Argo Ocean Data Analysis

Analyzing the impact of ocean water temperature and salinity on fish population near the Gulf of St-Laurent.

# Project Summary

1. Summary

Our proposed project is to use Argo data (a data set of ocean floats that collect variables such as salinity and temperature in oceans around the world), map the change of salinity and temperature in the Estuary and Gulf of St. Lawrence, and identify potential correlations to fish populations totals in the same area.

*Why study the Estuary and Gulf of St. Lawrence?*

The Gulf of St. Lawrence is an incredibly diverse and complex marine and estuary ecosystem and is one of the largest of its kind in the world. The area is made up of freshwater from the Canadian Shield, the Great Lakes basin and the St. Lawrence River system emptying out into the Atlantic Ocean, where it combines with the cold Labrador Current from the Arctic and the warm Gulf Stream from the tropics. All of these currents merge in a semi-enclosed and mostly shallow area, creating the perfect conditions for incredible diversity of life. The Gulf of St. Lawrence is the most important source of fish on the Atlantic side for the commercial fishing industry in both the U.S. and Canada, both top exporters of fish and seafood in the world[[1]](#footnote-0). By studying and reporting about this ecosystem and the species that depend on it we hope to better inform others about its critical importance.

1. Questions

* How has ocean properties such as temperature, salinity, etc., changed over a period of ten years in the Estuary and Gulf of St. Lawrence?
* How have fish populations (pelagic primarily, but also possibly demersal fish populations as well) changed in the Estuary and Gulf of St. Lawrence within the same time period?
* Is there a meaningful correlation between temperatures/salinity/other factors and population that merits further study?

# Data Sets

1. Data Set 1

**Name**: Argo float data and metadata from Global Data Assembly Centre (Argo GDAC) –Atlantic Ocean, 2009-2019

**Short description**: Argo is an international program that collects information about Earth’s oceans using a fleet of robotic instruments that drift with the ocean currents and move between the surface and mid-water level. These floats capture information such as salinity, temperature, oxygen levels and depth. For the scope of this project, we will use the Atlantic Ocean dataset from 2009 to 2019.

**Size**: approx. 20 GB

**Location**: <ftp://usgodae.org/pub/outgoing/argo>

**Format**: netCDF files

**Access method**: FTP and from relational database post processing.

1. Data Set 2

**Name**: Pelagic fish species abundance in the Estuary and Gulf of St. Lawrence between 2009 and 2018

**Short Description**: The Canadian Department of Fisheries and Ocean has been conducting annual multidisciplinary surveys of the Northern and Southern Gulf of St. Lawrence to capture information on groundfish and invertebrates’ abundance, spatial distribution and diversity. The pelagic species represented in the dataset are: Arctic Cod, Atlantic Argentine, Atlantic Herring, Atlantic Mackerel, Atlantic Soft Pout, Capelin, Lumpfish, Pollock, Rainbow Smelt, Sand Lances, Silver Hake, Threespine Stickleback and White Barracudina.

**Size**: approx. 1,800 lines of data (0.2MB)

**Location**: <https://open.canada.ca/data/en/dataset/f1fc359c-0ed1-4045-a421-adef2497b68d>

**Format**: CSV/REST API

**Access method**: HTTP download and from relational database post processing.

1. Data Set 3

**Name**: Gulf of St. Lawrence shapefile

**Short Description**: Shapefile of the Gulf and Estuary of St. Lawrence for visualization purposes.

**Size**: <1MB

**Location**: https://www.marineregions.org/gazetteer.php?p=details&id=4290

**Format**: Shapefile

**Access method**: Download and open and manipulate with Python’s shapely and/or geopandas.

# Data Manipulation

1. Initial processing:

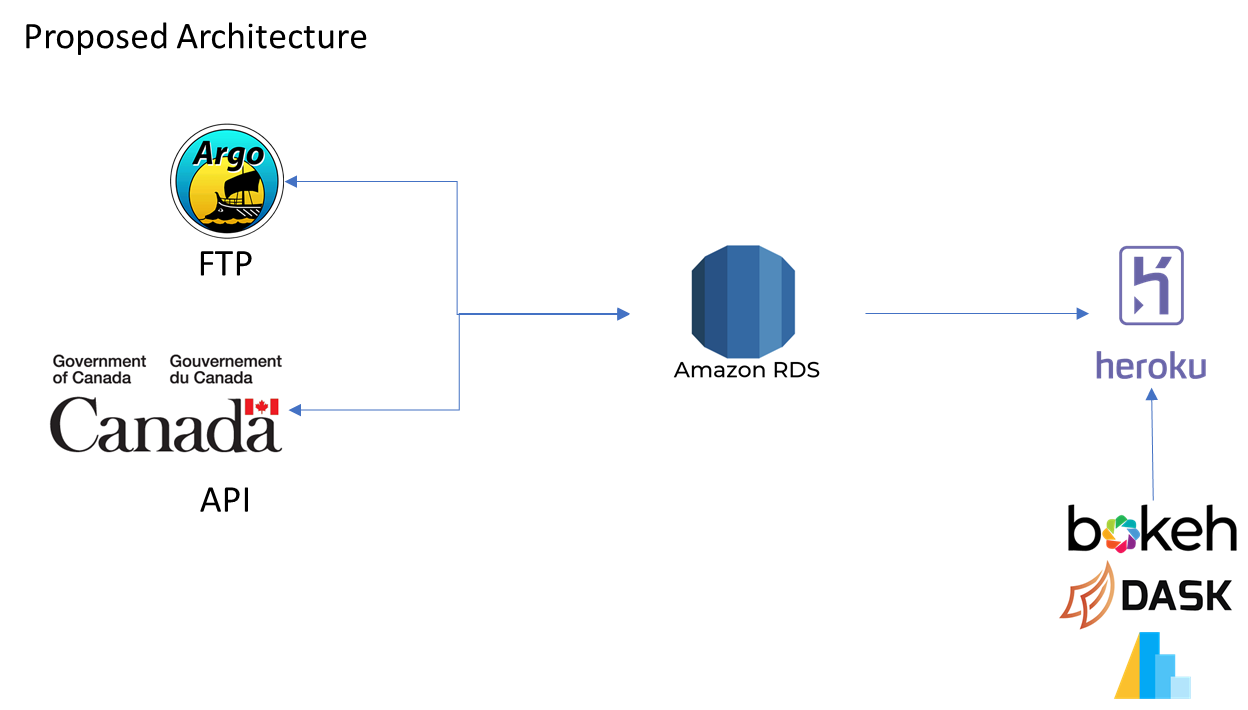
* The Argo Atlantic Ocean data will need to be queried and filtered to retrieve data from close to the Estuary and Gulf of St. Lawrence. NetCDF files will need to be transformed into CSV files (leveraging Python’s netCDF4 library) to be stored in a table in a relational database. We plan on using Amazon RDS (Postgres) for our project, with Heroku postgres as a backup if RDS does not meet our needs.
* The pelagic fish data will need to be downloaded (as a zip file), cleaned, and then uploaded as a table in our relational database.

1. Data manipulation:

* We will first need to define an outer boundary to our dataset as the Argo data set is not limited to the Estuary and Gulf of St. Lawrence. Once that is defined, we can filter both Argo and fish data within these boundaries and further subdivide as necessary. We can use predefined shapefiles from existing map sources (e.g. Canada’s coastal map data) or use the limits of our smaller data sets to define boundaries. Shapefiles typically rely on latitude and longitude to define map areas.
* Fish and float data are not necessarily in the same exact location and depth, so we will need to define shapefiles and depth bins on which we can join our datasets.
* Since the fish data is only provided for a subset of a year, we cannot compare it with the entirety of our ocean data for the same given year. We will need to compute annual average temperature, salinity, oxygen level and their changes compared to the previous year. We can also compute the year-to-year increase/decrease of populations in the pelagic fish data.
* Since we may have areas where we have fish data but no ocean property data or vice versa, we may also need to interpolate data for these areas based on data values from adjacent areas.
* Then we can join our data sets on shapefiles and depth bins.

1. Output:

* With the joined dataset, we can create several maps to compare variations in ocean data at different depths with fish populations. To this end, we plan on using a combination of Dask and Bokeh (or Altair) to build an interactive dashboard to visualize the impact of ocean data variation on each fish population over time (i.e. do increases/decreases in temperature, salinity and oxygen level impact pelagic fish population, and if so, do they impact them in the same way?). This dashboard would be made publicly available via the Heroku platform, which we can connect directly to our Github repository.



# Visualizations

Maps are a must when dealing with geographic data and would be included in our final presentation and report. An interesting visualization would be to map overall changes in temperature and salinity over the 10-year period (2009-2019) our dataset covers with overall changes in fish population to visualize trends and provide correlation coefficients. For example, a one-year variation of +0.01°C may not have a significant impact, but an overall increase of +0.5°C over 10 years may drastically impact the Arctic Cod population. This would answer one question as to whether fish populations are correlated to specific ocean variables, and if so, how strong the correlation may be. Our limitation on visualizations will be largely related to the visualization package we will rely on.

# Contributions

Team members will split the work evenly across all portions of the project – both Claire-Isabelle Carlier and Sharon Sung have approximately the same level of experience with all parts of the project and will employ an asynchronous pair-programming setup where we will jointly work and review all project parts together. Work is shared via a GitHub project folder (located here: <https://github.com/CICarlier/argo_ocean_data_analysis>).

1. <https://www.dfo-mpo.gc.ca/stats/trade-commerce/world-mondial/export/wxv1517-eng.htm> [↑](#footnote-ref-0)