

Description of all Data Files as part of BCO-DMO dataset 885637, “Doliolid Distribution Synthesis” (PI: Adam Greer)

(1) **File name: DoliolidAbundances-All.zip**

(ZIP Archive (ZIP), 1.43 MB)

MD5:ba60e1a49b9cf4e3f131ede3accadf1e

URL: <https://datadocs.bco-dmo.org/data/305/DoIMICROBE/885637/1/data/DoliolidAbundances-All.zip>

Description:

This folder contains the calculated concentrations and average oceanographic variables in each bin from both automated (using computer vision) and manually verified data.

File names and descriptions:

- **CA2018-dolconc.csv** = Northern California in summer 2018.
- **CA2019-dolconc.csv** = Northern California in summer 2019.
- **FI2014-dolconc.csv** = Straits of Florida in summer 2014.
- **FI2015-dolconc.csv** = Straits of Florida in summer 2015.
- **GOM-ECORR2016-dolconc.csv** = Northern Gulf of Mexico Eastern corridor, south of Perdido Bay, FL in 2016.
- **GOM-MCORR2016-dolconc.csv** = Northern Gulf of Mexico Middle corridor, south of Mobile Bay, AL in 2016.
- **med_dols_summary_1mbin.csv** = Mediterranean Sea in July 2013 – abundances are binned to 1 meter cubed.
- **OR2018-dolconc.csv** = Oregon coast in summer 2018.
- **OR2019-dolconc.csv** = Oregon coast in summer 2019.
- **SD2010-dolconc.csv** = Offshore of San Diego in October 2010.

Column names, descriptions, and units:

- **rjul** = rounded julian time; units = percentage of a day.
- **sdol** = solo doliolid abundance - gonozooid life stage; units = individuals per cubic meter; (can also be labeled "Gonozooid").
- **buds** = doliolids with buds abundance - phorozoid life stage; units = individuals per cubic meter; (can also be labeled "Phorozoid").
- **nurse** = doliolid nurse abundance - nurse life stage; units = individuals per cubic meter; (can also be labeled "Nurse").
- **dolconc** = total doliolid concentration including all life stages; units = individuals per cubic meter; (can also be labeled as "totdol").
- **mdepth** = mean depth for bin where concentration was calculated; units = meters.
- **msal** = mean salinity for bin where concentration was calculated; units = practical salinity units.

- **mtemp** =mean temperature for bin where concentration was calculated; units = degrees Celsius.
- **mfluor** =mean chlorophyll-a fluorescence for bin where concentration was calculated; units = voltage (relative measure).
- **moxy** = mean dissolved oxygen for bin where concentration was calculated; units = milligrams per Liter.
- **mirr** = mean irradiance (PAR) for bin where concentration was calculated; units = microEinsteins per meter squared.
- **mlat** = mean latitude for bin where concentration was calculated; units = degrees.
- **mlon** = mean longitude for bin where concentration was calculated; units = degrees.

(2) File name: DoliolidEnvironment-Automated.zip

(ZIP Archive (ZIP), 45.69 MB)

MD5:824fd92da94ffa2471c39f5fb16778e1

URL: <https://datadocs.bco-dmo.org/data/305/DoIMICROBE/885637/1/data/DoliolidEnvironment-Automated.zip>

Description:

This folder contains the raw data from the different ecosystems generated with computer vision algorithms where each row is an individual doliolid and all of the oceanographic parameters associated with it.

File names and descriptions:

- File "**env_visufront_cc.csv**" contains some different data columns that were used in analysis.
- The following files contain the automated identification of objects (i.e., segments) detected in the images, as predicted by a convolutional neural network (CNN). The columns contain a file name for the segment and what the algorithm predicted was its identification. All are predicted to be doliolids of different life stages. Additional data columns are not important or used in the analysis because these segments were merged to the sensor data from the vehicle using the time stamp of the segments in the file name (see processing code in R):

- **205_doliolids_reduced.Rdata**

- **213_doliolids_reduced.Rdata**

- **401_doliolids_reduced.Rdata**

- **425_doliolids_reduced.Rdata**

- **all_joined_3W_doliolid_unbinned_90FTclass.Rdata**

- all_joined_10W-7W_doliolid_unbinned_90FTclass.Rdata
- dols_ORCA_Mo.Rdata
- PTS03_085_T029_MCORR_ft_ess_tunicate_doliolid.csv
- PTS03_085_T029_MCORR_ft_ess_tunicate_doliolid_juvenile_buds.csv
- PTS03_085_T029_MCORR_ft_ess_tunicate_doliolid_nurse.csv

Column names, descriptions, and units for file "env_visufront_cc.csv":

- **transect** = transect label. In this case, all transects were cross-current (cc).
- **yo** = the sequential number of profiles or tow-yos through the water column.
- **yo_type** = describes whether the vehicle was going up or down.
- **datetime** = full date and time when data were collected; in ISO8601 format (YYYY-MM-DDThh:mm:ssZ); Z indicates UTC.
- **dist** = distance along the transect; units = kilometers (km).
- **lon** = longitude; units = degrees East.
- **lat** = latitude; units = degrees North.
- **press** = pressure; units = decibars (dbar).
- **depth** = depth; units = meters (m).
- **temp** = temperature; units = degrees Celsius.
- **sal** = salinity; units = practical salinity units.
- **fluo** = chlorophyll-a fluorescence; units = voltage (relative measure).
- **oxy** = dissolved oxygen; relative units.
- **irrad** = irradiance; units = micro Einsteins per cm squared.
- **dens** = density calculated from temperature and salinity; units = kilograms per cubic meter (kg/m³-1000).
- **hor_vel** = horizontal velocity; units = millimeters per second (mm/s).
- **vert_vel** = vertical velocity; units = millimeters per second (mm/s).
- **pitch** = pitch of the vehicle (negative is pointing down, positive is upward); units = degrees.
- **roll** = roll of the vehicle to the left or right; units = degrees.
- **heading** = heading of the vehicle (0 is north, 180 is south); units = degrees.
- **datetime_utc** = full date and time in ISO8601 format (YYYY-MM-DDThh:mm:ssZ); Z indicates UTC.
- **period** = categorical variable describing the time of day.

Column names, descriptions, and units for files starting with "PTS03":

- **LATITUDE_DEG** = latitude in decimal degrees; units = degrees North.
- **LONGITUDE_DEG** = longitude in decimal degrees; units = degrees East.
- **DEPTH_M** = depth; units = meters.
- **TEMPERATURE_DEG_C** = temperature; units = degrees Celsius.
- **SALINITY** = salinity; units = practical salinity units.

- **FLUORESCENCE_VOLTS** = fluorescence; units = volts.
 - **PAR_uE_PER_M2** = irradiance (PAR); units = microEinsteins per meter squared ($\mu\text{E}/\text{m}^2$).
 - **O2_mg_PER_L** = Oxygen; units = milligrams per liter (mg/L).
-

(3) File name: DoliolidEnvironment-Manual.zip

(ZIP Archive (ZIP), 2.07 MB)

MD5:f3f43db97bf0cf5468028e90521ce441

URL: <https://datadocs.bco-dmo.org/data/305/DolMICROBE/885637/1/data/DoliolidEnvironment-Manual.zip>

Description:

This folder contains the manually verified doliolid identifications in the Gulf of Mexico and Southern California. Southern California had individual “casts” analyzed, so the times for these chunks of data are also contained in this folder. In these files, each row is an individual animal, and the associated data are similar to full descriptions of the oceanographic data collected.

File names and descriptions:

726dolenv3cats.csv = doliolids manually classified into 3 life stages from July 26, 2016. This is for the eastern corridor (ECORR) in the northern Gulf of Mexico. Each row is an individual.

dateTime_ranges_transect_2_3.csv = contains the start and stop times for each profile in the San Diego dataset. This is required to accurately measure the volume sampled. Two transects were analyzed (transects 2 and 3). Each profile has a sequential cast number, with a start and stop time in hours:minutes:second and Julian time.

doliolids_transects2_3_presence_only.csv = contains the counts of pelagic tunicates (i.e., doliolids) in the images, along with the corresponding oceanographic data. Column names are similar to other datasets. Count is the number per full image frame that was converted to an abundance per unit volume in the R code. Taxon and group are simply different levels of classification. In this study, we were interested in the doliolid taxon.

ManualDolAbundance.csv = doliolids abundances from July 26 in one cubic meter bins. The columns correspond to the abundance of the different life stages in that particular bin, along with the average environmental parameters associated with that bin. Rjul = rounded Julian time for the bin, msal = mean salinity for that bin. The rest column names follow the same pattern of being the mean “m” for the bin. Totdol = the sum of

the abundances of the 3 life stages.

Column names, descriptions, and units:

File "726dolenv3cats.csv":

- **nearjul** = nearest julian time corresponding to the oceanographic data; units = percentage of a day.
- **julian** = julian time calculated from the image time stamp of individual organisms; units = percentage of a day.
- **ID** = identification of the image region of interest (ROI) - category.
- **Junk** = A column with the word Plankton in the file name of ROI.
- **label** = file name with the date and timestamp of the first frame in the stack YYYYMMDDHHMMSS.SSS.tif.
- **slice** = The frame number in the stack of images - will be 1-430.
- **xloc** = The x coordinate of the top left of the bounding box for the ROI; will be between 0 and 2047 (the size of the image); units = pixels.
- **yloc** = The y coordinate of the top left of the bounding box for the ROI; will be between 0 and 2047 (the size of the image); units = pixels.
- **width** = the width of the ROI (bounding box); units = pixels.
- **height** = the height of the ROI (bounding box); units = pixels.
- **ID2** = more specific ID to the life stage for the doliolids.
- **utc** = timestamp in GMT; format: HHMMSS.SSS.
- **timestamp** = UNIX timestamp in seconds from January 1, 1970; units = seconds.
- **alt** = distance of the vehicle from the benthos (will be 999.9 if the bottom cannot be detected - see altok); units = meters (m).
- **temp** = temperature measurement from the vehicle CTD (SBE49); units = degrees Celsius.
- **depth** = depth of the vehicle (SBE49); units = meters (m).
- **fluor** = chlorophyll-a fluorescence measurement from the vehicle (ECO FL-RT); units = voltage (relative measurement).
- **fvel** = forward velocity of the vehicle (measured by doppler velocity log); units = meters per second (m/s).
- **heading** = heading of the vehicle (0 is north, 180 is south); units = degrees.
- **o2** = incorrect measurement of oxygen using the voltage.
- **o2volts** = dissolved oxygen measurement; units = dissolved oxygen measurement. (can be used to generate correct oxygen).
- **par** = irradiance (PAR) measured from the vehicle; units = microEinsteins per meter squared ($\mu\text{E}/\text{m}^2$).
- **parvolts** = voltage from the PAR sensor; units = voltage (relative measurement).

- **pitch** = pitch of the vehicle (negative is pointing down, positive is upward); units = degrees.
- **roll** = roll of the vehicle to the left or right; units = degrees.
- **salinity** = salinity measurement from the vehicle CTD; units = practical salinity units.
- **lat** = latitude measured by the ship GPS; units = degrees North.
- **lon** = longitude measured by the ship GPS; units = degrees East.
- **vertvel** = vertical velocity of the vehicle; units = meters per second (m/s).
- **altok** = value of 1 or 0 depending on if the benthos can be detected (1 means the altitude data are good).
- **secs** = number of seconds from the start of the day; units = seconds.
- **hdist** = horizontal distance calculated using seconds and mean fvel; units = meters (m).

File "dateTime_ranges_transect_2_3.csv":

- **transect** = the transect name.
- **cast** = the cast number for datasets where only the up or down casts were analyzed.
- **begin_dateTime** = time at beginning of a cast.
- **end_dateTime** = time at end of a cast.
- **begin** = time at beginning of a cast in seconds; units = seconds.
- **end** = time at end of a cast in seconds; units = seconds.
- **julbeg** = time at beginning of a cast in julian; units = percentage of day.
- **julend** = time at end of a cast in julian; units = percentage of day.

File "doliolids_transects2_3_presence_only.csv":

- **transect** = the transect name.
- **cast** = the cast number for datasets where only the up or down casts were analyzed.
- **down.up** = whether or not the vehicle was going up or down.
- **dateTime** = full date and local time; format = YYYY-MM-DD HH:MM:SS.sss.
- **taxon** = taxonomic classification.
- **count** = total number of organisms (doliolids) present within an image; units = number per image.
- **group** = taxonomic group.
- **depth** = depth of the vehicle (SBE49); units = meters (m).
- **lat** = latitude measured by the ship GPS; units = degrees North.
- **long** = longitude measured by the ship GPS; units = degrees East.
- **temp** = temperature measurement from the vehicle CTD (SBE49); degrees Celsius.
- **salinity** = salinity measurement from the vehicle CTD; units = practical salinity units.

- **fluoro** = chlorophyll-a fluorescence measurement from the vehicle (ECO FL-RT); units = voltage (relative measurement).
- **heading** = heading of the vehicle (0 is north, 180 is south); units = degrees.
- **vertical.vel** = vertical velocity of the vehicle; units = meters per second (m/s).
- **pitch** = pitch of the vehicle (negative is pointing down, positive is upward); units = degrees.
- **swRho** = density of the water, calculated from temp and salinity on CTD; units = kilograms per cubic meter (kg/m³).

File "ManualDolAbundance.csv":

- **rjul** = rounded julian time; units = percentage of a day.
- **Gonozoid** = solo doliolid abundance - gonozoid life stage; units = individuals per cubic meter.
- **Nurse** = doliolid nurse abundance - nurse life stage; units = individuals per cubic meter.
- **Phorozoid** = doliolids with buds abundance - phorozoid life stage; units = individuals per cubic meter.
- **mdepth** = mean depth for bin where concentration was calculated; units = meters (m).
- **msal** = mean salinity for bin where concentration was calculated; units = practical salinity units.
- **mtemp** = mean temperature for bin where concentration was calculated; units = degrees Celsius.
- **mfluor** = mean chlorophyll-a fluorescence for bin where concentration was calculated; units = voltage (relative measurement).
- **moxy** = mean dissolved oxygen for bin where concentration was calculated; units = milligrams per liter (mg/L)
- **mpar** = mean PAR for the bin.
- **mlat** = mean latitude for the bin.
- **mlon** = mean longitude for the bin.
- **totdol** = the sum of the abundances of the 3 life stages.

(4) File name: DolTransectLocations.csv

(Comma Separated Values (.csv), 1000 bytes)

MD5:afd36ea1e5c3143a3c09321eb43523b1

URL: <https://datadocs.bco-dmo.org/data/305/DolMICROBE/885637/1/data/DolTransectLocations.csv>

Description:

This file contains the Start and Stop latitudes and longitudes for all of the transects analyzed in this study. This was used to generate the map in the manuscript (Greer et al., 2022). Note that all longitudes in the Western hemisphere should be negative. This applies to all data except those collected in the Mediterranean Sea.

Column names, descriptions, and units:

- **Location** = transect name.
- **Start Lat** = start latitude; units = degrees North.
- **Start Lon** = start longitude; units = degrees East.
- **Stop Lat** = stop latitude; units = degrees North.
- **Stop Lon** = stop longitude; units = degrees East.

(5) File name: GulfofMexicoDoliolidImages-ManuallyClassified.zip

ZIP Archive (ZIP), 580.46 MB
MD5:7931bc3ca446d550cbc6f7fac016f375

URL: <https://datadocs.bco-dmo.org/data/305/DoIMICROBE/885637/1/data/GulfofMexicoDoliolidImages-ManuallyClassified.zip>

Description:

This folder contains actual images identified to 3 different life stages of doliolids from the northern Gulf of Mexico.

(6) File name: GulfPhysicalOceanographicData.zip

(ZIP Archive (ZIP), 16.33 MB)
MD5:0a52b48d3a9d923168ace5caed39ece5

URL: <https://datadocs.bco-dmo.org/data/305/DoIMICROBE/885637/1/data/GulfPhysicalOceanographicData.zip>

Description:

This folder contains oceanographic sensor data from the towed imaging system, as well as linearly interpolated data used to make plots of the doliolid distributions in the paper.

File names and descriptions:

072516physical.csv = sensor data collected on July 25, 2016 at the middle corridor (MCORR) just south of Mobile Bay, AL. This is the raw sensor data.

072616physical.csv = sensor data collected on July 26, 2016 at the eastern corridor (ECORR) just south of Perdido Bay, FL. This is the raw sensor data.

These sensor datasets were interpolated across the entire transect for temperature and salinity:

725salinityinterp.csv
725tempinterp.csv
726salinityinterp.csv
726tempinterp.csv

Column names, descriptions, and units:

Physical data (072516physical.csv, 072616physical.csv):

- **julian** = Julian time; units = percentage of a day.
- **utc** = timestamp in UTC; format: HHMMSS.SSS.
- **UNIX_timestamp** = UNIX timestamp in seconds from January 1, 1970; units = seconds.
- **ALTITUDE_M** = distance of the vehicle from the benthos (will be 999.9 if the bottom cannot be detected - see altok); units = meters (m).
- **CTD_TEMPERATURE_DEG_C** = water temperature; units = degrees Celsius.
- **DEPTH_M** = depth; units = meters (m).
- **FLOR** = fluorescence intensity; units = volts.
- **FORWARD_VELOCITY_M_PER_S** = forward velocity of the vehicle; units = meters per second (m/s).
- **HEADING_DEG** = heading of the vehicle (0 is north, 180 is south); units = degrees.
- **O2_MG_PER_L** = erroneous calculation of oxygen (correct calculation in the R code).
- **O2_VDC** = voltage from the oxygen sensor.
- **PAR_UE_PER_M2** = PAR; units = micro Einsteins per cm squared.
- **PAR_VDC** = voltage from the PAR sensor.
- **PITCH_DEG** = pitch of the vehicle (negative is pointing down, positive is upward); units = degrees.
- **ROLL_DEG** = roll of the vehicle to the left or right; units = degrees.
- **SALINITY_PPT** = salinity; units = practical salinity units.
- **TS_LATITUDE_DEG** = latitude of the ship's position in degrees.
- **TS_LONGITUDE_DEG** = longitude of the ship's position in degrees.
- **VERTICAL_VELOCITY_M_PER_S2** = vertical velocity of the vehicle; units = meters per second (m/s).
- **ALTITUDE_OK** = binary variable indicating whether or not the altitude reading is valid (1) or invalid (0).

Salinity data (725salinityinterp.csv, 726salinityinterp.csv):

- **dist** = distance along the transect; units = meters (m).
- **depth** = depth of the vehicle; units = meters (m).
- **salinity** = interpolated salinity value in that grid cell; units = practical salinity units.

Temperature data (725tempinterp.csv, 726tempinterp.csv):

- **dist** = distance along the transect; units = meters (m).
 - **depth** = depth of the vehicle; units = meters (m).
 - **temp** = interpolated temperature value in that grid cell; units = degrees Celsius.
-

(7) File name: LiteratureReview-PSEM.zip

(ZIP Archive (ZIP), 1.28 MB)

MD5:459707a6b0d37bb15c4f3928e7c29b49

URL: <https://datadocs.bco-dmo.org/data/305/DoIMICROBE/885637/1/data/LiteratureReview-PSEM.zip>

Description:

This folder contains a compilation of data from two open access databases used to make some calculations in the published manuscript. The folder also contains the data frame used to run the piecewise structural equation models (PSEM).

File names and descriptions:

COPEPOD2012_DoIPaper_locations.csv = has the location of the crustacean biomass estimates from the COPEPOD database. The column names are as follows: lon and lat of the estimate, Biomass of crustaceans in mg of Carbon per meter cubed, dataset = the dataset from which this estimate was acquired.

JeDI2015_DoIPaper_locations.csv = has the location of pelagic tunicate estimates from the JEDI database. These columns include the approximate latitude and longitude of the estimate.

psem_df.csv = contains all of the automated data with binned abundances derived from the CNN output and used in the piecewise structural equation model. Headers are similar to other files described.

Column names, descriptions, and units:

File "**COPEPOD2012_DoIPaper_locations.csv**":

- **lon** = longitude of the biomass estimate; units = degrees East.
- **lat** = latitude of the biomass estimate; units = degrees North.
- **Biomass.mgCm3** = biomass of crustaceans in milligrams of Carbon per meter cubed.
- **dataset** = the dataset from which this estimate was acquired.

File "**JeDI2015_DoIPaper_locations.csv**":

Columns C-H are all different levels of taxonomic accuracy for the estimate, including, phylum, class, order, family, genus, species.

- **numericDensity_number.m3** = the number per meter cubed.

- **Biomass.mgC.m3** = the biomass of the group in mg C per meter cubed.
 - **Loc** = the general area where the estimate is made.
 - **dataset** = an indicator of the estimate being from the JeDI dataset.
-

(8) File name: R-Scripts.zip

(ZIP Archive (ZIP), 32.87 KB)

MD5:b47e4d46a0ff0c1e3d3d2293e9c811f2

URL: <https://datadocs.bco-dmo.org/data/305/DoIMICROBE/885637/1/data/R-Scripts.zip>

Description:

This folder contains the R-scripts used to conduct all analyses synthesizing these datasets. For all of these scripts, comments are provided in the script to elaborate on what is happening as far as the inputs and outputs.

File names and descriptions:

DolBinning.R = Takes all individual doliolid data and merges it with the oceanographic sensor data to get counts per unit volume and patchiness statistics.

Dol-ManualvAuto.R = compares the manual and automated data collected in the Gulf of Mexico on July 26, 2016.

ORCA_dols_Mo_fileloading.R = reads in the files from the automated analysis and converts them to a common format for comparison.

PSEM_doliolid.R = code for running the piecewise structural equation model using the psem_df.csv file
