

# Ocean distribution of eulachon, a threatened anadromous fish, uncovered with environmental DNA

Owen R. Liu, Ole Shelton, Krista Nichols, Eric Ward, Elizabeth M. Phillips, Jen Zamon, Ana Ramón-Laca, Abigail Wells, Ryan P. Kelly

## Supplementary Information

All data and code required to recreate the Results in this study are publicly available online on Zenodo (<https://doi.org/10.5281/zenodo.14721414>). This supplement describes the files in the Zenodo archive, but interested readers should refer directly to the Methods section of the main text, as well as the modeling scripts themselves, for the most detailed step-by-step description of data construction, model fitting, and model validation. The main modeling script (`eulachon_sdms_final.Rmd`) is written in a Markdown format that can be directly rendered to html or pdf, and is extensively commented.

The files below are presented in alphabetical order, and organized into a few categories:

- Raw data: datasets in original, unaltered form
- Intermediate dataset: datasets compiled in various steps of data processing that are then further processed, summarized, or joined to create final datasets
- Final dataset: a clean dataset used in the final modeling
- Code: R and RMarkdown scripts used for data processing, modeling, and plotting

**Note on file paths:** All of the data and files used in this study are available in the Zenodo archive, and were originally run in an RStudio project environment using relative path names for pointing to files. Consequently, interested users of this code may have to ensure these file paths align with their own file organization system, or alter the file paths accordingly.

File Name	File Size	Description
<a href="#">calc_river_discharge.R</a> md5:ba2a90b64189f50854afef14a5d182fb	5.1 kB	Code: Create river discharge covariate data, from RC4USCoast
<a href="#">cmems_mod_glo_bgc_my_0.25_P1M-m_1709245106655.nc</a> md5:eeaf068a4040173b86dbb540f3096d05	106.3 kB	Raw data: Subset of GLORYS ocean model data, originally downloaded from the Copernicus Marine Data Store
<a href="#">cmems_mod_glo_bgc_my_0.25_P1M-m_1709245354119.nc</a> md5:3dc3bc65ab21d5515646d41587dc551a	157.4 MB	Raw data: Subset of GLORYS ocean model data, originally downloaded from the Copernicus Marine Data Store

<a href="#">cmems_mod_glo_phy_my_0.083deg_P1M-m_1708197434600.nc</a> md5:743f57c264fb4f8a5ff8c34eba0d5e01	36.8 MB	Raw data: Subset of GLORYS ocean model data, originally downloaded from the Copernicus Marine Data Store
<a href="#">cmems_mod_glo_phy_myint_0.083deg_P1M-m_1708197567621.nc</a> md5:74f2eb97853f8c4796433344e97ec341	29.4 MB	Raw data: Subset of GLORYS ocean model data, originally downloaded from the Copernicus Marine Data Store
<a href="#">create_5km_prediction_grid.R</a> md5:fd739994cf3604a0e7e15d0808f00ad6	5.4 kB	Code to create the 5km grid (25 square km cells) that is used for model projection and prediction maps
<a href="#">CTD meta 2021.csv</a> md5:ed2ad642d5eff0715f3f651209e842f	14.9 kB	Raw data: Metadata and location covariates for field samples collected in 2021
<a href="#">CTD_hake_data_10-2019.csv</a> md5:0d92876fd83642eae0946e6821201a34	104.4 kB	Raw data: Metadata and location covariates for field samples collected in 2019
<a href="#">d_obs_ready_for_sdmTMB.rds</a> md5:51e1a77f1925ea494dba995c6bde2dd6	4.0 MB	Processed dataset ready for sdmTMB modeling, with all candidate covariates matched and included. This is the dataframe that is used in models.
<a href="#">eDNA_glorys_bgc_matched.rds</a> md5:6bf1a8fb031efd8da4e6d55296ccf7c7	3.1 MB	Intermediate dataset: eDNA observations matched to biogeochemical GLORYS data
<a href="#">eDNA_glorys_matched.rds</a> md5:bab51e7c1c59d84d1df03093b998e162	3.3 MB	Intermediate dataset: eDNA observations matched to physics GLORYS data (temperature, salinity)
<a href="#">edna_krill_matched_5nn.rds</a> md5:0f9b6dfbdaf3fd1d8b2c3092149861fc	119.6 MB	Intermediate dataset: eDNA observations matched to 5 nearest neighbors from the krill/euphausiid observations dataset
<a href="#">edna_krill_metrics.rds</a> md5:314d25fca6f28c83c0d2ae5c40e9d46c	3.3 MB	Intermediate dataset: eDNA observations matched to all derived variables from the krill/euphausiid observations dataset

<a href="#">eDNA_river_influence_matched.rds</a> md5:7d95697cb4b2d012dc80794d97a66062	2.9 MB	Intermediate dataset: eDNA observations matched to the compiled river discharge data
<a href="#">eulachon qPCR 2019 and 2021 samples clean.rds</a> md5:4fbefab3b7b6999969a0f59531d24604	2.4 MB	Intermediate dataset: eDNA observations cleaned and joined across years, but with no covariates
<a href="#">eulachon qPCR 2019 and 2021 standards clean.rds</a> md5:e6ed10e5906d0ba4f8ab9b1d5305d029	69.0 kB	Final dataset: eDNA standards (samples of known concentration that are used to calibrate the relationship between DNA concentration and PCR cycle threshold)
<a href="#">eulachon qPCR 2019 joined cleaned 11_15_2023.rds</a> md5:a38c904f18af12835f0d5cde5467848c	2.2 MB	Intermediate dataset: eulachon eDNA field samples from 2019
<a href="#">eulachon qPCR 2019 standards cleaned 11_20_2023.rds</a> md5:d1ad53167f669377ea7ba6c7ebc3fa25	105.1 kB	Intermediate dataset: eulachon eDNA standards from 2019
<a href="#">eulachon qPCR 2021 joined cleaned.rds</a> md5:e9dc64226e0e0d1be8dd8ee3b59ba5c5	2.3 MB	Intermediate dataset: eulachon eDNA field samples from 2021
<a href="#">eulachon qPCR 2021 standards cleaned 11_20_2023.rds</a> md5:6a5d8f804810acd690a79d736961fc5e	77.0 kB	Intermediate dataset: eulachon eDNA standards from 2021
<a href="#">eulachon_sdms_final.Rmd</a> md5:5b856311b655ceffcee759fdcd479b7e	36.3 kB	<b>Code: Main modeling script and description</b>
<a href="#">fivekm_grid.tif</a> md5:d96f5b3fd82bd952e358cb3c6a184794	230.8 kB	Raw data: 5km project grid raster
<a href="#">Hake eDNA 2019 qPCR results 2020-01-04 standards.csv</a> md5:0e5b1c184884d4b39dd52f99d675fc9b	74.6 kB	Raw data: 2019 qPCR standards
<a href="#">Hake eDNA 2019 qPCR results 2021-01-04 results.csv</a> md5:5908b2326aa0b909f27e03876333501e	872.9 kB	Raw data: 2019 qPCR field samples

<a href="#">Hake eDNA 2019 qPCR results 2023-02-10 sample details.csv</a> md5:b92f7296a7aeb37aceb019f3f2eb139a	206.3 kB	Raw data: 2019 qPCR field samples metadata
<a href="#">Hake eDNA 2021 qPCR results 10.13.2023.csv</a> md5:185796436f2904c8a2078b0b89de9c60	771.4 kB	Raw data: 2021 qPCR field samples
<a href="#">Hake eDNA 2021 qPCR sample details 10.16.2023.csv</a> md5:f83b56258e4f3139f459050a6f05b7d8	232.9 kB	Raw data: 2021 qPCR field samples metadata
<a href="#">Hake eDNA 2021 qPCR standards 10.13.2023.csv</a> md5:ba4188a579a809de77450fc9c0a56772	65.9 kB	Raw data: 2021 qPCR standards
<a href="#">HakeSurvey_KrillNASC_0.5nmix10m_14-150m_2019-2023_AllCombine.csv</a> md5:49d6a8133b917618046fb515a7e7356d	27.8 MB	Raw data: 2019 and 2021 krill NASC data
<a href="#">HakeSurvey_KrillNASC_0.5nmix10m_cells_2019-2023.csv</a> md5:67af0d95c1eb46078716ac885bf8a270	40.5 MB	Raw data: 2019 and 2021 krill NASC data
<a href="#">join_eulachon_qPCR_data.R</a> md5:98166691773cc26db79f8d9e681b70ca	6.9 kB	Code: join cleaned 2019 and 2021 eulachon qPCR samples
<a href="#">krill data construction.Rmd</a> md5:42adf14f7a3b9063328047d01b0a90f8	25.6 kB	Code: build candidate covariates from the krill NASC dataset
<a href="#">krill_k2_k4_k5_matched_to_pred_grid.rds</a> md5:41a4f711bc376990c460f919924feb05	4.4 MB	Intermediate dataset: krill covariates joined to the 5km prediction grid
<a href="#">KrillKDE_2019_2023.csv</a> md5:82ee61c7896011b4035321530cfda556	48.9 MB	Raw data: 2019 and 2021 krill kernel density estimation from Phillips et al. 2022
<a href="#">match_GLORYS_bgc.R</a> md5:f32c2d4dc67ca3127cb65f38a4bf4cd1	2.9 kB	Code: match GLORYS data to qPCR field samples
<a href="#">match_GLORYS_temp_salinity.R</a> md5:46adc7874547b19936f68711a31b53d1	3.4 kB	Code: match GLORYS data to qPCR field samples

<a href="#">mclim_19902022_disc.nc</a> md5:1f13f6d955a1062bcde57efc68466f93	80.2 kB	Raw data: River discharge from RC4USCoast
<a href="#">plotting_utils.R</a> md5:816fbd7f5c07d81917d5c24b5f371b6d	13.7 kB	Code: Helper code to create consistent diagnostic figures, plots, and maps
<a href="#">prediction_grid_5km_3depths_no_covars.rds</a> md5:1f208e7641093057764290bf0e9ea579	1.3 MB	Intermediate dataset: 5km grid duplicated across 3 depth categories, used for projecting the model at multiple depths
<a href="#">prediction_grid_5km_sdmTMB_no_covars.rds</a> md5:57226117564134bb91fe45f22b407b8d	433.4 kB	Intermediate dataset: 5km grid (surface only)
<a href="#">prediction_grid_5km_sdmTMB_with_covars.rds</a> md5:28955cec5722414d0419035135387c3f	7.9 MB	Final dataset: 5km grid with all calculated covariates added
<a href="#">process 2019 eulachon data.Rmd</a> md5:28f24f644e9c98595f36d1f8347e5162	11.7 kB	Code to clean and QA/QC 2019 eulachon qPCR data
<a href="#">process 2021 eulachon data.Rmd</a> md5:2c3f7317b1060de4dd9c368085bdaddc	10.8 kB	Code to clean and QA/QC 2021 eulachon qPCR data
<a href="#">river_influence_grid_matched.rds</a> md5:14549549fece8023c065d59bbf014c35	300.1 kB	Intermediate dataset: river outflow metric matched to the 5km grid
<a href="#">rivers_mean_discharge_MarMay.rds</a> md5:4505884d7f3e912feed2b1ba434890db	1.6 kB	Intermediate dataset: river outflow metric summarized as March-May means
<a href="#">weighted_mean_NGDC_depths_for_5km_gridcells.csv</a> md5:bd321f38efe278b01be77ec9134d89d2	1.8 MB	Raw data: bottom depths associated with the 5km prediction grid