**The Data**

**u:** The east-west component of the current velocity, representing movement in the eastward (positive u) or westward (negative u) direction. It's a measure of how fast the water is moving across the Earth's surface in the horizontal plane from west to east or vice versa.

**v:** The north-south component of the current velocity, representing movement in the northward (positive v) or southward (negative v) direction. Like ‘u’, it measures the speed of water moving in the vertical plane from north to south or the opposite.

Together, u and v components allow for the representation of the direction and speed of sea surface currents at a given point. By combining these two components, one can calculate the total velocity and direction of the current at that point.

The magnitude (or speed) of the current can be calculated from the u and v components using the Pythagorean theorem. If you consider u and v as the legs of a right triangle, the magnitude of the current (hypotenuse of the triangle) is the square root of the sum of the squares of u and v.

The direction of the current, measured as an angle from the north (often in degrees or radians), can be calculated using the arctangent of v over u, which gives the angle θ.  
  
**Dimensions:**

**time:** Hourly Intervals (amount depends on period)

**lat:** 52 latitude points, indicating the geographic scope in the north-south direction.

**lon:** 43 longitude points, indicating the geographic scope in the east-west direction.

**Original Boundaries:** *DOUBLE CHECK*  
Latitude Range in the Dataset: 35.74470138549805 to 36.88019943237305

Longitude Range in the Dataset: 13.676799774169922 to 15.380399703979492

**Updated for Lagrangian:**lon\_min, lat\_min, lon\_max, lat\_max = 14.15, 35.79, 14.81, 36.3

**For simulation to prevent stuck boundaries:**

extent = [14.15, 14.71, 35.79, 36.2]

**Final polygon for predictions:**

polygon\_coordinates = [

(14.6, 35.87),

(14.32, 36.04),

(14.32, 36.09),

(14.6, 36.09),

(14.6, 35.87)

]

**Lagrangian Model Simulation**

* Mention pre-processing of the data, how the current data was merged and the length (no. of days), as of writing is 6 months of 2023.
* Simulation can be for same length used to train AI model or for brevity just simulate for only 7 days.
* Originally the idea was to use and combine wind and current data, but we decided to only use the sea surface currents data. This decision was based on two factors. Firstly, sea surface debris is proven to be more affected by sea surface currents instead of wind as seen in *[https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0111913]… In this study, researchers analysed data collected during five expeditions across all major ocean basins between 2014 and 2015 using a variety of methods including visual surveys, net sampling, and acoustic imagery. They found that "the majority of microplastics were concentrated in subtropical gyres" and that "wind-driven processes played only a minor role in determining the spatial distribution of microplastics." This finding suggests that while wind does contribute to some extent, it has less significance compared to ocean currents when discussing the distribution of sea surface debris.* Although it must be said that the wind data would make this more accurate in a real-life scenario. Secondly, we decided to do this to lessen the amount of work that had to be done.
* Original simulation was for 3 years but after some consideration, this proved to be too long and pointless. This decision was originally taken since initially we wanted to make predictions for a month but after some consideration this was not possible to implement and therefore we decided to simulate the Lagrangian model for a shorter period and make a prediction for a total of 1 day instead.
* First step was to create a land-sea mask. This was done by using a shapefile, This was then rasterised, and land is represented by 1 and sea by 0. Then it was converted to NETCDF format and then added within the grid boundaries to match with the boundaries of the data set. Show outputted image here. Can also mention how shapefile was used as map of Malta.
* Important to mention that due to the radars only north part of the island there is data available.
* Created the Lagrangian pipeline, and custom kernels (explain more in detail how they are deleted at boundaries and how the land sea mask reflects or deletes.. find reference for probability of being deleted or reflected).
* From simulations we noticed how particles were getting stuck at boundaries. This was found out to be since we were using all the dataset there was no data at the borders, so the particles were not affected. This was solved by using a bit smaller boundary (by like 00.1).
* Originally I was going to have multiple particles random but then it was more realistic to have around 50 (1 particle represents a group of debris) starting from very close proximity to see clusters of debris how they move.
* Simulation run for period of dataset, dt (elaborate) saved every hour? Elaborate as well.

**AI Model**

* Original was to train on 3 years single model and then predict 1 months. This was unfeasible. Then the next test was to train model on 7-day dataset and predict 1 day but making one model was hard due to the dimensionality of the data (4-d) and the predicted results were not accurate at all. Therefore, we decide to move to a smaller portion of the data (smaller boundaries) to predict the ‘u’ and ‘v’ components individually and then combine the predictions into one file to make the simulation from it.

**Future Work**

* Incorporate wind into the simulation and the prediction to make the model more accurate in real life scenarios.

**Miscellaneous**

* Put Adam in Acknowledgments.