Prediction of Origin and Drift of Marine Drift

Charanya Manivannan[1], Shivaani Suseendran[1], Jovina Virgin[1] and Dr. Vani K.[1]

[1]Department of Information Science and Technology, College of Engineering, Guindy, Chennai, India

*charanyamanivannan@gmail.com

Abstract

*Marine debris is in a constant state of motion, so identifying the origin of marine debris is hard. Marine debris' position keeps changing depending on various factors like wind, water speed, weather. Predicting origin of marine debris, and its trajectory in the following weeks, until collection can be done is necessary. For predicting debris origin, data provided by Florida University is used, containing waste (tons) exported by countries, waste destination, beached waste (tons) and origin of beached marine debris of countries. The probable origin countries of marine debris are found using the probability of marine debris beaching, waste (tons) exported by countries, and the distance between marine debris and countries. The nearest countries to the debris along with the countries whose waste reaches these neighboring countries and the destination countries of debris released from these neighboring countries are identified. The origin-values for these countries are calculated by giving more importance to distance than to countries' waste exports, because debris exported is heavily skewed. So, top 10 origin countries are identified by finding top 10 origin values. For predicting marine debris trajectory, open-source buoy data is used. Earth is split into 1*1 latitude-longitude grid and 52 such transit matrices are created for each week of a year. The probability of the buoy being in each cell of the grid during each week is calculated, and corresponding transit matrices of all buoys are averaged to get 52 transit matrices that depict all buoy data. To find the trajectory, the transit matrix corresponding to the week after the marine debris is found is used. Location of debris in the next week is the neighboring cell of debris location in the transit matrix with the highest probability.This is verified using existing buoy data.This process is repeated using subsequent transit matrices to get the trajectory.*

Keywords: Debris origin, Debris Trajectory, Probability, Transit Matrices.

Introduction

Management of marine debris is becoming increasingly difficult because of the unprecedented increase in the quantity of marine debris. The constant motion of marine debris adds another facet to the problem of marine debris management. Tracking its path is crucial for predicting spread, managing environmental impacts, and coordinating clean-up efforts effectively. This paper predicts the drift patterns of marine debris on water bodies and also investigates the origins of marine debris to understand where the marine debris comes from and pathways of pollution. Additionally, the path that the marine debris can take in future weeks is predicted, if the current location of marine debris is provided. By analyzing these drift patterns, it aims to anticipate the movement of debris over time, facilitating targeted cleanup efforts and enhancing overall waste management strategies.

Literature Review

The research done by Sushen et al.(2023) used segmentation to detect marine debris. It utilized multiple datasets for training, validation, and evaluation, involving a large training dataset with marine debris and non-debris patches made of refined Floating Objects, MARIDA test images and S2 Ship images, quality-focused validation dataset made from Refined Floating Objects dataset and MARIDA dataset for balanced patches. Preprocessing steps included data curation, balancing examples and rasterization. MARIDA scenes were re-downloaded with Sen2COR atmospheric correction, resulting in 12 bands. The preprocessing steps for floating objects in the study involved data sampling strategies to crop image patches centered on marine debris annotations and extract negative examples from random points within Sentinel-2 scenes. The Label Refinement Module used Otsu-threshold and random walk segmentation algorithms to refine annotated masks, repeating the process 25 times and combining all the results. These addressed uncertainty in debris borders from manual annotation, while iterative addition of negative examples depicted non-debris classes, improving classification accuracy. 2 segmentation models, UNet and UNet++ were trained and evaluated using scenes from Plastic Litter Project, MARIDA test dataset and Accra and Durban regions. The models trained with refined data was able to predict marine debris better than the models trained with the original data with no refinement.

Research done by Erik P Chassinet et al. (2021) used particle tracking simulations to provide an estimate of where marine litter released into water bodies by a given country go to, and where the marine litter found on a coastline of a country could have come from. The study used OceanParticles v2.1.15 which used outputs from ocean circulation models to create customizable particle tracking simulations. The authors released 32,500 particles every year along the global coastline during the years 2010-2019. After release, motion of these particles by ocean currents and joint effects of wind and wave is monitored and final results are published.

To track the path of debris, a paper by Erik et al. (2021) proposes a method using transit matrices. The methodology used in predicting the trajectory of particles in the ocean involves mapping buoy trajectories onto a $1° \times 1°$ grid resulting in transit matrices P_b, which provide the probability of particle movement between grid boxes over a 60-day period, by using probability of particle moving out of said grid boxes. This method incorporates observational data from the Global Drifter Program, which tracks buoys released since the early 1980s to study the accumulation of marine debris in the global ocean. Tracer in the grid box of the transit matrix P_b is scaled to the local population density within 200 km from the coast. By iteratively applying the matrix equation v_t + 60 days = v_tP_b, the trajectory of 10 particles is predicted. The study identified six major garbage patches in the ocean, one in each of the five subtropical basins and an additional patch in the Barents Sea.

Methodology

Marine debris is detected in any sentinel image using UNET, a segmentation model trained by an extensive dataset. Once marine debris is identified, the location of the marine debris is extracted from the sentinel image. Based on this location, the country which could have been the origin of the marine debris is predicted. Moreover, the trajectory that the marine debris will take in the following weeks is predicted, so collection efforts can be planned accordingly. The process is explained figuratively in Figure 1.

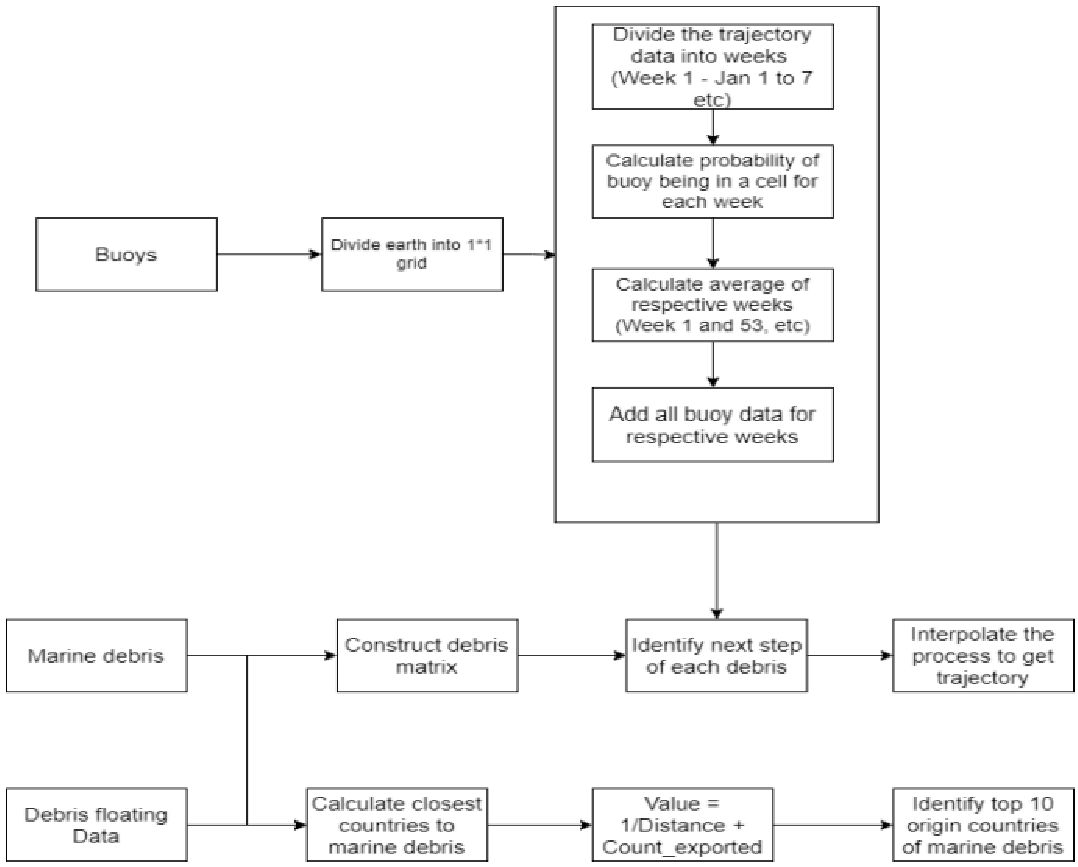


Figure 1. Architecture of prediction of origin and trajectory of marine debris

Marine Debris Detection

The MARIDA dataset and Floating Object Dataset are preprocessed and fed into the UNet segmentation model for detecting marine debris. The Marida Dataset undergoes filtering while the Floating Object Dataset is refined. It has 15 classes of images including Marine Debris, Dense Sargassum, Sparse Sargassum, Natural Organic Material, Ship, Clouds. Out of these 15 classes, only 3 classes are used for the scope of this project - Natural Organic Material, Foam, and Marine Debris. For every image, a new mask is created with only the required classes. The Floating Object Dataset consists of masks which are crudely drawn and do not represent the marine debris accurately. They are refined to better depict the presence of marine debris. The mask is rasterized so preprocessing can be done. The refining process following the study done by Sushen et al. (2023) is carried out using different combinations of Water Seed Probability, Object Seed Probability, Random Walker. The outputs of each of these refining operations are then combined using OR operation.The original mask is stacked in the beginning and the result is written on the new mask. Flip, rotation and noise addition operations are performed on the dataset and the final dataset is split into test, train and validation datasets, resulting in 52,608 patches of images containing both debris and non-debris. A UNet segmentation model is then trained using these datasets.

Once marine debris is detected using the UNet model, the location of the marine debris and the date the image was taken is extracted from the sentinel image. The country that could have been the origin of this marine debris, and the path that the marine debris might take, is predicted, so targeted debris management can be carried out.

Prediction of Origin of Marine Debris

Prediction of the origin of the marine debris is done using marine drifter data globally available. Marine Drifter Data released as part of the study done by Chassinget et al. (2021) is used to predict the origin of marine debris. It contains data collected over a period of 10 years and it consists of tons of wastes exported by each country, where the waste goes, and the number of tons of wastes which never reaches shores, or sinks. Basic preprocessing steps involved removing null values, dividing into multiple tables to better analyse and use the data. Visualization is attempted to provide a visual on the waste exports and imports of all countries present in the dataset. Visualizing the entire data proved to be impossible because of the count of waste exported by all countries. So, data pertaining to India is filtered. Only the data which contains India as the source or destination of the data is kept for visualization. This data is then visualized using matplotlib.animations to better understand the flow of marine debris to and from India.

The steps involved in predicting the origin country of marine debris is detailed below.

The center of each country is identified using geopandas dataset. The dataset contains the centroids of all countries. The list of countries present in the data published by the study is filtered. The distance between the centroid of each country and the marine debris location is calculated using the Haversian formula shown in Figure 2. The Haversian formula is used because it accounts for the spherical surface of the Earth. The goal here is to find the neighboring countries of the point where the marine debris was found.

$$d = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) (4.30)$$

Figure 2. Haversian formula to calculate distance

If the distance calculated is less than 500 km, the country is accepted as a neighboring country. From the dataset published by the study undertaken by Chassinget et al. (2021), the data pertaining to the neighboring countries are filtered - data where the neighboring countries are a source or a destination of marine debris. This is done because the marine debris can either be from the neighboring countries or on the way to the neighboring countries. So, such countries where the wastes found on their coasts could have originated from one of the neighboring countries, or countried whose wastes reach the neighboring countries are filtered. The distance between the filtered countries and the marine debris is identified. The origin value for each of the filtered countries is calculated as mentioned in Figure 3.

$$OriginValues = 0.5 * Countofwasteexported + 2/Distance$$

Figure 3. Formula for calculating Origin Values

The higher the origin value, the higher the probability of the country being the origin of the marine debris. Because there is a huge discrepancy in the count of waste exported by the countries, the top 10 countries which could be the origin of marine debris are 48 identified.

Prediction of Trajectory of Marine Debris

The path that the marine debris could take is predicted using buoy data. The path is identified by predicting where marine debris would be at consecutive weeks from the date that the sentinel image was taken. Using these locations, the path is drawn which can help in collection efforts.

A buoy is a floating device that can be allowed to float, while collecting information like sea surface temperature, and water velocity along with the current latitude and longitude of the buoy. The buoy data is collected every 6 hours as long as the buoy is able to send data. Buoy data is available as part of the Global Drifter Program. Each buoy is identified using its WMO. The buoys which have a trajectory surrounding India are filtered. Buoys collect data like Sea Surface Temperature, Current Latitude, Current Longitude, Time, and Type of Buoy. Out of these data, only Latitude, Longitude and Time are used for our prediction.

The Earth is divided into a 1*1 grid where each cell depicts the corresponding latitude and longitude. The Earth's latitudes range from -90 to 90, while longitudes range from -180 to +180. So, a grid of 181*361 is initialized with zeros. Each cell depicts the corresponding latitude and longitude. For example, cell (0,0) depicts (-90,-180) as (latitude,longitude). The weeks of the year are standardized to get uniform data, with week 1 being Jan 1 to Jan 7, week 2 being Jan 8 to Jan 15, and so on. For each buoy, the probability of the buoy moving to each cell in the grid in each week is calculated. This matrix is called a transit matrix. At least 52 such transit matrices are calculated, corresponding to each week of the year. If there's no buoy data for a week, which could happen in cases where buoy was alive from the middle of the year, or if the buoy died abruptly, the transit matrix for that week remains a matrix filled with zeros. If buoy data is present for more than a year, transit matrices for all available weeks are calculated and the respective weeks, like week 1 and week 53 are averaged. The matrices of corresponding weeks of all buoys like week 1 of buoy 1, buoy 2 are added together to get a comprehensive set of 52 matrices where each matrix shows the probability of any buoy moving into the latitude longitude represented by the cell of the matrix in a certain week of the year.

The current position of the marine debris is mapped onto a debris matrix similar to the buoy's transit matrix, where the cell depicting the current position of the marine debris is 1 while all the other cells are zero. The week in which the date that the sentinel image was taken belongs to, is calculated. In the corresponding week's transit matrix, the values of neighboring cells of the current position are obtained. The neighboring cell which has the highest probability is the next location of the debris. This is because that was the cell which most of the actual buoys moved into. This process is then repeated with the next week's transit matrix. These steps are repeated to get the trajectory of the debris. The trajectory is then visualized on a world map with pointers depicting where the marine debris could be in subsequent weeks. This interpolation can be performed for as many weeks as is required so collection can take place.

Results and Discussion

The goal of this paper is to predict the origin country of marine debris and the path that it could take in the upcoming weeks so collection efforts can be made easier. In the scope of this paper, marine debris is identified using sentinel images by using a segmentation model. Once marine debris is identified, the location and the date when the sentinel image is taken is extracted from the metadata of the sentinel image. Based on the location where the marine debris is found, the top 10 probable origin countries are identified. This is because the count of waste exported by each country is heavily skewed, based on the population density of different countries. The next step is to predict the next locations of the marine debris using global drifter data, A path is drawn through these locations. so collection of these marine debris can be targeted and made effective.

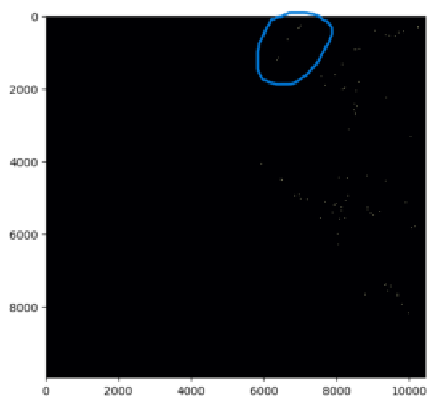


Figure 4. Original mask

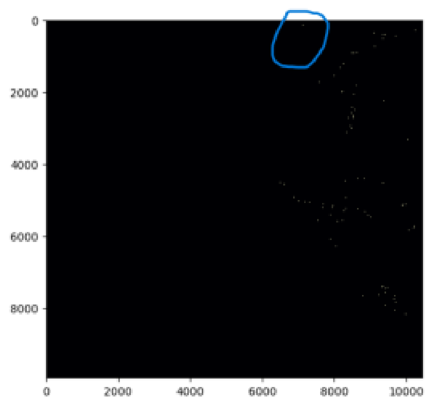


Figure 5. Refined mask of Floating Object

The detection of marine debris is done using a UNet segmentation model trained on a dataset consisting of 2 datasets - the MARIDA dataset and the Floating Object dataset. The MARIDA dataset is filtered to contain only the required 3 classes out of all 15 available classes. The Floating Object dataset consist of crude hand drawn masks to depict marine debris. The masks present in this dataset undergoes refining so that they can better represent actual marine debris area. Figure 5. shows the original mask of one image and Figure 6 shows the refinement of the same mask with the changes highlighted.

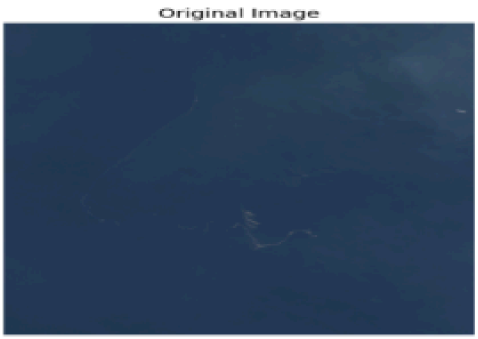
Finally, the two datasets are combined and undergo preprocessing steps like flipping, adding noise. They are then split into train, test and validation datasets before being fed into the UNet model. This split along with the composition of the train, test and validation datasets is shown in Figure 7.

```
Dataset Composition total

Train
MARIDA : 2149
Floating Objects : 39194

Validation
MARIDA : 1256
Floating Objects : 3006

Test
MARIDA : 918
Floating Objects: 6085


Dataset Composition debris/non-debris

Train
Floating Objects: 19587/19607
MARIDA: 1062/1087

Validation
Floating Objects : 1503/1503
MARIDA : 687/569

Test
Floating Objects : 3040/3045
MARIDA : 316/602
```

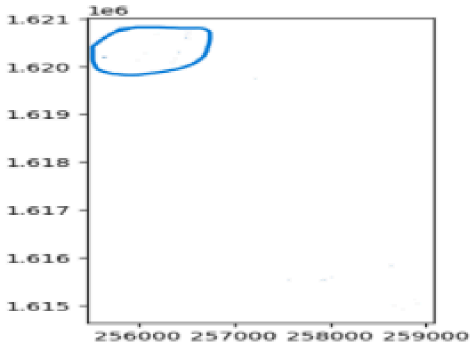
Figure 7. Composition of test, train and validation datasets

The UNet model was trained on these images and was evaluated using the metrics - Intersection over Union (IoU), Dice Score and Accuracy. The performance of the model was adequate for the scope of this project which focused on finding the origin and trajectory of the marine debris once the marine debris is calculated. The model gave an IoU Score of 0.773, Dice Score of 0.713, and an Accuracy of 0.703. These metrics are tabulated in Table 1.

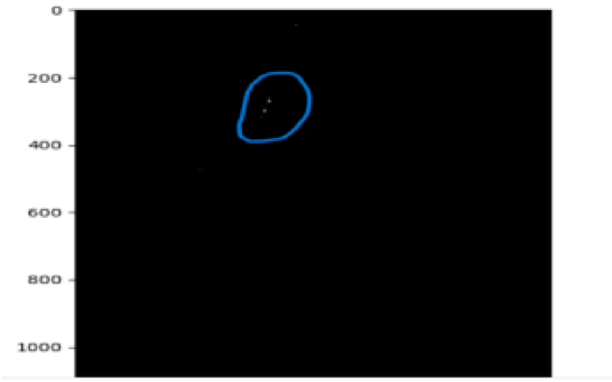
Figure 8 shows the segmentation output of the trained UNet model. Figure 8(a) shows the input image while Figure 8(b) shows the marine debris present in the image. Figure 8(c) shows the actual output of the model.



(a) Input



(b) Mask



(c) Output

Figure 8. Input and Output of UNet model

Once marine debris is detected on a sentinel image, the location and date that the sentinel image was taken is extracted from the image. Figure 9 shows the latitude longitude extracted from the metadata of the sentinel image used in Figure 8.

Location (lon, lat): (253.14994833314526, 89.31337271258325)

Figure 9. Location of marine debris

The origin of the marine debris is calculated using the data published by Chassinet et al. (2021) which contains waste exported by each country and their destinations, waste reaching shores of different countries and the source of these wastes, as shown in Figure 10.

Country N.	Tons Expc	End in the	End in the	To	Count	From	Count
albania	11718	3257	8462				
				Tons to Egypt	3197	From Albania	60
				Tons to Libya	2314	From Montenegro	11
				Tons to Italy	1479	From Algeria	4
				Tons to Israel	391	From Greece	4
				Tons to Syria	303	From Croatia	3
				Tons to Lebanon	185	From Ukraine	3
				Tons to Tunisia	140	From Italy	1
				Tons to Croatia	117	From Tunisia	0
				Tons to Turkey	93	From Libya	0
				Tons to Cyprus	86	From Spain	0
				Tons to Albania	60		
				Tons to Greece	44		
				Tons to Northern Cy	37		
				Tons to Palestina	8		
				Tons to Montenegrc	7		
				Tons to Akrotiri and	1		
				Tons to Bosnia and I	1		

Figure 10. Data for origin prediction of a sample country, Albania

Based on this data, visualization of all countries was attempted, to see the disparity in the amount of waste exported by different countries. This visualization is shown in Figure 11. Further, the flow of waste to and from India is also visualized in a .gif format to understand where wastes in coasts of India come from, and which country it goes to.

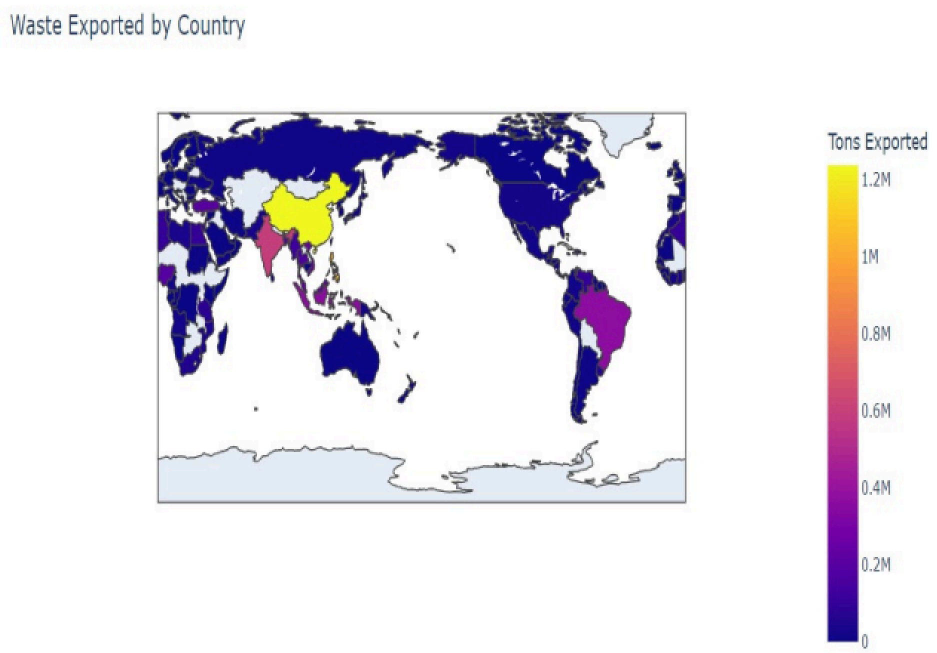


Figure 10. Visualisation of Waste exported by each country

As seen in Figure 10, China is the leading exporter of wastes exporting around 1.2M tons in the given data. It is also the only exporter exporting more than 1M tons of waste based on the available data.

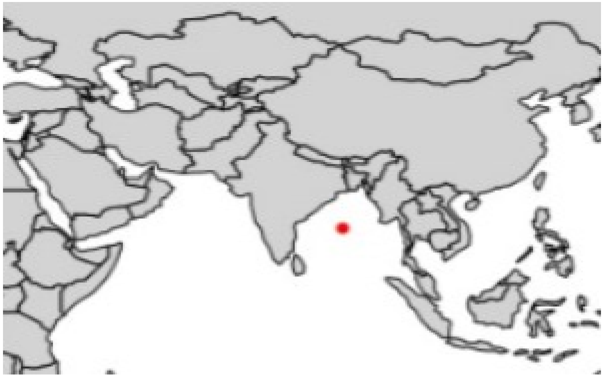


Figure 11. Sample location of marine debris to find origin

For a sample location near India as shown in Figure 11, the probable origin countries of marine debris in that location is predicted. The closest countries to this location are identified using the Haversian formula. These are the neighboring countries to this point. These countries and those countries whose waste reaches these countries are filtered from the dataset. Distances between the debris location and the centroids of these countries are calculated and stored as shown in Figure 12. This is done because the waste could either be from the coast of the neighboring countries or from other countries on their path to the coast of the neighboring countries. Since there's a huge difference in the amount of waste exported by different countries as seen above, origin values are calculated by emphasizing distance over waste count. The country with a higher origin value has a greater chance of being the origin of the marine debris. Because origin values are greatly skewed towards the count of waste exported even after giving more weightage to distance, identifying the top country alone isn't valuable. So, the top 10 probable origin countries are identified and the origin values are shown in Figure 12.

From the values in Figure 12, the most probable countries that the marine debris could have come from includes China, India, Philippines, Pakistan, Vietnam considering their distance from the place where the marine debris is assumed to be, and the amount of waste they export which aligns with expectations.

	Country	Distance	Tons Exported	End in the Ocean	End in the beach	Probability_Beach	Probability_Ocean	Values
41	china	2867.050674	1239253.0	226454.0	1012800.0	0.817267	0.182734	185.889694
43	philippines	3788.921916	1027988.0	333743.0	694245.0	0.675343	0.324657	154.199520
34	india	1247.472249	591942.0	174845.0	417097.0	0.704625	0.295375	88.795308
14	brazil	15773.515921	365897.0	117727.0	248170.0	0.678251	0.321749	54.884867
4	indonesia	3762.010623	360322.0	108329.0	251993.0	0.699355	0.300645	54.049629
31	vietnam	1964.153090	255453.0	55389.0	200064.0	0.783173	0.216827	38.320496
44	malaysia	2686.477289	212533.0	66979.0	145554.0	0.684854	0.315146	31.881811
35	bangladesh	1011.192084	175902.0	43896.0	132006.0	0.750452	0.249548	26.390245
28	thailand	1396.681163	165392.0	29356.0	136035.0	0.822500	0.177493	24.812380
30	myanmar	1120.521513	126574.0	36194.0	90380.0	0.714049	0.285951	18.990562
13	uruguay	15885.294016	62882.0	29614.0	33268.0	0.529054	0.470946	9.432615
11	south africa	8337.779848	53158.0	41905.0	11253.0	0.211690	0.788310	7.974300
42	taiwan	3583.355914	52501.0	15657.0	36844.0	0.701777	0.298223	7.876545
21	mozambique	6784.325497	43859.0	15129.0	28730.0	0.655054	0.344946	6.579587
36	pakistan	2526.295838	32604.0	9500.0	23104.0	0.708625	0.291375	4.892579

Figure 12. Origin Value Calculation for the sample point

The prediction of the path that the marine debris will take in the upcoming weeks uses buoy data. Buoys are strategically placed in water bodies to collect data, and they continue sending data every six hours for as long as they are alive. Figure 13 shows the passage of buoys over time. This is a subset of all buoy data that is available.

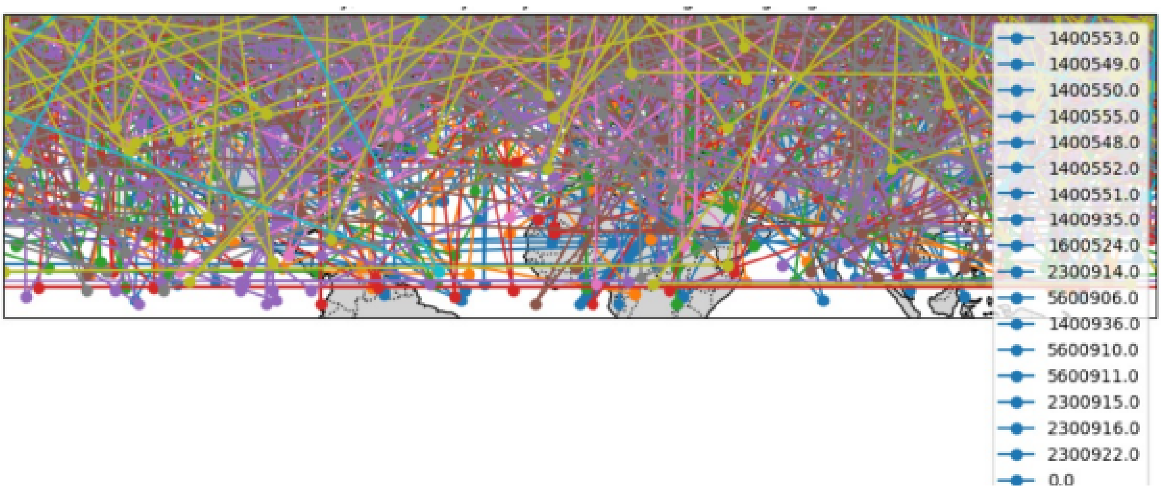


Figure 13. Movement of Buoys over time

With this buoy data, transit matrices are created for each buoy. A transit matrix represents a 181*361 matrix where each cell represents a corresponding latitude longitude, meant to identify points on the surface of the Earth. The value in each cell is the probability of the buoy moving into the position represented by that cell, in that week's time. For example, If Transit Matrix 1's value at cell position (90,180) is 0.5 for a buoy, the probability that the buoy moves to (0,0) (latitude,longitude) point in Week 1 of that year - January 1 to January 7 is 0.5. Figure 14 shows the expanded view of 1 transit matrix with the probabilities.

0.	0.	0.	0.	0.	0.
0.	0.	0.08928571	0.18714286	0.08928571	0.01785714
0.	0.	0.	0.	0.	0.
0.	0.	0.	0.	0.	0.

Figure 14. Expanded transit matrix

The weeks of the year are standardized. This means that Week 1 is January 1 to January 7, regardless of when the buoy was released or when it died. Week 52 is December 25th to December 31st. If there's no data for a week, the transit matrix is a matrix where values of all cells are zero. If data is present for more than 52 weeks, transit matrices are created for all weeks. Corresponding weeks' transit matrices are averaged to better reflect the movement of the buoys. That is, Week 1 and week 53 are averaged. Week 2 and Week 54 are averaged and so on. Figure 15 shows the data of 1 buoy with WMO 1401752. This data is stored as a 2 dimensional dictionary with the WMO as primary key and week numbers as secondary keys.

```
{1401752: {0: array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]]), 1: array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]]), 2: array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
```

Figure 15. Transit matrices of a sample buoy

The week in which the sentinel image with marine debris was taken is calculated. To predict the position of debris on a future date, the respective week's transit matrix is used.

For the location of the marine debris, the neighboring cells in the transit matrix are used. The probabilities in the neighboring cells of the transit matrix are identified. The cell with the highest probability is the next position of the debris. This process is repeated until the trajectory is identified.

Figure 10. The predicted path of the marine debris

For the sentinel image discussed in Figure 8 and Figure 9, the path that the marine debris will take in the next 3 weeks is shown in Figure 10.

Conclusion

Marine debris management is hard because locations in the middle of water bodies can not be accessed everyday for marine debris identification or collection. Marine debris identification is difficult because of their locations and the booming count of marine waste. Collection proves to be harder. Added to the costs in reaching locations of marine debris is the fact that the marine debris will keep on moving and will not stay stagnant like debris in land. If marine debris is identified today, and the collection efforts are organized for a week later, the marine debris would have floated somewhere else due to wind and water movement. Thus, predicting the trajectory that the marine debris will take becomes vital. The trajectory of the marine debris is predicted using real time buoy data and the trajectory can be predicted for as long as required. This will prove immensely useful to organize targeted collection efforts. The trajectory represents a point on the Earth's surface for each week, so collection efforts can be focused in and near that area for that week. Finding the origin of the marine debris can help in obtaining much needed information about the waste exporting pattern of countries. This can help in identifying the cause of marine debris and managing debris before it reaches water bodies. Future work will involve finding the trajectory of the marine debris after taking into account the chance that the marine debris will sink instead of floating until it can be collected. Further, identifying 1 origin country from the top 10 identified in this study can help.

References

Marc Rußwurm, Sushen Jilla Venkatesa, Devis Tuia (2023). Large-scale detection of marine debris in coastal areas with Sentinel-2. IScience, Volume 26, https://doi.org/10.1016/j.isci.2023.108402

Erik van Sebille, Matthew H England and Gary Froyland (2012). Origin, dynamics and evolution of ocean garbage patches from observed surface drifters. Environmental Research Letters Volume 7http://dx.doi.org/10.1088/1748-9326/7/4/044040

Eric P. Chassignet,Xiaobiao Xu, Olmo Zavala-Romero (2021). Tracking Marine Litter With a Global Ocean Model: Where Does It Go? Where Does It Come From?. Frontiers in Marine Science. Volume 8. https://doi.org/10.3389/fmars.2021.667591