Cleaned species occurrence data from 2005 to 2025 from GBIF as part of a workflow to assemble species and community temperature indices for Port Fourchon, LA in 2006, 2016, 2022 and 2023

Website: https://www.bco-dmo.org/dataset/991175
Data Type: Synthesis
Version: 1
Version Date: 2026-01-08

Project
» CAREER: Integrating Seascapes and Energy Flow: learning and teaching about energy, biodiversity, and ecosystem function on the frontlines of climate change (Louisiana E-scapes)

Contributors	Affiliation	Role
Nelson, James	University of Georgia (UGA)	Principal Investigator
Leavitt, Herbert	University of Georgia (UGA)	Student
Thomas, Alexander	University of Georgia (UGA)	Student
York, Amber D.	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
This dataset is part of a workflow to assemble species temperature indices and community temperature indices for estuarine fauna of Port Fourchon, LA based on drop sampling studies from 2006, 2016, and 2022/2023. This step assembles GBIF occurrence records for all taxa detected in the Port Fourchon nekton surveys, cleans spatial artifacts, and delivers per-taxon CSVs ready for thermal niche extraction. The workflow expands harmonized taxon names into GBIF queries, submits authenticated bulk downloads, filters out problematic coordinates and land points, and documents all DOI citations needed for downstream analyses and reporting.

Table of Contents

- Coverage
- Dataset Description
  - Methods & Sampling
  - Data Processing Description
  - BCO-DMO Curation Notes
  - Problem Description
- Data Files
- Supplemental Files
- Related Publications
- Related Datasets
- Parameters
- Project Information
- Funding

Coverage

Location: Port Fourchon, Louisiana
Spatial Extent: **N**:29.168 **E**:-90.16 **S**:29.095 **W**:-90.244
Temporal Extent: 2005 - 2025

Dataset Description

This is one of four datasets that were produced with the "Fourchon Nekton Turnover Workflow" (v1.0.0, doi: https://doi.org/10.5281/ZENODO.18165331). The datasets and supplemental data produced by this workflow

have had minor modifications to enhance the interoperability of the data (See more in section "BCO-DMO Processing"). The workflow contains the exact formats of the data files produced and used by the workflow scripts. It also contains scripts, configurations, readme files, and input/output files for four stages listed below. Each workflow stage corresponds to a dataset in the "Related Datasets" section.

"Fourchon Nekton Turnover Workflow" steps with corresponding dataset IDs:

- "1_raw_data" = includes raw drop-sampling data corresponding to BCO-DMO dataset 991168 (doi: 10.26008/1912/bco-dmo.991168.1)
- "2_gbif_workflow" = includes GBIF species observation data corresponding to metadata in BCO-DMO dataset 991175 (doi: 10.26008/1912/bco-dmo.991175.1)
- "3_CTI_calculations" = includes community temperature index (CTI) data corresponding to BCO-DMO dataset 941250 (doi: 10.26008/1912/bco-dmo.941250.1)
- "4_species_of_interest" = includes the results of a species pool analysis identifying species of interest corresponding to BCO-DMO dataset 991182 (doi: 10.26008/1912/bco-dmo.991182.1)

The workflow release (v1.0.0) contains data and scripts used to run analyses and produce figures for publication Leavitt, H; Thomas, A; Doerr, J; Johnson, D; Nelson, J. (In press) Resilient Nekton Composition in the Face of Climate-Driven Foundation Species Shifts. Ecology. Accepted 2025-11-14

This dataset maintains the original source data licenses and rightsHolders specified at GBIF within the main dataset table 991175_v1_port-fourchon_cleaned-occurrence.csv. See the data user agreement and citation guidelines at GBIF:
GBIF Data User Agreement: https://www.gbif.org/terms/data-user
GBIF data usage and citation guidelines: https://www.gbif.org/citation-guidelines

Methods & Sampling

We downloaded all human observations and preserved specimens for each of the species observed from 2005 to the access date in 2025 from the Global Biodiversity Information Facility (GBIF)(Accessed on 6/6/2025). See query parameters in supplemental file "gbif_source_metadata.csv."

Species occurrence records were downloaded from GBIF using the Interface to the Global Biodiversity Information Facility API (rgbif API, https://github.com/ropensci/rgbif; doi: 10.5281/zenodo.1045299) and subjected to a multi-stage quality-control workflow prior to analysis. See Supplemental File "gbif_source_metadata.csv" for citations for each derived data query. Only post-2005 human observation records with valid geographic coordinates and no flagged geospatial issues were retained. Records lacking coordinates or returned as empty downloads were excluded.

Data Processing Description

We further filtered occurrences to retain only records with an occurrence status of “present” and removed fossil and captive/living specimen records. Spatial accuracy was enforced by removing records with coarse coordinate precision (>0.01 degrees) or large coordinate uncertainty, using a threshold of 1,000 m for most taxa and a relaxed threshold of 30,000 m for threatened taxa to avoid disproportionate data loss. Records with known placeholder or erroneous uncertainty values (e.g., 301, 999, 9999, 3036 m) and records with zero latitude or longitude were also removed.

To reduce common spatial artifacts, we applied automated coordinate cleaning using the *CoordinateCleaner* framework (Zizka et. al, 2023), excluding records falling within 2 km buffers of country centroids, capital cities, and known biodiversity institutions. Finally, duplicate records were removed based on identical longitude, latitude, species key, and dataset key combinations. The resulting cleaned datasets were saved for downstream analyses, and all GBIF download DOIs or citations were archived to ensure reproducibility and data provenance.

This dataset corresponds to Step "2_gbif_workflow" of the study's processing workflow 'Fourchon Nekton Turnover Workflow', doi: 10.5281/zenodo.18165331). See "Description" and "BCO-DMO Processing" sections for context about the relationship between the workflow files and the data as published at BCO-DMO.

Workflow README for Step "2_gbif_workflow" : Step 2: GBIF Downloads and Cleaning

Abstract

This step assembles GBIF occurrence records for all taxa detected in the Port Fourchon nekton surveys, cleans spatial artifacts, and delivers per-taxon CSVs ready for thermal niche extraction. The workflow expands harmonized taxon names into GBIF queries, submits authenticated bulk downloads, filters out problematic coordinates and land points, and documents all DOI citations needed for downstream analyses and reporting.

Purpose: fetch occurrence data for target taxa, apply coordinate/data-quality filters, and save cleaned per-taxon CSVs with citation logs.

Primary script

- gbif_download_BCODMO.R: annotated script that expands pooled taxa, submits GBIF downloads, cleans coordinates (CoordinateCleaner), and logs citations/DOIs.

Inputs

- ../1_raw_data/outputs/presence_pivot_merged_sp.csv: taxon list used to build download targets.
- Environment variables: GBIF_USER, GBIF_EMAIL, GBIF_PWD for GBIF API access.

Outputs

- gbif_downloads/clean_csvs/<taxon>_clean.csv: cleaned occurrences per taxon.
- gbif_citations.txt: citations/DOIs for all downloads.

Software

- R >= 4.3 with rgbif, tidyverse, terra, sp, sdmpredictors, CoordinateCleaner.

Run order

1. Ensure Step 1 outputs are present.
2. Set GBIF credentials in the environment.
3. Run gbif_download_BCODMO.R; verify cleaned CSVs and gbif_citations.txt.

BCO-DMO Curation Notes

Version 1 (2026-01-08):

Data from the processing workflow were prepared and published at BCO-DMO after reorganization into datasets with minor changes performed to meet the required conventions implemented by BCO-DMO designed for interoperability, standardization, and a variety of data access methods.

Main data table "991175_v1_port-fourchon_cleaned-occurrence.csv" was created by merging submitted files. Submitted folder "Fourchon Nekton Turnover Workflow/clean_csvs/" contained 75 csv files named with _clean.csv (one per species, a file per GBIF query, see gbif_source_metadata.csv). These were imported into the BCO-DMO data system, concatenated and additional columns added (GBIF_query_doi, AphiaID, LSID). Empty columns removed, columns subselected. The original format of the "Fourchon Nekton Turnover Workflow/clean_csvs/" is available in the Workflow package (doi: 10.5281/zenodo.18165331).

Submitted data files for this dataset correspond to the study's outputs in workflow (doi: 10.5281/zenodo.18165331) step 2:

The supplemental file "gbif_source_metadata.csv" contains additional information about the source GBIF data that was obtained by using the GBIF DOIs contained within workflow file "Fourchon Nekton Turnover Workflow/2_gbif_workflow/gbif_citations.txt" and adding additional information provided by the following APIs on 2026-01-09:

GBIF_API = "https://api.gbif.org/v1"
WORMS_API = "https://www.marinespecies.org/rest"

Problem Description

Note about values in dataset: eventDate and identifiedDate are typed as string type not date type due to inconsistent formats in the column (some report just year, some full datetime). identifiedDate includes quality issues (e.g. 1764-01-01 which has eventDate in 2005).

Note that emoji characters are included in observedBy, identifiedBy, and rightsHolder columns which were left as-is in the dataset (does not meet standard BCO-DMO conventions for allowed characters).

[table of contents | back to top]

Data Files

File	
991175_v1_port-fourchon_cleaned-occurrence.csv	(Comma Separated Values (.csv), 100.23 MB) MD5:3c9e62eefc2a62b75508dfbc9bbfab00
Primary data file for dataset ID 991175, version 1. This file concatenates workflow files "clean_csvs/_clean.csv." See "BCO-DMO Processing" section for minor differences between the workflow files and the BCO-DMO hosted file.	

[table of contents | back to top]

Supplemental Files

File	
gbif_source_metadata.csv	(Comma Separated Values (.csv), 39.20 KB) MD5:5a46cd44f647e172484afad124e6e3f5

Additional GBIF source data information based on the GBIF dois included in contents of workflow file gbif_ciations.txt (https://doi.org/10.5281/zenodo.18165331).

The information added here was provided by the following APIs:
GBIF_API = "https://api.gbif.org/v1"
WORMS_API = "https://www.marinespecies.org/rest"

This metadata table includes columns:

name_in_data, The scientific name as used in workflow file gbif_citations.txt
GBIF_query_doi, the GBIF query DOI as provided in gbif_citations.txt
GBIF_TAXON_KEY, The GBIF backbone taxonKey extracted from the stored GBIF download predicate that identifies the taxon used in the query
AphiaID, The WoRMS AphiaID resolved from the GBIF taxonKey by matching the GBIF canonical scientific name and authorship to WoRMS records
LSID, The WoRMS Life Science Identifier (LSID) associated with the AphiaID, providing a persistent identifier for the taxon in WoRMS
GBIF_totalRecords, The total number of occurrence records returned by the GBIF download at the time it was created
GBIF_downloadLink, The direct GBIF URL for downloading the occurrence data associated with the query
GBIF_data_type, The GBIF download request type reported in the download metadata (for example, occurrence)
GBIF_format, The file format of the GBIF download reported in the download metadata (for example, SIMPLE_CSV)
GBIF_access_date, The GBIF download creation timestamp, used here as the access date for the query results
GBIF_query_parameters, A human-readable representation of all query constraints used to generate the GBIF download, derived from the stored GBIF request predicate
GBIF_occurrence_citation, The reference text including the query DOI for the GBIF derived dataset used.

[table of contents | back to top]

Related Publications

Chamberlain, S., Oldoni, D., Geffert, L., Desmet, P., Barve, V., Ram, K., Blissett, M., Waller, J., McGlinn, D., Ooms, J., Steven (Siwei) Ye, Oksanen, J., Marwick, B., , John, Sumner, M., & Sriram. (n.d). ropensci/rgbif: rgbif (no version cited) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.1045299 https://doi.org/10.5281/zenodo.1045299
Software

Global Biodiversity Information Facility (2024) Occurrence download formats :: Technical Documentation. https://techdocs.gbif.org/en/data-use/download-formats
Methods

Leavitt, H; Thomas, A; Doerr, J; Johnson, D; Nelson, J. (In press) Resilient Nekton Composition in the Face of Climate-Driven Foundation Species Shifts. Ecology. Accepted 2025-11-14
Results

Zizka, A. (2017). CoordinateCleaner: Automated Cleaning of Occurrence Records from Biological Collections [dataset]. In CRAN: Contributed Packages. The R Foundation. https://doi.org/10.32614/cran.package.coordinatecleaner https://doi.org/10.32614/CRAN.package.CoordinateCleaner
Software

[table of contents | back to top]

Related Datasets

IsRelatedTo

Nelson, J. (2026) **Drop Sampling Data from Port Fourchon, Louisiana collected in 2006, 2016, 2022 and 2023.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2026-01-08 doi:10.26008/1912/bco-dmo.991168.1 [view at BCO-DMO]
Relationship Description: Datasets that are part of the same workflow (doi: 10.5281/zenodo.18165331) for a study to be published: Leavitt, H; Thomas, A; Doerr, J; Johnson, D; Nelson, J. (In press) Resilient Nekton Composition in the Face of Climate-Driven Foundation Species Shifts. Ecology.

Nelson, J. (2026) **Results of a species pool analysis identifying species of interest responding to climate changes in Port Fourchon, LA in 2006, 2016, 2022 and 2023.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2026-01-08 doi:10.26008/1912/bco-dmo.991182.1 [view at BCO-DMO]
Relationship Description: Datasets that are part of the same workflow (doi: 10.5281/zenodo.18165331) for a study to be published: Leavitt, H; Thomas, A; Doerr, J; Johnson, D; Nelson, J. (In press) Resilient Nekton Composition in the Face of Climate-Driven Foundation Species Shifts. Ecology.

Nelson, J., Leavitt, H., Thomas, A. (2026) **Community Temperature Index Calculations for Port Fourchon, Louisiana Drop Sampling data from 2006 to 2023.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2026-01-08 doi:10.26008/1912/bco-dmo.941250.1 [view at BCO-DMO]
Relationship Description: Datasets that are part of the same workflow (doi: 10.5281/zenodo.18165331) for a study to be published: Leavitt, H; Thomas, A; Doerr, J; Johnson, D; Nelson, J. (In press) Resilient Nekton Composition in the Face of Climate-Driven Foundation Species Shifts. Ecology.

IsPartOf

heleavitt. (2026). *heleavitt/Workflow-for-Leavitt_et_al_Resilient-Species-Nekton-Composition-in-the-Face-of: Workflow for Resilient Nekton Composition in the Face of Climate-Driven Foundation Species Shifts* (Version v1.0.0) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.18165331 https://doi.org/10.5281/zenodo.18165331

IsDerivedFrom

GBIF.org (n.d.). Global Biodiversity Information Facility: Home Page. https://www.gbif.org/

[table of contents | back to top]

Parameters

Parameter	Description	Units
GBIF_query_doi	DOI representing each GBIF database query. Each species was queried indevidually. DOI was generated by the rgbif API	unitless
gbifID	Unique GBIF key for the occurrence.	unitless
datasetKey	The UUID of the GBIF dataset containing this occurrence.	unitless
occurrenceID	An identifier for the dwc:Occurrence (as opposed to a particular digital record of the dwc:Occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the dwc:occurrenceID globally unique.	unitless
kingdom	The kingdom name (excluding authorship) for the kingdom from the GBIF backbone matched to this occurrence.	unitless
phylum	The phylum name (excluding authorship) for the phylum from the GBIF backbone matched to this occurrence.	unitless
class	The class name (excluding authorship) for the class from the GBIF backbone matched to this occurrence.	unitless
order	The order name (excluding authorship) for the order from the GBIF backbone matched to this occurrence.	unitless
family	The family name (excluding authorship) for the family from the GBIF backbone matched to this occurrence.	unitless
genus	The genus name (excluding authorship) for the genus from the GBIF backbone matched to this occurrence.	unitless
species	The species name (excluding authorship) for the species from the GBIF backbone matched to this occurrence.	unitless
infraspecificEpithet	The infraspecific name part of the species name from the GBIF backbone matched to this occurrence.	unitless
taxonRank	The taxonomic rank of the most specific name in the scientificName.	unitless

scientificName	The scientific name (including authorship) for the taxon from the GBIF backbone matched to this occurrence. This could be a synonym, see also acceptedScientificName.	unitless
verbatimScientificName	Scientific name as provided by the source.	unitless
verbatimScientificNameAuthorship	The authorship information for the dwc:scientificName formatted according to the conventions of the applicable dwc:nomenclaturalCode.	unitless
countryCode	The 2-letter country code (as per ISO-3166-1) of the country, territory or area in which the occurrence was recorded.	unitless
locality	The specific description of the place.	unitless
stateProvince	The name of the next-smaller administrative region than country (state, province, canton, department, region, etc.) in which the occurrence occurs.	unitless
occurrenceStatus	A statement about the presence or absence of a Taxon at a Location. For definitions, see the GBIF occurrence status vocabulary.	unitless
individualCount	The number of individuals present at the time of the Occurrence.	unitless
publishingOrgKey	The UUID of the organization which publishes the dataset containing this occurrence.	unitless
decimalLatitude	The geographic latitude (in decimal degrees, using the WGS84 datum) of the geographic centre of the location of the occurrence	unitless
decimalLongitude	The geographic longitude (in decimal degrees, using the WGS84 datum) of the geographic centre of the location of the occurrence.	decimal degrees
coordinateUncertaintyInMeters	The horizontal distance (in metres) from the given decimalLatitude and decimalLongitude describing the smallest circle containing the whole of the Location.	meters
coordinatePrecision	A decimal representation of the precision of the coordinates given in the decimalLatitude and decimalLongitude.	decimal degrees

elevation	Elevation (altitude) in metres above sea level. This is not a current Darwin Core term.	meters
elevationAccuracy	The value of the potential error associated with the elevation. This is not a current Darwin Core term.	meters
depth	Depth in metres below sea level. This is not a current Darwin Core term.	meters
depthAccuracy	The value of the potential error associated with the depth. This is not a current Darwin Core term.	meters
eventDate	The date-time during which an Event occurred. For occurrences, this is the date-time when the event was recorded. Not suitable for a time in a geological context.	unitless
day	The integer day of the month on which the Event occurred.	day
month	The integer month in which the Event occurred.	month
year	The four-digit year in which the event occurred, according to the Common Era calendar.	year
taxonKey	A taxon key from the GBIF backbone for the most specific (lowest rank) taxon for this occurrence. This could be a synonym, see acceptedTaxonKey.	unitless
speciesKey	A taxon key from the GBIF backbone for the species of thisoccurrence.	unitless
AphiaID	AphiaID of the species matched to this occurance. Used by WorMS as a unique species identifier to query and link data	unitless
LSID	Life Science Identifyer (LSID) of the species matched to this occurrence. This is used by the ZooBank registry as a unique digital signature for a taxa	unitless
basisOfRecord	The values of the Darwin Core term Basis of Record which can apply to occurrences.	unitless
institutionCode	The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record.	unitless

collectionCode	The name, acronym, coden, or initialism identifying the collection or data set from which the record was derived.	unitless
catalogNumber	An identifier (preferably unique) for the record within the data set or collection.	unitless
recordNumber	An identifier given to the dwc:Occurrence at the time it was recorded. Often serves as a link between field notes and a dwc:Occurrence record, such as a specimen collector's number.	unitless
identifiedBy	A list (concatenated and separated) of names of people, groups, or organizations who assigned the Taxon to the occurrence. Note (may contain non-standard characters, emojis, etc. as some are usernames from identification platforms like iNaturalist).	unitless
dateIdentified	The date on which the subject was determined as representing the Taxon. ISO 8601 Date.	unitless
license	License for the original source data at GBIF for the occurrence record. See GBIF data usage and citation guidelines: https://www.gbif.org/citation-guidelines	unitless
rightsHolder	Rights holder for the original source data at GBIF for the occurrence record. Note (may contain non-standard characters, emojis, etc. as some are usernames from identification platforms like iNaturalist).	unitless
recordedBy	A person, group, or organization responsible for recording the original occurrence. Note (may contain non-standard characters, emojis, etc. as some are usernames from identification platforms like iNaturalist).	unitless
typeStatus	A list (concatenated and separated) of nomenclatural types (type status, typified scientific name, publication) applied to the occurrence.	unitless
establishmentMeans	Statement about whether an organism or organisms have been introduced to a given place and time through the direct or indirect activity of modern humans.	unitless
lastInterpreted	The time this occurrence was last processed by GBIF's interpretation system Pipelines.	unitless
mediaType	The media type given as Dublin Core type values, in particular StillImage, MovingImage or Sound.	unitless

issue	A specific interpretation issue found during processing and interpretation of the record.	unitless

[table of contents | back to top]

Project Information

CAREER: Integrating Seascapes and Energy Flow: learning and teaching about energy, biodiversity, and ecosystem function on the frontlines of climate change (Louisiana E-scapes)

Website: http://www.nelsonecolab.net/career

Coverage: Saltmarsh ecosystem near Port Fourchon, LA

NSF Award Abstract:
Coastal marshes provide a suite of vital functions that support natural and human communities. Humans frequently take for granted and exploit these ecosystem services without fully understanding the ecological feedbacks, linkages, and interdependencies of these processes to the wider ecosystem. As demands on coastal ecosystem services have risen, marshes have experienced substantial loss due to direct and indirect impacts from human activity. The rapidly changing coastal ecosystems of Louisiana provide a natural experiment for understanding how coastal change alters ecosystem function. This project is developing new metrics and tools to assess food web variability and test hypotheses on biodiversity and ecosystem function in coastal Louisiana. The research is determining how changing habitat configuration alters the distribution of energy across the seascape in a multitrophic system. This work is engaging students from the University of Louisiana Lafayette and Dillard University in placed-based learning by immersing them in the research and local restoration efforts to address land loss and preserve critical ecosystem services. Students are developing a deeper understanding of the complex issues facing coastal regions through formal course work, directed field work, and outreach. Students are interacting with stakeholders and managers who are currently battling coastal change. Their directed research projects are documenting changes in coastal habitat and coupling this knowledge with the consequences to ecosystems and the people who depend on them. By participating in the project students are emerging with knowledge and training that is making them into informed citizens and capable stewards of the future of our coastal ecosystems, while also preparing them for careers in STEM. The project is supporting two graduate students and a post-doc.

The transformation and movement of energy through a food web are key links between biodiversity and ecosystem function. A major hurdle to testing biodiversity ecosystem function theory is a limited ability to assess food web variability in space and time. This research is quantifying changing seascape structure, species diversity, and food web structure to better understand the relationship between biodiversity and energy flow through ecosystems. The project uses cutting edge tools and metrics to test hypotheses on how the distribution, abundance, and diversity of key species are altered by ecosystem change and how this affects function. The hypotheses driving the research are: 1) habitat is a more important indirect driver of trophic structure than a direct change to primary trophic pathways; and 2) horizontal and vertical diversity increases with habitat resource index. Stable isotope analysis is characterizing energy flow through the food web. Changes in horizontal and vertical diversity in a multitrophic system are being quantified using aerial surveys and field sampling. To assess the spatial and temporal change in food web resources, the project is combining results from stable isotope analysis and drone-based remote sensing technology to generate consumer specific energetic seascape maps (E-scapes) and trophic niche metrics. In combination these new metrics are providing insight into species' responses to changing food web function across the seascape and through time.

This project is jointly funded by Biological Oceanography and the Established Program to Stimulate Competitive Research (EPSCoR).

This award reflects NSF's statutory mission and has been deemed worthy of support through evaluation using the Foundation's intellectual merit and broader impacts review criteria.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-2418012

[table of contents | back to top]