

# Distribution of Six Aquatic Species in the Southern Gulf of Saint Lawrence, Canada

Statistics for Data Science

David Fishman<sup>1</sup>

<sup>1</sup>dayjfish@gmail.com

April 2024

# 1 Summary

## 1.1 Data source

Canadian oceans provide an important source of livelihood to Canadian in addition to fulfilling many important ecological roles. The Canadian Government has a mandate to ensure the Oceans remain ecologically healthy as well as economically viable. The Department of Fisheries and Oceans (DFO) is tasked with conducting regular stock assessments surveys to monitor the populations of fish and other aquatic organisms, including crustaceans. These surveys are carried out on either Canadian Coast Guard Science vessels or contracted out through privately owned fishing vessels. Across Canada, fishing is conducted using a variety of gear types. In the southern Gulf of St. Lawrence, fishing is primarily carried out using bottom trawls. In addition to biomass data, oceanographic measurements, such as water salinity, temperature and dissolved oxygen, are often collected. The ultimate goal of these surveys is to estimate population indices for species of ecological and economic interest. Many of the data from stock assessment surveys have been made available via the Open Government Portal [2].

The importance of having robust models for predicting the distribution of aquatic organisms cannot be overstated. From a resource extraction point of view, modelled distributions can help direct fishing efforts; resulting in more efficient fisheries. From an environmental conservation point of view, distributions modelling can help to identify which geographic areas should be targets for conservation efforts. The goal of this analysis is to compute models that can be used to predict the probability of occurrences of species of interest in the Southern Gulf of St. Lawrence (sGSL). Specifically, I employ a logistic regression approach, using latitude, longitude and water depth (i.e., elevation) to predict the probability of occurrence of the following six species:

- American plaice (*Hippoglossoides platessoides*)
- Atlantic cod (*Gadus morhua*)
- Atlantic herring (*Clupea harengus*)
- Redfish unidentified (*Sebastes sp.*)
- American lobster (*Homarus americanus*)
- Snow crab (*Chionoecetes opilio*)

The primary dataset used originating from the Government of Canada’s Open Government Portal [2], is a dataset called **NAFO Division 4T ground-fish research vessel trawl survey (September Survey) dataset** [3]. This dataset contained information biological information (i.e., species caught, specimen counts and specimen weights), fishing information (i.e., gear type used, fishing vessel) and geospatial information (i.e., latitude and longitude). Bathymetric data for the sGSL was obtained from the **General bathymetric Chart of the Oceans (GEBCO)** [4] website.

## 1.2 Analysis

The material in the course modules on Logistic Regression provided a useful starting point for this analysis. This analysis had a strong spatial component and it took me lots of time to figure out how to properly re-project a geographical dataset from one coordinate reference system to another. The operations on the large data arrays for elevation and probability of occurrence were done using the **Xarray** python package [1]. This was my first time using this package and while it was very impressive and comprehensive, it was also somewhat overwhelming. It took me a long time to figure out how to extract the data from a data array, manipulate it and create a new data array. A combination of persistence and stack overflow helped me to overcome these challenges.

## 1.3 Conclusions

The distribution models of the six species demonstrated different patterns of occurrence; some being more heterogeneous than others. Species like American plaice and Atlantic cod were generally observed to be ubiquitous in the sGSL, especially in the shallower waters. For those species, there is a much lower probability of occurrence in the deeper waters of the St. Lawrence seaway. The modeled distributions of Atlantic herring and Snow crab seemed to be more influenced by latitude-longitude than by elevation. Redfish and American lobster had sharp signals of presence-absence, i.e., a relatively narrow range of inflection in the sigmoid  $\text{logit}(p)$  vs.  $p$  plots. The distribution of both of these species seemed heavily driven by elevation, albeit inversely from one another. Redfish were predicted to occur primarily in deep waters and lobster were predicted to occur primarily in shallow waters. This in fact matches reality very closely, as can be observed by the overlaid points on the heatmaps in Figure 1 below.

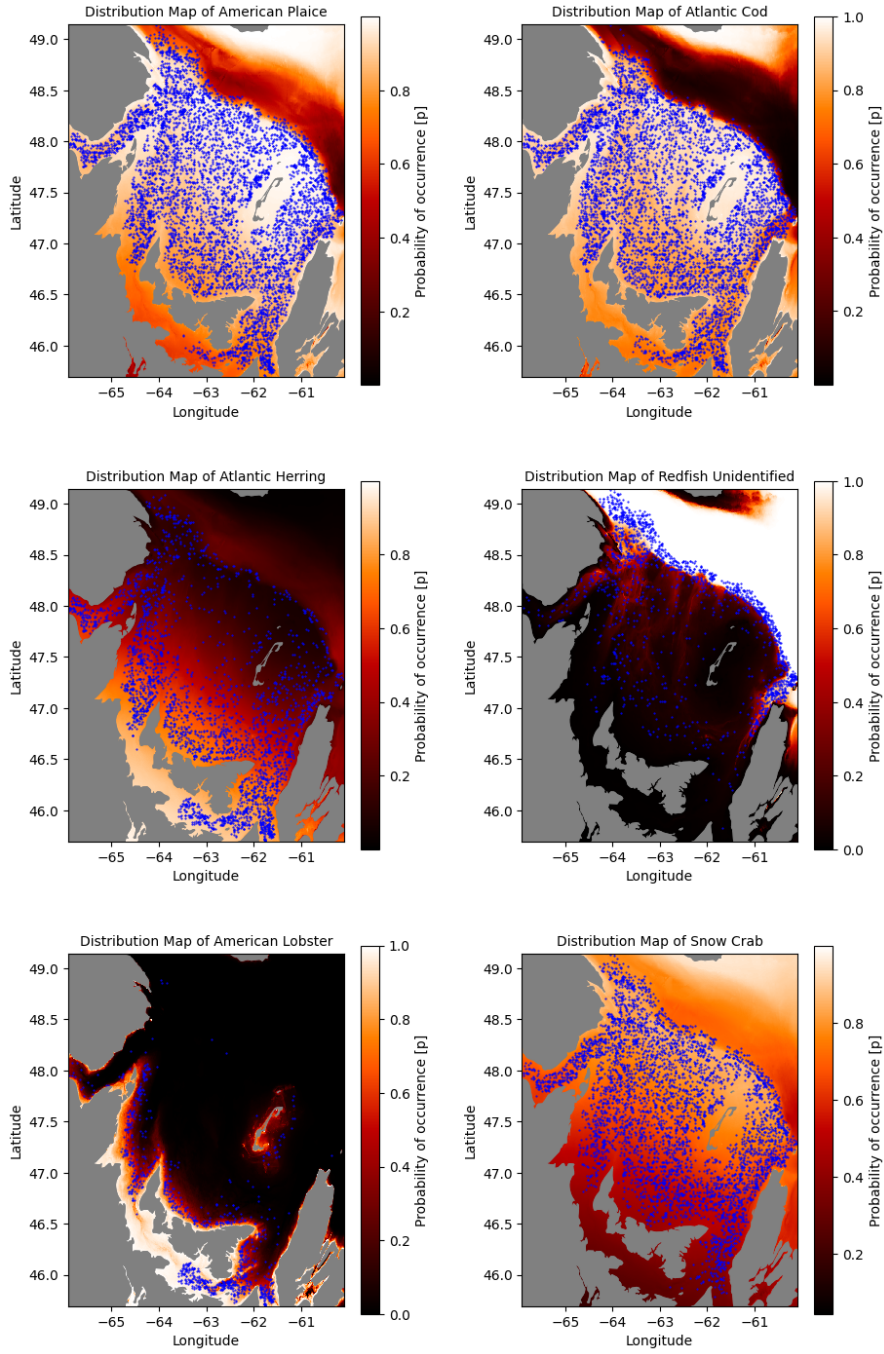


Figure 1: Distribution maps for all six species. Each map displays the inferred estimated probability for each point on the map and is depicted using a heatmap themed color mapping. The true occurrences for each species are overlain in order to provide a visual sense of model's performance.

## References

- [1] NumFOCUS. Xarray. <https://docs.xarray.dev/en/stable/>, 2024. Accessed on 2024-04-13.
- [2] Government of Canada. Open government portal. <https://open.canada.ca/>, 2023. Accessed on 2023-12-11.
- [3] Government of Canada. Nafo division 4t groundfish research vessel trawl survey (september survey) dataset. <https://open.canada.ca/data/en/dataset/1989de32-bc5d-c696-879c-54d422438e64>, 2024. Accessed on 2024-04-13.
- [4] The General Bathymetric Chart of the Oceans (GEBCO). Gebco gridded bathymetry data download. <https://download.gebco.net/>, 2024. Accessed on 2024-04-13.