Blue Ocean Gear

Buoy Location Prediction and Temperature Analysis

Aurora Peng, Brinda Sapra, Hazel Chui, Zoey Jiao
December 9th, 2022

Introduction

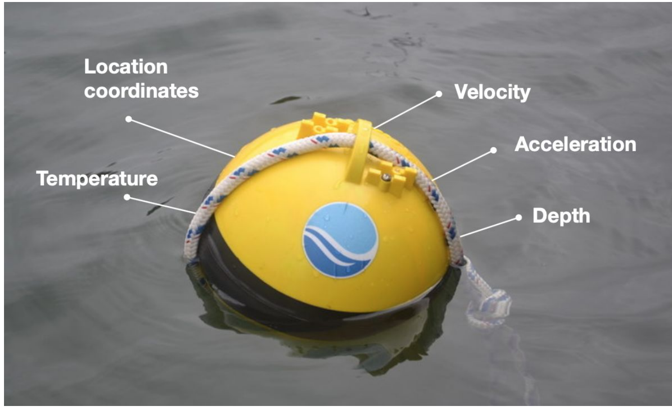
Background on BOG

- Ghost fishing gear one of the most harmful forms of marine plastic as it catches wildlife, subjecting them to a slow and painful death through exhaustion and suffocation.
- Fishing gear accounts for **10%** of ocean debris: between **500,000 to 1 million** tons of fishing gear are discarded or lost in the ocean every year.
- Blue Ocean Gear, founded in California, has developed a smart buoy system that tracks lost fishing gear in the ocean.

Introduction

How fishermen use BOG Smart Buoys

- Trap (hook, pot)
- Fishery (Massachusetts, Maine, New Brunswick)



Project Objective

- **Goal**
  - Create a dashboard allowing users to generate reports of buoy activity in near real-time
    - Location prediction of drifting buoys to ensure faster recovery of lost gear
    - Establishing baseline comparisons for buoys’ temperature sensors
- **Research questions**
  - How do Smart Buoys’ sensor readings compare to readings from external data sources?
  - Where will a Smart Buoy ‘drift’ in the ocean if its gear breaks free?

Research Question #1: Temperature Sensor Validation

Temperature Sensor Validation

Motivation:

- Accurate temperature readings allow fishermen to fish more efficiently
- BOG hopes to validate their sensor readings against third-party datasets to identify any systematic errors/ discrepancies
  - Third-party datasets studied: Copernicus satellites and Canadian weather stations
- But the problem set is a huge dataset if we just want to merge two datasets together. So we aim to use the batch dataframe to reduce the memory and be more efficient

Temperature Sensor Validation

Geographical locations/ fisheries:

Bounding Box of Maine
Bounding Box of Massachusetts
Bounding Box of New Brunswick/Gulf of St. Lawrence

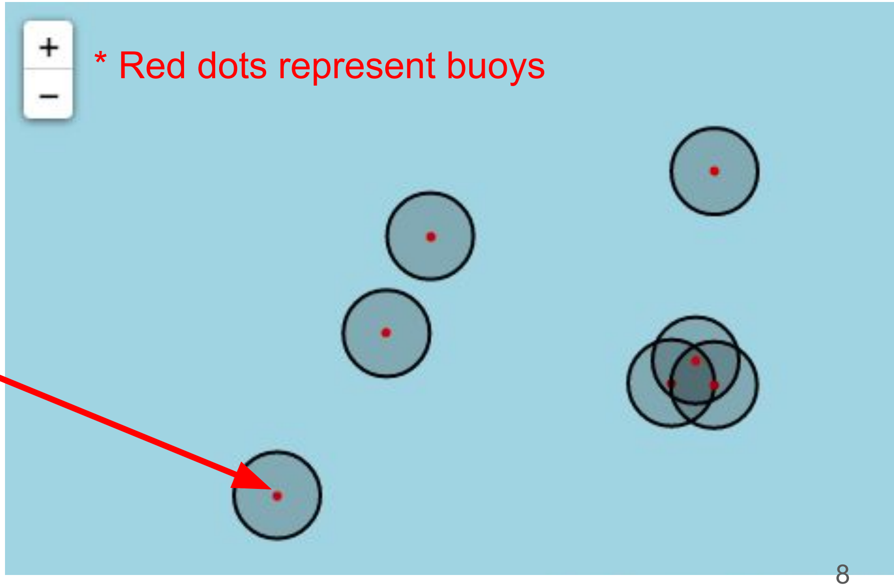
	BOG reading	Copernicus satellites	Canadian weather stations
Range of time	**2021-03-10** to **2022-10-04**	**2022-04-11** to **2022-10-04**	**2022-03-09** to **2022-10-05**
Frequency	On average every 1 min to 60 mins	60 mins	10 minutes

Methodology

- Segmenting buoy readings into deployments
  - A deployment is one continuous sequence of latitude/longitude coordinates of a buoy **from its entry to its leaving of the water**
- Writing scripts to query APIs for sea surface temperatures recorded by Copernicus satellites and Canadian weather stations
- Matching deployment readings by:
  - (a) For the Copernicus satellites:

1. Draw **buffers** around each buoy location

2. Calculate the **min, max, and average temperature** of the satellite data within the buffers



Methodology - Continued

- Matching deployment readings by:
  - (b) For the nearest neighboring Canadian weather station:
    - Dividing buoy data with timestamp with different batch data and calculate the difference in temperature within the same timestamp
    - Calculate the distance between each batch's buoy location with concatenated Canadian station with weather station ECCC-MSC and DFO-MEDS
    - Return the temperature for the nearest station and compare the temperature validation

Comparison results by fishery (Copernicus)

Fig. 1. **'Maine'**: BOG and Copernicus data comparison

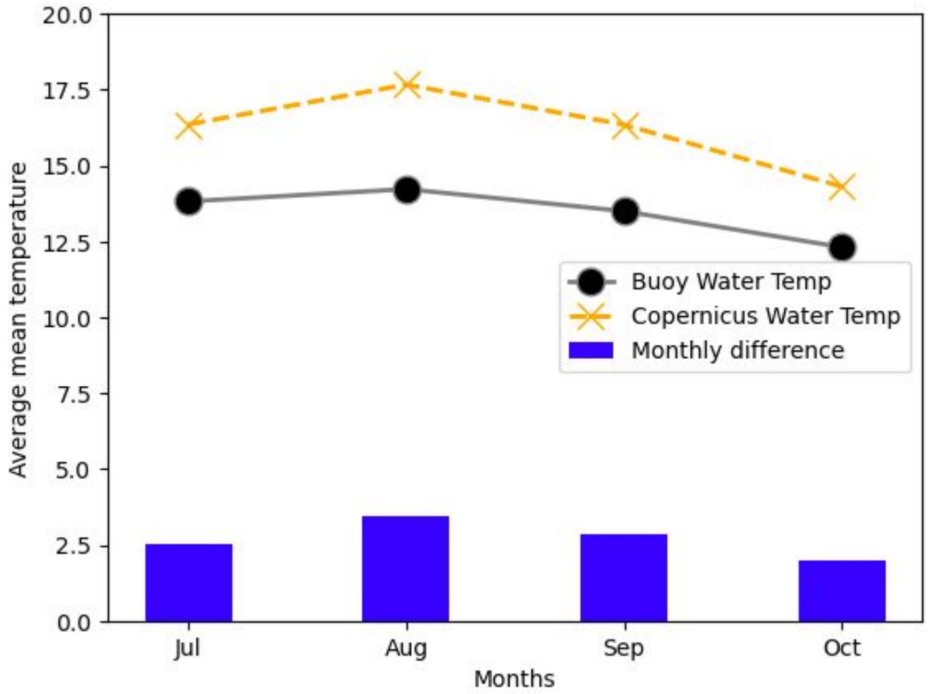
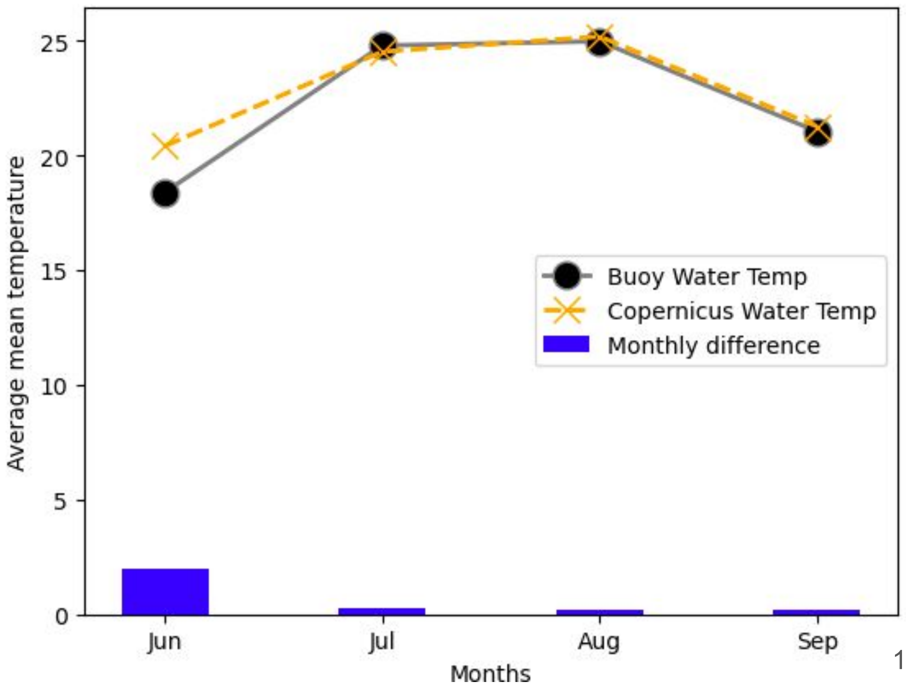
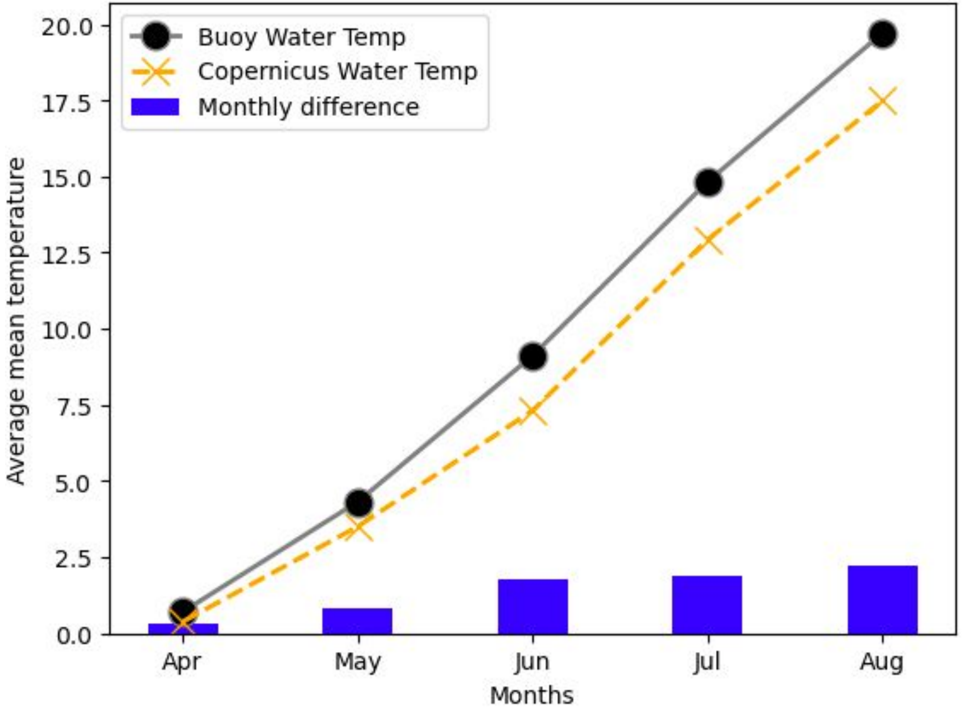


Fig. 2. **'Massachusetts'**: BOG and Copernicus data comparison



Comparison results by fishery (Copernicus)

Fig. 3. **'New Brunswick/Gulf of St. Lawrence'**: BOG and Copernicus data comparison



Summary:

- Maine (Fig. 1) had the largest temperature discrepancies between Buoy reported data and Copernicus satellites data
- Overall, temperature differences between BOG buoy readings and Copernicus satellites data were small

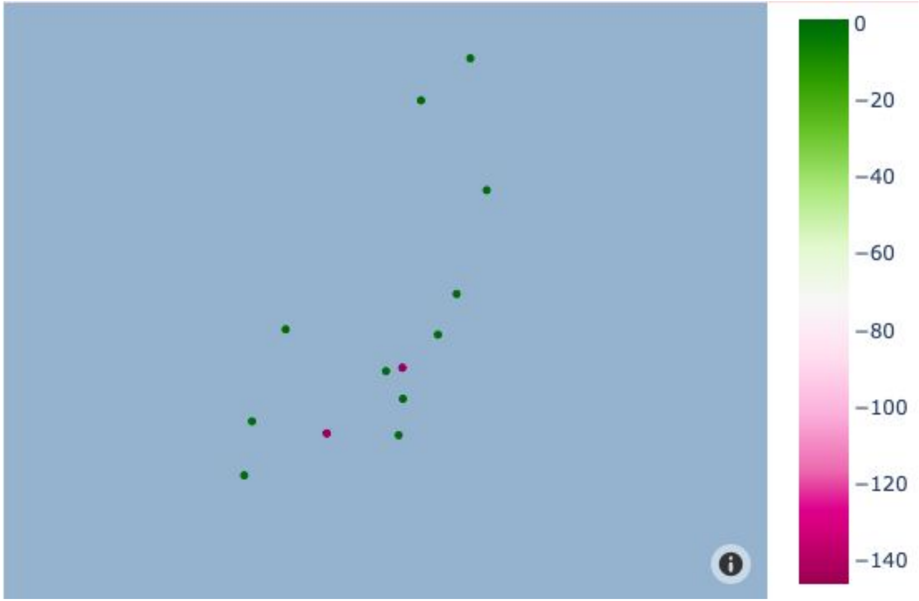
Outlier examination

An example of outlier: Buoy 481

Water_temperature_mean = **-146 °C**

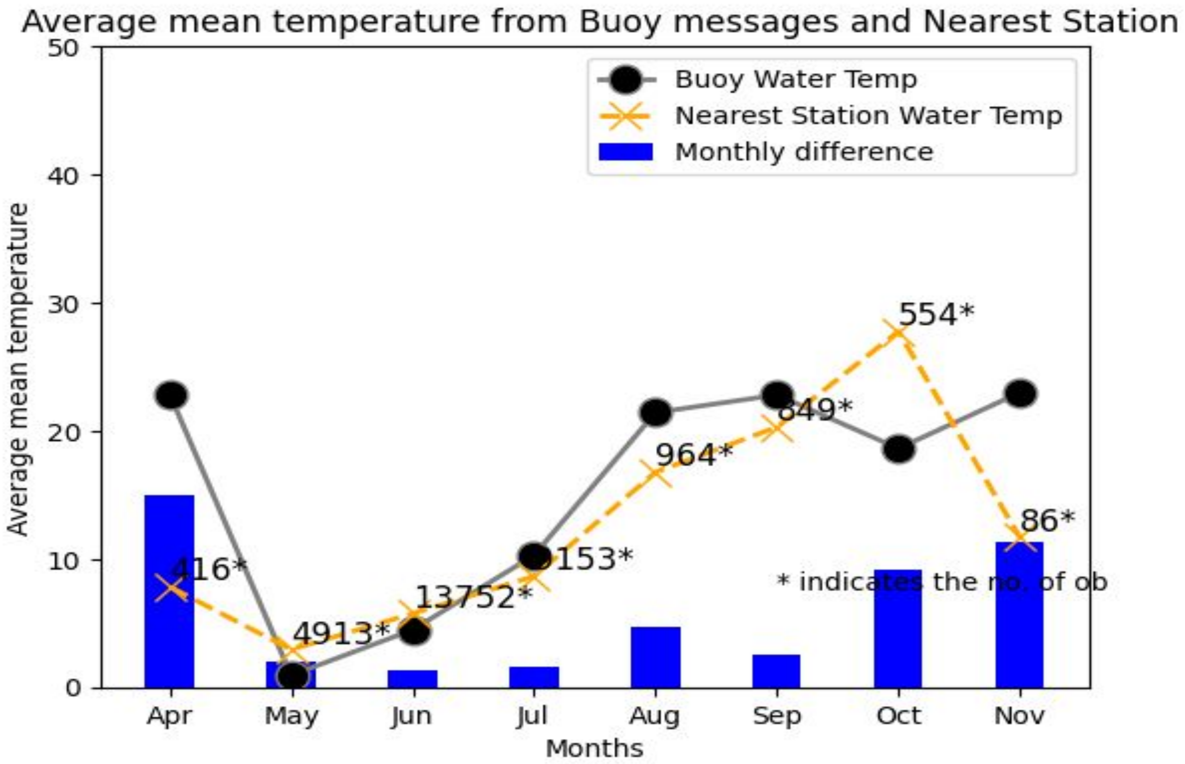
buoy_id	report_datetime	latitude	longitude	water_temperature_mean
481	2022-05-01 23:01:00	47.411736	-62.588898	-146.0
481	2022-05-02 22:41:20	47.411945	-62.588543	-146.0
481	2022-05-03 00:17:17	47.420850	-62.610760	-146.0
481	2022-05-03 02:20:45	47.421024	-62.610610	-146.0
481	2022-05-03 14:20:36	47.420803	-62.611397	-146.0

Fig. 4. Outlier buoy (id: 481) locations (pink dots)



Results: Comparing Canadian weather stations and BOG buoy readings within the same day

Fig. 3. **'Buoy Temperature mean from nearest station'**: Average mean temperature from Buoy messages and Canadian Station



Caveats of analysis and next steps

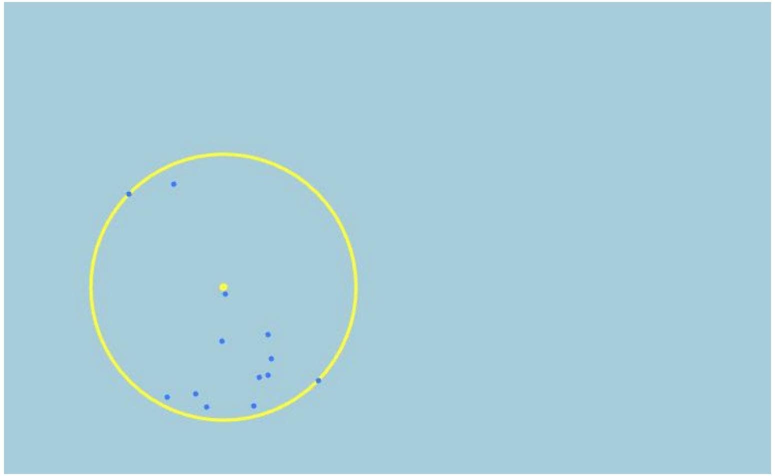
- Although BOG buoy readings and third-party datasets seemed to be similar in terms of water temperature, we have not run statistical analysis to demonstrate the significance of the differences
- Currently we only compared data for 6 months. A longer timeframe may be needed to show whether there are systematic differences in particular months or fisheries
- For next steps:
    - Run regression analysis to identify whether differences are statistically significant
    - Perform a more detailed analysis to identify potential factors causing minor differences between BOG buoy readings and external datasets

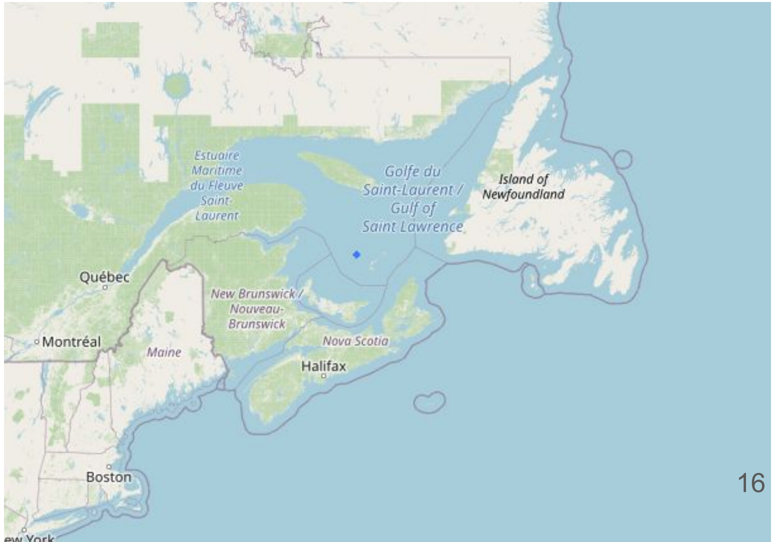
Research Question #2: Location Prediction

- Motivation
  - How buoys typically move (e.g., smallest enclosing circle, by fishery) – Zoey
  - Highlighted examples of buoys in the dataset that drifted (due to breaking free, or being attached to a different type of fishing gear) - Zoey
  - Problem of lost/ghost fishing gear and how it affects wildlife
  - Giving fishermen the ability to recover lost gear more quickly
- Methodology
  - Generating synthetic data using simulations (not enough data in BOG dataset) - Brinda
  - Pre-processing data for modeling - Brinda for Normalization, Zoey for masking rows, Hazel for transformation
  - Configuring LSTM model - Brinda
  - Training model on simulation data - Brinda
- Results
  - Performance of latitude/longitude versus distance/bearing
  - Ability of model to predict next trajectory of drifting BOG buoy, without further fine-tuning
- Caveats of analysis and next steps

Understanding Buoy Movements

1. **How buoys typically move by fishery**
   a. Buoy deployment reading with smallest enclosing circle
   b. Average Swing Radius (circle radius) - after removing outliers
      i. Massachusetts: 6.182km
      ii. Maine:0.124km
      iii. New Brunswick/Gulf of St. Lawrence: 0.948km

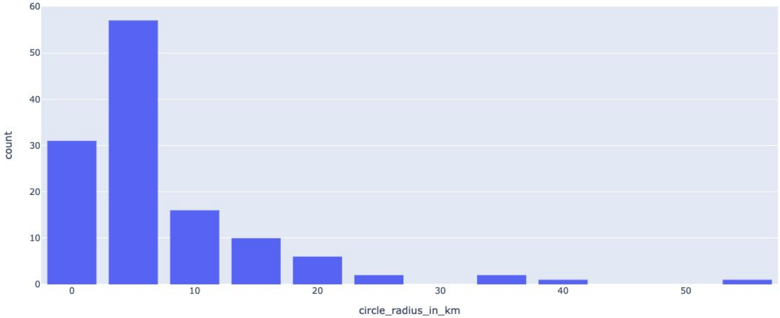


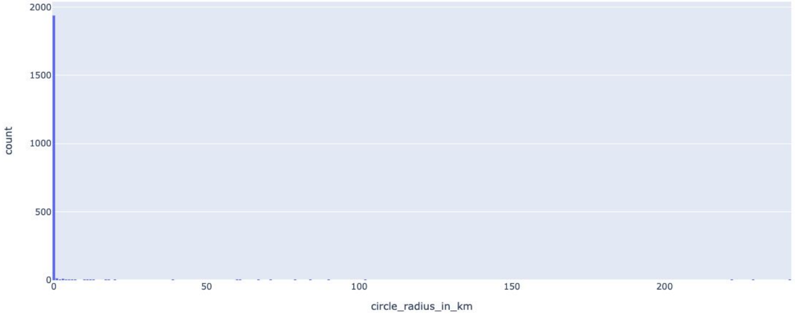


Understanding Buoy Movements

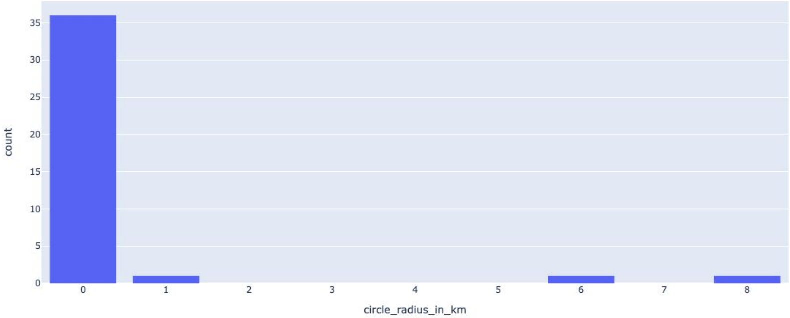
1. Highlighted examples of buoys in the dataset that drifted (large radius)

Massachusetts





New Brunswick/Gulf of St. Lawrence



Maine

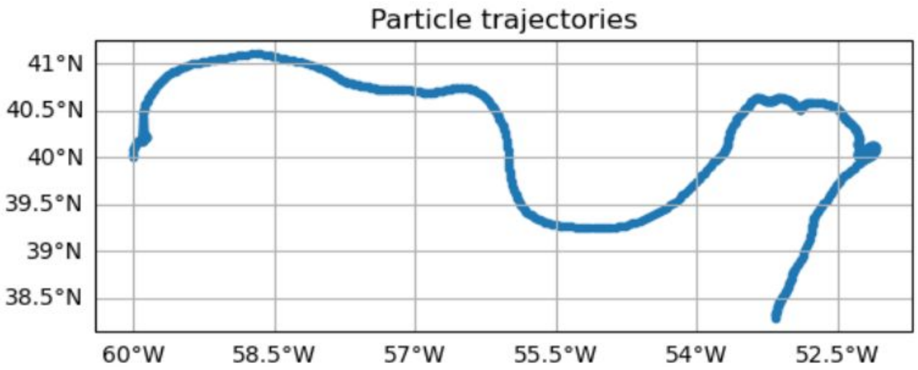
BUOY LOCATION PREDICTION MODEL

Synthetic Data: Why do we need it?

1. **To predict the next location where the buoy would “drift”** in the ocean, we need historical data to train the model about the buoy’s past drifting trajectory
2. It would have been ideal to have years of “trajectory data” for each of our client’s buoys
3. **Trajectory data** simply means if we track a given buoy which is deployed in the water over several months, what are the latitude and longitude coordinates that it would drift to, based on the ocean currents and ocean temperature
4. Trajectory data for the buoys is available **only for ~2 years** which is insufficient for the machine learning model to learn the pattern of a buoy’s movement
5. Therefore, we simulate the ocean environment using Copernicus data to **artificially increase the number of training samples**
6. Copernicus dataset that was previously discussed is used for the purpose of generating synthetic data which we queried using their service for sea surface current velocities (horizontal/eastward/"U" and vertical/northward/"V") that had been created from oceanic general circulation, wave, and tidal models

Methodology: Synthetic Data Generation

1. We use a Python package **OceanParcels** (MIT license) which simulates the movement of particles in the ocean over time by using ocean current velocities as input data.
2. We simulate **500 particles** in the ocean for every fishery
3. Each particle is intended to **resemble the movement of 1 buoy** at different time stamps, thereby generating a particle's trajectory
4. A **trajectory ID** uniquely identifies the multiple locations at which a single particle has "moved" as a result of the ocean's current velocities and horizontal ocean movement.



The plot shows the trajectory for 1 particle which tells us the location coordinates of the moving particle after every 10 minutes for the next ~550+ timestamps.

Synthetic Data: Methodology

1. Fieldset Class
   a. A **region** is generated using NetCDF files where every point in that region is associated with ocean current speeds and specific temperatures
   b. This creates a boundary for the region within which the “synthetically” generated particle will move
2. Particle Set Class
   a. For each fishery, a list of **random** latitude-longitude pair coordinates are generated which declare the **starting location** of a particle being simulated.
   b. We ensure that these lat, lon ranges are bounded by the region in which our client’s fisheries are defined.
3. Advection Kernel Class
   a. The particle **“moves”** across the ocean to different locations as a result of the type of movements defined within this class.
   b. How often does the particle move from one location to another? This is defined based on the minimal time difference between deployment readings from our client’s buoy deployment data.

Methodology: Data Pre-Processing

1. The goal of pre-processing is to streamline the data quality for synthetic data generated from the OceanParcels simulation so it can be used as input for our location prediction model.
2. Normalization
   a. Features of synthetic data (latitude, longitude and temperature) are scaled using min-max scaling technique such that each feature has values between 0 and 1
3. Feature Engineering
   a. To transform the time data to the position of the hand of a clock, making it easier for the network to observe trends that depend on the time of day
   b. To convert each latitude/longitude coordinate pair of a buoy to a distance value and a bearing value relative to the coordinate pair of the previous location

4. Standardization

   a. Convert datetime object to timestamp, calculate intervals from starting point to each subsequent timestamp object
   b. Mask all rows of latitude, longitude and temperature (or other variables of interest) with -999 if the datetime does not fall on a specific time interval)

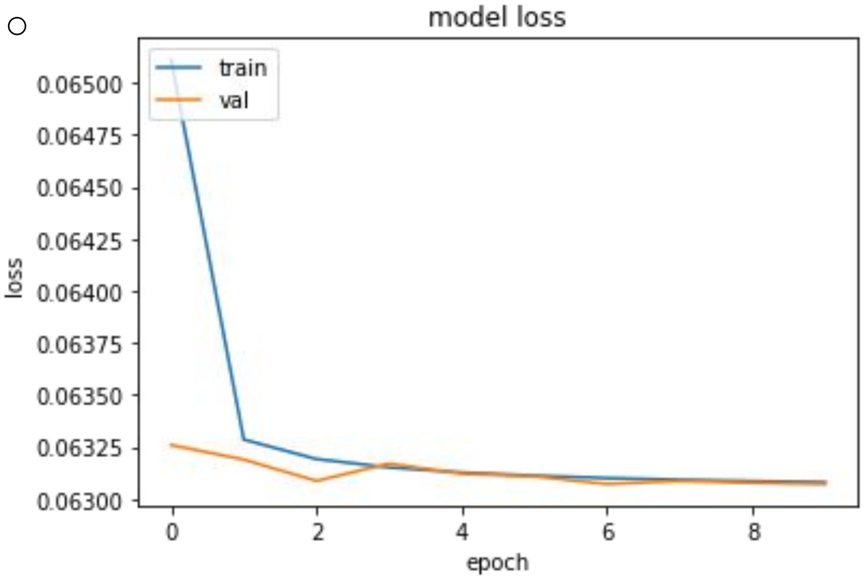
Location Prediction Model

1. Goal: Predict the “next sequence of locations” where a batch of particles would drift to, given the previous trajectories for a batch of particles
2. We use a Long Short Term Memory Model
   a. A type of neural network model commonly used for time series data
   b. Especially helpful to use this model when the input to the model is a “sequence” of data points relating to the same entity and the output of the model is also a “sequence” of data points
   c. In our case, we input a sequence of particles (i.e. sequence of trajectory IDs)

LSTM Model Architecture

1. We experiment with a 2 layer LSTM model using Keras package in Python
2. Model Input:
   a. A batch of trajectories with each trajectory consisting of trajectory ID, time, longitude, latitude and temperature upto 100 time steps of historical trajectory data for each of the particles in that batch
3. Model Output:
   a. For each particle in the batch, the model spits out the next 10 locations where that particle (or the buoy) would drift to
4. Loss Function:
   a. The function used to define these errors is Mean Squared Error i.e. the model will aim to minimize the mean of squared errors that occur on each particle in the training data
5. Hyperparameter Tuning:
   a. We experiment with different combination of parameters with the goal of selecting that combination of parameters that minimize the Loss Function described above.
   b. These hyperparameters include: learning rate, optimizer, batch size, number of epochs, and length of input sequences

Results

- Performance of latitude/longitude
- Ability of model to predict next trajectory of drifting BOG buoy, without further fine-tuning
- 

Caveats of analysis and next step

- Running LSTM model on Massachusetts and New Brunswick Fishery
- Compare the latitude and longitude model and distance bearing model results
- Synthetic Data
  - How closely the synthetic data will mimic the trajectory of the actual data of buoy trajectories
  - Time interval (hourly frequency)
- Ground-Truthing
  - It would be an interesting verification to assess whether the model's prediction are correct
  - This can be done by comparing real-time information about the actual location of the buoy's drifting trajectory with the predicted trajectory spit out from the model
- Anomaly detection

Impact

- By analyzing external datasets (i.e., Copernicus satellites & Canadian weather stations), we can establish baseline comparisons for buoys' temperature sensors. This may serve as a basis for any fine-tuning of the buoys in the future
- The fishermen who use Blue Ocean Gear's smart buoys can track their fishing gear in the ocean so now they will not have to worry about finding the lost buoys as they would have that information on their fingertips with the help of the locations of drifting buoys predicted by LSTM model

Thank you slide

We would like to thank:

- Blue Ocean Gear
- University of Chicago Data Science Institute
- 11th Hour Project

Blue Ocean Gear

THE UNIVERSITY OF CHICAGO
DATA SCIENCE INSTITUTE

11th Hour Project
The Schmidt Family Foundation