Tradeoffs between phenology and geography constraints in response to climate change across species life cycles

Data Policy Compliance

Identify any published data policies with which the project will comply, including the NSF OCE Data and Sample Policy as well as other policies that may be relevant if the project is part of a large coordinated research program (e.g. GEOTRACES).

The project investigators will comply with the data management and dissemination policies described in the NSF Award and Administration Guide (AAG, Chapter VI.D.4) and the NSF Division of Ocean Sciences Sample and Data Policy.

Pre-Cruise Planning

If the proposed project involves a research cruise, describe the cruise plans. (Skip this section if it is not relevant to your proposal.) Consider the following questions:

1. **How will pre-cruise planning be coordinated? (e.g. email, teleconference, workshop)**
2. **What types of sampling instruments will be deployed on the cruise?**
3. **How will the cruise event log be recorded? (e.g. the Rolling Deck to Repository (R2R) event logger application, an Excel spreadsheet, or paper logs)**
4. **Will you prepare a cruise report?**

The project does not involve research cruises.

Description of Data Types

Provide a description of the types of data to be produced during the project. Identify the types of data, samples, physical collections, software, derived models, curriculum materials, and other materials to be produced in the course of the project. Include a description of the location of collection, collection methods and instruments, expected dates or duration of collection. If you will be using existing datasets, state this and include how you will obtain them.

The types of data to be produced during the project include derived and simulation data.

Derived data

1. Long-term ichthyoplankton (fish eggs and larvae) and juvenile fish collections conducted by the National Oceanographic and Atmospheric Administration (NOAA). We will analyze synthesized data from five regional NOAA programs: 1. the Ecosystem-Fisheries-Oceanography Coordinated Investigations (Eco-FOCI) program for the Gulf of Alaska, since the 1970s, coverage mostly concentrated in Spring (March-June), eggs and larvae; 2. the California Cooperative Fisheries Oceanography Investigations (CalCOFI) for the southern California Current, since the 1950s, quarterly cruises, eggs and larvae, 3. the Estuarine and Ocean Ecology program for the Newport Hydrographic line (NH-line, North California Current), bi-weekly sampling since the mid 1996, eggs and larvae, 4. the Rockfish Recruitment and Ecosystem Assessment Survey (RREAS) for the Central California Current, Spring sampling since the mid 1980s, juveniles, 5. the pre-recruit survey for the North California Current (Oregon), spring sampling (May-June since the mid-2000s, pelagic juveniles. Our project team include NOAA collaborators from each of these regions who can help us to access and prepare the the data for the analyses described in the proposal. Repository: BCO-DMO

2. Station oceanographic CTD data collected during above mentioned ichthyoplankton surveys using a SeaBird SBE CTD package; processing done using SeaBird's SeaSave software; data will include standard environmental measurements (such as pressure, temperature, salinity, fluorescence, and oxygen). Additional station samples when available include Chlorophyll-a and zooplankton. File types for CTD: Raw (.con, .hdr, .hex, .bl) and processed and .cnv, .asc, .btl) ASCII files. File types for Chl-a and zooplankton when available: .csv. Repository: BCO-DMO.

Simulation data

Historical and downscaled climate projections for the northeast Pacific (i.e., from Baja California to Aleutian Islands) using the Regional Ocean Modeling System (ROMS). The ROMS domain will extend from Baja California to the Aleutian Islands with ~5km horizontal resolution and ~40 terrain-following vertical levels. The historical simulation will extend back to 1980 and the downscaled climate projections to 2100. Projections will be based on Shared Socio-economic Pathway 5-8.5 from three earth system models. The three earth system models will be selected so as to provide a representative scope of projected ocean conditions across all models in the CMIP6 ensemble. This approach has successfully worked for the California Current and other marine ecosystems with CMIP5 members GFDL-ESM2M (Dunne et al., 2013), IPSL-CM5A-MR (Seferian et al., 2013) and HadGEM2-ES (Collins et al., 2011), as illustrated in Muhling et al. (2017), Asch et al. (2018) and Lotze et al. (2019). These earth system models will be used to prescribe ROMS boundary conditions and surface atmospheric forcing via a time-varying delta downscaling method. The ROMS output will consist of daily physical fields (x,y,z,t) for sea surface height, temperature, salinity and ocean currents, amounting to ~?? TB of data. Model input (initial/boundary conditions and surface forcing) and output will be saved and archived on a dedicate storage server at UCSC that will be purchased in the first year of the project.

Data and Metadata Formats and Standards

Identify the formats and standards to be used for data and metadata formatting and content. Where existing standards are absent or deemed inadequate, these formats and contents should be documented along with any proposed solutions or remedies. Consider the following questions:

1. **Which file formats will be used to store your data?**
2. **What type of contextual details (metadata) will you document and how?**
3. **Are there specific data or metadata standards that you will be adhering to?**
4. **Will you be using or creating a data dictionary, code list, or glossary?**
5. **What types of quality control will be used? How will data quality be assessed and flagged?**

The synthesized ichthyoplankton and in-situ oceanographic data used for analyses will be placed in comma-separated-values in plain ASCII format, which are readable over long time periods by different software packages. The final data file will contain dates for each observation (time, day, month and year), geographic coordinates, oceanographic variables (temperature, salinity, etc), and the abundance of ichthyoplankton or juvenile of fish standardized either by volume filtered or area swept. The final data product will occupy 5-10 GB, once oceanographic information is also included. Quality flags will be assigned according to the CalCOFI ((https://oceanview.pfeg.noaa.gov/CalCOFI/calcofi_info.html) and Eco-FOCI (https://www.ecofoci.noaa.gov/field-operations) field operation manuals. Metadata will be prepared in accordance with BCO-DMO conventions (i.e. using the BCO-DMO metadata forms) and will include detailed descriptions of collection and analysis procedures.

ROMS is compliant with the NetCDF Climate and Forecast (CF) 1.0 metadata conventions. As such, all input and output files associated with ROMS historical simulation and downscaled climate projections are already following an accepted standard for processing and sharing model information. The conventions define metadata that provide a definitive description of what the data in each variable represents, and the spatial and temporal properties of the data. This enables users of data from different sources to decide which quantities are comparable, and facilitates building applications with powerful extraction, regridding, and display capabilities." (http://cf-pcmdi.llnl.gov).

Feng et al. (2019) recently developed a metadata checklist to maximize the reproducibility of ecological niche models (https://github.com/shandongfx/ENMchecklist). We plan to utilize this tool to ensure the reproducibility of our results. This checklist will be appended as a supplemental document to all publications developed based on this project.

Data Storage and Access During the Project

Describe how project data will be stored, accessed, and shared among project participants during the course of the project. Consider the following:

1. **How will data be shared among project participants during the data collection and analysis phases? (e.g. web page, shared network drive)**
2. **How/where will data be stored and backed-up?**
3. **If data volumes will be significant, what is the estimated total file size?**

The investigators will store project derived and simulated data (including spreadsheets, ASCII files, NetCDF, images) on laboratory computers that are backed up by the respective University's central IT organization. OSU has G Suite account for cloud services, data storage, and sharing among project investigators. Personal computers in all laboratories are backed up daily using Apple Time Machine to an onsite external hard drive, and weekly to an offsite hard drive. ECU uses Crash Plan to regularly back up all data on lab computers to the cloud.

UCSC will purchase a dedicated storage server in the first year of the project to save and archive ROMS model output (historical simulations and downscaled climate projections). All investigators will have accounts on the storage server to facilitate access and sharing of model output. Model configurations and forcing files will be regularly backup on other local storage servers so that all ROMS simulations can be easily recreated.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution

Describe mechanisms for data access and sharing, and describe any related policies and provisions for re-use, re-distribution, and the production of derivatives. Include provisions for appropriate protections of privacy, confidentiality, security, intellectual property, or other rights or requirements. Consider the following:

1. **When will data be made publicly available and how? Identify the data repositories you plan to use to make data available.**
2. **Are the data sensitive in nature (e.g. endangered species concerns, potential patentability)? If so, is public access inappropriate and how will access be provided? (e.g. formal consent agreements, restricted access)**
3. **Will any permission restrictions (such as an embargo period) need to be placed on the data? If so, what are the reasons and what is the duration of the embargo?**
4. **Who holds intellectual property rights to the data and how might this affect data access?**
5. **Who is likely to be interested in re-using the data? What are the foreseeable re-uses of the data?**

Data sets derived and simulated by the science party will be made available through the BCO-DMO data system within two-years from the date of collection. The project investigators will work with BCO-DMO data managers to make project data available online in compliance with the NSF OCE Sample and Data Policy. Data, samples, and other information collected under this project can be made publically available without restriction once submitted to the public repositories.

Data produced by this project may be of interest to physical, chemical, and biological oceanographers, and climate scientists interested in species distribution models (SDMs). We will adhere to and promote the standards, policies, and provisions for data and metadata submission, access, re-use, distribution, and ownership as prescribed by the BCO-DMO Terms of Use (http://www.bco-dmo.org/terms-use).

Plans for Archiving

Describe the plans for long-term archiving of data, samples, and other research products, and for preservation of access to them. Consider the following:

1. **What is your long-term strategy for maintaining, curating, and archiving the data?**
2. **What archive(s) have you identified as a place to deposit data and other research products?**

R2R will ensure that the original underway measurements are archived permanently at NCEI and/or NGDC as appropriate. BCO-DMO will also ensure that project data are submitted to the appropriate national data archive. The PI will work with R2R and BCO-DMO to ensure data are archived appropriately and that proper and complete documentation are archived along with the data.

Roles and Responsibilities

Describe the roles and responsibilities of all parties with respect to the management of the data. Consider the following:

1. **If there are multiple investigators involved, what are the data management responsibilities of each person**
2. **Who will be the lead or primary person responsible for ultimately ensuring compliance with the Data Management Plan?**

Each PI will be responsible for sharing his/her subset of data among the project participants in a timely fashion. Lorenzo Ciannelli will be responsible for analyzing the ichthyoplankton and in-situ oceanographic sampling data from the Gulf of Alaska region, while Rebecca Asch will be responsible for analyzing the ichthyoplankton and in-situ oceanographic sampling data from the California Current region. Mercedes Pozo will oversee the ROMS projections. The Lead PI, L. Ciannelli, will coordinate the overall data management and sharing process and will submit the project data, including derived and simulated data, and metadata to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) who will be responsible for forwarding these data and metadata to the appropriate national archive.