxonomic codes used in the Fish Restoration Program macrozooplankton (500  $\mu\text{m}$  mesh) dataset

**DOP\_Meso** Taxonomic codes used in the Directed Outflow Project mesozooplankton (150  $\mu\text{m}$  mesh) dataset

**DOP\_Macro** Taxonomic codes used in the Directed Outflow Project macrozooplankton (500  $\mu\text{m}$  mesh) dataset

**YBFMP** Taxonomic codes used in the Yolo Bypass Fish Monitoring Program zooplankton dataset

**Lifestage** Plankton lifestage

**Taxname** Current scientific name

**Level** Taxonomic level of the taxa

**Phylum** Phylum

**Class** Class

**Order** Order

**Family** Family

**Genus** Genus

**Species** Species

**Intro** Introduction year for non-native species

**EMPstart** First year the Environmental Monitoring Program starting counting this taxa

**EMPend** Last year the Environmental Monitoring Program counted this taxa

**FMWTstart** First year the Fall Midwater Trawl starting counting this taxa

**FMWTend** Last year the Fall Midwater Trawl counted this taxa

**twentymmstart** First year the 20mm Survey starting counting this taxa

**twentymmend** Last year the 20mm Survey counted this taxa

**twentymmstart2** First year the 20mm Survey restarted counting this taxa

**DOPstart** First year DOP starting counting this taxa

**DOPend** Last year DOP counted this taxa

### See Also

[Zoopdownloader](#), [Zoopsynter](#), [zooper](#)

---

LCD\_Taxa

*Apply LCD approach for "Taxa" option*

---

### Description

Sums to least common denominator taxa, one taxonomic level at a time

### Usage

```
LCD_Taxa(
  Data,
  Taxalevel,
  Groupers = c("Genus_g", "Family_g", "Order_g", "Class_g", "Phylum_g"),
  Response = "CPUE"
)
```

### Arguments

Data	Zooplankton dataset including columns named the same as the Groupers, a Taxname column, "CPUE", and no other taxonomic identifying columns.
Taxalevel	The value of Groupers on which to apply this function.
Groupers	A character vector of names of additional taxonomic levels to be removed in this step. This vector can include Taxalevel and, if so, it will be removed from the vector within the function so Taxalevel is preserved.
Response	Which response variable(s) would you like for the zooplankton data? Choices are "CPUE" (catch per unit effort) and "BPUE" (carbon biomass per unit effort ( $\mu\text{g}/\text{m}^3$ )). Defaults to Response = "CPUE".

### Details

This function is designed to work on just one Taxalevel at a time. To apply to multiple Taxalevels, use [map](#) or [apply](#) functions to apply across taxonomic levels.

**Value**

A tibble with sums calculated for each unique value in `Data$TaxaLevel`. Sums will be excluded for grouping taxa that only contain 1 unique Taxname.

**Author(s)**

Sam Bashevkin

**See Also**

[Zoosynther](#), [crosswalk](#), [zoopComb](#)

**Examples**

```
## Not run:
library(dplyr)
df <- zoopComb%>%
  mutate(dplyr::across(tidyselect::all_of(c("Genus", "Family", "Order", "Class", "Phylum")),
    list(g=~if_else(.%in%completeTaxaList, ., NA_character_)))%>%
  select(-Phylum, -Class, -Order, -Family, -Genus, -Species, -Taxlifestage)
family_sums <- LCD_Taxa(df, "Family_g")

## End(Not run)
```

---

SourceTaxaKeyer	<i>Unique taxa by lifestage combinations present in each source and size class</i>
-----------------	--

---

**Description**

Computes a dataframe with all unique taxa by lifestage combinations present in each source and size class

**Usage**

```
SourceTaxaKeyer(Data, Crosswalk)
```

**Arguments**

Data	Zooplankton dataset. Must have a column named Source with the names of the source datasets and a column named SizeClass with the names of the zooplankton size classes.
Crosswalk	Crosswalk table (e.g., <a href="#">crosswalk</a> ) with columns named "Phylum", "Class", "Order", "Family", "Genus", "Taxname", "Lifestage", and column names corresponding to each unique value of <code>paste(data\$Source, data\$SizeClass, sep="_")</code> .



**Value**

a tibble with the complete taxonomic information for each combination of source and size class.

**Author(s)**

Sam Bashevkin

**See Also**

[Zoopsynther](#), [crosswalk](#), [zoopComb](#)

**Examples**

```
SourceTaxaKey <- SourceTaxaKeyer(Data = dplyr::filter(zoopComb, Source!="YBFMP"),
Crosswalk = crosswalk)
```

---

startDates	<i>Start dates</i>
------------	--------------------

---

**Description**

First dates sampled by each survey and size class

**Usage**

```
startDates
```

**Format**

a tibble with 14 rows and 3 columns.

**Source** Abbreviated name of the source dataset. "EMP"=Environmental Monitoring Program, "FRP"=Fish Restoration Program, "FMWT"= Fall Midwater Trawl, "STN"= Townet Survey, "20mm" =20mm survey, "DOP" = Directed Outflow Project Lower Trophic Study, and "YBFMP"= Yolo Bypass Fish Monitoring Program.

**SizeClass** Net size class. Micro corresponds to 43 (EMP) or 50 (YBFMP)  $\mu\text{m}$  mesh, Meso corresponds to 150 (FRP and DOP) or 160 (EMP, FMWT, STN, 20mm, YBFMP)  $\mu\text{m}$  mesh, and Macro corresponds to 500 (FRP and DOP) - 505 (EMP, FMWT, STN)  $\mu\text{m}$  mesh. However, prior to 1974 EMP macrozooplankton were sampled with a 930  $\mu\text{m}$  mesh net.

**Startdate** Date first sample was collected.

**See Also**

[Uncountedyears](#), [Zoopsynther](#)

---

stations	<i>Station locations</i>
----------	--------------------------

---

**Description**

Latitudes and longitudes for each zooplankton station.

**Usage**

```
stations
```

**Format**

a tibble with 387 rows and 4 columns

**Source** Abbreviated name of the source dataset

**Station** Sampling station name

**Latitude** Latitude in decimal degrees

**Longitude** Longitude in decimal degrees

**See Also**

[Zoopdownloader](#), [zooper](#)

---

stationsEMPEZ	<i>EMP EZ Station locations</i>
---------------	---------------------------------

---

**Description**

Latitudes and longitudes for EMP EZ stations on each sampling date from 2004 to present.

**Usage**

```
stationsEMPEZ
```

**Format**

a tibble with 491 rows and 4 columns

**Date** Date sample was collected

**Station** Sampling station name

**Latitude** Latitude in decimal degrees

**Longitude** Longitude in decimal degrees

**See Also**

[Zoopdownloader](#), [zooper](#)

---

Taxnamefinder	<i>Finds all of the lowest-level (i.e. counted) taxonomic names within a vector of taxa</i>
---------------	---

---

### Description

Helps filter the zooplankton dataset by returning a set of lowest-level taxa (i.e. the level taxa were recorded at when counted in plankton samples) within a vector of taxa (which can include taxa from any taxonomic level).

### Usage

```
Taxnamefinder(Crosswalk, Taxa)
```

### Arguments

Crosswalk	Crosswalk table (such as <a href="#">crosswalk</a> ) with columns named "Phylum", "Class", "Order", "Family", "Genus", "Species", and "Taxname." "Taxname" corresponds to the full scientific name of the taxonomic level assigned to the plankter when recorded in the dataset.
Taxa	A character vector of taxa you wish to select. These taxa can be from any taxonomic level present in the list above. If using the built-in data and crosswalk, they should be present in the <a href="#">completeTaxaList</a> .

### Value

A character vector of scientific names contained within the vector of Taxa provided.

### Author(s)

Sam Bashevkin

### See Also

[completeTaxaList](#), [Zoopsynther](#)

### Examples

```
Taxnames <- Taxnamefinder(crosswalk, c("Calanoida", "Cyclopoida"))
```

---

Uncountedyears      *Detect dates when a species was not counted*

---

### Description

Detects years when a species was present in the system (i.e., post-invasion for invasive species) but not counted in each zooplankton survey

### Usage

```
Uncountedyears(Source, Size_class, Crosswalk, Start_year, Intro_lag)
```

### Arguments

Source	String with the name of the source dataset (e.g., Source="EMP").
Size_class	String with the name of the desired zooplankton size class (e.g., Source="Meso").
Crosswalk	Crosswalk table like <a href="#">crosswalk</a> or another table in the same format.
Start_year	First year the Source survey started sampling zooplankton.
Intro_lag	Number of years buffer after a species is introduced when we expect surveys to start recording them. Effectively adds Intro_lag years to the introduction year of each species.

### Details

This function is designed to work on one source and size class at a time. To apply across multiple IDs, use the [map](#) or [apply](#) functions.

### Value

A tibble with columns for the Taxlifestage, Taxname, Lifestage, Source, Sizeclass, and then a list-column of years in which that particular taxon was not counted in the specified study and size class. Taxa that were counted in all applicable years are not included in the tibble.

### Author(s)

Sam Bashevkin

### See Also

[Zoopsynther](#), [crosswalk](#), [startDates](#)

**Examples**

```

require(purrr)
require(dplyr)
require(lubridate)

datasets<-zooper::zoopComb%>%
  mutate(names=paste(Source, SizeClass, sep="_"))%>%
  select(names, Source, SizeClass)%>%
  filter(Source%in%c("EMP", "FMWT", "twentymm"))%>%
  distinct()

BadYears<-map2_dfr(datasets$Source, datasets$SizeClass, ~ Uncountedyears(Source = .x,
  Size_class = .y,
  Crosswalk = zooper::crosswalk,
  Start_year = zooper::startDates%>%
  filter(Source==.x & SizeClass==.y)%>%
  pull(Startdate)%>%
  year(),
  Intro_lag=2))

```

undersampled

*Taxa undersampled in each size class***Description**

A table listing the taxonomic names and life stages of plankton undersampled by each net mesh size (i.e. size class)

**Usage**

undersampled

**Format**

a tibble with 27 rows and 3 columns

**SizeClass** The size class of zooplankton intended to be capture be each net mesh size. Micro corresponds to 43  $\mu\text{m}$  mesh and Meso corresponds to 150-160  $\mu\text{m}$  mesh.

**Taxname** The scientific name of taxa undersampled by the corresponding mesh size class

**Lifestage** The lifestage of each taxa undersampled by the corresponding mesh size class

**See Also**

[Zoopsynther](#), [zooper](#)

---

Zoopbiomass	<i>Convert CPUE to biomass This function converts zooplankton CPUE to carbon biomass (Carbon biomass per unit effort (<math>\mu\text{g}/\text{m}^3</math>)) for taxa with conversion equations.</i>
-------------	---

---

### Description

Convert CPUE to biomass This function converts zooplankton CPUE to carbon biomass (Carbon biomass per unit effort ( $\mu\text{g}/\text{m}^3$ )) for taxa with conversion equations.

### Usage

```
Zoopbiomass(
  Zoop,
  ZoopLengths,
  Biomass_mesomicro = zooper::biomass_mesomicro,
  Biomass_macro = zooper::biomass_macro
)
```

### Arguments

Zoop	Zooplankton count dataset
ZoopLengths	Zooplankton length dataset for macrozooplankton.
Biomass_mesomicro	The micro and meso zooplankton biomass conversion table. The default is <a href="#">biomass_mesomicro</a>
Biomass_macro	The macro zooplankton biomass conversion table. The default is <a href="#">biomass_macro</a>

---

zoopComb	<i>Combined zooplankton dataset</i>
----------	-------------------------------------

---

### Description

All source zooplankton datasets combined into one tibble.

### Usage

```
zoopComb
```

**Format**

a tibble with 3,615,105 rows and 14 columns.

**Source** Abbreviated name of the source dataset. "EMP"=Environmental Monitoring Program, "FRP"=Fish Restoration Program, "FMWT"= Fall Midwater Trawl, "STN"= Townet Survey, "20mm" =20mm survey, "DOP" = Directed Outflow Project Lower Trophic Study, and "YBFMP"= Yolo Bypass Fish Monitoring Program.

**SizeClass** Net size class. Micro corresponds to 43-50  $\mu\text{m}$  mesh, Meso corresponds to 150-160  $\mu\text{m}$  mesh, and Macro corresponds to 500-505  $\mu\text{m}$  mesh. However, prior to 1974 EMP macrozooplankton were sampled with a 930  $\mu\text{m}$  mesh net.

**Volume** Volume ( $m^3$ ) of water sampled

**Lifestage** Zooplankton life stage

**Taxname** Scientific name

**Phylum** Phylum

**Class** Class

**Order** Order

**Family** Family

**Genus** Genus

**Species** Species

**Taxlifestage** Combined Taxname and Lifestage

**SampleID** Unique ID of the zooplankton sample. This key and SizeClass link to the zoopEnvComb dataset

**CPUE** Catch per unit effort (number  $m^{-3}$ )

**BPUE** Carbon biomass per unit effort ( $\mu\text{g}/m^3$ )

**Details**

Note that EMP Macro samples with QAQC flags (any value of AmphipodCode other than "A") have had their Amphipod CPUE set to NA in this integrated dataset. For more information on the source datasets see [zooper](#).

**See Also**

[Zoopdownloader](#), [Zoopsynter](#), [zooper](#)

Zoopdownloader

*Downloads and combines zooplankton datasets collected by the Inter-agency Ecological Program from the Sacramento-San Joaquin Delta***Description**

This function downloads all IEP zooplankton datasets from the internet, converts them to a consistent format, binds them together, and exports the combined dataset as .Rds R data files and/or an R object. Datasets currently include "EMP" (Environmental Monitoring Program), "FRP" (Fish Restoration Program), "FMWT" (Fall Midwater Trawl), "STN" (TowNet Survey), "20mm" (20mm survey), "DOP" (Directed Outflow Project Lower Trophic Study), and "YBFMP" (Yolo Bypass Fish Monitoring Program).

**Usage**

```
Zoopdownloader(
  Data_sets = c("EMP_Meso", "FMWT_Meso", "STN_Meso", "20mm_Meso", "FRP_Meso",
    "EMP_Micro", "FRP_Macro", "EMP_Macro", "FMWT_Macro", "STN_Macro", "DOP_Meso",
    "DOP_Macro"),
  Biomass = TRUE,
  Data_folder = tempdir(),
  Save_object = TRUE,
  Return_object = FALSE,
  Return_object_type = "List",
  Redownload_data = FALSE,
  Download_method = "auto",
  Zoop_path = file.path(Data_folder, "zoopforzooper"),
  Env_path = file.path(Data_folder, "zoopenvforzooper"),
  Crosswalk = zooper::crosswalk,
  Stations = zooper::stations
)
```

**Arguments**

Data_sets	Datasets to include in combined data. Choices include "EMP_Meso", "FMWT_Meso", "STN_Meso", "20mm_Meso", "FRP_Meso", "YBFMP_Meso", "EMP_Micro", "YBFMP_Micro", "FRP_Macro", "EMP_Macro", "FMWT_Macro", "STN_Macro", "DOP_Macro", and "DOP_Meso". Defaults to including all datasets except the two YBFMP datasets.
Biomass	Whether to add carbon biomass (carbon biomass per unit effort ( $\mu\text{g}/\text{m}^3$ )) to the dataset (where conversion equations and required data are available). Defaults to Biomass = TRUE
Data_folder	Path to folder in which source datasets are stored, and to which you would like datasets to be downloaded if you set Redownload_data = TRUE. If you do not want to store every source dataset, you can leave this at the default tempdir(). If you do not wish to redownload these datasets every time you run the function,



you can set this to a directory on your computer and run the function in the future with `Redownload_data = FALSE`, which will load the source datasets from `Data_folder` instead of downloading them again.

<code>Save_object</code>	Should the combined data be saved to disk? Defaults to <code>Save_object = TRUE</code> .
<code>Return_object</code>	Should data be returned as an R object? If <code>TRUE</code> , the function will return the full combined dataset. Defaults to <code>'Return_object = FALSE'</code> .
<code>Return_object_type</code>	If <code>Return_object = TRUE</code> , should data be returned as a combined dataframe ( <code>Return_object_type = "Combined"</code> ) or a list with component "Zooplankton" containing the zooplankton data and component "Environment" containing the environmental data ( <code>Return_object_type = "List"</code> , the default). A list is required to feed data into the <code>Zoopsynther</code> function without saving the combined dataset to disk.
<code>Redownload_data</code>	Should source datasets be redownloaded from the internet? Defaults to <code>Redownload_data = FALSE</code> .
<code>Download_method</code>	Method used to download files. See argument <code>method</code> options in <a href="#">download.file</a> . Defaults to <code>"curl"</code> .
<code>Zoop_path</code>	File path specifying the folder and filename of the zooplankton dataset. Defaults to <code>Zoop_path = file.path(Data_folder, "zoopforzooper")</code> .
<code>Env_path</code>	File path specifying the folder and filename of the dataset with accessory environmental parameters. Defaults to <code>Env_path = file.path(Data_folder, "zoopenvforzooper")</code> .
<code>Crosswalk</code>	Crosswalk table to be used for conversions. Must have columns named for each unique combination of source and size class with an underscore separator, as well as all taxonomic levels Phylum through Species, Taxname (full scientific name) and Lifestage. See <a href="#">crosswalk</a> (the default) for an example.
<code>Stations</code>	Latitudes and longitudes for each unique station. See <a href="#">stations</a> (the default) for an example.

### Details

Note that EMP Macro samples with QAQC flags (any value of `AmphipodCode` other than "A") have had their `Amphipod CPUE` set to NA in this function. For more information on the source datasets see [zooper](#).

### Value

If `Return_object = TRUE`, returns the combined dataset as a list or tibble, depending on whether `Return_object_type` is set to `"List"` or `"Combined"`. If `Save_object = TRUE`, writes 2 .Rds files to disk: one with the zooplankton catch data and another with accessory environmental parameters.

### Author(s)

Sam Bashevkin

**See Also**

[Zoopsynther](#), [crosswalk](#), [stations](#), [zooper](#)

**Examples**

```
## Not run:
Data <- Zoopdownloader(Data_folder = tempdir(), Return_object = TRUE,
  Save_object = FALSE, Redownload_data = TRUE)

## End(Not run)
```

---

 zoopEnvComb

*Environmental data*


---

**Description**

Accessory environmental data from the combined zooplankton dataset. Not all datasets report all environmental parameters.

**Usage**

```
zoopEnvComb
```

**Format**

a tibble with 44,690 rows and 20 columns

**Source** Abbreviated name of the source dataset. "EMP"=Environmental Monitoring Program, "FRP"=Fish Restoration Program, "FMWT"= Fall Midwater Trawl, "STN"= Townet Survey, "20mm" =20mm survey, "DOP" = Directed Outflow Project Lower Trophic Study, and "YBFMP"= Yolo Bypass Fish Monitoring Program.

**Year** Year sample was collected

**Date** Date sample was collected

**Datetime** Date and time sample was collected, if time was provided

**TowType** Sample collection method identifying whether each tow was a surface, bottom, oblique, or vertical pump sample

**Tide** Tidal stage

**Station** Station where sample was collected. This is the key that links to the stations dataset

**Chl** Chlorophyll concentration in  $\mu\text{g/L}$

**Secchi** Secchi depth in cm

**Temperature** Temperature in  $^{\circ}\text{C}$ .

**BottomDepth** Total depth of the water column in m

**TurbidityNTU** Water turbidity in Nephelometric Turbidity Units

**TurbidityFNU** Water turbidity in Formazin Nephelometric Units

**Microcystis** Intensity of Microcystis bloom coded qualitatively from 1-5 where 1 = absent, 2 = low, 3 = medium, 4 = high, 5 = very high

**pH** Water pH

**DO** Dissolved oxygen in mg/L

**SalSurf** Surface salinity in PPT

**SalBott** Bottom salinity in PPT

**Latitude** Latitude in decimal degrees

**Longitude** Longitude in decimal degrees

**AmphipodCode** Code indicating sample quality for EMP macro amphipod samples (A=valid, B=questionable [veg/algal bloom in net], C=not valid [error in lab], D=suspect [possible missing data])

**SampleID** Unique ID of the zooplankton sample. This is the key that links to the zoopComb dataset

### See Also

[Zoopdownloader](#), [Zoopsynther](#), [zooper](#)

---

zooper

*zooper: A package for integrating zooplankton datasets from the Sacramento San Joaquin Delta*

---

### Description

This package contains functions, lookup tables, and 2 built-in pre-combined datasets (one with the zooplankton data and another with the environmental data).

### zooper functions

- [Zoopsynther](#)
- [Zoopdownloader](#)
- [Taxnamefinder](#)
- [SourceTaxaKeyer](#)
- [LCD\\_Taxa](#)
- [Uncountedyears](#)

### zooper lookup tables

- [crosswalk](#)
- [undersampled](#)
- [stations](#)
- [completeTaxaList](#)
- [startDates](#)

**zooper pre-combined (with the [Zoopdownloader](#) function) zooplankton datasets. These may be out of date**

- [zoopComb](#)
- [zoopEnvComb](#)

### Source datasets

**Environmental Monitoring Program (EMP)** The EMP zooplankton survey is run by the California Department of Fish and Wildlife. Zooplankton were first collected in 1972. It samples monthly at 17 fixed stations, 2 floating entrapment zone stations, and 3 stations in Carquinez Strait and San Pablo Bay that are only sampled during high outflow and low salinity conditions. EMP samples using micro (43  $\mu\text{m}$ ), meso (160  $\mu\text{m}$ ), and macro (505  $\mu\text{m}$ ) zooplankton nets. Note that additional Amphipod data with quality issues (e.g., vegetation in net) are available in the EMP data publication. [Data are available here.](#)

**20-mm Survey** The 20-mm survey is run by the California Department of Fish and Wildlife. Zooplankton were first collected in 1995. Zooplankton are collected concurrently with fish samples at 41-55 fixed open-channel stations per year. Samples are collected twice per month between March and July. Only Mesozooplankton are collected with a 160  $\mu\text{m}$  mesh net. [Data are available here.](#)

**Fall Midwater Trawl (FMWT) and Summer Towntnet Survey (STN)** The FMWT and STN are run by the California Department of Fish and Wildlife. FMWT samples are collected monthly between September and December from a subset of the 122 fixed open-channel stations. STN samples are collected monthly between June and August from 40 fixed open-channel stations. Macrozooplankton have been collected since 2007 with a 505  $\mu\text{m}$  mesh net while mesozooplankton have been collected since 2005 with a 160  $\mu\text{m}$  mesh net. [Data are available here.](#) Supplemental sampling from the Suisun Marsh Salinity Control Gate study data are also included and [those data can be found here.](#)

**Fish Restoration Program (FRP)** FRP is run by the California Department of Fish and Wildlife. Zooplankton were first collected in 2015. Samples are collected monthly between March and December in shallow-water habitats near marshes. FRP samples with meso (150  $\mu\text{m}$ ) and macro (500  $\mu\text{m}$ ) zooplankton nets. [Data are available here.](#)

**Directed Outflow Project Lower Trophic Study (DOP)** The Directed Outflow Project Lower Trophic Study is run by ICF for the United States Bureau of Reclamation. Zooplankton were first collected in fall 2017. 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. Three sampling stations per region are randomly selected for 5 regions (Suisun Bay, Suisun Marsh, Lower Sac. River, Cache Slough, Sac Ship Channel). 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 ( $\leq 10$  feet) and channel ( $> 10$  feet) habitat. Channels are sampled at the surface and, if deeper than 20 feet, also at the bottom 1/2 to 1/3 of the water column. DOP samples with meso (150  $\mu\text{m}$ ) and macro (500  $\mu\text{m}$ ) zooplankton nets. [Data are available here.](#)

**Yolo Bypass Fish Monitoring Program (YBFMP)** YBFMP is run by the California Department of Water Resources. Zooplankton were first collected in 1999. Samples are from 2 sites, one in the Yolo Bypass and one in the Sacramento River. In 1999 and 2000, samples were collected for a couple of months each year, and then it increased to roughly winter/spring

from 2001-2010. Since 2011, samples are collected twice monthly during most of the year, and weekly when the bypass is inundated. YBFMP samples with micro (50  $\mu\text{m}$ ) and meso (150  $\mu\text{m}$ ) zooplankton nets. [Data are available here.](#)

### Author(s)

**Maintainer:** Samuel M Bashevkin <sam.bashevkin@waterboards.ca.gov> ([ORCID](#))

Authors:

- Rosemary Hartman <Rosemary.Hartman@water.ca.gov> ([ORCID](#))
- Karrin Alstad <Karrin.Alstad@wildlife.ca.gov>
- Catarina Pien <cpien@usbr.gov> ([ORCID](#))

### See Also

Useful links:

- <https://github.com/InteragencyEcologicalProgram/zooper>
- Report bugs at <https://github.com/InteragencyEcologicalProgram/zooper/issues>

---

Zoopsynter

*Integrates zooplankton datasets collected by the Interagency Ecological Program from the Sacramento-San Joaquin Delta*

---

### Description

This function returns an integrated zooplankton dataset with taxonomic issues resolved, according to user-specifications, along with important caveats about the data. It requires the output of the [Zoopdownloader](#) function to run. This can be provided either as a list or paths to saved .Rds files generated by the [Zoopdownloader](#) function. The function defaults to loading pre-packaged combined datasets (which may be outdated).

### Usage

```
Zoopsynter(
  Data_type = NULL,
  Zoop = zooper::zoopComb,
  ZoopEnv = zooper::zoopEnvComb,
  Zoop_path = NULL,
  Env_path = NULL,
  Sources = c("EMP", "FRP", "FMWT", "STN", "20mm", "DOP"),
  Size_class = c("Micro", "Meso", "Macro"),
  Time_consistency = FALSE,
  Intro_lag = 2,
  Response = "CPUE",
  Taxa = NULL,
  Date_range = c(NA, NA),
```

```

Months = NA,
Years = NA,
Sal_bott_range = NA,
Sal_surf_range = NA,
Temp_range = NA,
Lat_range = NA,
Long_range = NA,
Reload_data = F,
Redownload_data = F,
All_env = T,
Shiny = F,
Crosswalk = zooper::crosswalk,
Undersampled = zooper::undersampled,
CompleteTaxaList = zooper::completeTaxaList,
StartDates = zooper::startDates,
...
)

```

## Arguments

Data_type	What type of data are you looking for? This option allows you to choose a final output dataset for either community (Data_type = "Community"; the default) or Taxa-specific (Data_type = "Taxa") analyses. NOTE: If you set Data_type="Community" we do not recommend utilizing the Taxa argument. See below for more explanation of this argument.
Zoop	Zooplankton data. You must provide the "Zooplankton" element from the list returned from <a href="#">Zoopdownloader</a> (Save_object = FALSE, Return_object = TRUE, Return_object_type="List"). The default argument provides the built-in (and possibly outdated) version of this combined dataset. If you instead wish to provide paths to saved datasets from the <a href="#">Zoopdownloader</a> function, set Data_list = NULL and provide Zoop_path.
ZoopEnv	Accessory environmental data. You must provide the "Environment" element from the list returned from <a href="#">Zoopdownloader</a> (Save_object = FALSE, Return_object = TRUE, Return_object_type="List"). The default argument provides the built-in (and possibly outdated) version of this combined dataset. If you instead wish to provide paths to saved datasets from the <a href="#">Zoopdownloader</a> function, set Data_list = NULL and provide Env_path.
Zoop_path	If you wish to save time by saving the combined zooplankton datasets returned from the <a href="#">zoopdatadownloader</a> to disk, provide here the path to the combined zooplankton dataset on disk. You must also set Data_list = NULL.
Env_path	If you wish to save time by saving the combined zooplankton datasets returned from the <a href="#">zoopdatadownloader</a> to disk, provide here the path to the combined accessory environmental data on disk. You must also set Data_list = NULL.
Sources	Source datasets to be included. Choices include "EMP" (Environmental Monitoring Program), "FRP" (Fish Restoration Program), "FMWT" (Fall Midwater Trawl), "STN" (Townet Survey), "DOP" (Directed Outflow Project), and "20mm" (20mm survey). The YBFMP datasets cannot be used in this function

due to taxonomic and life stage issues with that dataset. Defaults to Sources = c("EMP", "FRP", "FMWT", "STN", "20mm", "DOP").

Size_class	Zooplankton size classes (as defined by net mesh sizes) to be included in the integrated dataset. Choices include "Micro" (43 $\mu\text{m}$ ), "Meso" (150 - 160 $\mu\text{m}$ ), and "Macro" (500-505 $\mu\text{m}$ ). Defaults to Size_class = c("Micro", "Meso", "Macro").
Time_consistency	Would you like to apply a fix to enforce consistent taxonomic resolution over time? Only available for the Community option.
Intro_lag	Only applicable if Time_consistency = TRUE. How many years after a species is introduced should we expect surveys to start counting them? Defaults to 2.
Response	Which response variable(s) would you like for the zooplankton data? Choices are "CPUE" (catch per unit effort) and "BPUE" (carbon biomass per unit effort ( $\mu\text{g}/\text{m}^3$ )). Defaults to Response = "CPUE".
Taxa	If you only wish to include a subset of taxa, provide a character vector of the taxa you wish included. This can include taxa of any taxonomic level (e.g., Taxa = "Calanoida") to include only calanoids. NOTE: we do not recommend you use this feature AND set Data_type="Community". This is better suited to selecting higher-level taxa. If you wish to just include one or a few species, it would be faster to just filter the output of <a href="#">Zoopdownloader</a> to include those taxa. Defaults to NULL, which includes all taxa.
Date_range	Range of dates to include in the final dataset. To filter within a range of dates, include a character vector of 2 dates formatted in the yyyy-mm-dd format exactly, specifying the upper and lower bounds. To specify an infinite upper or lower bound (to include all values above or below a limit) input NA for that infinite bound. Defaults to Date_range = c(NA, NA), which includes all dates.
Months	Months (as integers) to be included in the integrated dataset. If you wish to only include data from a subset of months, input a vector of integers corresponding to the months you wish to be included. Defaults to Months = NA, which includes all months.
Years	Years to be included in the integrated dataset. If you wish to only include data from a subset of years, input a vector of years you wish to be included. Defaults to Years = NA, which includes all months.
Sal_bott_range	Filter the data by bottom salinity values. Include a vector of length 2 specifying the minimum and maximum values you wish to include. To include all values above or below a limit, utilize Inf or -Inf for the upper or lower bound respectively. Defaults to Sal_bott_range = NA, which includes all bottom salinities.
Sal_surf_range	Same as previous, but for surface salinity.
Temp_range	Same as Sal_bott_range but for surface temperature.
Lat_range	Latitude range to include in the final dataset. Include a vector of length 2 specifying the minimum and maximum values you wish to include, in decimal degree format. Defaults to Lat_range = NA, which includes all latitudes.
Long_range	Same as previous, but for longitude. Don't forget that Longitudes should be negative in the Delta!

Reload_data	If set to <code>Reload_data = T</code> runs the <a href="#">Zoopdownloader</a> function to re-combine source datasets. To include local versions of the datasets without re-downloading them from online, set <code>Reload_data = TRUE</code> and <code>Redownload_data = FALSE</code> . Defaults to <code>Reload_data = FALSE</code>
Redownload_data	Should data be re-downloaded from the internet? If set to <code>Redownload_data = TRUE</code> , runs <a href="#">Zoopdownloader</a> ( <code>Redownload_data=Redownload_data</code> , <code>Zoop_path=Zoop_path</code> , <code>Env_path=Env_path</code> , ...). Defaults to <code>Redownload_data = FALSE</code> .
All_env	Should all environmental parameters be included? Defaults to <code>All_env = TRUE</code> .
Shiny	Is this function being used within the shiny app? If set to <code>Shiny = TRUE</code> , outputs a list with the integrated dataset as one component and the caveats as the other component. Defaults to <code>Shiny = FALSE</code> .
Crosswalk	Crosswalk table to be used for conversions. Must have columns named for each unique combination of source and size class with an underscore separator, as well as all taxonomic levels Phylum through Species, Taxname (full scientific name) and Lifestage. See <a href="#">crosswalk</a> (the default) for an example.
Undersampled	A table listing the taxonomic names and life stages of plankton undersampled by each net mesh size (i.e. size class). See <a href="#">undersampled</a> (the default) for an example.
CompleteTaxaList	Character vector of all taxonomic names in source datasets. Defaults to <a href="#">completeTaxaList</a> .
StartDates	Tibble with the starting dates of each source dataset. Defaults to <a href="#">startDates</a> .
...	Arguments passed to <a href="#">Zoopdownloader</a> .

## Details

This function combines any combination of the zooplankton datasets (included as parameters) and calculates least common denominator taxa to facilitate comparisons across datasets with differing levels of taxonomic resolution. For more information on the source datasets see [zooper](#).

## Value

An integrated zooplankton dataset.

## Data type

The `Data_type` parameter toggles between two approaches to resolving differences in taxonomic resolution. If you want all available data on given Taxa, use `Data_type="Taxa"` but if you want to conduct a community analysis, use `Data_type = "Community"`.

Briefly, `Data_type = "Community"` optimizes for community-level analyses by taking all taxa x life stage combinations that are not measured in every input dataset, and summing them up taxonomic levels to the lowest taxonomic level they belong to that is covered by all datasets. Remaining Taxa x life stage combos that are not covered in all datasets up to the phylum level (usually something like Annelida or Nematoda or Insect Pupae) are removed from the final dataset. However, some taxa x life stage combos are retained if they are taxonomic levels higher than species that are counted in some surveys, and a lower taxonomic level within this group is counted in all surveys.



For example, if we had 3 surveys where surveys A and B count *Pseudodiaptomus forbesi*, *Pseudodiaptomus marinus*, and *Pseudodiaptomus* spp. (UnID) but survey C only counts *P. forbesi* and *P. marinus* then the *Pseudodiaptomus* spp. (UnID) category would be retained after applying the community approach.

Data\_type = "Taxa" optimizes for the Taxa-level user by maintaining all data at the original taxonomic level (but it outputs warnings for taxa not measured in all datasets, which we call "orphans"). To facilitate comparisons across datasets, this option also sums data into general categories that are comparable across all datasets and years: "summed groups." The new variable "Taxatype" identifies which taxa are summed groups (Taxatype = "Summed group"), which are measured to the species level (Taxatype = "Species"), and which are higher taxonomic groupings with the species designation unknown: (Taxatype = "UnID species").

### Author(s)

Sam Bashevkin

### See Also

[Zoopdownloader](#), [Taxnamefinder](#), [SourceTaxaKeyer](#), [crosswalk](#), [undersampled](#), [zoopComb](#), [zoopEnvComb](#), [zooper](#)

### Examples

```
MyZoops <- Zoopsynther(Data_type = "Community",
  Sources = c("EMP", "FRP", "FMWT"),
  Size_class = "Meso",
  Date_range = c("1990-10-01", "2000-09-30"))
```

---

zoop_urls	<i>Extract latest EDI files This function extracts the latest version of a zooplankton EDI package and the list of files from that package</i>
-----------	--

---

### Description

Extract latest EDI files This function extracts the latest version of a zooplankton EDI package and the list of files from that package

### Usage

```
zoop_urls(Sources)
```

### Arguments

Sources Source datasets to be included. Choices include "EMP" (Environmental Monitoring Program), "FRP" (Fish Restoration Program), "FMWT" (Fall Midwater Trawl), "STN" (Townet Survey), "DOP" (Directed Outflow Project), and "20mm" (20mm survey). The YBFMP datasets cannot be used in this function due to taxonomic and life stage issues with that dataset. Defaults to Sources = c("EMP", "FRP", "FMWT", "STN", "20mm", "DOP").

**Value**

A list with the files and/or URLs for each source dataset

**Author(s)**

Sam Bashevkin

# Index

## \* **Taxonomy**

Commontaxer, 4  
LCD\_Taxa, 7  
SourceTaxaKeyer, 8  
Taxnamefinder, 11

## \* **datasets**

biomass\_macro, 2  
biomass\_mesomicro, 3  
completeTaxaList, 5  
crosswalk, 5  
startDates, 9  
stations, 10  
stationsEMPEZ, 10  
undersampled, 13  
zoopComb, 14  
zoopEnvComb, 18

## \* **download**

Zoopdownloader, 16

## \* **integration**

Zoopdownloader, 16  
Zoopsynther, 21

## \* **synthesis**

Zoopdownloader, 16  
Zoopsynther, 21

## \* **zooplankton**

Taxnamefinder, 11  
Zoopsynther, 21

## \* **zooplankton**

Commontaxer, 4  
LCD\_Taxa, 7  
SourceTaxaKeyer, 8  
Zoopdownloader, 16

apply, 4, 7, 12

biomass\_macro, 2, 3, 14  
biomass\_mesomicro, 3, 3, 14

Commontaxer, 4  
completeTaxaList, 5, 11, 19, 24

crosswalk, 4, 5, 8, 9, 11, 12, 17–19, 24, 25

download.file, 17

LCD\_Taxa, 7, 19

map, 4, 7, 12

SourceTaxaKeyer, 4, 8, 19, 25  
startDates, 9, 12, 19, 24  
stations, 10, 17–19  
stationsEMPEZ, 10

Taxnamefinder, 5, 11, 19, 25

Uncountedyears, 9, 12, 19  
undersampled, 13, 19, 24, 25

zoop\_urls, 25

Zoopbiomass, 14

zoopComb, 8, 9, 14, 20, 25

Zoopdownloader, 7, 10, 15, 16, 19–25

zoopEnvComb, 18, 20, 25

zooper, 5, 7, 10, 13, 15, 17–19, 19, 24, 25

zooper-package (zooper), 19

Zoopsynther, 4, 5, 7–9, 11–13, 15, 18, 19, 21