Plan Overview

A Data Management Plan created using DMPonline

Title: Improving estimates of climate-driven body size changes and range shifts in fishes by accounting for fine-scale spatial heterogeneity

Creator:Max Lindmark

Principal Investigator: Max Lindmark

Data Manager: Max Lindmark

Project Administrator: Max Lindmark

Contributor: Alessandro Orio, Sean C. Anderson

Affiliation: Swedish University of Agricultural Sciences

Funder: Swedish Research Council

Template: Swedish Research Council Template

ORCID iD: 0000-0002-3841-4044

Project abstract:

Fish are generally predicted to “shrink” (grow faster but attain smaller adult body sizes) when the oceans warm. However, observations from natural systems show high variability among species, and this could be because body size responses are usually evaluated as averages over large spatial scales. This overlooks spatial heterogeneity in temperature, warming trends, and the distribution of species. Species also tend to shift their geographical range to track temperatures they are adapted to, meaning they do not necessarily experience warming. This project aims to advance our understanding of climate impacts on body size and the spatial distribution of fishes by explicitly accounting for spatial heterogeneity and local-scale responses using state-of-the-art geostatistical models. We will quantify the degree of warming species experience and contrast that to the overall ecosystem warming, evaluate the role of fishing and other potentially confounding variables, and estimate at which spatial scales body size trends are correlated. The project starts with local and regional case studies to resolve specific interim aims and ends with a global analysis of size and range shifts using data collated throughout the project. This will provide a unique test of the generality of body size changes in the ocean due to climate change on a global scale, and how range shifts may allow species to avoid warming, which is crucial for improving climate impact assessments and fisheries management.

ID: 122769

Start date: 01-01-2023

End date: 31-12-2026

Last modified: 05-06-2023

Grant number / URL: 2022-01511

Improving estimates of climate-driven body size changes and range shifts in fishes by accounting for fine-scale spatial heterogeneity

General Information

Project Title

Improving estimates of climate-driven body size changes and range shifts in fishes by accounting for fine-scale spatial heterogeneity

Project Leader

Max Lindmark

Registration number/corresponding, date and version of the data management plan

2022-01511, 2023.06.01, V2

Version

v.1

Date

24-04-2023

Description of data – reuse of existing data and/or production of new data

How will data be collected, created or reused?

We will use trawl survey data stemming from annual or bi-annual scientific surveys from regions in the Atlantic and Pacific Ocean that are publicly available (e.g., on the ICES database DATRAS for northeast Atlantic). Environmental data will be acquired from the same surveys and as gridded predictions from regional oceanographic models, also freely available online (e.g., the datahost Copernicus, https://www.copernicus.eu/en/access-data)

What types of data will be created and/or collected, in terms of data format and amount/volume of data?

We will process numerical trawl data and gridded environmental data from the abovementioned databases for analysis. This typically includes cleaning, filtering and summarizing data. We expect the processed data ready for analysis will be in the order of ~few MBs in size per species.

Documentation and data quality

How will the material be documented and described, with associated metadata relating to structure, standards and format for descriptions of the content, collection method, etc.?

Information on content and collection is available on each trawl survey data base, which will be cited in our project. Metadata for specific datasets will be included as downloaded in the project folder. Processing of data is documented in annotated R-scripts detailing the steps for downloading and processing raw data from the databases, following largely the Tidyverse style guide for coding to facilitate reproducability (https://style.tidyverse.org/). Readme files in the project directories will explain project structure and file structures, wheras specific R-scripts provide more detailed information. Git, an open source software, will be used for version control. R-scripts and data will be hosted on GitHub repositories.

How will data quality be safeguarded and documented (for example repeated measurements, validation of data input, etc.)?

Data on databases are already quality checked before submission by each member organization. During processing, we will implement additional standard quality controls (valid data entries, including positive length, age and catch data coordinates within the survey domain etc.), which will be documented in annotated R-scripts.

Storage and backup

How is storage and backup of data and metadata safeguarded during the research process?

Quality controlled raw data exist on public databases. Code to processes these data and output data prepared for analysis is hosted on GitHub (an Internet hosting service for software development and version control using Git). Locally, code and data are stored on a Dropbox server (professional account) for continuous synchronisation before pushed to GitHub.

How is data security and controlled access to data safeguarded, in relation to the handling of sensitive data and personal data, for example?

This project does not handle such data.

Legal and ethical aspects

How is data handling according to legal requirements safeguarded, e.g. in terms of handling of personal data, confidentiality and intellectual property rights?

This project does not handle such data.

How is correct data handling according to ethical aspects safeguarded?

This project does not handle such data.

Accessibility and long-term storage

How, when and where will research data or information about data (metadata) be made accessible? Are there any conditions, embargoes and limitations on the access to and reuse of data to be considered?

All information is available from Day 1, as the code and data are hosted on public HitHub repositories. There will be no limitations, embargoes, or conditions for reuse of code and data. The data downloaded from databases are allowed to be hosted, shared and archived as long as they are cited.

In what way is long-term storage safeguarded, and by whom? How will the selection of data for long-term storage be made?

Before publication in Scientific Journal Articles, we will archive the GitHub repositories on Zenodo (Zenodo is a general-purpose open repository developed under the European OpenAIRE program and operated by CERN) and link the Zenodo DOI to the GitHub project and the journal article. I.e., it is permanently safeguarded when the GitHub repository is given a persistent digital object identifier (DOI) via Zenodo. We aim to include the data needed to reproduce the analysis, i.e., going from raw data to scientific results, nothing more and nothing less. We will then pass the DOI to SLU Archive for archiving also at SLU.

Will specific systems, software, source code or other types of services be necessary in order to understand, partake of or use/analyse data in the long term?

The project uses the open-source R (a programming language for statistical computing and graphics supported by the R Core Team and the R Foundation for Statistical Computing), which also is the most used software for data processing and analysis in our field. Each project repository will have a .Readme file with instructions for reproducing the results and which versions of the software and additional packages are used.

How will the use of unique and persistent identifiers, such as a Digital Object Identifier (DOI), be safeguarded?

We will use Zenodo to get a DOI (the GitHub repository at the time of journal article publication) for each work package.

Responsibility and resources

Who is responsible for data management and (possibly) supports the work with this while the research project is in progress? Who is responsible for data management, ongoing management and long-term storage after the research project has ended?

Max Lindmark as the PI of the project and lead author on all papers, and thus also responsible for data management and data file system.

What resources (costs, labour input or other) will be required for data management (including storage, back-up, provision of access and processing for long-term storage)? What resources will be needed to ensure that data fulfil the FAIR principles?

No extra resources are needed. The workflow of open code for all steps, from reading data from databases, to processing and analysis, to archiving on Zenodo, is already integrated in common workflow within the project.