Predictive modelling of sea debris around Maltese coastal waters

Mark Dingli

Supervisor: Dr Kristian Guillaumier

May 2024

Submitted in partial fulfilment of the requirements for the degree of B.Sc. IT Artificial Intelligence.

L-Università ta' Malta
Faculty of Information & Communication Technology

Abstract

The accumulation of sea surface debris around the coastal waters of Malta, presents numerous ecological and environmental challenges that negatively affect both marine ecosystems and human activities. This is exacerbated by the absence of an effective system that can predict their movement, making it more challenging to address and mitigate this issue effectively.

The primary objective of this project was to develop a system that can predict dispersion patterns of sea surface debris around Malta's coast. To achieve this, we developed a comprehensive machine learning and physics-based pipeline. This pipeline uses historical sea surface current data to predict future conditions, while also having the ability to visualise the movement of debris.

Central to this system is the integration of LSTM and GRU models, trained to predict the next 24 hours of sea currents within a specific area. These predictions were subsequently utilised by the Lagrangian model to visualise the movement of surface debris, offering insights into future dispersion patterns.

A comparative evaluation was conducted for both models, examining the accuracy of their predictions and the quality of the simulations generated by the Lagrangian model, based on these predictions. The results indicated that the LSTM model outperformed the GRU model. This was evidenced by the LSTM's enhanced precision in forecasting the movements of sea surface currents, thereby providing a more reliable basis for the subsequent simulation of debris dispersal patterns.

Overall, this project offers a novel approach to addressing the challenge of sea surface debris around Malta. By harnessing the power of machine learning in tandem with a physics based Lagrangian model, we have established a framework that not only predicts sea surface currents with notable accuracy, but also visualises the movement of surface marine debris, allowing us to make more informed decisions about our environment and our effect on it.

Acknowledgements

I would like to extend my deepest gratitude to several individuals whose support and guidance were invaluable in the completion of this project.

Foremost, I express my heartfelt appreciation to my supervisor, Dr Kristian Guillaumier, for his unwavering guidance, encouragement, and insightful critiques throughout the whole project. His expertise and dedication were instrumental in navigating the complexities of this project and in pushing the boundaries of my academic capabilities.

My sincere thanks also go to Dr Adam Gauci, who not only provided the essential data for this project but also offered his expertise and support. His contributions have been pivotal in enriching the quality of this work.

Additionally, I owe a profound debt of gratitude to my family, especially my parents and sister. Their unwavering support and love have been the cornerstone of my resilience throughout this journey. Their belief in my potential and constant encouragement has been instrumental in empowering me to pursue my goals.

Lastly, I must express profound appreciation to my girlfriend, Ilenia, and my close friends. Their understanding, patience, and encouragement have provided me with the motivation needed to persevere through the challenges of this project.

To all mentioned, and to those who contributed, I am eternally thankful. Your roles in this academic endeavour have left an indelible mark on both the project and my personal growth.

Contents

List of Figures

List of Tables

No table of figures entries found.

List of Abbreviations

FYP	Final year project
LSTM	Long short-term memory
GRU	Gated recurrent unit
NetCDF	Network common data form
RNN	Recurrent neural network
ANN	Artificial Neural Networks
AI	Artificial Intelligence
RMSE	Root Mean Squared Error
MAE	Mean Absoulut error
NaN	Not a number

Note that the List of Abbreviations should be sorted on the acronym list.

1 Introduction

This project is an integration of machine learning techniques with a physics based Lagrangian model [1] to address the environmental issues of sea debris. At the core of this project is a pipeline that harnesses historical data to forecast future conditions, specifically predicting the next 24 hours of sea surface currents. These predictions serve as inputs for a Lagrangian model [1], enabling it to simulate the movement of surface marine debris. Finally, a comparative evaluation of both LSTM and GRU models is conducted, focusing on their predictive accuracy and the quality of the visualizations. This project introduces an approach of merging machine learning with a physics-based model, offering valuable insights to marine conservation efforts and improving decision-making for managing marine debris around the Maltese Islands.

1.1 Problem Definition

Sea surface debris around the coastal waters of Malta presents a significant environmental challenge. Predominantly composed of plastics, which constitute 82% of all man-made floating items encountered in the Mediterranean sea [2], this debris endangers marine life, disrupts ecological balances, and undermines the ecological integrity of coastal areas [3]. This problem is further aggravated by the lack of an effective system that can predict and forecast the movement of this surface debris, since as of writing, there exists no system that adequately addresses this challenge specifically for the coastal areas around Malta. This further underscores the need for a system that can accurately predict and visualise the dispersion patterns of sea surface debris.

1.2 Motivation

The geological characteristics of the Mediterranean sea makes it difficult for surface debris to escape the area naturally, leading to the accumulation of sea surface debris [4]. The current absence of a predictive system tailored to the coastal regions of Malta impedes effective interventions to mitigate environmental harm. This gap opens an opportunity for the implementation of a system that through the application of Machine Learning and physics-based modelling, aims to address an urgent ecological issue, which is widely recognised as a global crisis [5]. By fulfilling this need, the project aims to provide accurate predictions that can guide effective cleanup operations and inform strategies for long-term marine conservation around the surrounding waters of Malta.

1.3 Aims and Objectives

The aim of this project is to create a system enhanced with Machine Learning for simulating and predicting the movement of marine debris in the coastal waters of Malta, thereby supporting marine conservation efforts. To achieve this aim, the following objectives have been identified:

01. Data integration: To preprocess and integrate the sea surface currents datasets ensuring compatibility and consistency for input into both models.
02. Lagrangian model development: To utilize develop a Lagrangian physics-based model for simulating the movement of surface marine debris, employing historical data to ensure accurate simulations.
03. AI models development: To develop and fine-tune both LSTM and GRU models for the prediction of future sea surface currents. These models will serve as a crucial component of the forecasting system, leveraging their respective strengths in time series data processing to ensure robust and accurate predictions.
04. Integrating the AI models with the Lagrangian model: To integrate the model's predictions into the Lagrangian model. This integration aims to create future simulations and visualisations of marine debris movement, enhancing the project's predictive capabilities for marine conservation.
05. Comparison of AI models: To conduct a comparative evaluation of both LSTM and GRU models, focusing on their predictive accuracy and the quality of the final visualizations.

1.4 Proposed Solution

This project aims to develop an integrated pipeline for predicting and simulating the movement of marine debris around Malta's coastal waters. The process begins with the preprocessing of the sea surface currents datasets that will be used as input for the subsequent modelling stages. A Lagrangian model will be developed to visualise the debris movement. This approach is designed to clarify both the expected input from the AI models and the expected nature of the ensuing visualizations. The core of the solution involves developing and fine-tuning two types of machine learning models: LSTM and GRU. These models will undergo extensive testing to determine the optimal architecture and hyperparameters, aiming to accurately predict sea surface currents for a future 24-hour period.

Upon establishing the predictive models, the pipeline integrates these predictions into the Lagrangian model, transforming the predicted data into dynamic visualisations of future debris movement. The project culminates in a comparative analysis of the LSTM and GRU models, evaluating their effectiveness through various metrics, including their predictive accuracy and the quality of the generated visualisations. By analysing the results and visualisations, this project aims to provide actionable insights for effective cleanup operations and strategies for long-term marine conservation around the coastal waters of Malta.

1.5 Summary of Results

To force Word to automatically update the cross-referencing, select the entire document by pressing CTRL-A on your keyboard, followed by F9.

1.6 Document Structure

The remainder of this document is organised into the following chapters:

Background: Here, the foundational elements of the project are discussed. This chapter includes a thorough overview of the utilized datasets, an explanation of the Lagrangian model's principles and capabilities, and an insight into the Machine Learning models.

Literature Review: In the literature review, we will delve into existing research and findings relevant to marine debris, the use of Lagrangian models, and the application of different AI models in environmental forecasting, establishing the scientific grounding for the project’s methodologies.

Methodology: This section details the processes undertaken in implementing the FYP. It includes the steps involved in data integration, the development and integration of the Lagrangian and AI models, and the comparative evaluation of the AI models.

Evaluation: A comprehensive outline of the strategies employed to test and evaluate the effectiveness and reliability of the implementation is presented in this section. This will be followed by the presentation and discussion of the results.

Conclusion: This FYP is concluded by summarizing conducted work, revisiting the aims and objectives, acknowledging any encountered limitations, highlighting obtained results, and finally suggesting any proposals for future work.

2 Background and Literature Review

In this chapter, we provide a comprehensive background for the project, while also presenting an overview of pertinent academic papers and literature to underpin the project's scientific foundations.

2.1 Background

This section is divided into five distinct segments where we collectively provide detailed information regarding this project. First, we begin with an exploration of the impacts of marine debris on ecosystems and also delve into the specific datasets used in this project. Then, we discuss the intricacies of the physics based Lagrangian Model, provide an explanation of time series analysis, and finally round off with an exposition on deep learning models, specifically LSTMs and GRUs.

2.1.1 *The impact of marine debris on ecosystems*

The environmental and ecological impact of marine debris, particularly in coastal and marine ecosystems, has been extensively researched, as evidenced by [6] and [7]. Several studies in this area reveal significant negative effects, ranging from harm to marine wildlife due to ingestion and entanglement [8], to the disruption of natural habitats [9]. The impact on coastal ecosystems extends beyond the environment, affecting economic sectors reliant on marine health, such as tourism and fishing as discussed in [9]. Further research delves into the long-term ecological consequences, highlighting the urgent need for effective management and mitigation strategies as seen in [10]. These studies collectively emphasize the critical nature of addressing marine debris for ecosystem sustainability and conservation.

2.1.2 *The Dataset*

The dataset forms the backbone of any project, with its selection and preprocessing being crucial for creating subsequent models. In this project we utilize a single type of dataset which is provided by the Department of Geosciences at the University of Malta.

This dataset consists of sea surface currents velocity data, recorded in hourly increments across four years, spanning from January 2020 to December 2023. These data points are derived from a model generated by high-frequency (HF) radar systems [11], located on the northern parts of the Maltese islands and southern Sicily. The locations of these radar systems, depicted in Figure 2.1 and identified from [12], provide a temporal snapshot of the sea surface currents movements.

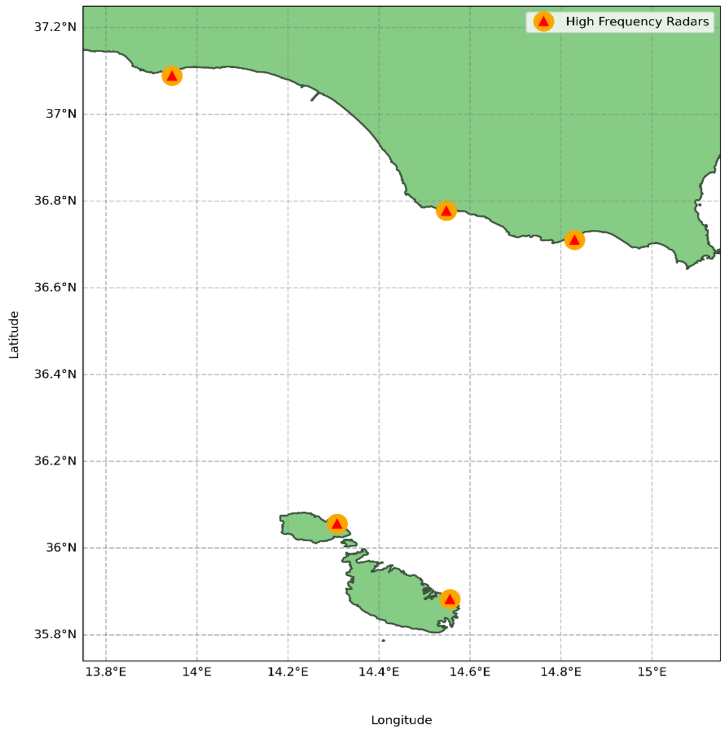

Figure 2.1 High Frequency Radars Locations

The data is composed of several variables including longitude, latitude, and time, coupled with eastern and northern sea current velocities; denoted as 'u' and 'v'. The variable 'u' signifies the east-west current component, indicating the velocity at which surface currents travel horizontally, either eastwards (positive 'u') or westwards (negative 'u'). Similarly, 'v' represents the north-south current component, denoting vertical movement towards the north (positive 'v') or south (negative 'v'). The data's geographical scope is defined within the boundaries of 14.15° to 14.81° longitude and 35.79° to 36.30° latitude. This coverage translates into a grid of 52 latitude points by 43 longitude points, for a total of 180 data points, as detailed in Figure 2.2. The dataset is in NetCDF format [13], a commonly used standard for climate and meteorological data, ensuring compatibility with the Lagrangian Model employed in the simulations.

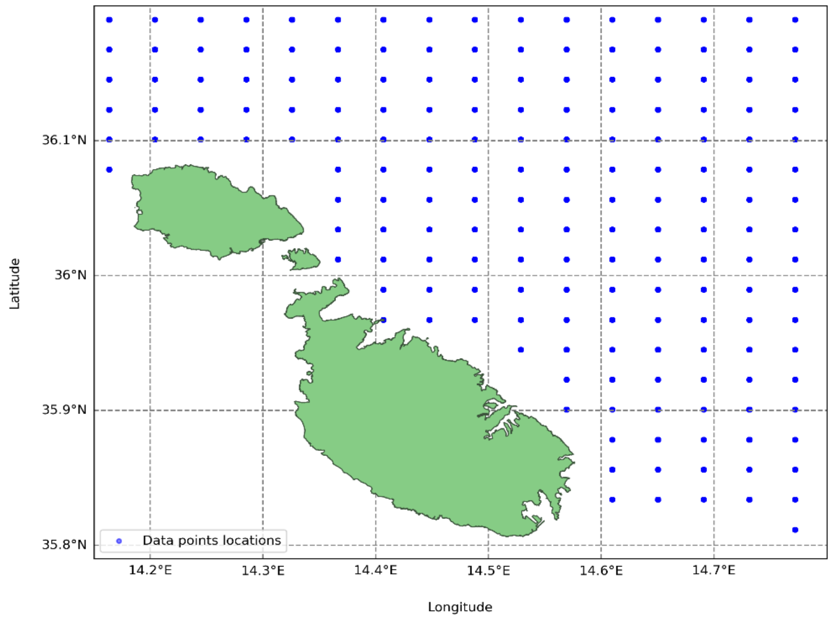

Figure 2.2 Radar Data Points Locations for Dataset

This dataset is integral to the project, providing comprehensive environmental parameters essential for the subsequent development of both the physics-based simulations and the Machine Learning models.

2.1.3 Physics-Based Lagrangian Model

The practice of tracking ocean surface movements in a Lagrangian framework dates back to the earliest days of oceanography, with early methods involving observing the drift of ships or the paths of specially designed floats to document current movement, as outlined by [14]. The physics based Lagrangian model [1] plays a pivotal role in environmental simulations. By offering a dynamic method to trace individual particle trajectories within fluid mediums, the model ensures precise tracking of the particle's temporal movement. Its broad applicability spans from localized studies, to complex, global-scale environmental systems, underlining its adaptability. This is evident in its varied applications, which include tracking oil spills diffusion [15], mapping floating plastic debris [16], simulating jellyfish migrations [17], analysing smoke dispersion [18], and many other.

The physics based Lagrangian model [1] operates by representing particles within a fluid medium, tracking their position and properties as they move with the fluid's flow. The model calculates the trajectory of each particle by integrating the velocity field of the fluid, which may vary in time and space. This approach enables the simulation of dispersal patterns of particles, such as marine debris by accounting for both advection and diffusion processes. Advection represents the movement of particles by the flow of the fluid [19]. Diffusion, on the other hand, models the dispersion of particles through random motion [19]. This is done by employing techniques such as random walks or Gaussian distributions. This inclusion of randomness enhances the realism of the simulation.

To facilitate these simulations, several Python toolkits like OceanParcels [20], PyGnome [21], and Flexpart [22] have been developed, each specifically designed for simulating the movement of particles using the Lagrangian model framework. These toolkits enable the customization and execution of particle tracking simulations, leveraging data on ocean currents, wind fields, and other environmental phenomena.

OceanParcels [20] is distinguished by several features that make it suitable for our project. One of its notable capabilities are custom kernels. These are user-defined functions that allow for tailored simulation scenarios at each time step. Through custom kernels, users can implement complex behaviours and interactions of particles within the fluid, such as particle reflection or response to environmental variables like temperature and wind. Another significant feature is particle initialization. This feature enables the creation of particles at specific locations, times, and with distinct properties, allowing for more detailed and accurate simulations.

All these attributes render OceanParcels [20] as an optimal choice for this project. By integrating these features, this toolkit facilitates the development of comprehensive simulations. This is crucial for understanding and predicting the movement of marine debris, thereby enhancing our strategies for marine conservation and debris management.

2.1.4 Time Series Modelling

Time series modelling is a technique used to predict future data points by analysing the trends, cycles, and patterns in a series of data points collected over an interval of time [23]. Its main focus is on analysing historical data to uncover the underlying structure of the dataset, which can then be used to forecast future trends. This method is particularly powerful for its ability to incorporate the sequence and time dependence within the dataset. By examining how values are interconnected over time, time series models can forecast future values based on the inherent temporal dynamics present in the historical data [24]. This form of predictive modelling assumes that past patterns are indicative of future behaviours, making it an indispensable tool in a variety of fields ranging from weather forecasting [25] to stock market predictions [26].

While time series modelling is a powerful tool for forecasting future data, it also encounters several limitations. Time series data often exhibit seasonality and trends, which can complicate the forecasting process [27]. Outliers, missing sequences of data, and anomalies can also significantly impact the accuracy of forecasting models, requiring careful identification, and handling. The capacity of these models to integrate external influential factors and variables is also somewhat limited, often necessitating the integration of additional features for enhanced predictive accuracy [28]. Additionally, time series models require significantly more data for training, which can be a challenge in situations where data is limited [28]. These challenges highlight the importance of adopting a methodical approach to time series modelling, emphasizing the need to carefully consider the specific context and characteristics of the data being analysed when utilizing time series models for effective forecasting.

In the context of this project, we harness time series modelling to predict sea surface current velocities. Accurate predictions require a detailed analyses of the data sequences to discern patterns that could forecast future predictions. The historical hourly data of surface currents form a time series, which is inherently continuous but sampled at discrete intervals. To address this, deep learning models, a subset of artificial neural networks, are employed due to their proficiency in handling vast amounts of sequential data and their capacity to learn complex temporal patterns [29]. Through training on past sea surface current data, these models are equipped to predict future values.

2.1.5 Deep Learning Models

Deep learning is a subset of machine learning that harnesses the power of ANNs to interpret and predict data through multiple layers [30]. Deep learning has revolutionized the way we approach complex problems by being able to detect intricate patterns from different types of data [30]. In the context of this FYP, deep learning models are pivotal in analysing and predicting the dynamic and complex patterns of sea surface current velocities. By employing models specific to sequential data processing like LSTM and GRU networks, this project aims to accurately predict the dispersion of marine debris around Malta's coastal waters, addressing both the temporal dynamics and spatial complexities inherent in sea current movements.

LSTM networks, a specialized type of RNN, are designed to address the challenge of learning long-term dependencies, overcoming the limitations faced by traditional RNNs, notably the vanishing gradient problem [31]. This challenge inhibits RNNs from effectively learning and retaining information over long sequences. LSTMs employ a unique architecture, characterized by a system of gates, namely the input, forget, and output gates. These gates collectively decide which information should be stored, discarded, or passed through, based on the relevance to the task at hand [32]. Memory cells within LSTMs retain information over long intervals, making them adept at managing sequences where understanding past context is crucial for future predictions [32]. This capability is pivotal for predicting sea surface currents, as demonstrated in this FYP. Their ability to remember previous information for extended durations without degradation makes them ideal for capturing the underlying patterns in historical data of sea surface currents, which is crucial for accurate prediction and subsequent debris dispersion simulation.

GRU networks are another variant of RNNs that aim to solve the vanishing gradient problem [31] but with a more simplified structure compared to LSTMs. GRUs simplify the LSTM model by combining the input and forget gates into a single update gate and merging the cell state and hidden state [32], [33]. This reduction in complexity leads to a model that is faster to train without significantly compromising the model's ability to capture dependencies in a sequence [32]. In the context of this project, GRUs are employed alongside LSTMs to forecast sea surface currents. Their efficiency and effectiveness in handling time series data render them adapt at predicting the movements of marine debris, offering a comparative perspective to the LSTM's performance.

In conclusion, LSTMs and GRUs distinguish themselves primarily through their structure and information processing: LSTMs offer a more detailed gating mechanism that excels in managing long-term dependencies, while GRUs provide a streamlined architecture that enables quicker training without significantly sacrificing performance

[33]. Their inherent capabilities make them exceptionally suited for time series modelling, where understanding and predicting sequential data patterns is crucial [29], thereby making them highly applicable to the objectives of this FYP. It is for these reasons that both models were leveraged in this project, utilizing their strengths to predict future sea surface current velocities from historical data. Their performances were also compared against one another, aiding in the accurate simulation of marine debris dispersion around Malta's coastal waters.

2.2 Literature Review

This section outlines the structure of the literature review, which is divided into three distinct subsections, each focusing on a critical aspect of marine debris dispersion and the methodologies employed to predict and simulate it. The first subsection delves into studies that forecast the movement and accumulation of marine debris. The second subsection highlights research that applies machine learning techniques to predict sea surface currents. The final subsection explores the integration between AI models predictions and physics-based models. The goal is to provide an overview of current methodologies in the field.

2.2.1 Prediction of Marine Debris Dispersal

The prediction of marine debris dispersal has significant impact on marine ecosystems. This is why researchers have explored multiple methodologies to understand and forecast the movement and accumulation zones of debris in marine environments. The variation in these approaches reflects the complexity of the problem, encompassing various methods that aim to capture the dynamic nature of marine debris movement. Through the implementation of numerical simulations [34], [35], deep learning techniques [36], and advanced simulation tools [37], [38], the field continues to evolve, seeking more accurate and efficient ways to predict debris dispersal patterns.

E. van Sebille et al. [34] delve into marine debris dispersal, focusing on the crucial role of numerical simulations in predicting and understanding the dispersion of marine debris. This study utilizes a number of numerical simulations that leverage various physical oceanographic phenomena to model the movement of floating marine debris. Central to their approach is the use of extensive datasets, capturing various environmental factors such as the velocity and direction of ocean currents, wind patterns, and wave dynamics. These variables are crucial for determining the dispersal patterns of marine debris.

In [34], numerical simulations, specifically, Eulerian and Lagrangian frameworks are employed. The Eulerian approach models plastics as tracers within a grid, focusing on the interaction between fluid and particle phases, incorporating turbulence through diffusivity parameterization. Conversely, the Lagrangian framework, preferred for its three-dimensional transport analysis, traces virtual particles using pre-computed velocity data, integrating stochastic terms to reflect the turbulence's impact on dispersion patterns. Both these methods highlight the significant influence of environmental phenomena have on debris movement, especially in nearshore processes. However, accurately simulating coastal dynamics and beaching patterns remains challenging.

Aligning with [35], E. van Sebille et al. [34] highlight the need for enhanced models that better capture surface interactions. Experiments conducted within [34] and [35] include the deployment of drifters and buoys equipped with GPS tracking, enabling the researchers to validate their simulation results against real-world data. These findings underscore the importance of integrating numerical simulations with empirical data to refine model accuracy and forecast reliability. Such efforts demonstrate the versatility and efficiency of numerical simulations and ultimately contribute to more effective mitigation and management strategies for marine pollution.

Deep Learning techniques such as computer vision have also emerged as powerful tools in addressing environmental challenges, notably in the management and mitigation of marine debris dispersal, as evidenced in [36]. These techniques offer approaches to interpret large datasets, enabling more precise and effective solutions to combat marine pollution.

This is illustrated in [36], where the authors utilize deep learning and object detection methodologies combined with remote sensing, to automate the identification and classification of marine debris across extensive coastal areas. Specifically, their research targets a stretch of 1900km along the Hawaiian coastline.

The research outlined in [36] carries out an evaluation of three distinct object detection models, demonstrating an analysis of each model's ability to tackle the complexities involved in detecting marine debris. The research is based on an extensive dataset comprising of 1587 image chips, which together contain 10,703 individual debris labels across various categories. The inclusion of data augmentation techniques further enhances the quality and reliability of the analysis.

Among the key findings, the Single Shot MultiBox Detector (SSD) paired with a MobileNet-v2 feature extractor (SS-MN) [39] stands out for its performance, achieving an average precision rate of 72%. This metric, along with other indicators employed in the [36], reinforces the practical viability and efficiency of leveraging deep learning for environmental monitoring.

The research also acknowledges some challenges and limitations, notably in maximizing recall rates to ensure minimal oversight of debris objects. Despite this, [36] stands as an effective demonstration of how deep learning and computer vision can be strategically deployed to tackle the issue of marine debris dispersal.

In our review of methodologies for modelling and simulating surface marine debris dispersal, we have come across numerous studies that employ the OceanParcels toolkit [20] to tackle issue of simulating marine debris. The work by MS. Yuniarti et al. [37] provides an insightful analysis of microplastic distribution patterns, which originate from the Seto Inland Sea and extend throughout Japanese waters. OceanParcels [20] is used to simulate these trajectories. A similar approach is employed in [38], where OceanParcels [20] and the Regional Ocean Modelling System (ROMS) activated with its built-in Lagrangian model, facilitated the tracking of river plume dispersal, highlighting the toolkit's versatility across different marine environments.

Utilizing Python-based OceanParcels [20], [37] and [38] delve into the trajectories of particles within marine environments, each focusing on a distinct subject under real-world conditions, offering insights into how these different particles navigate through varied marine environments throughout the year. The area simulated by MS. Yuniarti et al. [37], spanning latitudes 28° to 55°N and longitudes 120° to 160°E, was strategically selected to optimize the study's focus, similar to the approach in [38], where a specific polygon area on the northeast coast of Australia was identified.

In [37], data from the Hybrid Coordinate Ocean Model (HYCOM) provided the necessary current velocities for the simulations, while statistical analysis employing the RMSE method validated this data against in-situ observations. The validation process confirmed the suitability of the data for simulation inputs, with RMSE values indicating a close match to observed data, thus improving the accuracy of the simulation results. [38] used ROMS with a built-in Lagrangian model, incorporating wind fields from global models and recorded river volume discharges, akin to how [37] utilized HYCOM data to provide current velocities for the simulations.

The findings from [37] reveal that microplastic dispersion exhibits significant seasonal variations, with distinct pathways and accumulation zones becoming apparent in different seasons. The observed distribution patterns align with those documented within the review of previous studies by MS. Yuniarti et al. [37], affirming the reliability of the simulation approach utilized. Furthermore, this enables visualizations that illustrate the dispersal patterns of microplastics, enhancing the spatial and temporal dynamics of marine debris movement.

Both [37] and [38] address the challenges associated with tracking large numbers of particles across vast marine areas. These challenges highlight the need for advanced computational resources and methodologies to accurately simulate marine dispersal patterns. The insights from these studies are pivotal in demonstrating the importance of such approaches, which are essential for the development of effective mitigation strategies against marine pollution. This not only advances our understanding of debris trajectories but also establishes a standard for the application of advanced simulation tools like OceanParcels [20].

2.2.2 Machine Learning models for predicting sea surface currents

As highlighted in [40], ocean currents are a fundamental phenomenon within ocean hydrodynamics, having a significant influence on various marine processes. Numerous studies have turned to machine learning to unravel the intricacies of sea surface currents, crucial for understanding marine debris dispersion. By leveraging different algorithms, these studies offer new perspectives on marine environmental monitoring, demonstrating the potential of machine learning to provide accurate predictions of sea surface currents.

Dauji et al. [41] harness the capabilities of ANNs for the task of predicting ocean currents across multiple depths, not just the sea surface. Ali et al. [42] employ a similar approach of predicting ocean currents across multiple depths by using LSTM networks. These studies propose time series models to overcome the constraints inherent in numerical models, which necessitate extensive external information, substantial computational resources, and often struggle with noise and gaps in data.

Both studies highlight the challenge of accurately forecasting ocean currents in different regions. [41] focuses on two locations within the North Atlantic and North Pacific oceans. This dataset comprises of hourly records of current velocity and direction. These measurements were taken at depths of 18.3m and 460m, representing shallow and deep-water situations. On the other hand, Ali et al. [42] conduct their study in the Gulf of Mexico. The dataset includes measurements at 50 different depth levels, reaching down to 3000 meters below the surface, and spans horizontally from 88.5° W to 85° W and 24.65° N to 27° N.

In addressing the challenges posed by the data, [41] set up a feed-forward back-propagation network ANN architecture, which is recognized for its efficiency. The consideration of additional inputs, specifically currents from lower depths, was explored but ultimately showed no significant improvement to the model's prediction accuracy. In [42] a deep learning approach was employed using LSTM networks, chosen for their ability to handle long-term dependencies in data. Both studies explored the optimum length of past data segments for input, underscoring the temporal dynamics of sea currents.

[41] and [42] encountered and addressed several limitations. One limitation was the initial underprediction of extreme values in [41]. This issue was tackled by introducing methods for scaling target extreme values during training. Moreover, due to the high cost and complexity of collecting sea current data, both studies faced limitations in the availability of long-term observations.

The performance of the ANN models in [41] was evaluated quantitatively and qualitatively, showing high correlation coefficients and low error metrics (RMSE and MAE) across various testing durations and prediction intervals. The study also compared the ANN model performance with past works and the random walk model. Notably, the models maintained high performance for currents at both shallow and deep-water layers and were effective across different forecasting durations. The ANN models outperformed traditional forecasting methods, marking a significant improvement in predictive accuracy. The performance of the LSTM models was evaluated using similar error metrics, including RMSE, Peak Signal to Noise Ratio (PSNR), and Structural Similarity (SSIM).

The studies [41], [42], validate the potential of deep learning models as formidable tools for real-time prediction of ocean currents. Their success in surpassing traditional forecasting methods advocate for the integration of deep learning for the prediction of sea currents.

In [43], Zulfa et al. investigated the potential of LSTM networks for predicting the velocity and direction of sea surface currents in Labuan Bajo, Indonesia. Given Labuan Bajo's significance as a pivotal point for trade and tourism, the study aimed to improve maritime navigation and safety through precise forecasts of sea currents, highlighting the crucial need for dependable prediction techniques.

To conduct this study, Zulfa et al. [43] utilized a dataset consisting of hourly sea surface current velocities collected by the Perak Maritime Meteorology Station II. This dataset, comprised of 24 data points, captures the sea surface current velocities at a single geographical point. Before applying any predictive modelling, the data underwent preliminary preprocessing, which included normalization using the Min-Max method. This step was crucial for adjusting the data values to a common scale, thereby facilitating the subsequent training of the predictive model.

The choice of LSTM as the predictive model was driven by its proven effectiveness in handling time-series data, making it particularly suited for forecasting tasks such as predicting sea surface current velocities. Zulfa et al. [43] faced certain limitations, particularly the challenge of applying LSTM to short-term datasets. These models typically excel with long-term data, benefiting from extensive datasets to learn underlying patterns effectively.

In evaluating the predictive performance of the LSTM model, the Mean Absolute Percentage Error (MAPE) metric was utilized. MAPE measures the accuracy of forecasted values compared to actual values, calculating the average of absolute percentage errors. The study achieved low MAPE values for the U and V components of sea surface currents—14.15% and 8.43%, respectively. These results were obtained with an LSTM model configured with 50 hidden layers, a batch size of 32, and a learning rate drop period of 150, indicating a good level of accuracy in the predictions.

The research concluded that using LSTM networks with specific parameter configurations, serves as a reliable tool for predicting the velocity and direction of sea surface currents. However, [43] also suggest that further exploration into methods more suited to short-term data or the inclusion of seasonal variations and tidal factors in the analysis could enhance predictive accuracy.

Bayindir [40] has a similar approach to [43], where the focus is also on using LSTMs to predict ocean currents velocities. This choice is motivated by the LSTM's capability to capture long-term dependencies in sequential data, a common characteristic of ocean current velocities.

The study uses a dataset collected by the National Oceanic and Atmospheric Administration (NOAA) in Massachusetts Bay, covering the period from November 2002 to February 2003, with measurements taken at 23.5m depth and recorded at intervals every 3 minutes and 44 seconds. This dataset, consisting of the current speed in two directions (u and v), undergoes preprocessing to standardize the data, ensuring zero mean and unit variance.

The methodology section stands out by providing a clear and concise explanation of how LSTM networks operate, including their sequence-to-sequence regression capability, which is central to predicting future states of ocean currents. Bayindir evaluates the LSTM model's performance by employing the RMSE error metric. This metric offers a quantitative measure of the model's accuracy, providing a direct comparison between the predicted and actual current speeds.

The results from [40] demonstrate the LSTM model's ability to make accurate predictions, with significant improvements observed when the model incorporates real-time data updates. Initially, even without these updates, the LSTM model shows a strong capacity for predicting ocean currents, suggesting it can make reliable forecasts within a few future time steps. This is highlighted by the model's predictions exhibiting a higher frequency peak compared to the actual observed data, indicating a solid baseline accuracy.

However, the research further reveals that when the model is refined with observed values, essentially updating it with real data, the accuracy of predictions markedly increases. This aspect underscores a common hurdle in machine learning and deep learning applications, where the quantity and quality of historical data can significantly impact the accuracy of the predicted values.

H.-M. Choi et al. [44], develop LSTM networks to predict sea surface temperatures (SSTs) near the Korean Peninsula. The aim is to mitigate the impacts of rising SSTs due to global warming on marine ecosystems and aquaculture. The LSTM models demonstrate promising results in predicting SSTs and identifying high water temperature events with high accuracy for short-term forecasts.

[44] acknowledges limitations such as decreased prediction accuracy for longer-term forecasts and a reliance solely on SST data without considering other environmental factors. The evaluation of the model's performance was done through metrics like R2, RMSE, MAPE, and F1 scores. These metrics collectively assess the model's accuracy and its capability in classifying high water temperature events. The results indicate promising accuracy, particularly for short-term predictions up to four days in advance. However, the model's accuracy decreases for longer prediction windows, highlighting a critical area for improvement.

In H.-M. Choi et al. [44], the goal of predicting sea surface temperatures (SSTs) across a grid of 1519 data points near the Korean Peninsula closely reflects the task we face in forecasting sea surface current velocities for various data point locations, which will subsequently be utilized as inputs into a Lagrangian model. By training a predictive model on a 12-year dataset for each data point and then mapping the predictions for subsequent days, it showcases a structured methodology for accurate environmental prediction. This process can be effectively applied to predict conditions across multiple locations in a marine area, paving the way for other applications.

2.2.3 Model integration with Physics-Based Lagrangian Model

In [45], J. Mansui et al. set out to explore the dispersal of floating macro litter across the Mediterranean Sea by employing a two-stage modelling approach to achieve their objective. Initially, they utilized the NEMO Oceanic General Circulation Model, configured specifically for the Mediterranean basin, to simulate the sea state and velocity fields necessary for the drift simulations. This model configuration allows for a fine-scale representation of the region's oceanic conditions, which is crucial for accurate simulation of ocean currents and phenomena. The velocity outputs from this model are taken as daily averages, providing a consistent long-term description of the surface conditions.

Following the generation of these velocity fields, the second stage involves Lagrangian simulations. These simulations use the data from the Oceanic General Circulation Model to simulate the movement of virtual particles that mimic the behaviour of floating macro litter at the sea surface. To specifically observe the surface transport pathways of the floating macro litter, the method focuses solely on surface movements without considering vertical dynamics or windage. The integration of these two models enables a comprehensive simulation of debris movement and accumulation.

[45] produced significant results, demonstrating seasonal and regional variations in floating macro litter distribution across the Mediterranean Sea. These findings were visualized to illustrate accumulation zones, offering a dynamic view of how marine debris disperses over time. The results aligned well with empirical data from previous studies, reinforcing the model's validity and effectiveness. J. Mansui et al. conclude that the integration of Lagrangian simulations with the OGCM offers a powerful framework for predicting marine litter distribution, highlighting its accuracy and reliability.

In [46], the primary focus is to evaluate floating marine litter within the Northwest Pacific region. This is done by employing an approach that integrates different models with a physics based Lagrangian model, aiming to enhance the understanding and management of marine litter trajectories.

The methodology employed in this report is noteworthy for the integration of Eulerian models with Lagrangian particle tracking to predict and analyse the behaviour and dispersion patterns of marine litter. Eulerian models provide crucial data on ocean currents and winds by solving fluid dynamics equations on a fixed grid, establishing the environmental values that will be later utilized in the pipeline. Building upon this, the Lagrangian model simulates the trajectories of individual particles as they navigate through the ocean's dynamic conditions like currents and winds, which are determined by the Eulerian outputs. Through various case studies conducted within the Northwest Pacific region, the report demonstrates the effectiveness of this combined methodology in predicting not only the movement but also the deposition areas of marine litter, providing valuable insights into effective management and mitigation strategies.

[46] also addresses some challenges inherent in these models. A significant challenge highlighted is that the integration of wind-induced leeway drift poses discrepancies between observed and modelled trajectories, particularly under conditions of strong winds. Overall, The results derived from the applied models are largely successful, providing visual maps and simulations that depict litter trajectories and accumulation zones. These visualizations serve as crucial tools for understanding the impact of physical factors like currents and winds on the distribution of debris.

2.2.4 Summary

3 Methodology

This section delves into the detailed approaches taken to achieve the project's objectives, highlighting the reasoning behind each decision, and offering perspectives on the design and implementation of each objective.

3.1 Data Integration and preprocessing

The pre-processing of raw data files is a crucial first step in any project. In this FYP, we dealt with raw historical NetCDF [13] data spanning a total of four years, from January 2020 to December 2023 in hourly increments. The data was split into multiple folders and subfolders for each day, necessitating a robust method to merge and preprocess the data without interfering with its temporal and spatial dimensionality.

To address this, we developed a framework that allows us to specify the start and end dates for the required merging of the sea surface currents data. The framework then merges the individual files along the time dimension, creating a single comprehensive dataset that encompasses all relevant data across the specified interval. This merged dataset is not only more manageable but also streamlined for any subsequent processes.

A key feature of our framework is the preservation of the geographical boundaries and temporal aspects of the data. The dataset maintains the latitude and longitude ranges, ensuring the spatial integrity of the data is uncompromised. Similarly, checks were performed to ensure the time remained consistent, preserving the temporal integrity of the data.

Upon analysing the data, a substantial number of NaN values were discovered. These NaNs are likely due to the proximity of the data to the coast, where high-frequency radars often struggle to capture all the data accurately. We decided not to address these NaN values at this stage, as each project objective requires specific handling of missing data, details of which will be explored in the respective sections.

This preprocessing framework was utilized in every section of the project, from the Lagrangian simulations, the AI models' training and also the project's evaluation. This underscores the critical role of preprocessing throughout the whole project.

3.2 Lagrangian Model Development

The OceanParcels [20] was utilized for this part of the project. The first step before proceeding with the simulation is to open previously pre-processed sea surface current dataset. Next, the shapefile [] of Malta is loaded and used to create the land-sea mask. The land-sea mask in Figure 3.1 is generated by rasterizing [] the coastline shapefile. This mask distinguishes between land and sea areas, allowing for accurate particle tracking near coastal regions. The mask is saved as a NetCDF file and added within the grid boundaries to match with the boundaries of the data set. The Malta shapefile was also used for all subsequent visualisations. These coastline boundaries are essential for defining the simulation area and implementing land-sea interactions.

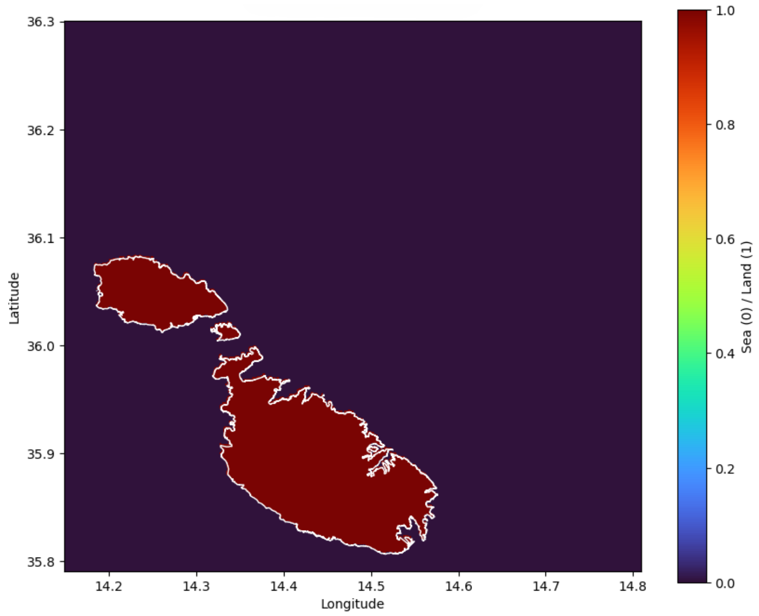

Figure 3.1 Land-sea mask of Malta

Next, a FieldSet is created from the sea surface current dataset. It serves as the simulation environment, defining the velocity fields that drive particle movement. The land-sea mask is also incorporated into the FieldSet as an additional field, providing the necessary data for reflecting or deleting particles upon reaching the coastline. Next, the simulation particles are initialized near a specified location 36.0475 and 14.5417 with random offsets to simulate a dispersed release as seen in figure 3.2. The particles represent the objects of interest, such as sea surface debris, whose movements are to be simulated. Originally, we were 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 simulate how clusters of debris move.

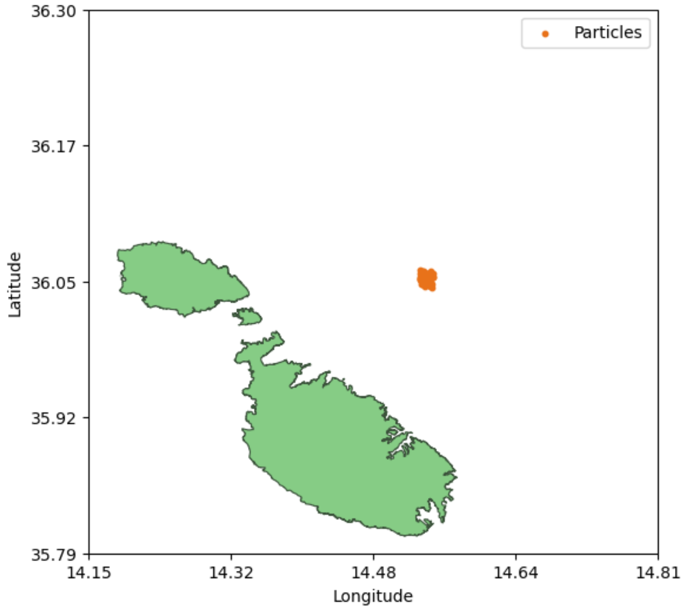

Figure 3.2 Initial Particle Locations

Next, a very important part is the implementation and creation of custom kernels. Custom kernels are defined to introduce specific behaviours into the simulation, modelling realistic scenarios that particles may encounter:

- *CheckOutOfBounds*: Removes particles from the simulation if they move beyond the defined boundaries. This prevents particles from straying outside the simulation domain.
- *CheckError*: Deletes particles encountering computational errors, specifically those with state codes above 50. This ensures the simulation proceeds without disrupted or incorrect particle data.
- *ReflectiveParticle*: Enhances particle tracking by recording additional variables such as previous positions and elapsed time since the start of the simulation.
- *UpdateElapsedTime*: Increases each particle's elapsed time by the absolute value of its time step, offering a time-dependent analysis feature by showing how long a particle has been in the system.

- *UpdatePreviousPosition*: Captures the current position of particles before they move. This is crucial for the reflection logic, ensuring accurate position tracking.
- *ReflectOnLand*: Applies a reflection behaviour when particles encounter land, as indicated by the land-sea mask, preventing them from passing through terrestrial areas. Additionally, introduces a probability that particles may 'beach' or adhere to the land, symbolizing deposition, or entrapment. []

The simulation is then executed, and the output of the simulation is visualized to assess the movement and dispersion of particles over time. This visualization provides insights into the particles' trajectories and how they interact with the environment, particularly in response to ocean currents. The timestep of the Lagrangian simulation is set at every 10 minutes, this means… This is saved as a GIF file, providing a dynamic and accessible way to observe the simulated dispersion over time.

There were also some limitations encountered during this. 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 0.1). 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 of 7 days and make a prediction for of 1 day instead. The 7-day simulation was done just to experiment and understand exactly how the Lagrangian model works and how the input data for the Lagrangian needs to be. So in this case the pre processing code was used to merge the files from 1st jan 2023 till 7th jan 2023.

When it came to handling the missing data, when we tried to interpolate the data (for the 7 days) we got very different visualisation results for the data with nans and data that was interpolated. We also got very similar results, where the particle movement was always the same even if we tried different dates. Basically, the interpolated data (tried both linear and spline) gave us the same result regardless of the time frame of the data! The decision remains that nan values will not be removed for the Lagrangian Simulation! The decision not to remove nan values for the

Lagrangian Simulation can be informed by the fact that interpolation, whether linear or spline, has been consistently homogenizing the data, leading to non-representative or misleading results regarding particle movements across different timeframes. This uniformity post-interpolation can mask the true variability and dynamic nature of the original data, which is critical in accurately simulating and understanding the drifters' behaviour and movement in varying oceanic conditions. Hence, retaining nan values preserves the integrity of the original dataset, ensuring that the Lagrangian Simulation more accurately reflects the real-world conditions and variations that the debris would encounter.

The overall goal of this model was to help us understand better and make better decisions when it comes to the implementation of the AI models. Goal was to explore and see how it works so we know what is to be expected from the AI model predictions. This was an exploration.

4 Evaluation

4.1 Writing the Evaluation Chapter

The evaluation component of an FYP is critical.

- *One has to make sure and explain why all tests used to evaluate the system are relevant, using evidence from the literature about similar systems, and justifying any deviations from standard approaches.*
- *Demonstration that system works as intended (or not, as the case may be).*
- *Include comprehensible summaries of the results of all critical tests that have been made.*
- *The student must also critically evaluate the system in the light of these tests results, describing its strengths and weaknesses.*
- *Ideas for improving it can be carried over into the Future Work section.*
- *Comparison of practical with theoretical results and their interpretation.*
- *Comparison with published work when available.*

5 Conclusion

5.1 Writing the Conclusions Chapter

The Conclusions section should be a summary of the project and a restatement of its main results, i.e. what has been learnt and what it has achieved. An effective set of conclusions should not introduce new material. Instead, it should draw out, summarise, combine, and reiterate the main points that have been made in the body of the report and present opinions based on them.

5.2 Writing the Future Work Chapter

Whether by the end of the project all the original aims and objectives have been completed or not, there is always scope for future work. Also, the ideas will have evolved during the project beyond the original target. The Future Work section is for expressing these ideas.

References

References

[1] C. Kehl *et al*, "Efficiently simulating Lagrangian particles in large-scale ocean flows — Data structures and their impact on geophysical applications," *Comput. Geosci.*, vol. 175, pp. 105322, 2023. Available: https://www.sciencedirect.com/science/article/pii/S0098300423000262. DOI: 10.1016/j.cageo.2023.105322.

[2] G. Suaria and S. Aliani, "Floating debris in the Mediterranean Sea," *Mar. Pollut. Bull.*, vol. 86, *(1)*, pp. 494-504, 2014. Available: https://www.sciencedirect.com/science/article/pii/S0025326X14004056. DOI: 10.1016/j.marpolbul.2014.06.025.

[3] M. Compa *et al*, "Risk assessment of plastic pollution on marine diversity in the Mediterranean Sea," *Sci. Total Environ.*, vol. 678, pp. 188-196, 2019. Available: https://www.sciencedirect.com/science/article/pii/S0048969719318984. DOI: 10.1016/j.scitotenv.2019.04.355.

[4] J. Mansui *et al*, "Predicting marine litter accumulation patterns in the Mediterranean basin: Spatio-temporal variability and comparison with empirical data," *Prog. Oceanogr.*, vol. 182, pp. 102268, 2020. Available: https://www.sciencedirect.com/science/article/pii/S0079661120300069. DOI: 10.1016/j.pocean.2020.102268.

[5] P. G. Ryan, "A brief history of marine litter research," in *Marine Anthropogenic Litter*, M. Bergmann, L. Gutow and M. Klages, Eds. 2015, Available: https://doi.org/10.1007/978-3-319-16510-3_1. DOI: 10.1007/978-3-319-16510-3_1.

[6] P. R. Pawar, S. S. Shirgaonkar and R. B. Patil, "Plastic marine debris: Sources, distribution and impacts on coastal and ocean biodiversity," 2016.

[7] S. Katsanevakis, "Chapter 2 - marine debris, a growing problem: Sources, distribution, composition, and impacts," in Anonymous New York: Nova Science Publishers, 2008, pp. 53-100.

[8] D. W. Laist, "Impacts of marine debris: Entanglement of marine life in marine debris including a comprehensive list of species with entanglement and ingestion records," in *Marine Debris: Sources, Impacts, and Solutions*, J. M. Coe and D. B. Rogers, Eds. 1997, Available: https://doi.org/10.1007/978-1-4613-8486-1_10. DOI: 10.1007/978-1-4613-8486-1_10.

[9] C. M. Rochman *et al*, "The ecological impacts of marine debris: unraveling the demonstrated evidence from what is perceived," *Ecology*, vol. 97, *(2)*, pp. 302-312, 2016. . DOI: 10.1890/14-2070.1.

[10] P. Agamuthu *et al*, "Marine debris: A review of impacts and global initiatives," *Waste Manag. Res.*, vol. 37, *(10)*, pp. 987-1002, 2019. . DOI: 10.1177/0734242X19845041.

[11] J. Harlan *et al*, "The Integrated Ocean Observing System High-Frequency Radar Network: Status and Local, Regional, and National Applications," *Marine Technology Society Journal*, vol. 44, *(6)*, pp. 122-132, 2010. Available: https://www.ingentaconnect.com/content/mts/mtsj/2010/00000044/00000006/art00017. DOI: 10.4031/MTSJ.44.6.6.

[12] Anonymous.*"Portus 3.0."* portus.research.um.edu.mt. https://portus.research.um.edu.mt/?p=13.833 (accessed Apr 06, 2024).

[13] Anonymous.*"UNIData | NETCDF."* unidata.ucar.edu. https://www.unidata.ucar.edu/software/netcdf/ (accessed Apr 06, 2024).

[14] E. van Sebille *et al*, "Lagrangian ocean analysis: Fundamentals and practices," *Ocean Modelling*, vol. 121, pp. 49-75, 2018. Available: https://www.sciencedirect.com/science/article/pii/S1463500317301853. DOI: 10.1016/j.ocemod.2017.11.008.

[15] S. A. Lonin, "Lagrangian model for oil spill diffusion at sea," *Spill Science and Technology Bulletin*, vol. 5, *(5-6)*, pp. 331-336, 1999. Available: http://inis.iaea.org/search/search.aspx?orig_q=RN:32026786.

[16] L. C. -. Lebreton, S. D. Greer and J. C. Borrero, "Numerical modelling of floating debris in the world's oceans," *Marine Pollution Bulletin*, vol. 64, *(3)*, pp. 653-661, 2012. Available: https://dx.doi.org/10.1016/j.marpolbul.2011.10.027. DOI: 10.1016/j.marpolbul.2011.10.027.

[17] M. N. Dawson, A. S. Gupta and M. H. England, "Coupled biophysical global ocean model and molecular genetic analyses identify multiple introductions of cryptogenic species," *Proceedings of the National Academy of Sciences*, vol. 102, *(34)*, pp. 11968-11973, 2005. Available: https://doi.org/10.1073/pnas.0503811102. DOI: 10.1073/pnas.0503811102.

[18] D. Hertwig *et al*, "Development and demonstration of a Lagrangian dispersion modeling system for real-time prediction of smoke haze pollution from biomass burning in Southeast Asia," *Journal of Geophysical Research. Atmospheres*, vol. 120, *(24)*, pp. 12605-12630, 2015. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/2015JD023422. DOI: 10.1002/2015JD023422.

[19] R. G. Williams and M. J. Follows, *Ocean Dynamics and the Carbon Cycle: Principles and Mechanisms.* 2011Available: https://www.cambridge.org/core/product/31EF28FEF48A172FF746B3E654F9455A. DOI: 10.1017/CBO9780511977817.

[20] Anonymous.*"OceanParcels."* oceanparcels.org. https://oceanparcels.org (accessed Apr 06, 2024).

[21] Anonymous.*"PyGNOME."* gnome.orr.noaa.gov. https://gnome.orr.noaa.gov/doc/pygnome/index.html (accessed Apr 06, 2024).

[22] I. Pisso *et al*, "The Lagrangian particle dispersion model FLEXPART version 10.4," 2019. Available: http://hdl.handle.net/11250/2634384. DOI: 10.5194/gmd-12-4955-2019.

[23] R. Adhikari and R. K. Agrawal, "An Introductory Study on Time Series Modeling and Forecasting," vol. abs/1302.6613, 2013. Available: https://api.semanticscholar.org/CorpusID:17070340.

[24] T. Raicharoen, C. Lursinsap and P. Sanguanbhokai, "Application of critical support vector machine to time series prediction," in . DOI: 10.1109/ISCAS.2003.1206419.

[25] S. Raksha *et al*, "Weather forecasting framework for time series data using intelligent learning models," in . DOI: 10.1109/ICEECCOT52851.2021.9707971.

[26] A. Chatterjee, H. Bhowmick and J. Sen, "Stock price prediction using time series, econometric, machine learning, and deep learning models," in . DOI: 10.1109/MysuruCon52639.2021.9641610.

[27] P. Wang *et al*, "Interval time series forecasting: A systematic literature review," vol. 43, *(2)*, pp. 249-285, 2024. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3024. DOI: 10.1002/for.3024.

[28] S. Jadon, J. Milczek and A. Patankar, "Challenges and approaches to time-series forecasting in data center telemetry: A survey," Cornell University Library, arXiv.org, Ithaca, Feb 11,. 2021.

[29] A. Alsharef *et al*, "Time Series Data Modeling Using Advanced Machine Learning and AutoML," vol. 14, *(22)*, 2022. . DOI: 10.3390/su142215292.

[30] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," vol. 2, *(6)*, pp. 420, 2021. Available: https://doi.org/10.1007/s42979-021-00815-1. DOI: 10.1007/s42979-021-00815-1.

[31] S. Hochreiter, "The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions," vol. 6, *(2)*, pp. 107-116, 1998. Available: http://www.worldscientific.com/doi/abs/10.1142/S0218488598000094. DOI: 10.1142/S0218488598000094.

[32] M. J. Hamayel and A. Y. Owda, "A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms," vol. 2, *(4)*, pp. 496, 2021. . DOI: 10.3390/ai2040030.

[33] P. T. Yamak, L. Yujian and P. K. Gadosey, "A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting," pp. 49–55, 2020. Available: https://doi.org/10.1145/3377713.3377722 http://dx.doi.org/10.1145/3377713.3377722. DOI: 10.1145/3377713.3377722.

[34] E. van Sebille *et al*, "The physical oceanography of the transport of floating marine debris," *ERL*, vol. 15, *(2)*, pp. 23003-32, 2020. Available: https://iopscience.iop.org/article/10.1088/1748-9326/ab6d7d. DOI: 10.1088/1748-9326/ab6d7d.

[35] B. D. Hardesty *et al*, "Using Numerical Model Simulations to Improve the Understanding of Micro-plastic Distribution and Pathways in the Marine Environment," vol. 4, 2017. Available: https://search.proquest.com/docview/2307448993. DOI: 10.3389/fmars.2017.00030.

[36] W. R. Winans *et al*, "Large-area automatic detection of shoreline stranded marine debris using deep learning," vol. 124, pp. 103515, 2023. Available: https://www.sciencedirect.com/science/article/pii/S1569843223003394. DOI: 10.1016/j.jag.2023.103515.

[37] MS. Yuniarti *et al*, "Trajectory mapping of microplastics originating from the Seto Inland Sea, Japan," vol. 16, *(6)*, pp. 3138-3149, 2023. Available: http://www.bioflux.com.ro/aacl.

[38] S. Aijaz, F. Colberg and G. B. Brassington, "Lagrangian and Eulerian modelling of river plumes in the Great Barrier Reef system, Australia," vol. 188, pp. 102310, 2024. Available: https://www.sciencedirect.com/science/article/pii/S1463500323001506. DOI: 10.1016/j.ocemod.2023.102310.

[39] W. Liu *et al*, "SSD: Single shot MultiBox detector," in *Computer Vision – ECCV 20*, 16, .

[40] C. Bayindir, "Predicting the Ocean Currents Using Deep Learning," vol. 13, *(1)*, pp. 373-385, 2023. Available: https://jaem.isikun.edu.tr/web/images/articles/vol.13.no.1/34.pdf.

[41] S. Dauji, M. C. Deo and K. Bhargava, "Prediction of ocean currents with artificial neural networks," vol. 21, *(1)*, pp. 14-27, 2015. Available: https://doi.org/10.1080/09715010.2014.938133. DOI: 10.1080/09715010.2014.938133.

[42] A. Muhamed Ali *et al*, "A Deep Learning Model for Forecasting Velocity Structures of the Loop Current System in the Gulf of Mexico," vol. 3, *(4)*, pp. 953, 2021. . DOI: 10.3390/forecast3040056.

[43] I. I. Zulfa *et al*, "Prediction of Sea Surface Current Velocity and Direction Using LSTM," vol. 11, *(1)*, pp. 93-102, 2021. Available: https://doi.org/10.22146/ijeis.63669. DOI: 10.22146/ijeis.63669.

[44] H. Choi, M. Kim and H. Yang, "Deep-learning model for sea surface temperature prediction near the Korean Peninsula," vol. 208, pp. 105262, 2023. Available: https://www.sciencedirect.com/science/article/pii/S0967064523000127. DOI: 10.1016/j.dsr2.2023.105262.

[45] J. Mansui *et al*, "Predicting marine litter accumulation patterns in the Mediterranean basin: Spatio-temporal variability and comparison with empirical data," *Prog.Oceanogr.*, vol. 182, pp. 102268, 2020. Available: https://www.sciencedirect.com/science/article/pii/S0079661120300069. DOI: 10.1016/j.pocean.2020.102268.

[46] E. P. United Nations, "Review and Analysis of Floating Marine Litter Prediction Models in the NOWPAP Region - Technical Report No. 36," 2018. Available: https://wedocs.unep.org/20.500.11822/26237.