Institute for Marine and Atmospheric research Utrecht (IMAU)

FACULTY OF SCIENCE

Climate Physics

Beached Debris from Fisheries at the Galápagos Islands: can we identify sources?

Ocean Pathways, Source Identification, and an Investigation into the Predictability of Fishing Activity

Author:
Ina Nagler

Supervisors:
Stefanie Ypma
Erik van Sebille
Anna von der Heydt

May, 2022

Marine debris from fisheries is a major source of marine pollution and significantly endangers ecosystems. One possible moment to remove this debris is when it arrives at its shores. Being able to connect beached debris to its sources is a substantial step in encouraging mitigation measures. This research presents a framework that allows for the source identification of debris from fisheries at beaches. Being able to predict areas where fishing activity is taking place could help to anticipate where and when debris from fisheries might beach. This study presents an overview of current methods and practices for the prediction of fisheries and investigated the relationship between net primary production and fishing activity. The Galápagos Islands, which are known for their pristine ecosystems and endemic biodiversity, are threatened by debris beaching at their shores. Island authorities face the challenge of effectively removing this debris. Efficient mitigation measures could be enforced with knowledge of the sources of debris. We show that debris from fisheries contributes to the littering of debris on the islands' shores. An analysis of the beaching distribution of different categories of fishery sources demonstrates that debris from fisheries sailing under Ecuadorian, Chinese and Peruvian flags is most likely to beach. About 50% of the beached debris is released within the Exclusive Economic Zone of the Galápagos Islands, and 23 % alone stem from International Waters in the South East of the Islands. The review of existing literature concludes that oceanographic parameters like primary production are often successful in the prediction of fishery production. However, the analysis of the relationship between net primary production and fishing effort was not conclusive. This source identification frame is universally applicable and can be implemented for island nations and archipelagos worldwide. It allows for determining the beaching distribution of different categories of fishery sources. In this, it provides support for the implementation of policies reducing fishing debris pollution.

Contents

1 Introduction

Levels of marine debris are rapidly increasing on a global scale. This poses a complex and multi-dimensional predicament requiring fast and suitable measures to ensure protection for the marine environment. Marine debris, manufactured or processed solid material that is discarded into the marine environment (Law, 2017), harms marine food webs, ecosystem services, and coastal economies (Kershaw, 2016). Eriksen et al. (2014) estimate that of the total mass of floating marine debris located in the open ocean, 70% is Abandoned, Lost, or Discarded Fishing Gear (ALDFG). Already in 1930, researchers reported the entanglement of seals in rubber bands (O'hara et al., 1988). In 1988, O'hara et al. (1988) portrayed how the introduction of plastic fishing gear contributed to plastic being the "most common man-made objects sighted at sea". With the expansion of fishing activity in the last decades and the increasing use of synthetic materials, the quantity, distribution, and detrimental effects of ALDFG have presumably increased (Kuczenski et al., 2022). Marine debris endangers marine habitats in a multitude of ways. Marine organisms can entangle themselves or ingest it (Andrades et al., 2018; Page et al., 2004). The chance of entanglement is high since ALDFG is made to catch marine organisms (Kuczenski et al., 2022). Additionally, it damages in-use fishing gear by causing more gear loss, poses navigational hazards, and creates safety risks at sea (Kershaw, 2016).

There has been increasing international attention to the problem of ALDFG. The scale of impacts of debris from fisheries has instigated the United Nations (UN) to appeal to their member countries to take actions to reduce ALDFG (Richardson et al., 2019). This appeal is highlighted in the UN 2030 Agenda for Sustainable Development Goal 14, which calls for members to regulate destructive fishing practices and significantly reduce marine pollution (Virto, 2018). To minimize sea-based sources of pollution the International Convention for the Prevention of Pollution from Ships (MARPOL) Annex V has been created and signed by more than 150 countries (Organisation, 2014). It prohibits the discharge of all garbage into the sea. All ships flagged under countries that are signatories to MARPOL are subject to its requirements, regardless of where they sail. However, there is limited monitoring of MARPOL, and little information exists about illegal pollution activities by vessels at sea (Richardson et al., 2017).

Although plastic debris is an important and growing source of pollution with numerous impacts, our understanding of its sources and trends is deficient (van Sebille et al., 2020). Due to its spatial and temporal heterogeneity, the execution of observations of marine debris is complex and challenging (Ryan et al., 2009). The variety of ALDFG studies, concerning differing data types, fisheries, and geographic areas, have made it difficult to offer stringent estimates for fishing gear losses on a global scale. Richardson et al. (2021) conducted the first statistically rigorous analysis and estimated that 5.7% of all fishing nets, 8.6% of all traps, and 29% of all lines are lost each year. An erroneous global estimate stating that 640,000 tonnes of fishing gear are lost annually has been perpetuated by scientists, media, and international bodies such as the UN (Richardson et al., 2021). This can hinder support for what is perceived as an answered question. It is important to have accurate knowledge about the sources of fishing debris as these can assist in the implementation of source-reducing measures. They can inform management interventions and encourage mitigation initiatives at an effective scale to decrease the potential risks of plastics in the marine ecosystem.

To identify the sources of plastic debris, accurate knowledge of how plastic travels can be essential. This is especially the case when few observations are available or analysis of labels on debris does not allow for their origin to be identified (Schofield et al., 2020). Modeling approaches

using virtual floating marine debris transported on ocean surface currents can be useful to gain a better understanding of how plastic is transported. This understanding is vital to implementing efficient mitigation measures (van Sebille et al., 2020). Lagrangian particle tracking techniques have turned out to be an effective approach to simulating floating marine litter trajectories (Ruiz et al., 2022). They allow for insights into circulation patterns and can support the identification of accumulation zones. Using pathways and connectivity between the regions of interest allows for the analysis of the origin and fate of virtual floating particles.

An example of a region that faces the threat of marine debris is the Galápagos Islands. These constitute an isolated archipelago in the Pacific Ocean, 1000 kilometers west of Ecuador. The islands are renowned for their unique wealth of endemic species and today underpin one of the largest UNESCO World Heritage Sites and Marine Reserves on Earth. They are situated at the confluence of multiple ocean currents: the Humboldt Current from the south-east, the Panama Current from the north-east, the eastward-flowing Equatorial Counter Current, and the Pacific Equatorial Undercurrent. Upwelling brings nutrient-rich water to the surface and results in an area of extremely high productivity (P. J. Jones, 2013).

Impacts from marine pollution can have extensive ecological and socioeconomic consequences on the Galápagos Islands. Already stressed under the influence of global warming, and projections stating that the eastern equatorial Pacific will experience increasingly frequent and severe climate extremes in coming decades, the Islands can be further burdened by exceeding amounts of marine debris (J. S. Jones et al., 2021). Coastal clean-up efforts show that much of the plastic found in the Galápagos comes from mainland America, continental Asia, and fisheries in the Pacific Ocean (van Sebille et al., 2019). An estimate by the Galápagos Conservation Trust found that only less than 10% of the trash at the shores of the Galápagos Islands stems from the towns in Galápagos and the rest, about 90%, is either from fisheries operating in waters near the Galápagos or transported via ocean currents. An archaeological analysis of debris found at beaches indicates that maritime sources, especially fisheries, are likely a significant contributor (Schofield et al., 2020). Research to date has not yet determined which fisheries play a role in debris beaching at the islands. This study investigates different kinds of sources contributing to trash at the Galápagos shores.

The high productivity in the area around Galápagos Islands attracts various marine species. This results in a high presence of industrial fishing. Being able to predict areas where fishing activity may take place is a vital step for analyzing the pathways of debris being released from these fisheries. In this, it can help to predict where and when debris from fisheries might beach at the Galápagos Islands (Chávez-Castrillón et al., 2020). Already existing prediction methods often use proxies for primary production to predict fishing activity. Primary producers form the basis of the food chain. The assumption is that these primary producers attract consumers, which attract fish and the corresponding fisheries. This study attempts at examining the significance of the relationship between net primary production and fishing activity in the surrounding of the Galápagos Islands.

Being able to connect beached litter with its sources may be a first step toward preventing pollution and its consequences. This research aims at constructing a framework combining release locations and particle tracking, to determine where the debris originates. Connecting this source attribution framework with the predicting of fishing activity would allow ultimately clean up management to direct their efforts to specific times and locations.

This leads to the following three research questions:

1. Does the fishing industry contribute to the trash at Galápagos Islands?
2. Where does beached debris from fisheries originate?
3. Is there a relationship between fishing effort and net primary production?

My thesis is composed of five sections. Section 2 is concerned with the methodology considered and applied in this study. Here, the pathways of fishing debris toward the Islands are analyzed. A tool to calculate the conditional probability of origins of this debris at the shores of the Galápagos is introduced. To establish familiarity with and understanding of current research, a literature review on methods of the prediction of fishing activity is presented. Finally, an investigation into the relationship between fishing effort and net primary production is exhibited. The findings of this study and their discussion are shown in section 3. A conclusion, bringing together the main areas covered in this study, is presented in **??**. Section **??** discusses improvements on the set-up, gives an overview of studies that could advance the outcome of this research, and proposes next steps towards a predicting of the beaching of debris on beaches. Finally, the process of research is discussed and how it can be improved. Writing the code has been done following the principles presented by (Wilson et al., 2014).

2 Methods

2.1 Tracking of Fishing Debris

In this section, I will introduce the dataset containing information on fishing activity which we will use to investigate whether the fishing industry contributes to the debris at the Galápagos Islands. To do so, we obtain a fitting domain and assume that debris is released at fishing locations. We conduct particle tracking simulations where the particles are released at fishing locations and advected using the output of an ocean circulation model. The results from this simulation allow us to investigate the particle trajectories. Should these come close to the shores of the Galápagos Islands a beaching probability is determined. To be able to connect the number of particles beaching with the mass of the debris beaching, a framework that allows for the assessment of quantities of debris will be introduced.

2.1.1 Data on Fishing Effort

In 2001, the International Maritime Organisation (IMO) set objectives for the implementation of an automatic ship tracking system, the automatic identification system (AIS). Vessels carrying AIS transceivers broadcast to nearby vessels and can be tracked by base stations located along coast lines or, when out of range, through a growing number of satellites (Harati-Mokhtari et al., 2007). This system is primarily intended to allow ships to view marine traffic in their area and to be seen by that traffic. In order to achieve this, AIS transponders publicly broadcast information about a ship's identity, position, and course (McCauley et al., 2016). The mandatory use of AIS trackers has been progressively extended and the IMO's requirements to carry AIS have fostered compliance for the largest ocean-going vessel. It is important to note that numerous vessels, particularly fishing vessels, slip through the cracks in existing IMO policy (Malarky et al., 2018). Still, AIS presents a global feed of publicly available vessel locations and there provides the first tenable option to quantify the behavior of global fleets down to individual vessels (Natale et al., 2015).

The global footprint of fishing is poorly understood as oceans remain the least observed part of our world (McCauley et al., 2016). The use of AIS data allowed Kroodsma et al. (2018) to investigate apparent fishing efforts and deliver the first global assessment of commercial fishing activity. They found that while more than 50% of fishing occurs in just 0.5% of the ocean, it is also occurring throughout half of our oceans. Their study uncovered that global fishing is strongly affected by culture. For instance, they registered a drop in fishing activity at higher latitudes corresponding to Christmas vacation in Europe and North America. Some fleets did display seasonal affected behavior, but work weeks, holidays, and political closures showed to be more influential on the global temporal footprint of fishing (Kroodsma et al., 2018). Global Fishing Watch (GFW) published the dataset created by Kroodsma et al. (2018) and provides daily high-resolution global rasters of fishing effort online.

Gaps in AIS transmission that are a result of high vessel density or low satellite coverage, are removed by algorithms created by Global Fishing Watch (Malarky et al., 2018). When permitted access they also used regionally administered vessel monitoring systems (VMS) to close gaps in coverage or cross-reference where possible (McCauley et al., 2016). With convolutional neural networks, Kroodsma et al. (2018) can identify AIS positions that are characteristic of fishing activity. Their research finds that fishing activity is detected with $> 90\%$ accuracy. The resulting

dataset provides information on fishing location, duration of fishing, gear type, Maritime Mobile Service Identity (MMSI), and flag state.

Analysis was based on version 2.0 of the Apparent Fishing Effort dataset distributed covering the period 2012-2020. The data is handled via the python package `pandas`. The pandas library provides integrated routines for performing data manipulations and analysis on structured data sets (McKinney et al., 2011). It provides tools for data alignment, re-sampling, and aggregation.

To connect fishing activity with marine debris loss this report follows the assumption of (Kaandorp et al., 2020), that marine debris is lost during fishing activity. The dataset used here provides data on the duration of fishing, which can vary between a few minutes and 24 hours. To incorporate this duration parameter in the following study, the assumption is made that an increase in fishing duration results in an increase in the amount of debris released, see section 2.1.5. Debris released from fisheries includes ALDFG but also any manufactured or processed solid material that was discarded into the marine environment. Since plastics are the most abundant material collected in studies of marine debris fishing gear is mostly made from plastic for durability reasons, another assumption made in this research is that debris released is made from plastic and floats at the surface.

2.1.2 Domain

To choose a domain it is important to make sure that most pathways toward the Galápagos Islands are included. Since I am assuming that debris from fisheries floats, the paths of surface currents are most relevant to determining which domain to choose. Oceanic surface currents are the result of wind stress on the water surface. Several forces affect and sustain surface currents, including the Coriolis effect, the location of land masses, wind patterns creating waves, and periodic gravitational forces of the moon and sun generating tides (Cushman-Roisin et al., 2011). The large-scale currents that play a role around the Galápagos Islands are the Humboldt current, the North Equatorial Countercurrent, and the South Equatorial Current. With a subset of the GFW data multiple particle tracking runs, see section 2.1.3, of different domain sizes, are conducted. This is done to see whether most pathways toward the islands are caught within a domain. The most favorable domain is chosen and compared to the dataset used by van Sebille et al. (2019) to investigate pathways of microplastic ending up in the Galápagos Archipelago. Based on this analysis first, a 20° by 20° domain was chosen, measuring from 75° to 95° west and 10° north to 10° south. Later in the study, it was found that this domain was not large enough. The small size of the test dataset meant that it was not possible to account for pathways from the west, see section 3.1.2 for further elaboration. The final domain used for the analysis of research questions 1 and 2 was chosen to be 75° to 105° west and 10° north to 10° south.

Since only floating macroplastic is considered, the Surface and Merged Ocean Currents (SMOC) dataset which provides surface currents suffices for tracking their trajectories. The dataset is defined on a 1/12° grid with an hourly frequency. It is created by modeling systems from the Copernicus Marine Environment Monitoring Service (CMEMS) to reproduce the net velocity felt by a body at sea surface (Drillet et al., 2019). Three components are added up to create the total current used by this research, the hourly general-circulation surface current from Mercator PSY4 1/12° operational system, the time interpolated hourly wave currents from MFWAM 1/10° wave forecast system and the hourly tide currents from FES2014 tide model. Figure 1 shows the mean average total velocity of the SMOC dataset in the chosen domain. The large-scale currents mentioned above are visible. The Humboldt current follows the coast of Peru and leaves the

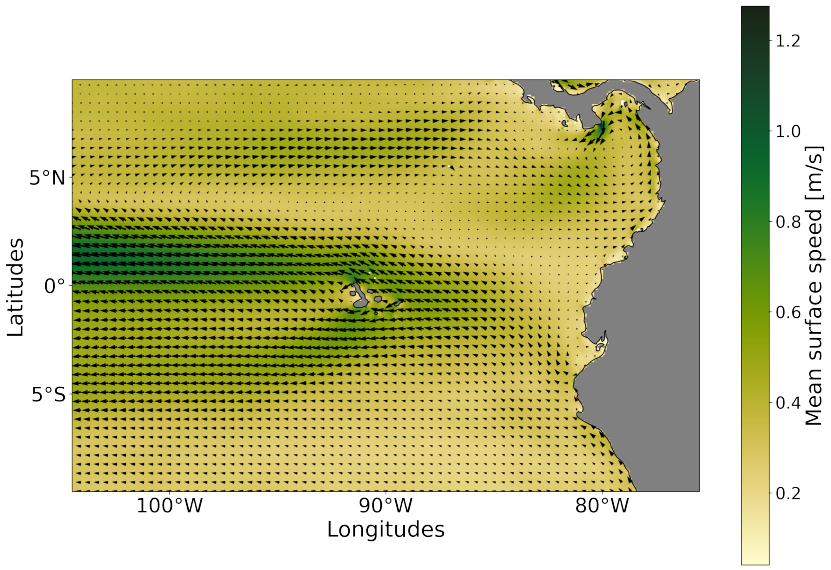


Figure 1: The mean average total velocity of the SMOC dataset in the domain [75°, 105°] west and [-10°, 10°] north, over the period 2016-2021 is depicted. The quiver plot shows the visualisation of horizontal velocity components.

coast at 5° north towards the Islands. The North Equatorial Countercurrent flows west-to-east at about 3-10° north. Finally, the South Equatorial Current flows east-to-west from about 5° north and about 5° south in this domain. In the entire Pacific basin, it can extend towards 20° south.

To handle the data the python library `xarray` is used. It is strongly inspired by `pandas` and implements data structures and provides an analytical toolkit for multi-dimensional labeled arrays. Since the fixed rank design used by `pandas` is inadequate for arbitrary rank arrays it can not be applied here. `xarray` uses the same model on which the network Common Data Form (netCDF) is built, and allows for the use and the manipulation of N-dimensional labeled arrays (Hoyer et al., 2017).

2.1.3 Lagrangian Analysis

Langrangian analysis is an ideal framework for the tracking of marine debris. In fluid dynamics, it can be regarded as the observation of fluid motion by following the individual fluid parcel (van Sebille et al., 2018). When applying the technique to marine debris, as they are advected by ocean currents. The particle velocity is the velocity given by the field at that time and location. The change in position of a particle, where $\vec{x}$ is the position of the particle at time t and $\vec{v}\left(\vec{x}\left(t\right), t\right)$, is computed by

$$\vec{x}\left(t + \Delta t\right) \; = \; \vec{x}\left(t\right) + \int_{t}^{t+\Delta t} \vec{v}\left(\vec{x}\left(\tau\right), \tau\right) d\tau. \tag{1}$$

Figure 2 shows how the flow velocity at the particle location is obtained through linear inter-

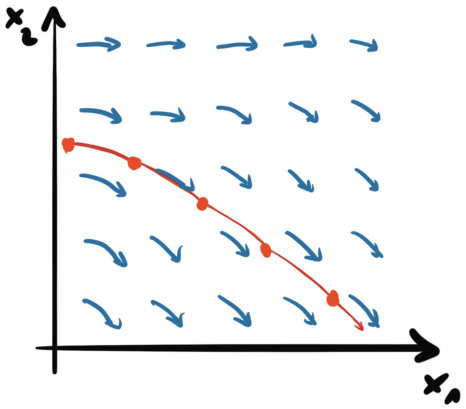


Figure 2: Visualisation on how the the flow velocity at the particle location is obtained through linear interpolation of the flow field data at each point (blue arrows). The particle follows the flow field (red line).

polation of the flow field data at each point and how the particle follows this flow field. Here, temporal discretization is done by the 4th-order Runge Kutta method which is a widely used method for solving the initial-value problems of differential equations.

To determine the paths of marine debris, Lagrangian simulations combined with source input data have been used before. Liubartseva et al. (2018) and Lebreton et al. (2012) investigated marine debris ejected during the course shipping routes. Marine debris from fisheries has been investigated by van Duinen et al. (2022), who used weekly averages of fishing duration and reverse modeling for source attribution of beach littering, and Kaandorp et al. (2020) who averaged fishing effort over the period 2012 to 2016 to investigate the marine debris budget in the Mediterranean. This research is unprecedented as it uses the daily locations of fishing efforts at points where particles are released. This allows for a more accurate analysis of the pathways involved, as these are subject to temporal variability, and a more precise investigation into the amounts of debris released. The release of debris is based on the data on apparent fishing effort provided by GFW, see section 2.1.1. Figure 3 shows all release positions over the domain and the period 2016 to 2021.

The simulation is being done by the use of OceanPARCELS (Probably A Really Computationally Efficient Lagrangian Simulator (Lange et al., 2017). OceanParcels uses 4th-order Runge Kutta integration to advect particles (Butcher, 2016). For operation, it is necessary to define the type of particles that are released. The particles are released on the day of fishing effort and since there is no exact time given for when fishing occurs all particles are released at midnight. This choice is based on the fact that most fishing in the domain is classified by squid jigging. Paulino et al. (2011) describe that most squid jigging vessels operate at night, using powerful lights to attract the squid. The particles are defined so that distance from the shore of the Galápagos is calculated at every step of every particle, and once a particle arrives within beaching distance a boolean variable is flipped to 'True'. Further explanation of this process can be found in Section 2.1.4. The integration time step is two hours. To ease computational effort sticky boundaries are implemented, and the particles are stopped in their motion once they reach the boundaries

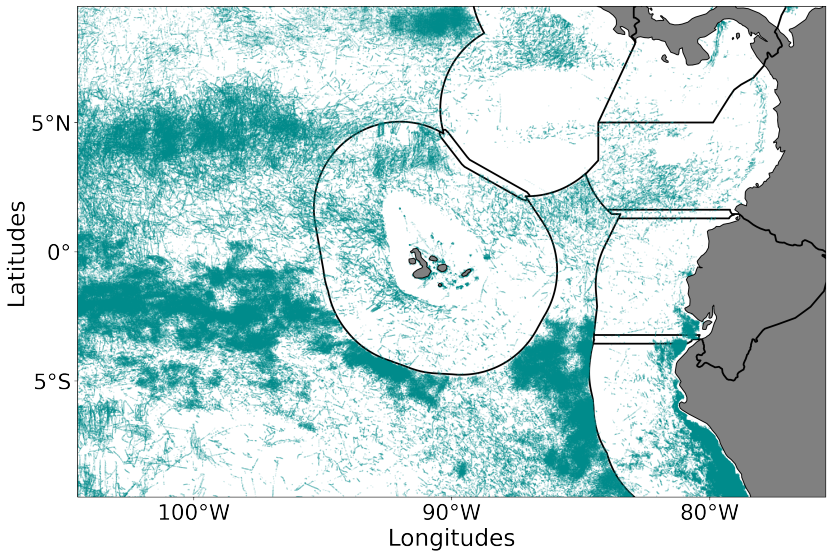


Figure 3: The regions of fishing effort over the domain in the domain[75°,95°] west and [-10°,10°] north, and the period 2012-2021 is depicted (cyan). The gray lines representing the borders of Exclusive Economic Zones (EEZ).

of the domain. After half a year particles are deleted. For execution, the velocity fields that the particles need to access are loaded. By compiling the kernels that encode the particle behavior at each timestep, the output is produced and written to a NetCDF file.

2.1.4 Beaching

The processes behind marine debris beaching are complex. Wind direction and speed play an important role, as do coast angle, aspect, and morphology (Hardesty et al., 2017)(Browne et al., 2015). Lagrangian models are mostly used in open-ocean domains and application to nearshore systems is intricate (van Sebille et al., 2020). It is unclear how the mentioned factors should be parameterized, and, besides, observations are scarce (Onink et al., 2021). A stochastic framework presented by Onink et al. (2021) assumes that on a global average, plastic beaching is primarily determined by surface currents and the location of plastic input.

With the results of the OceanParcels simulation, beaching can be post-processed. This study follows the approach from Onink et al. (2021). The OceanParcels simulation was defined so that once a particle enters the beaching grid, a variable containing information on beaching is set from false to true. The postprocessing code uses this information to calculate the beaching probability once a particle has been denoted to enter a beach cell. This probability is calculated in the following way:

$$p_b = 1 - \exp\left(-\frac{dt}{\lambda_b}\right). \tag{2}$$

Here, the characteristic timescale of beaching λ_B is the number of days that a particle must

spend within the beaching zone such that there is a 63.2% chance that the particle has beached. The integration time step is denoted as dt. The beaching grid was defined relative to the SMOC dataset. All SMOC ocean cells adjoining a land cell of the Galápagos Islands establish a beaching grid cell. Figure 4 shows the resulting beaching grid. The largest seven of Galápagos' 128 Islands and Islets, of which most have an area below 1 square kilometer, are detected. These are Pinta, Marchena, Isabela, Santiago, Fernandina, Santa Cruz, and San Cristobal, in Figure 4 denoted as Islands A-G, respectively.

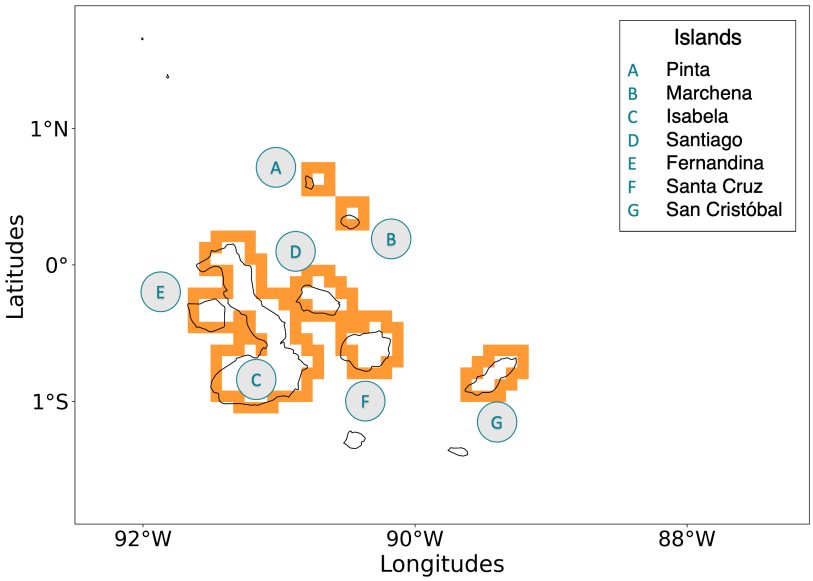


Figure 4: Here, the beaching grid defined by all ocean cells within the SMOC dataset adjoining a land cell, is shown. The Islands Pinta, Marchena, Isabela, San Salvador, Fernandina, Santa Cruz and San Cristòbal, are denoted by A to G, respectively.

To check how much the beaching parameter influences the beaching of particles, a sensitivity analysis is conducted by calculating the amount of beaching for varying λ_B. Note that resuspension has not been considered in this analysis, meaning that once a particle is considered as beached it is assumed to remain in place.

2.1.5 Framework for beached mass calculation

To receive information on the amount of debris arriving at the Islands, an analytical framework is derived. The aim is to represent the varying amount of time spent at fishing locations. A boat can spend from a few seconds to many hours at one location and it is assumed that meanwhile the amount of debris released increases. An assumption similar to Kuczenski et al. (2022)'s approach is made: the release of marine debris is linearly dependent on fishing duration. In this way, each particle is weighed by the amount of fishing time spend.

For relatability, a certain mass is assigned to each particle. Instead of mass, a different variable could be investigated, for example, volume, or surface area. This poses the question of how much debris is released per event of fishing. Richardson et al. (2019) explain how miscitation of an erroneous global estimate of an annual release of 640 000 tonnes of ALDFG has been perpetuated.

This estimate has the potential of obstructing efforts to understand the true scope and scale of the issue. Most estimates have been limited to specific gear types and geographic locations or are more than 30 years old. Kuczenski et al. (2022) created a framework that estimates that 48 000 tonnes of fishing gear alone are lost, not including situations of abandoned or improperly discarded gear. To allow for an example of how the mass could be calculated, the erroneous estimate, referred to by Richardson et al. (2019), is used as M_{tot} being 640 000 000 kg. The average fishing time spend per year globally, Δt_{tot}, is calculated by taking the average time of fishing effort provided by GFW. The total amount of beached debris M_{tot} is the related to the total time spent fishing Δt_{tot},

$$M_{tot} = \omega \cdot \Delta t_{tot}. \tag{3}$$

With the use of ω the mass m_i per particle can be determined

$$m_i = \Delta t_i \cdot \omega \tag{4}$$
$$. \tag{5}$$

This allows for the calculation of the total mass of beached particles M_b, where n_b is the number of beached particles, which is determined by the combination of particle tracking and beaching framework, section 2.1.3 and section 2.1.4,

$$M_b = \sum_{i \epsilon \mathcal{M}(n_b)} m_i. \tag{6}$$

Since these results are based on an erroneous estimate extreme caution needs to be put on the resulting numbers shown in section 3. With M_b known and analogously calculating the total mass of released particles, M_r, the fraction of beached weight can be calculated. In the following, n_r is the number of beached particles.

$$P = \frac{M_b}{M_r} \tag{7}$$
$$= \frac{\sum_{i \epsilon \mathcal{M}(n_b)} \Delta t_i}{\sum_{j \epsilon \mathcal{M}(n_r)} \Delta t_j}. \tag{8}$$

The fraction of the weighted amount of beached debris is, for the time being, more informative and sound, as the amount only depends on how long a fishing boat is fishing at each location and not on the scaling factor.

2.2 Origins

Plastic debris at shorelines has damaging implications for coastal and marine ecosystems, tourism, and fisheries. The Galápagos ecosystems are already impacted by the effect of warming,

food limitation, and heightened disease risk. Fast-growing levels of marine litter could exacerbate the stress they are already experiencing. Not only can plastic debris be ingested or cause entanglement, but it can also act as a surface for rafting organisms. This is of particular concern a the Galápagos islands, where marine ecosystems are highly vulnerable to the invasion of non-native species. By locating the origins of plastic pollution it might be possible to put mitigating measures in place. In this way, plastic pollution and its consequences might be prevented (J. S. Jones et al., 2021).

To determine where beached debris originates from, debris will be connected to the release region, fishing gear types, and nation flags. The resulting tool allows specifying where fishing debris originates from once it is found. The code created for this research is composed in a way that it can be easily adapted to other domains. In this way, my framework provides support for the implementation of policies to reduce the pollution by debris stemming from fisheries.

2.2.1 Categorical Variables

Global Fishing Watch provides information on nation flags and fishing gear. Attributing the sources of beached marine debris can be an important means on the path to decreasing the impact of plastic waste in the region. This research tries to create conditional distributions, which give the probability that a randomly selected item in a sub-population has a characteristic of interest. These require discrete characteristics that place observations into fixed groups.

Fishing gear can vary strongly in its characteristics. Material type, lifespan, and potential for ghost fishing and wildlife entanglements differ strongly for different kinds of fisheries (Richardson et al., 2021). It is therefore of interest to be able to distinguish gear types when looking at the origins of beached debris. Since some gear types are especially harmful, the results of this research can also help to further the implementation of the regulation of certain gear types. Around the Galápagos, three main categories of fishing gear types are employed. Squid jigging is mostly performed by placing overhead lights to illuminate the water and thereby attracting squid. The squid gather in the shaded area under the boat and are caught by the use of lures on fishing lines which are jigged up and down. Purse Seining is used for dense schools of single species of fish, like tuna and mackerel. A vertical net is used to surround the school of fish to then draw the bottom together. Set long lines consist of a mainline and vertical lines, with baited hooks at regular intervals. They are set in general, on or near the bottom. Their length can range from a few hundred meters to more than 50 kilometers (M. M. Jones, 1995).

The term nation flags refers to the country where a vessel is registered. Kroodsma et al. (2018) found that boats fishing principally fish within the Exclusive Economic Zones (EEZ) of the nation flag they are flying. An EEZ extends usually 200 nautical miles beyond a nation's territorial. In these regions, the coastal nation has jurisdiction over both living and nonliving resources. The research domain encompasses the EEZ from five countries: Costa Rica, Panama, Columbia, Ecuador, and Peru. In their study, Kroodsma et al. (2018) also found that 85% of fishing effort in international waters is conducted by fisheries sailing under a flag from China, Spain, Taiwan, Japan, and South Korea. This leads to the hypothesis that debris beaching at the Galápagos Islands will mostly stem from flags sailing under the mentioned nations.

Next to fishing gear and flag states, another important aspect of origins is the geographical location of the release of debris. Where does fishing effort occur most often? The fishing locations published by GFW need to be grouped into some form of categorical variables. At first,

an attempt is made to determine these regions via spatial clustering. Since this proves to be insufficient, a qualitative analysis is conducted additionally.

2.2.2 Spatial Clustering Analysis

To determine spatial patterns of fishing behavior Spatial Clustering Analysis is performed. It aims to partition spatial data into a series of meaningful subclasses determined by their location. Two methods were investigated both employing density-based analysis. These operate by detecting areas where data points are concentrated and where they are separated by areas of empty or sparse data.

Density Based Spatial Clustering Analysis (DBSCAN) is a way to identify classes in spatial databases. It relies on a density-based notion of clusters to discover groups of arbitrary shapes. The analysis starts with an arbitrary starting point, from where the point's ϵ-neighborhood, an area with a given radius ϵ, is retrieved. If it contains a sufficient amount of points, a cluster is started, otherwise, the point is labeled as noise. This point might later be found in a sufficiently sized ϵ-environment of a different point (Ester et al., 1996). DBSCAN is interesting because there is no need to specify the number of clusters to be found. It can find arbitrarily-shaped clusters and is robust to outliers as it has a notion of noise. The downside of this analysis is that border points that are reachable from more than one cluster can be part of either cluster, depending on the order the data are processed. It cannot cluster data sets well with large differences in densities and choosing a meaningful ϵ-neighborhood can be difficult (Schubert et al., 2017).

Hierarchical Density Based Spatial Clustering Analysis (HDBSCAN) performs the above explained DBSCAN over varying values for ϵ. The results are integrated to find a clustering that gives the best stability with respect to ϵ. This allows HDBSCAN to find clusters of varying densities (McInnes et al., 2017). The only parameter chosen up front is the minimum number of points required to form a cluster.

Scikit learn's `dbscan` is used. This implementation uses ball trees and kd-trees, which are algorithms that group data points in a hierarchical way to determine the neighborhood of points. This method moves points in sparse regions further away, to then perform single linkage clustering. Here, two clusters separated by the shortest distance are combined. Only if the distance exceeds a certain span, they are separated. In this way, clusters are formed. Single points are assigned as noise (Ester et al., 1996). For HDBSCAN the scikit learn contribution `hdbscan` library is employed. It performs DBSCAN over varying epsilon values and integrates the result to find a clustering that gives the best stability over ϵ (McInnes et al., 2017).

2.2.3 Bayesian Analysis

To answer the question of where the beached debris originates from, the conditional probability $p(R_i|S_{loc})$ was calculated. $p(R_i|S_{loc})$ describes the probability that a given location is the source of a plastic particle, given that this particle ends up at the studied beach. In order to do so, the categorical variables described above are used. Pierard et al. (2021) developed an approach to compute the probability that a piece of plastic found at sea comes from a particular source. Bayesian inference states the following relationship

$$p(R_i|S_{loc}) = \frac{S_{loc}|p(R_i)\, p(R_i)}{P(S_{loc})} \tag{9}$$

$p(S_{loc}|p(R_i))$ is the likelihood, which is the probability that a particle beaches at the specified location, given that it originates from a certain source location. $P(R_i)$ is the prior probability, in this case, the probability that a given location is a source of plastics. $P(S_{loc})$ is the probability of sampling a plastic particle in a specific location. $P(S_{loc})$ requires observations from sources that are not taken into account in equation 9. This can be circumvented by obtaining a normalizing constant that ensures that the sum of all posterior probabilities is one in each location (Downey, 2021):

$$\sum_{i=1}^{N} p(R_i|S_{loc}) = 1. \tag{10}$$

In this way $p(R_i|S_{loc})$ can be expressed as

$$p(R_i|S_{loc}) = \frac{p(S_{loc}|R_i)\, p(R_i)}{\sum_{i=1}^{N} p(S_{loc}|R_i)\, p(R_i} . \tag{11}$$

van Duinen et al. (2022) used backtracking and only used his fishing information intensity averaged per week and Pierard et al. (2021) used equally released and equal amounts of particles per source, allowing for an unweighted likelihood. In both cases, the likelihood only contains information on how particles have moved through the environment. In section 3.2, the results under these assumptions for different nation flags can be found. The calculated posterior probabilities over-represent large and under-represent small fractions of the distribution of fishing activity. Since each particle is released at a location where fishing activity occurs, the likelihood is already weighed by the respective categorical variables. To get the desired conditional probability $p(R_i|S_{loc})$, the distribution of beached debris is determined for each island and normalized by the amounts arriving there.

This allows me to set up a source identification framework. The combination of the GFW dataset, OceanParcels simulation, and beaching allows for an immediate determination of conditional probabilities of beaching. This research uses this linkage and sets up a framework to compute the probability that a given categorical variable is the source of a plastic particle, given that this particle ends up at the studied beach. It calculates directly the conditional probability $p(R_i|S_{loc})$. This is done by accounting for beached debris at the different Islands, A-G, and accounting for their categorical variables. The results are shown in section 3.2.

2.3 Fishing and Nutrients

The framework described in subsection 2.2 allows connecting debris released during fishing activity with littering on beaches. This source identification framework allows for efforts toward policies targeting fisheries and their littering behavior. Knowledge of when and where fishing debris will beach could facilitate and improve the efficiency of clean-up efforts. This could be attained by being able to predict fishing activity and having knowledge of the time scales involved

from fishing locations to beaching. In the following, a review of current predicting methods and their outcome is presented and an investigation into the relationship between Fishing effort and Net Primary Production (NPP) is conducted.

2.3.1 Review on Predicting Fisheries

Kroodsma et al. (2018) created a global dynamic footprint of fishing effort with a spatial and temporal resolution of two to three orders of magnitude higher than for previous data sets. They found that industrial fishing occurs in more than 55% of the ocean area. Global patterns of fishing have low sensitivity to short-term economic and environmental variation while showing a strong response to cultural and political events.

While no distinct global connection between environmental parameters and fishing effort has been found, regional investigations have come up with results indicating distinct relationships. The most common connections with fishing activity found are sea surface temperature (SST) and chlorophyll-a (CHL). Santos (2000) described how remote sensing methods are used as forecasting services and operational support to fishing fleets. For these purposes, mostly water temperature and its fluctuations were considered, but also CHL, a proxy for primary production, is often taken into account. Cimino et al. (2019) looked at spatial, temporal, and interannual variations in fishing efforts in Palau's EEZ. They compared it to oceanographic factors, flag state of the fishing fleet, and climate events. They report that environmental conditions were a predominant predictor of the seasonal variation in fishing effort. Vessels were surprisingly observed to fish in waters with low CHL concentrations.

The method Cimino et al. (2019) used was boosted regression trees, concurrent with most of the literature reviewed. Boosted trees are a statistical learning method that attains prediction and explanation for regression and classification analyses. Dealing with many types of response variables, loss functions, and predictors they can deal with missing values with minimal loss of information. They combine predictions from a collection of weak classifiers, with high prediction errors, to compose a strong classifier, with low prediction errors (De'Ath, 2007). Cimino et al. (2019) predicted whether fishing was present or absent and the amount of fishing effort. The amount was generally better predicted than presence/absence, with the best results explaining 47% of deviance. Scales et al. (2017) used boosted regression trees to model the relative probability of occurrence and catchability of broadbill swordfish, in the California Current System. By exploring the physical drivers of regional distribution, they found that water depth, STT, sea surface height (SSH), and the isothermal layer depth play a role. Soykan et al. (2014) use boosted regression trees to model the spatiotemporal distribution of fishing effort for two distinct fisheries. Their results suggest that it is possible to accurately predict fishing effort using temporally dynamic and temporally static predictor variables, among them being SST and NPP. They found that across a range of target species, fishing methods, and spatial scales, even a relatively short time series of fisheries data may suffice to accurately predict the location of fishing effort in the future.

Another method repeatedly employed is generalized additive models. These allow the modeling of non-linear data while maintaining explainability. An extension of these models are generalized additive mixed models, these allow for additionally incorporating of random effects. Both are widely used to model correlated and clustered responses (Groll et al., 2012). Lan et al. (2017) used longline fishery data to investigate the relationship between catch rates of yellowfin tuna and oceanographic conditions. The results suggest that SST, SSH anomaly, and CHL play a sig-

nificant role. Their cumulative deviances explained 33.6% of deviance, and the predicted relative catch rates exhibited a high correlation with observed catch rates. Using generalized additive mixed models to determine the probability of whale occurrence based on environmental variables Hazen et al. (2017) found a relationship between raised CHL levels and preferences for SST being in a season-dependent range. They compared their results with predictions from boosted regression trees, as these have fewer statistical assumptions and can predict when environmental layers are missing. However, these models performed poorly as predictions did not agree with known blue whale habitats.

Chávez-Castrillón et al. (2020) built a model forecasting the presence of CHL with spatial-temporal regression. The resulting model can predict CHL levels around the Galápagos Islands with a certainty of 75% to 82%, for five to one day ahead prediction, respectively. Should a strong correlation between fishing effort and nutrient data be found, this model could be implemented to forecast fishing.

It is important to note that the modeling and prediction of fishing habitats are extremely complex. Lehodey et al. (2008) describe how one species of fish shows different behavior according to their life phases. They used an ecosystem and population dynamics model to describe the spatial dynamics of tuna in the Pacific Ocean. Phillips et al. (2018) simulate the movement of either individual or groups of tuna. For this, they needed to take into account physical ocean currents, habitat-dependent stochastic movements, and representing active searching behaviors. As Hazen et al. (2017) describes whales have a preference for certain temperature ranges. An investigation into the kinds of fish being targeted and their migratory behavior should be taken into account.

The here presented literature review shows that oceanographic parameters, often proxies for primary production, are successful in the prediction of fishery prediction. Multivariate approaches are the most successful, and the inclusion of sea surface temperature is imperative. Boosted regression tree models are easy to implement and have been shown to accurately predict the location of fishing effort.

2.3.2 Nutrient Data

Predictor variables of the biological environment can be indicative of trophic linkages or fish food sources. Net Primary Production (NPP) corresponds to the amount of inorganic carbon which is converted into organic matter mostly, during photosynthesis. To explain NPP, a look at Gross Primary Production is advantageous. GPP is the amount of carbon biomass that primary producers create. Primary producers are organisms that produce complex organic compounds such as carbohydrates, fats, and proteins using carbon from simple substances such as carbon dioxide. This production is generally done by using photosynthesis or inorganic chemical reactions. Some fraction of this fixed energy is used by primary producers for cellular respiration and maintenance of existing tissues. The remaining amount of carbon biomass is referred to as NPP. It is a critical measure of marine ecosystem function since it represents the amount of carbon and energy that is available to organisms higher up the food chain, like fish (Field et al., 1998). In this way, it can be informative to compare fishing effort with NPP.

CMEMS Global Ocean Biogeochemistry Hindcast provides global daily fields covering the period 2016 – 2020. The three-dimensional fields cover chlorophyll, nitrate, phosphate, silicate, dissolved oxygen, and net primary production. It is defined on a standard grid at 1/4 degree and on 75 vertical levels (Perruche, 2018). It is created by the use of global biogeochemical simulations using

the Pelagic Interactions Scheme for Carbon and Ecosystem Studies (PISCES) model (Aumont et al., 2015). This model is forced by daily mean fields of ocean and sea ice produced at Mercator-Ocean, and atmosphere from the ERA-Interim3 reanalysis product. The output is provided in netCDF format and handled with `xarray`.

2.3.3 Histogram

To investigates the relationship between fishing effort and net primary production the CMEMS Global Ocean Biogeochemistry Hindcast is compared to the GFW dataset. For this, fishing activity needs to be transformed from listed data to two-dimensional data with the same spatial and temporal resolution as used for net primary production. The simplest way to achieve this is via a two-dimensional histogram.

Daily fishing effort is converted to a two-dimensional histogram with bin sizes of 1/4 degree lengths. `histogram2d` from the python library `numpy` allows performing this conversion. Points specified by their longitudinal and latitudinal coordinates are grouped into bins. These bins are weighed by the sum of fishing effort time spent within them. The bin size is with 1/4 degree chosen to be identical to nutrient data. These daily histograms are then converted into netCDF files and appended to the Ocean Biogeochemistry Hindcast files. Figure 5 shows the distribution of fishing effort for the entire period where the results are scaled to the square root of the original value. The scaling is done such that areas of little effort can be visualized.

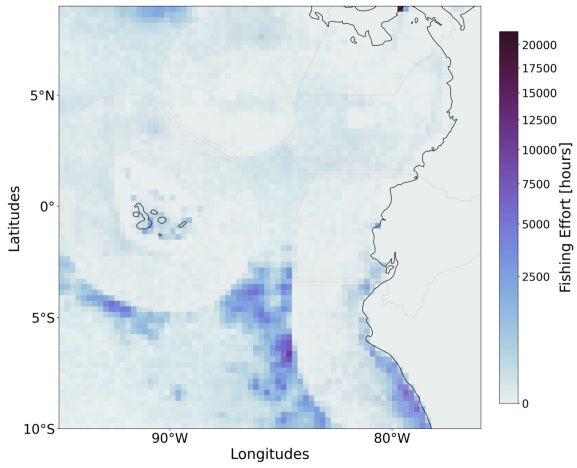


Figure 5: The histogram for fishing effort, weighted by fishing effort, for the period 2012-2021. To visualize areas of little effort, the values are scaled to $x^{-\frac{1}{2}}$.

2.3.4 Kernel Density Estimation

Daily fishing effort data can be very noisy. To remove noise Kernel Density Estimation (KDE) was examined. KDE estimates the underlying probability density function of a dataset. It produces smooth estimates by learning the shape of the density from the data automatically (Chen, 2017). It works by weighting the distances and amount of all the data points. If there are more data points nearby, the estimate is higher, indicating the probability of encountering a point

at that location. The smoothness of the estimate can be tuned via the bandwidth parameter. An incorrect choice of bandwidth parameter results in under smoothing or over smoothing. Figure 6 shows an illustrative example of how KDE works.

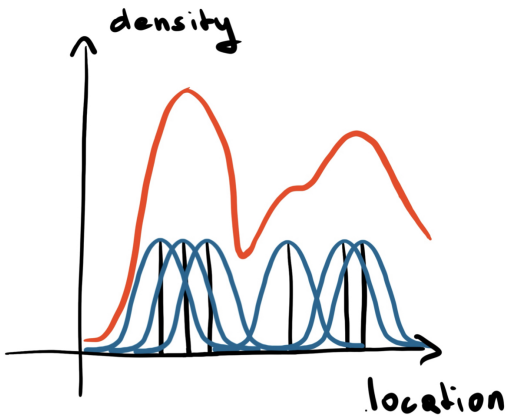


Figure 6: Kernel Density Estimation works by weighting the distances and amount of all the data points. If there are more points nearby, the estimate is higher, indicating that probability of encountering a point at that location.

The KDE is performed by using `gaussian_kde` from the python library `SciPy`. This method includes automatic bandwidth determination and uses a standard normal density function to describe the data.

2.3.5 Empirical Orthogonal Function Analysis

To find a relation between fishing effort and biogeochemical variables in a spatial and temporal sense, an Empirical Orthogonal Function (EOF) Analysis is conducted. EOF Analysis allows an alternative point of view about space-time relationships, especially in terms of correlation. It is generally used to summarize the information contained in large data sets. This is achieved by transforming to a smaller set of uncorrelated variables that can be more easily visualized and analyzed. The variables are ordered so that the first retains most of the variation of the original dataset (Messié et al., 2011).

In an EOF analysis, each variable is represented by a coordinate axis. The length of each coordinate is standardized by scaling to unit variance. The data is mean-centered by subtracting the variable averages from the data, such that the average point is the origin. The first principal component approximates the data best in the least squares sense. The second principal component is oriented such that it reflects the second largest source of variation in the data while being orthogonal to the first PC. Figure 7 shows the idea behind EOF.

By replacing p observed variables with a much smaller number k of uncorrelated principal components, the dimensionality of a spatial-temporal data set can be reduced by transforming it to a new basis in terms of variance (Hannachi et al., 2009). This transformation can be advantageous as often a large fraction of the total variance is contained within the first few spatial patterns.

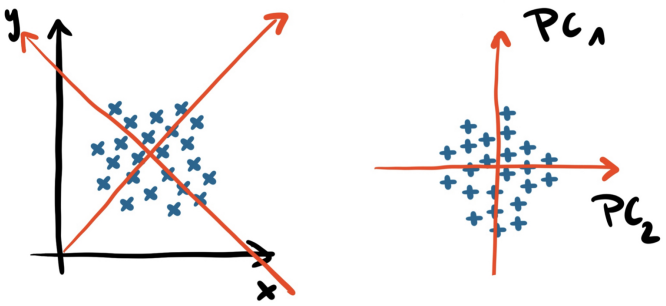


Figure 7: With Empirical Orthogonal Function Analysis, each variable is represented by a coordinate axis. The length of each coordinate is standardized by scaling to unit variance. The data is mean-centered by subtracting the variable averages from the data, such that the average point is the origin. The first principal component approximates the data best in the least squares sense. The second principal component reflects the second largest source of variation while being orthogonal to the first PC.

The others can be discarded whilst still maintaining the major variances present in the original data (Wheeler et al., 2004). The function `MultivariateEof` from the python package `eof`is used. It finds a set of orthogonal spatial patterns, known as EOFs, along with a set of associated uncorrelated time series or PCs.

2.3.6 Correlation of Time Series

To get an impression of fishing and net primary production, the correlation of timeseries is performed. To achieve this, the daily average occurrence of both parameters in set regions was investigated. These regions are based on the information gained by the qualitative analysis, where it was found that fishing behavior is very much contained within EEZs, see section 3.2. Based on this information, the time series are created by averaging the daily amounts of fishing and NPP in each EZZ, as well as for international waters. Fishing activity might be lagging to nutrient data since the trophic levels interesting to fisheries might take time to develop as NPP only accounts for the presence of primary producers. To investigate a lag of fishing effort behind NPP, the time-lagged cross correlation is calculated.

This is conducted by the `xarray` aggregation function `DataArray.mean`, which is based on `numpy`'s `mean` function. It calculates the arithmetic mean, the sum of the elements along the two spatial axes divided by the number of elements. For the analysis, it is relevant to take into account the varying sizes of the regions. Comparison between resulting values can be misleading. This effect does however not influence the computation of correlation, which is conducted by use of the resulting time series per region. The Pearson correlation r is a measure of global synchrony that reduces the relationship between two signals to a single value. Given paired data consisting of N pairs, r is defined as

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}\sqrt{(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}. \tag{12}$$

The `numpy` function `corrcoef` is used. It returns a matrix of correlations of x with x, x with y, y with x and y with y.

Cross correlation is computed by incrementally shifting one time series and repeatedly calculating the correlation between the two signals. Peak correlation is defined as the time where the correlation between the time series is maximal.

3 Results and Discussion

This study set out to answer the following three research questions.

1. Does the fishing industry contribute to the trash at Galápagos Islands?
2. Where does beached debris from fisheries originate?
3. Is there a relationship between fishing effort and net primary production?

Section 3.1, section 3.2 and section 3.3 set out to answer each of them consecutively.

3.1 Tracking of Fishing Debris

The first part of the study aimed to assess if debris from fisheries plays a role in littering at the Galápagos beaches. To achieve this I conducted a particle tracking simulation. Simulating fishing debris, particles were released at locations of fishing activity. First, to get an overview, the temporal and spatial variability of fishing activity is investigated. Then the results of the simulation and the beaching framework are presented and the major pathways are established. To investigate the beaching framework a sensitivity analysis was performed and finally, the results of the mass framework are discussed.

3.1.1 Fishing Activity

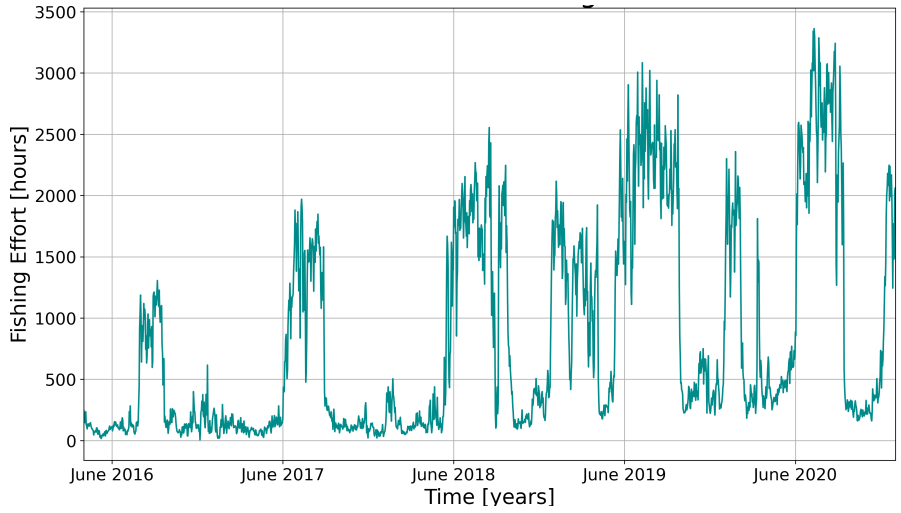


Figure 8: Daily fishing effort summed up over the domain [75°, 105°] West and [-10°, 10°] North, for the period April 2016 to January 2021.

Global Fishing Watch presents data on global fishing activity, in the form of date, duration, and location (see section 2.1.1). In this study, the amount of debris released is scaled to the duration

of fishing activity. It is, therefore, relevant to understand the major characteristics of fishing activity in the domain.

To get a general understanding of fishing effort and assess the temporal variability in the domain Figure 8 describes daily fishing effort summed over the entire domain. The x-axis represents time over the simulation period and the y-axis shows fishing effort in hours. The time series is depicted over the period April 2016 to January 2021, which is the overlapping time between the SMOC and GFW datasets. The most interesting aspect of this figure is that fishing effort increases over the years. The occurrence of annual periods of high fishing intensity is apparent, starting in June and lasting until September. There are peaks of high fishing effort starting in December 2018, 2019, and 2020. Fluctuations of smaller scale are discernable in periods of low fishing effort. The increase in fishing effort over the entire period is not to be mistaken for only a representation of an increase in fishing activity, since requirements to carry AIS transponders have increased over the last decade.

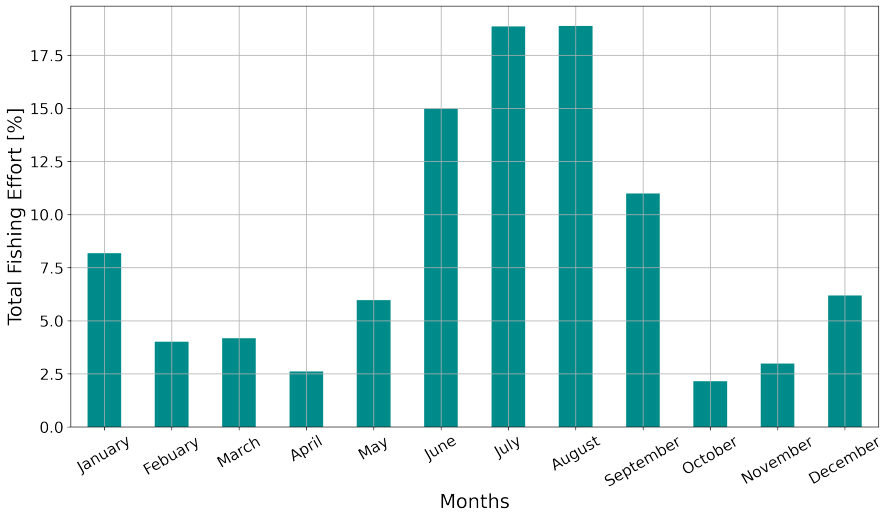


Figure 9: Fishing Effort per month summed over the entire period (April 2016 to January 2021) over the domain [75°, 105°] West and [-10°, 10°] North.

To focus on annual patterns, Figure 9 shows fishing effort per month summed over the entire period. The x-axis portrays the year in months, while the y-axis represents the amount of fishing in percent. Fishing Effort peaks in August to decrease drastically in the following month and result in the minimum in October. A second smaller peak occurs in January and a second trough in April. Finally, a smooth increase from March to August is observable.

Comparing figures 8 and 9, it can be seen that fishing effort is predominant in July and August, with more than 37 % occurring only in these two months. Several possible factors can explain these results. An annual harmonic is dominant for sea surface temperature variation in the eastern equatorial Pacific. It is characteristic of low temperatures starting in July until December (Li et al., 1996). The high presence of fishing activity is occurring simultaneously with the onset of low sea surface temperatures. However, the low temperatures continuing throughout December do not account for the drastic decrease in fishing activity in October. Other oceanographic factors could play a role. The relationship between nutrient presence and fishing activity is discussed in section 3.3. The variability of fishing activity might also be related to holidays and political

closures, considering that Kroodsma et al. (2018) state that these are significantly influential in determining the temporal footprint of fishing. The small-scale variations could, among other factors, be explained by variation in fuel price, as fuel represents, on average, 24% of costs.

Bias in these representations might stem from the incompleteness of the GFW dataset. Due to regulations, boats less than 15 meters are not required to carry AIS transponders and the dataset under-represents them. Small boats are very relevant for fishing debris around the Galápagos Islands. Approximately 97% of active vessels are smaller than 9.6 m long (Castrejón et al., 2020). Additional to only certain boats being covered, it has been established that boats are avoiding AIS registry (Malarky et al., 2018). Note that sometimes fishing duration can exceed 24 hours per day. An inquiry with Global Fishing Watch team provided the following explanation. Fishing hours are based on summing the hours associated with an MMSI's fishing positions in a day. Hours are assigned to each AIS position based on the time since the previous AIS position. In some cases, positions could be assigned time from the previous day (e.g. a position at 12:15 am that's assigned four hours will have 3:45 hours from the previous day). If that vessel then spends the whole day fishing, the number of fishing hours can exceed 24.

3.1.2 Beached Debris

Literature, for example the study by J. S. Jones et al. (2021), suggests that the currents most responsible for transport towards the Galápagos Islands is the Humboldt curren. Since these currents are directed westwards at the Islands, the initial choice for domain neglected transport from the west. Figure 10 shows this domain, 75°W-95°W and 10°N-10°S, and the trajectories of beached debris. The beaching depicted occurs over the period from April 2016 to January 2021. The white squares depict release locations.

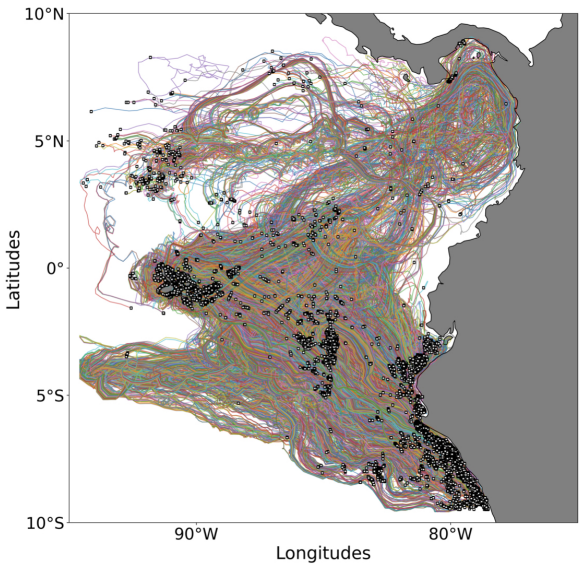


Figure 10: Trajectories of beached debris within initial choice of domain, [75°, 95°] West and [-10°, 10°] North. White squares represent release locations. This simulation covers the period from April 2016 to January 2021.

What stands out in figure 10 is an unexpected path of debris in the south of the Galápagos Islands. Debris is advected towards the domain's boundary in the west where it almost performs a 180° turn, to move towards the Islands, where it finally beaches. This effect vanished when extending the domain by 10° towards the west. It can thus be suggested that this branch is an artificial effect and connected with the choice of the domain of interest not being big enough. Particles coming from the east get stuck at the boundary and remain there. They are either deleted when they have reached the age of deletion or, once the direction of current changes they can return into the domain and then beach at the Galápagos Islands.

The trajectories of beached debris with the extended domain, 75°W-105°W, and 10°N-10°S, are shown in Figure 11. It illustrates the pathways of beached debris, from their point of release until they arrive at the Galápagos Islands. Release locations, based on positions of fishing activity, are depicted as white squares.

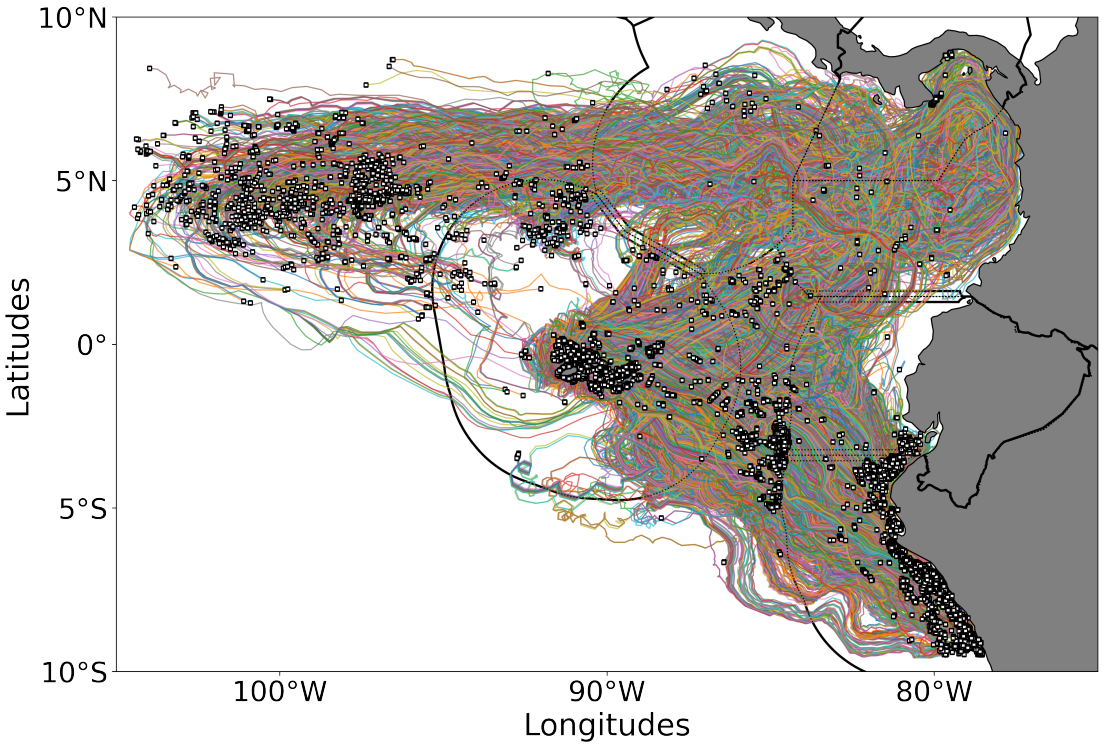


Figure 11: Trajectories of beached debris in the domain, [75°, 105°] West and [-10°, 10°] North. White squares represent release locations. This simulation covers the period from April 2016 to January 2021. Nations' boundaries and their EEZ regions are indicated by the black lines. The nation's present in this domain are, from north to south: Costa Rica, Panama, Colombia, Ecuador and Peru.

From figure 11 we can distinguish four separate release clusters. One is situated in front of the coast of Peru and one is located at the boundary of the Peruvian and Galápagos' EEZ. A lot of debris that is beached is released around the Islands. The fourth cluster is located in the northwest of the Islands. When comparing locations of fishing activity, Figure 3 shows that even though much debris is released in the southwest of the Islands, it does not contribute to beached

debris. This figure also shows that little fishing occurs in the EEZ of Panama and Columbia. This explains the small presence of release locations in these areas.

To further investigate the major pathways carrying fishing debris towards Galápagos, Figure 12 shows the trajectories from Figure 11 together with the average surface currents over the same period, see Figure 1.

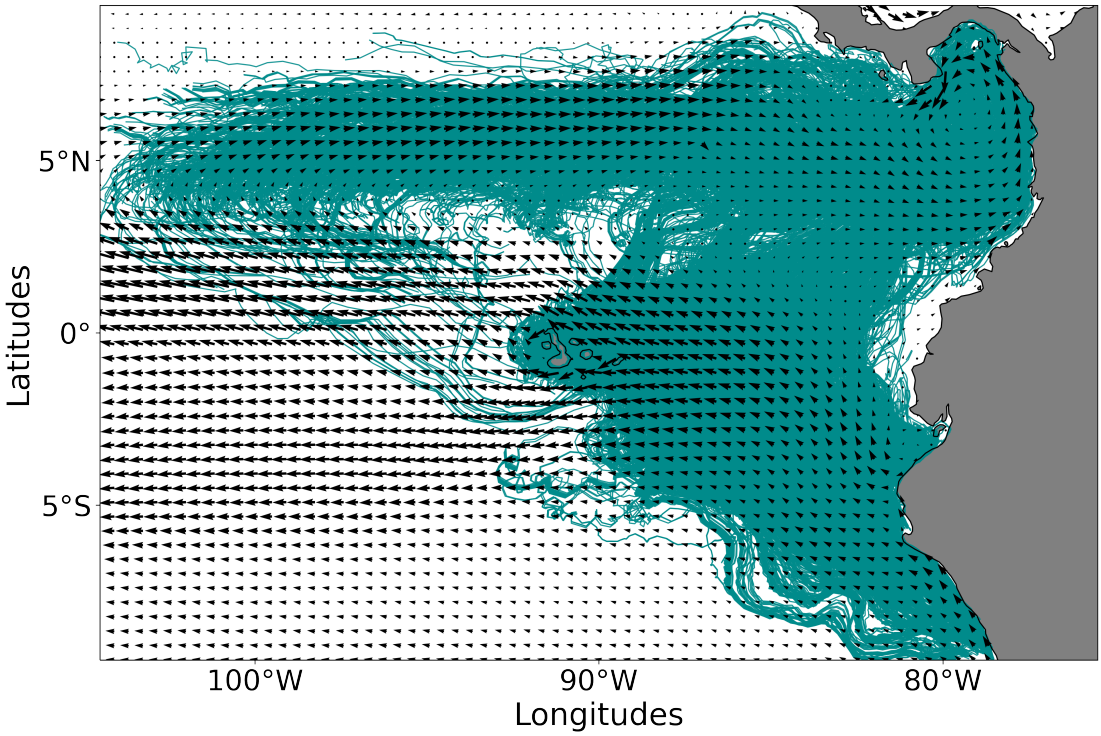


Figure 12: In cyan: Trajectories of beached debris. In black: Average surface current. Domain: [75°, 105°] West and [-10°, 10°] North. Period: April 2016 to January 2021.

Figure 12 accentuates the pathways responsible for the transport of debris. These are found where trajectories align with the arrows. The Humboldt Current follows the coast of Peru northwards. At about 5° south it separates from the mainland and directs itself towards the Galápagos Islands. The North Equatorial Countercurrent flows west-to-east around 3-10° north. The feature in front of the coast of Panama and Colombia is necessary for the debris being transported by the North Equatorial Countercurrent towards the Islands.

Number of released particles	1,238,834
Number of beached particles	17,528
Percent beached	1.41 %

Table 1: Number of released and beached particles for the extended domain.

The implementation of the beaching framework, following the approach of Onink et al. (2021), resulted in the beaching of 1.41 % of particles. Table **??** shows the number of particles beached and released. Section 3.1.4 shows the results weighted by the duration of fishing at each location.

The resulting data must be interpreted with caution, considering the choice of beaching parameters. The current approach is stochastic and does not take into account factors that affect beaching. More observations on beaching are highly required. The sensitivity analysis in section 3.1.3 investigates the variation of the beaching parameter.

The results presented here may be somewhat limited as particle transport was modeled only in 2 dimensions, neglecting upwelling and sinking. Another possible bias can stem from the fact that windage, particle degradation, and re-suspension of beached particles were not explicitly included. For example, a study by Ruiz et al. (2022) showed that the local distribution of fishing-related floating marine litter is significantly affected by windage. Note that all the percentages given in this thesis are based on the assumption that these processes do not play an important role. Another source of uncertainty is the resolution of hydrodynamic data. The simulation results can be improved by using an ocean circulation of a higher horizontal resolution and temporal frequency.

3.1.3 Sensitivity Analysis

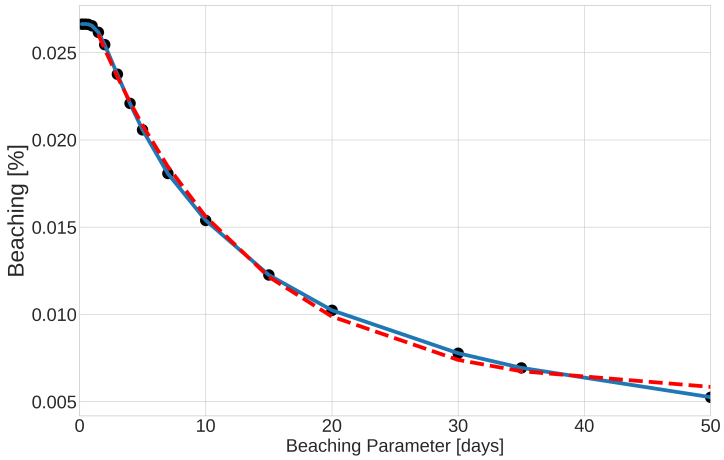


Figure 13: The dependence of the beaching of debris on the beaching parameter λ_B. On the x-axis the varying beaching parameter is assigned, the y-axis represents the fraction of beached debris. This analysis was done for the initial domain, [75°, 105°] West and [-10°, 10°] North.

Since we are interested in how much debris beaches an investigation into the choice of beaching parameter is necessary. The approach of Onink et al. (2021) accounts for the uncertainty of the ocean current data in land-adjacent ocean cells by parameterizing beaching as a stochastic process. This does not take into account local processes that can influence the amount of beached plastic on coastlines, like the coasts morphology, wind direction, speed, etc. Since there is no experimental study to base the value of the beaching parameter λ_B, a first step to shed light on the sensitivity of beaching processes is to calculate beaching for a range of possible values. This sensitivity analysis is shown in Figure 13. It demonstrates a clear exponential decrease for beaching with increasing beaching parameter. The maximum amount of beaching is capped at 2.66 percent, this implies that there is no more debris available to beach. When calculating the exponential fit for the function, excluding the first values, as these are physically constrained, the following exponential function is attained:

$$Y = 0.02 \times \exp(-0.08x) + 0.001. \tag{13}$$

Here 0.02 is the pre-exponential factor and 0.001 is the rate constant. The relative predictive power R^2 of this exponential model is 0.99.

The resulting data must be interpreted with caution, considering the choice of beaching parameters. The current approach is stochastic and does not take into account factors that affect beaching. More observations on beaching are highly required. The sensitivity analysis in Figure 13 shows the substantial influence of the beaching parameter λ_B on the amount of debris beached, with a variation between 2 and $0,5\%$.

3.1.4 Mass Calculation

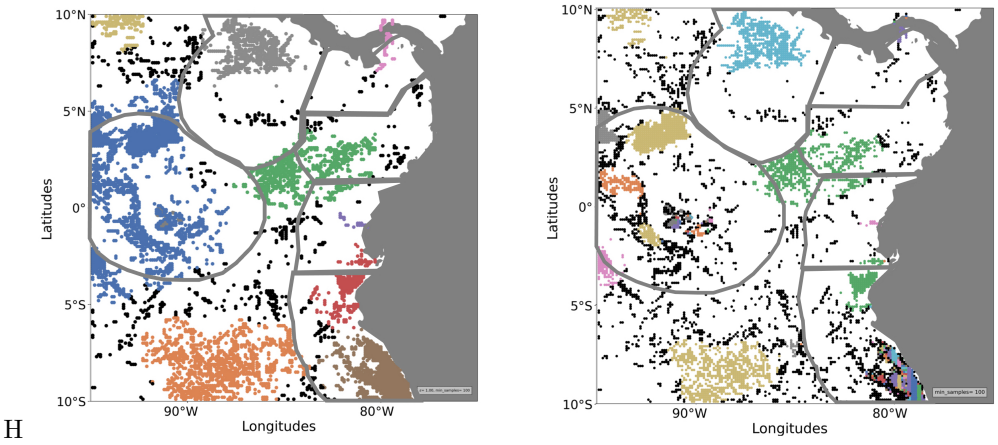
To determine the amount of debris beached, the released particles have been weighted by the fishing duration. The underlying assumption is that more debris is released with longer fishing activity. When taking into account the estimate of a loss of 640,000 tonnes of ALDFG per year a beached mass can be ascertained. To calculate the amount of fishing per year, the fishing duration over the entire dataset from 2012 to 2021 is taken, averaged, and divided by the number of years. This leads to a weighting factor ω, stating that 18 kg of fishing debris are released every hour globally. The total mass related to fishing effort that beached at the Galápagos Island in the period 2012-2021 is 350000 kg. This leads to an average beached mass of 40000 kg per year.

More precise statements are hard to make for lack of observations on several parameters, the most important being an estimate of the amount of fishing debris being released. A tentative glance ahead is done with implementing the weighting framework, see section 2.1.5, which allows for the calculation of the weight, or other quantities, of released debris. This requires information on the amount of fishing released related to the duration of fishing. In case this assumption proves to be unsound, this framework needs to be adjusted to a different kind of information. Since there is no information on the amount of released debris these results need to be considered with utmost care. It is mostly to be looked upon as an example for future calculations.

The fraction of beached debris weighed by fishing duration is 2.33 percent. This accounts for 0.32 percentage points less than before weighting. This implies that the duration of fishing activity at release locations of beached debris was shorter than the average fishing duration in the domain. This amount only depends on how long a fishing boat is fishing at each location and not on the scaling factor.

This research showed how important more research and information on debris from fisheries is. To make sound statements on release behavior the following questions need to be addressed: When is marine debris released? How is it released? What kind of debris is released? Does it make sense to account for fishing debris in mass, or should rather volume, surface area, harmfulness, or even toxicity be considered? First steps in this direction are undertaken, with research on derelict fishing gear to presenting actual estimates in mass (Kuczenski et al., 2022).

This study set out with the aim of assessing whether debris from fisheries contributes to the trash on the shores of the Galápagos Islands. This study found univocally that this is the case. There are clear pathways of debris leading toward the Islands. This fact might contribute to an incentive to reduce fishing-related littering. From a perspective of environmental concern but also for economic reasons.



H

Figure 14: Left: Density-Based Spatial Clustering of Applications with Noise (DBSCAN) of locations of fishing activity, with $\epsilon = 1$ and $min\_sample = 100$. Right: Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) of locations of fishing activity, with $min\_sample = 100$. Both figures cover the period 2018 and the domain [75°,95°]W and [-10°,10°]N. Groups of the same colour represent a determined clusters. Black dots represent noise, locations not affiliated with any of the clusters.

3.2 Origins

The second part of the study is concerned with the sources of debris. Being able to attribute origins to debris beaching, might help mitigation efforts and direct policy makers towards an effective approach to reducing debris from fisheries. Three different kinds of sources are investigated, nation flags that ships are sailing under, source regions, and the kind of fishing gear that is used. To do this sources must be categorized into fixed groups. Spatial clustering and a qualitative analysis were performed to determine source regions. The results from the framework by Pierard et al. (2021) are presented and I discuss why this approach is not applicable for this study. In the next step, the adjusted Bayesian analysis is presented. Here first the distribution per Island is investigated, and in the next steps, the analysis for sources is shown.

3.2.1 Clustering Analysis

Global Fishing watch presents fixed groups of variables of fishing gear and nation flags. Being able to categorize source regions, these clusters of concentrated fishing effort, can be advantageous for two reasons. These regions allow for the attribution of source regions of beached debris, and they can be applied for the comparison of fishing activity with oceanographic parameters. To find these regions two kinds of spatial cluster analysis, see section 2.2.2 were made and their results are presented in the following.

The left plot in Figure 14 shows DBSCAN for a minimum sample, min_sample, of 100 locations per cluster and an ϵ radius of 1. This analysis covers the all registered locations of fishing activity in the domain 75°W-95°W and 10°N-10°S in the year 2018. It is important to highlight that this is no temporal analysis but simply a spatial analysis. A group of fishing locations of the same

color represents a determined cluster. Black dots describe noise, these are locations not affiliated with any of the clusters. Multiple clusters are discernible, one below the south of the Galápagos EEZ, and one in front of the coast of Peru and Ecuador. The cluster from Colombia transgresses its EEZ boundaries. The cluster around Galápagos Islands does not differentiate between Fishing within the EEZ, the marine reserve, and international waters.

DBSCAN gives a great overview of data structuring, it is however difficult to implement as it depends on a careful choice of two parameters. The results shown here delivered the most satisfying findings, out of many different couplings of the parameters *min_sample* and ϵ. One unanticipated finding is the combination of multiple clusters within one cluster in the west of the Islands. This blue cluster crosses the boundaries of the EEZ. This behavior is not found for computations with smaller ϵ. However, the less satisfying couplings of parameters produced more noise or split groups of fishing that belonged to each other into multiple clusters. Larger values for *min_sample* covered even larger regions of the map and were less conclusive than the one shown in Figure 14. The period of one year was chosen since computational constraints did not allow for performing over longer periods.

To find clustering methods that can break up the big cluster in the west of the Islands HDBSCAN was performed. The right plot in Figure 14 depicts this clustering analysis for a minimum sample, *min_sample*, of 100 locations per cluster over the period 2018. It results in similar clusters as for DBSCAN. The cluster below the Islands EEZ, as well as the one transgressing Colombia's EEZ boundaries. You can see that this method can determine multiple clusters of fishing activity in the west of the islands. However, it creates many very spatially concentrated clusters on the coast of Peru. It also shows more noise on borders of clusters than DBSCAN.

This analysis is equally computationally expensive and more than one year of data burst the memory allowed for the computation. With HDBSCAN only one parameter is to be chosen, the minimum amount of data points within one cluster. To catch different clusters around the Galápagos Islands, the algorithm detects multiple clusters within one cluster in front of the coast of Peru. Computations for larger values of *min_sample* did not reduce the noise nor combine the many clusters at the coast of Peru into one. In further analysis, it might be interesting to conduct HDBSCAN as a temporal analysis by combining fishing effort into bins of months.

3.2.2 Qualitative Analysis

The results of the clustering analysis gave the first information on the spatial behavior of fishing activity in the domain in question. Most fishing effort does not occur in singular positions in space and time. The fishing fleets move to certain locations where they exhibit fishing activity for many consecutive days. During this activity, they don't stay fixed to one position but stay in motion. To incorporate temporal behavior a qualitative analysis was conducted. It was performed by creating an animation of fishing behavior over the period April 2016 until January 2021. This animation was watched multiple times and substantial behavior was denoted. The knowledge gained in the above the beaching analysis, Figure 10, and the cluster analysis, Figure 14 lays a good cornerstone for the investigation into spatial and temporal behavior.

Fishing activity showed to be mostly contained within EEZ boundaries. Either fisheries stayed within International Waters or remained within one EEZ. One exception is the fishing behavior in the EEZ of Colombia which regularly crosses the boundary. This behavior was also caught by both DBSCAN and HDBSCAN, see the green cluster in Figure 14. Disregarding this one

exception EEZs were chosen to be regional categorical variables. The union of world country boundaries and EEZ's dataset combines the boundaries of the world countries and the Exclusive Economic Zones of the world (FlandersMarineInstitute, 2016). It is provided as a `shapley` file and allows for simple distinguishing of different EEZ. The remaining international waters are divided into four regions, north and south of the equator and at 95°W. The resulting regions are depicted in Figure 15. The Galápagos Islands belong to Ecuador, and its EEZ belongs to Ecuador, in this research a distinction is made between the two.

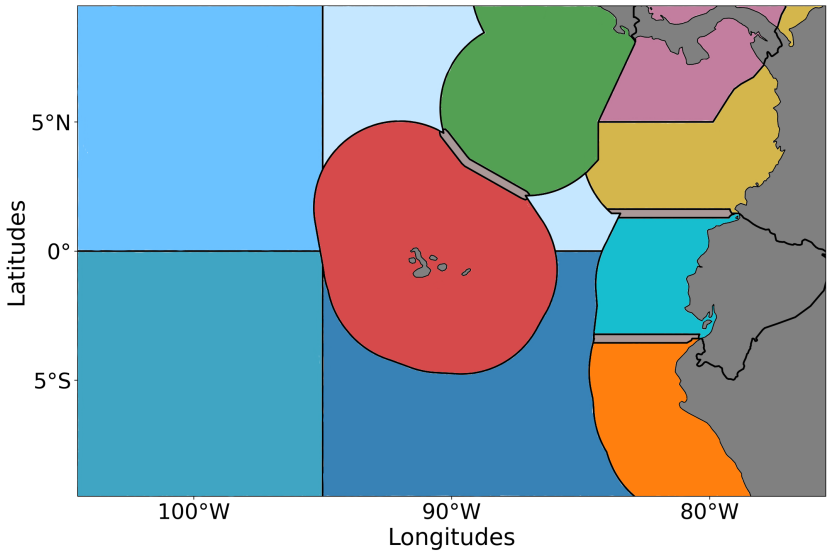


Figure 15: Here, the partitioning of source regions is shown. It is divided into EEZ regions, international waters were divided at 0°N and 95°W. The colours correspond to the conditional analysis of source region, shown in Figure 21.

This approach does not completely contain all clustering behavior, but it allows for an easy definition of source regions. It facilitates providing information for policymakers and stakeholders. The application onto other regions, but here it is effective and with the use of `shapley` files easy to implement computationally. The grey areas shwon between some EEZ in Figure 15 represent joint regimes. Debris released in these areas has been assigned to the southern EEZ of each area.

3.2.3 Bayesian Analysis

The approach from Pierard et al. (2021) to determine the conditional probability $p(R_i|S_{loc})$, that floating oceanic plastic comes is connected to certain river plastic emissions was highly successful. This study aimed at reproducing the same framework for sources of debris from fisheries, see section 2.2.3. Figure 16 shows the results for nation flags. Here the prior probability $P(R_i)$ was set as the distribution of fishing duration under nation flags. The likelihood $p(S_{loc}|R_i)$, the probability that debris beaches at the specified location, given that it originates from a certain source location, was set as the distribution of nation flags of beached debris at the Islands.

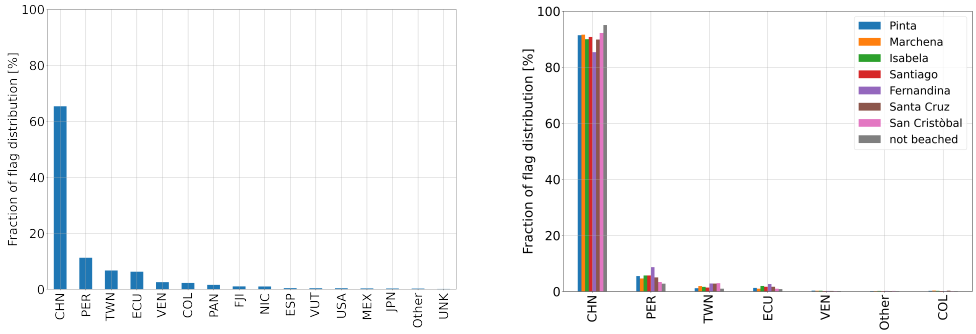


Figure 16: Left: Prior probability $P(R_i)$, the distribution of fishing duration under nation flags. Right: Supposed posterior probability $(R_i|S_{loc})$. This analysis was done for the initial domain, [75°, 105°] West and [-10°, 10°] North.

The left plot in Figure 16 depicts the representation of released debris per flag state in percent. China is most prevalent with more than 60% of fishing taking place under Chinese flags. All nation flags with a presence higher than 1% are represented in this figure, the remaining flags are summed up and combined within the label "Other". The right plot in Figure 16 shows the representation of the assumed posterior probability per flag state. The probabilities are given per Islands, see Figure 4, and debris that didn't beach. The results state that more than 80% of debris on each Island, as well as not beached debris, stems from fisheries under Chinese flags. Again flag representation lower than 1% is combined within the label "Other".

There is a clear discrepancy between release and posterior probabilities visible. 65% debris released stems from Chinese flags, while around 90% of posterior debris is accounted for by Chinese flags. Since "not beached" is included in this analysis, the entire flag distribution is represented in both cases. This shows that the calculated posterior probabilities over-represent large and under-represent small fractions of fishing activity. This points to wrong assumptions in the attempting Bayesian analysis. In reviewing the literature on this approach used by Pierard et al. (2021) difference in attaining the likelihood $p(S_{loc}|R_i)$ becomes clear. They used an equal amount of particles released per source, namely 100,000 particles per source cluster. In this way, the likelihood is not affected by the prior, which is the case in the here presented study. To attain a not affected likelihood, debris would have to be released at locations of clustered behavior where equal distribution of flags is present. The positive outcome of the application of this method is that I found that the assumed likelihood is nothing else but the distribution of debris per island and not beached normalized by the total amount of debris released. Investigating this distribution of debris per Island results in the conditional probability $p(R_i|S_{loc})$ that we have been looking for all along. The following sections present the results of this analysis.

3.2.4 Beaching per Island

The combination of Global Fishing Watch, OceanParcels, and beaching allows for the calculation of the distribution of debris stemming from fisheries. For each of the regions selected on the Galápagos Islands, this framework allows connecting the origins of fishing debris. This framework is applied in three different kinds of origins: nation flags, gear type, and source regions. The dataset presented by Global Fishing Watch gives information on the first two kinds of variables. Using a mask for EEZ regions allows for determining source regions. This framework is universally

applicable, while the choice of domain needs to be taken with care. The fraction of beached debris per Island is listed in table 2 and visualised in Figure 17.

Island A	Island B	Island C	Island D	Island E	Island F	Island G
6.7%	7.8%	47.8 %	6.5%	3.0%	8.9%	19.0 %

Table 2: Fraction of beached debris per Island

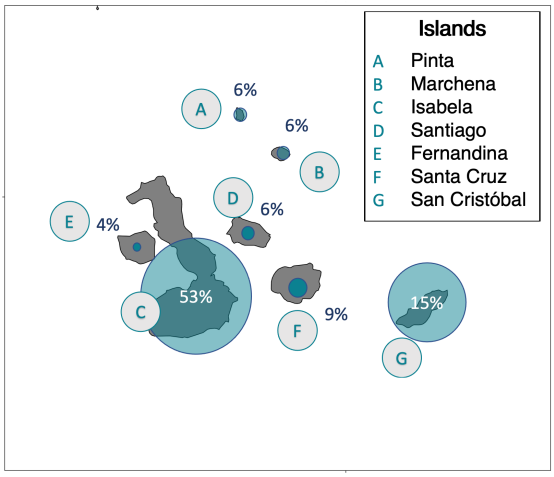


Figure 17: Representation of amount of debris arriving per Island in percent. The size of disk represent the amount of debris arriving.

Most debris from fisheries, 53%, arrives at Isabela (Island C). The second biggest amount, 15%, reach San Cristóbal (Island G). Figure 12 shows that San Cristóbal lies directly in the path of the Humboldt current. The smallest amount of beached debris is received by Fernandina (Island E). Figure 1 demonstrates that around this Island the smallest mean surface speed occurs. The Galápagos Islands shield its most westward Island from the strong currents coming from the east. Caution needs to be taken with this analysis, as the calculation of mean surface speed could hide strong variability in currents in this region. A temporal analysis of currents would need to be conducted to support this hypothesis.

In section 3.1.3 the variability of the beaching parameter λ_B on the amount of beaching was conducted. Here is an investigation into how the distribution of debris per Island is affected by a variation of λ_B. Figure 18 shows the resulting analysis. Pinta, Marchena, and Santa Cruz are most strongly affected by this variation. The proportion of debris decreased significantly when λ_B is increased. These islands are, compared to the other Islands, rather isolated and particles possibly do not spend as much time in their surroundings. The inverse effect is observed in the proportion of beached debris at San Cristòbal, which increased with a smaller beaching parameter. These effects need further inquiry, but a possible explanation for these results is that since this island is effectively blocking the path of the Humboldt current, particles can remain longer in its vicinity and therefore are very likely to beach.

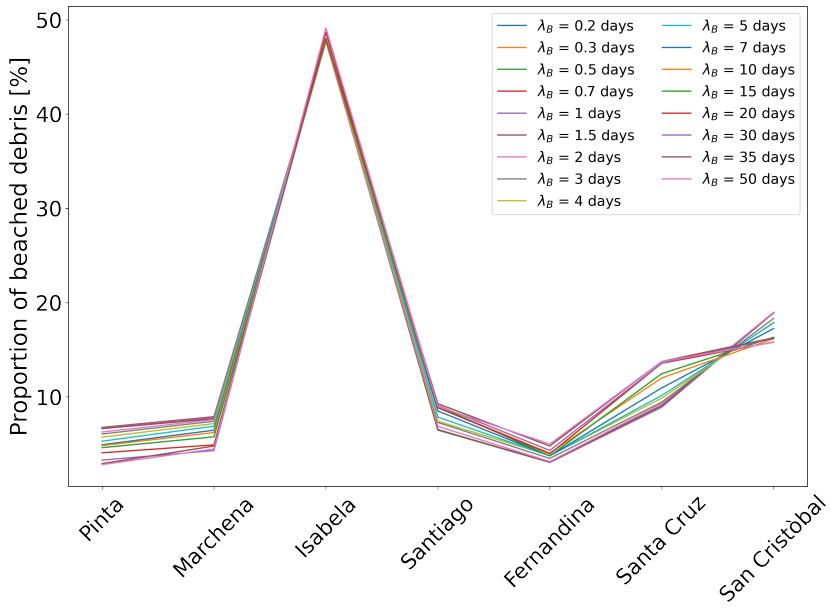


Figure 18: The sensitivity analysis of the fraction of debris arriving at each island. On the x-axis each island shown, the y-axis represents the proportion of beached debris. The beaching parameter λ_B is varied from 0.2 to 50 days.

3.2.5 Nation Flags

The results for the conditional analysis, the distribution of beached debris according to nation flags, in total and per Island is shown in Figure 19. The pie chart shows the distribution of beached debris according to nation flags. The map shows the beaching distribution for each Island with 'not beached' debris also included. Precise listing of the resulting fractions delivered in table 3.

Nation Flags	Total	Island A	Island B	Island C	Island D	Island E	Island F	Island G
CHN	22.5	21.8	34.1	12.7	19.5	14.3	42.0	45.5
PER	20.1	34.3	33.1	14.4	22.3	11.4	18.5	32.8
TWN	0.1	0.2	0.8	0.1	0.0	0.0	0.0	0.0
ECU	52.0	30.6	25.5	69.6	49.6	71.4	33.8	12.8
COL	1.7	4.2	1.7	0.8	1.3	0.5	1.3	4.6
VEN	0.6	2.4	1.4	0.2	2.6	0.0	0.6	0.1
PAN	0.2	1.6	0.3	0.1	0.0	0.0	0.4	0.2
NIC	0.5	0.4	0.1	0.5	0.7	0.5	1.4	0.3
ESP	0.9	2.3	1.9	0.4	2.7	0.1	0.9	1.1
MEX	0.8	1.9	0.3	0.4	0.5	0.0	1.0	2.1
USA	0.2	0.1	0.3	0.2	0.1	1.6	0.1	0.0
JPN	0.2	0.0	0.0	0.2	0.0	0.0	0.0	0.3

Table 3: Nation Flags: Detailed probabilities of beaching in total and per Island

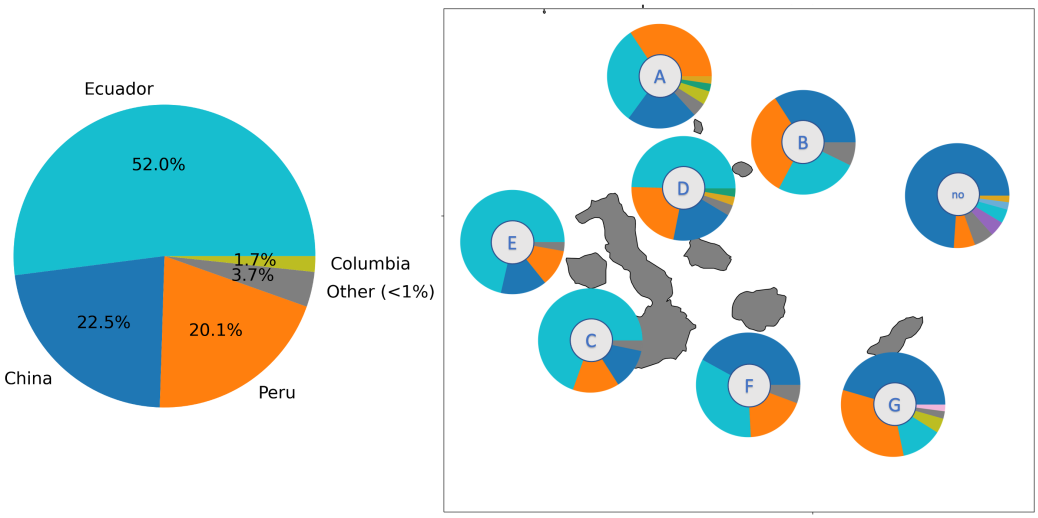


Figure 19: Here the distribution of beached debris according to nation flags is shown. The left pie chart represents the total distribution of beached debris at the Galápagos Islands. The distribution at each Island is shown at the map on the right.

The current study found that the total distribution of beached fishing debris is dominated by 52% stemming from Ecuadorian flags. Another important finding is that almost a quarter of debris stems only from ships sailing under Chinese flags. The distribution per Island indicates that the islands in the west catch most debris from Ecuadorian flags, while debris from Chinese flags dominates the islands in the east. Peruvian flags account for most debris on the northernmost island, Pinta, and more than one-quarter of the debris in the eastern islands, Marchena and San Cristòbal. Mexico and Spain contribute significantly to debris on Pinta (Island A) and San Cristòbal. These data must be interpreted with caution because they are based on the assumptions that all fisheries release the same amount of debris, and that this release is scaled by the duration of fishing. It is important to bear in mind that "nation flag" refers to the country where a vessel is registered. The degree of oversight and the conditions for registration by flag states vary widely (Goodman, 2009).

3.2.6 Gear Types

The results for the conditional analysis, the distribution of beached debris according to fishing gear types, in total and per Island is shown in Figure 19. The map shows the beaching distribution for each Island with 'not beached' debris also included. Precise listing of the resulting fractions delivered in table 3.

Interestingly, 'Fishing' is observed to account for 26% of total beached debris. 'Fishing' is attributed to a boat when the convolutional neural network model of GFW is not able to attribute a certain gear type. 21% of debris arriving at the Islands stems from squid jigging. Note that

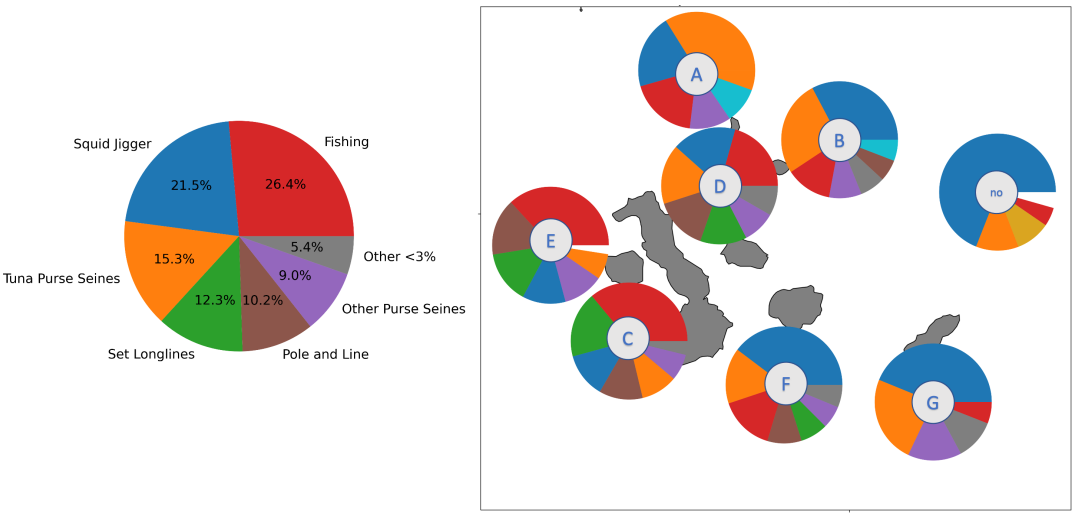


Figure 20: Here the distribution of beached debris according to gear types is shown. The left pie chart represents the total distribution of beached debris at the Galápagos Islands. The distribution at each Island is shown at the map on the right.

Gear Type	Total	Island A	Island B	Island C	Island D	Island E	Island F	Island G
Squid jigger	21.5	20.5	32.7	12.4	17.8	12.0	39.7	43.8
Tuna Purse Seines	15.3	33.9	26.6	10.1	16.6	7.1	15.4	24.2
Drifting Longlines	1.7	3.3	4.3	1.1	4.2	0.3	1.1	1.8
Fishing	26.4	18.7	12.8	36.0	20.6	36.9	15.3	6.0
Other Purse Seines	9.0	11.5	9.0	7.3	9.1	11.3	6.3	14.9
Purse Seines	2.8	5.5	5.8	1.8	4.0	1.4	3.0	3.7
Set Longlines	12.3	2.2	1.8	18.2	12.9	14.6	7.6	0.3
Pole and Line	10.2	3.5	6.0	12.1	14.7	15.7	9.5	4.55
Set Gillnets	0.9	1.0	1.0	0.8	0.1	0.8	2.1	0.9

Table 4: Gear Types: Detailed probabilities of beaching in total and per Island

squid jigging is the dominant form of fishing within the domain, accounting for 79% of all fishing activity. Closer inspection of the data shows that when combining the different kinds of purse seining, it accounts for 25% of debris arriving at the Islands. When looking at the distribution per Island it becomes apparent that 'Fishing' accounts for most debris on the westernmost Islands, Isabela, Santiago, and Fernandina. Within the entire domain 'Fishing' accounts for only 5% of debris. If we now turn to squid jigging, the data suggest that it is accountable for 44% of fishing debris beaching San Cristoòbal, Marchena, and Santa Cruz (Islands B, F, and G). Tuna purse seining contributes 34 % of beached fishing debris on Pinta (Island A). With an increase in observations on the release of different kinds of gears, information on the characteristics of different gear types could be added to the here presented framework. For example, the fact that

annually 5.7% of all fishing nets are lost, while 29% of all lines are released, could be included in the framework. Also, the size, mass, and floating behavior could be accounted for.

3.2.7 Source Regions

Figure 21 shows the distribution of beached debris according to source regions and the conditional probabilities at each Island. The pie chart shows the distribution of release regions, being the EEZs and international waters divided at the equator and 95°N, see Figure 15. The map shows the beaching distribution for each Island with 'not beached' debris also included. Precise listing of the resulting fractions delivered in table 5.

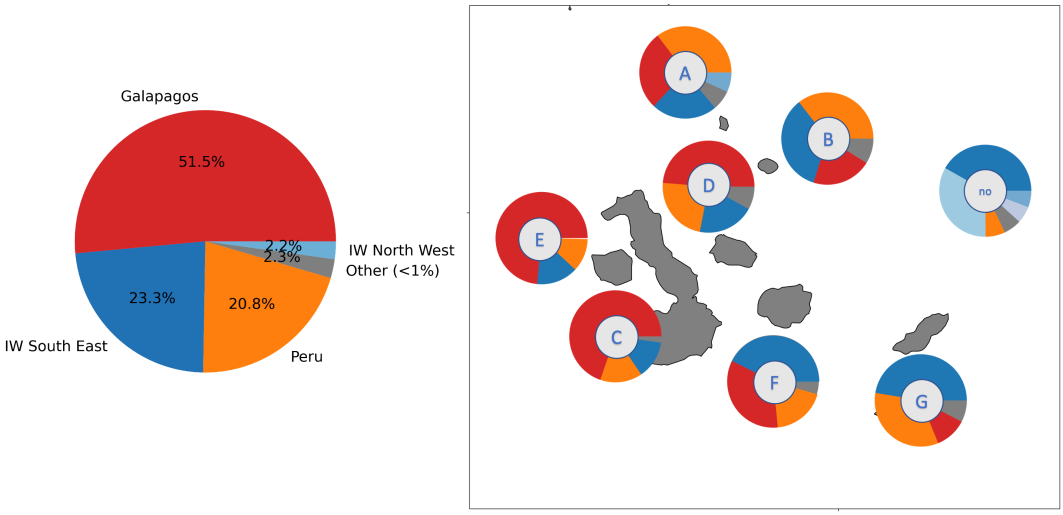


Figure 21: Here the distribution of beached debris according to source regions is shown. The left pie chart represents the total distribution of beached debris at the Galápagos Islands. The distribution at each Island is shown at the map on the right.

The results of this study indicate that the overall beaching distribution of source regions at the Galápagos Islands is dominated by 52% from within the EEZ of the Galápagos Islands. What stands out in this analysis is that 23% of debris from fisheries stems from international waters from the South East of the domain. Fishing activity in the EEZ of Peru contributes 21% to fishing debris arriving at the Islands. Figure 11 suggests that a lot of debris at the Islands is transported from the northwest of the domain. This study reveals that only 2.2 percent of debris comes from this region. This suggests that transport by the North Equatorial Countercurrent plays a minor role. When taking a closer look at the distributions at each Island it becomes clear that these fractions vary significantly. The westernmost Islands, Isabel, Santiago, and Fernandina (Islands C, D, and E), receive most fishing debris released within the Galápagos EEZ. Debris released in the Peruvian EEZ accounts for the biggest fraction at the two northernmost islands, Pinta and Marchena (Islands A and B). On the south-eastern Islands, Santa Cruz and San Cristòbal (Islands F and G), more than 40% of debris is accounted for by fishing activity in international

Nation Flag	Total	Island A	Island B	Island C	Island D	Island E	Island F	Island G
Ocean SE	23.3	23.2	34.6	13.3	20.0	14.6	42.6	47.4
Ocean NW	2.2	6.9	4.6	1.0	2.5	0.1	2.6	3.8
Ocean NE	0.7	1.2	0.6	0.4	0.8	0.0	0.7	1.6
Ocean SW	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Peru	20.8	35.3	35.6	14.8	23.3	11.6	19.0	33.8
Galápagos	51.5	27.7	21.0	69.7	48.7	73.5	34.0	11.2
Costa Rica	0.1	1.2	0.1	0.0	0.1	0.0	0.1	0.1
Colombia	0.3	2.1	1.4	0.1	0.2	0.2	0.4	0.2
Panama	0.6	0.6	0.8	0.4	3.7	0.1	0.3	0.6
Ecuador	0.6	1.8	1.3	0.2	0.7	0.0	0.4	1.4

Table 5: Source Regions: Detailed probabilities of beaching in total and per Island

waters in the southeast of the domain. The fact that a large fraction of debris stems from China and Peru, and that Islands in the east of the Galápagos receive a larger fraction from these regions, supports the hypothesis that most debris is carried towards the Islands via the Humboldt current.

These findings may be somewhat limited by the bias of the GFW dataset, previously discussed in section 3.1.1. The most important being that boats less than 15 meters are very likely to be not accounted for. This is important to bear in mind since Castrejón et al. (2020) found that approximately 97% of active vessels around the Galápagos are smaller than 9.6 meters long. The establishment of the here presented framework set the first steps towards developing a system that allows for reliable source identification of the amount of beached fishing debris. Many factors have not been taken into account. One of them is the assumption that more debris is released with loner fishing duration. On the other hand, the size of boats has not been taken into account. Another factor of uncertainty is the assumption that debris is released during fishing activity. And since there is no global, or local, estimate on the release of fishing debris, no mass can be assigned to the found beaching distributions. An extensive inquiry into existing literature did not yield information on these factors. The effect of each of them can be easily implemented into the framework, but without an observational background, it remains obtuse.

The previous section has presented a source identification framework that allows determining the beaching distribution of different categories of fishery sources. It found that the majority of debris is released within the EEZ of the Galápagos, International waters in the southeast, and the EEZ of Peru. The study suggests that squid jigging and purse seining are the gear types most responsible for the beaching of debris from fisheries. Ecuadorian, Chinese and Peruvian flags play contribute to most debris arriving at the Islands. With these results, this framework provides support for the implementation of policies to reduce the pollution by debris stemming from fisheries. It also supports the investigation and mapping of ocean pathways and it was created to be easily applied to island nations and archipelagos worldwide.

3.3 Nutrients

The final part of this study investigates the relationship between fishing effort and net primary production. To conduct a comparison fishing activity needs to be transformed from listed data to fields with the same spatial and temporal resolution. First Kernel Density was examined as a measure to achieve the transformation. Results show that the application of two-dimensional histograms turns out to be more efficient. With this information, an Empirical Orthogonal Function Analysis is performed to study possible patterns of spatial variability. The final step is to calculate the correlation and cross-correlation of the average regional timeseries of fishing effort and net primary production.

3.3.1 Kernel Density Estimation

A Kernel Density Estimation of the daily fishing effort was performed to introduce smoothness to the daily histograms of fishing efforts which proved to show a tendency to be noisy. The idea was that this would improve comparability to net primary production, as this data shows smooth behavior.

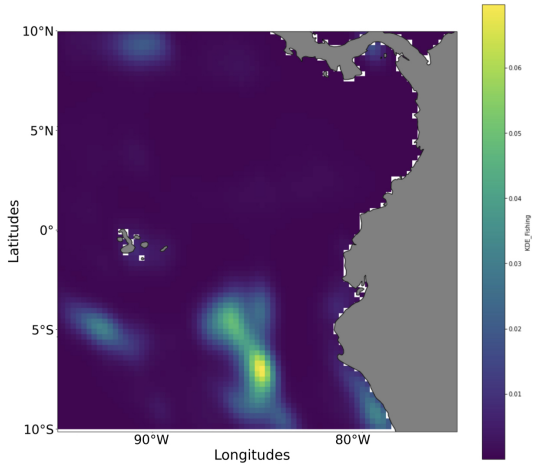


Figure 22: Kernel Density Estimation of fishing effort over the period April 2016 to January 2021.

Figure 22 shows the results of the KDE for the entire dataset. The figure succeeds in representing the general structure of the fishing distribution. However introduced on daily data, it shows a tendency of over smoothing smaller efforts. It proved difficult to choose a fitting smoothness since the data varies strongly. For this reason, the simpler approach, histograms were chosen. The use of KDE could be employed for averages over certain periods, for example, monthly averages. In this way, it could be used to replace the histogram when comparing to nutrient data. This option should be considered with caution since in this case the right kernel function and smoothing parameter would still have to be determined.

3.3.2 Empirical Orthogonal Function Analysis

To investigate the relationship an between fishing effort, net primary production and chlorophyll-a in a spatial and temporal sense, an EOF Analysis was conducted.

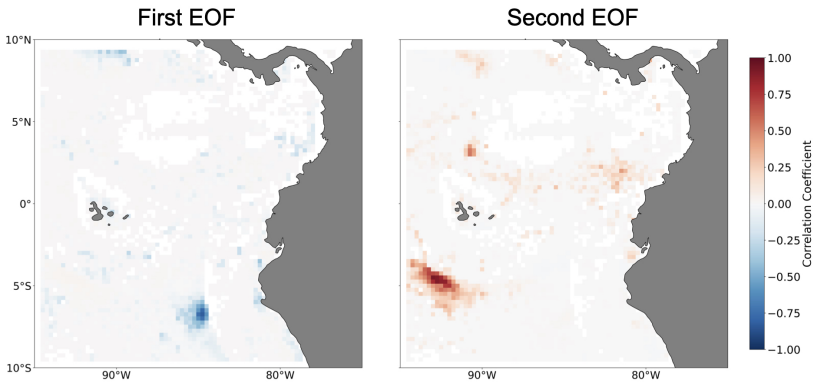


Figure 23: Empirical Orthogonal Function Analysis of fishing effort, net primary production and chlorophyll-a over the period 2012 to 2021 is shown. Left: First EOF expressed as correlation. Right: Second EOF expressed as correlation.

Figure 23 shows the EOF analysis for the first two EOFs.These two EOFs are calculated for the entire period. A strong negative correlation is visible before the coast of Peru and a strong positive correlation in the southwest of Galápagos EEZ. The relating time series is depicted in Figure 24, showing two strong events.

EOF-Analysis turns the input spatial-temporal data set into a set of maps representing patterns of variance. These EOFs are basis functions in terms of variance. The time series for each map are the coefficients for each basis function at a given time. These determine the contribution of that map to the original data set at each time. The first few PCs account for the largest variation in the original field. This proves unsuitable for this research as I am mostly interested in variables coinciding over long periods and not in their biggest contribution towards variance. It is important to note that the sign of correlation might be inverted due to the analysis. So negative correlation could also be an indication of a positive correlation.

The maps show regions of high correlation in international waters and associated peaks in variance within associated time series. These are both locations known for high intensity of fishing effort. One occurred in June 2019 and one occurred in June 2020. The periods are also known for the high intensity of fishing within international waters, see Figure 25. In conclusion, one can say that the EOF does show expected results but these are not useful for this analysis.

3.3.3 Correlation of Time Series

Time series of daily averaged amounts of fishing activity and net primary production within the domains' EEZ and international waters were created. To investigate a possible relationship

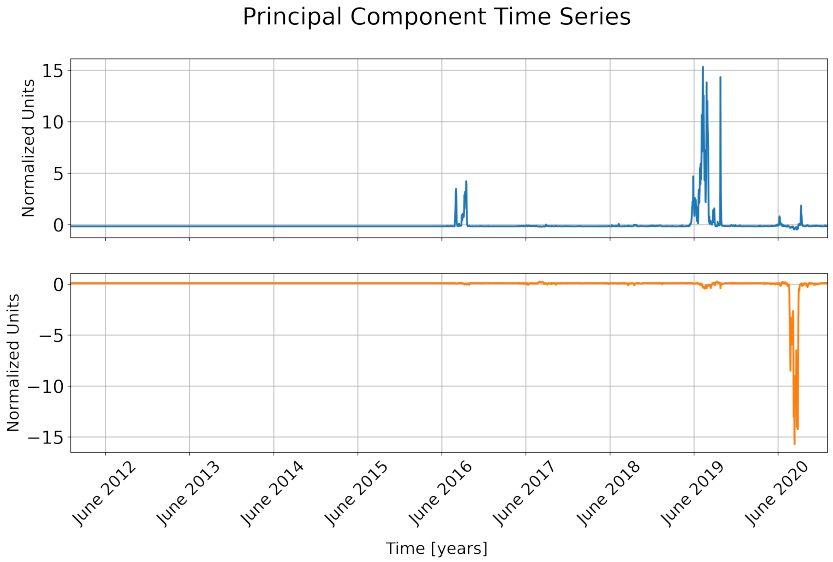


Figure 24: Principal Component Time Series of the first (upper) and second (lower) Principal component.

between the two parameters their correlation and cross-correlation were calculated

Galápagos	Costa Rica	Panama	Colombia	Ecuador	Peru	International Waters
-0.067	0.168	-0.272	-0.033	0.271	0.028	0.2

Table 6: Correlation of timeseries for fishing effort and net primary production per region

The comparison of time series for fishing effort and NPP are shown in Figure 25. The resulting Pearson correlation is shown in Table 6. In the Galápagos EEZ, a very small anticorrelation of -0.067 is found. This is visible in the plot as well, NPP shows a seasonal signal with low values in fall. Fishing effort, on the other hand, shows an increase during the entire period. There are apparent peaks in fishing effort while NPP dips in the middle of 2017, beginning and end of 2018, end of 2020. The time series within the Costa Rican EEZ show a positive correlation of 0.168. This is reflected when looking at the time series. During the boreal winter months is a visible decrease in both fishing effort and NPP. There is very little average fishing compared to others. This can be explained by the large size of the EEZ and the occurrence of little fishing effort, see Figure 3. Fishing effort shows some periodicity, with increased effort in late fall. Panama exhibits a negative correlation of -0.272. This is visible in in Figure 25, where, relative to other regions, strongly anticorrelated behavior is exhibited. Both time series exhibit seasonal signals. Additionally, fishing effort increases over time. The timeseries from the Colombian EEZ seem to coincide at the beginning, but this signal gets lost over time. A negative correlation of -0.033 is produced. In the beginning, it looks like the signal coincides. A seasonal signal in NPP is visible, with peaks in the spring months. The strongest correlation compared to other EEZ can be observed in Ecuador with 0.271. This is probably due to peaking values at the end of the period in both NPP and fishing effort. It is also visible that fishing effort increases over time. Peru accounts for 16.9% of fishing effort in the entire domain. There are strong seasonal signals for fishing effort and NPP. However, there are two peaks of fishing effort per year, but only of

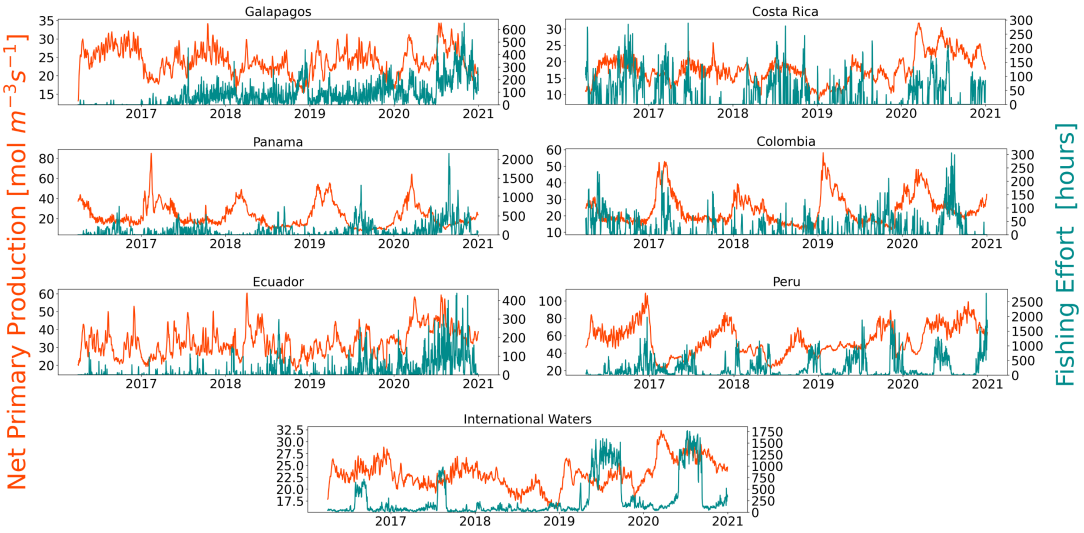


Figure 25: The time series of the averaged fishing effort and net primary production per each domain. This analysis was done for the domain [75°, 105°] West and [-10°, 10°] North, over the period April 2016 to January 2021. The corresponding correlation values are shown in Table 6.

NPP. The correlation results in 0.028. In International Waters the correlation is 0.2, this is also where 68 % of fishing takes place.

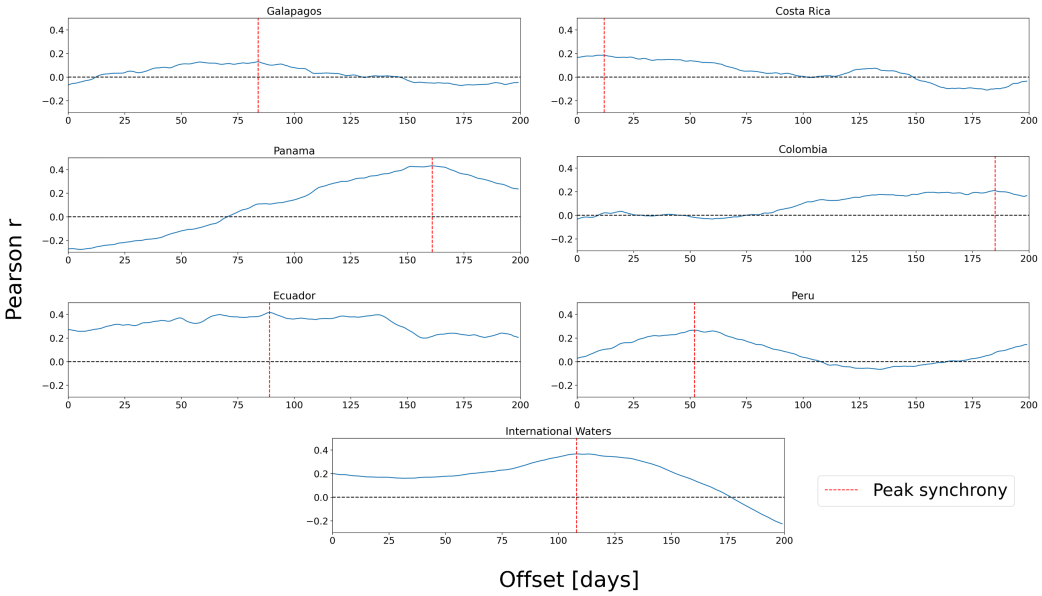


Figure 26: The cross-correlation of time series of the averaged fishing effort and net primary production per each domain. This analysis was done for the domain [75°, 105°] West and [-10°, 10°] North, over the period April 2016 to January 2021.

The cross-correlation between the time series is shown in Figure 26. The x-axis is the offset in days, while the y-axis shows the Pearson correlation. The red line indicates the maximum correlation. The average maximum offset is 99 days with a standard deviation of 55. This large standard deviation indicated that no common lag between net primary production and fishing effort was found.

A clear relationship between fishing and net primary production has not been found, this may be due to a multitude of factors. Kroodsma et al. (2018) notes that fishing depends on all kinds of variables, including holidays, fuel price, and distance harbour. Cimino et al. (2019) found a negative relationship between fishing effort and chlorophyll but attributed this inconsistency to different factors. It needs to be considered that cross-correlations can indicate spurious relationships, because of autocorrelation. Fishing effort increases strongly over time due to an increase in AIS registries. This artificial trend makes the investigation of correlations difficult. However, some signals indicate commonality and a different choice for spatial regions and the creation of time series could produce better results.

The literature review suggested that a relationship between fishing and environmental parameters can exist on a local scale. The most prevalent ones are SST and CHL, a proxy from primary production. Overall, there seems to be some evidence to indicate that the use of boosted regression trees can be a valuable tool in predicting fishing effort. Since some commonality between fishing and net primary production has been found an extension of this analysis could yield satisfying results. When including multiple variables into the analysis, like logistic factors mentioned before and oceanographic factors like sea surface temperature, and sea surface height a possible relationship could be found.

4 Conclusion

This study set out to determine whether debris from fishing activity played a role in the pollution of Galápagos islands' shores. A simulation of this kind of debris confirmed that fishing litter contributes to the pollution of the islands' beaches. An investigation into its trajectories has demonstrated that the Humboldt current plays a leading role in transporting debris toward the islands. However, the results of this study are limited since debris was considered a passive tracer floating on the ocean, and the influence of degradation, windage, and resuspension was not taken into account. The second aim of this study was to investigate the origins of beached debris. A framework was created that allows determining the source distribution of beached fishing debris. This source identification has been made possible by the connection of a dataset on apparent fishing activity presented by Global Fishing Watch, a Lagrangian particle simulator, OceanParcels, and a stochastic beaching approach, similar to Onink et al. (2021). The findings of this framework have shown that the largest fraction of beached fishing debris stems from within the Galápagos Exclusive Economic Zone. However, a significant number arrive from international waters. This makes it hard for single nation states to invoke regulations. A key policy priority should therefore be to find international measures that enforce management plans for debris on fisheries. The main weakness of this study is the lack of observations on the behavior of the release of fishing and missing knowledge of the amounts of debris discarded. The here presented findings study are based on a modeling approach. It is not yet possible to validate findings with the classification of actually observed debris on the islands. The third aim of this investigation was to assess the relationship between fishing effort and net primary production. It was not possible to establish a clear connection but the net primary production can potentially be used as a weak predictor in prediction models using multiple variate approaches. These results lay the groundwork for future research into the prediction of debris from fisheries.

Here I outline the several questions that remain to be answered. Further, I highlight missing observations that pose limitations for the presented study. Finally, I outline with which steps this study could have been improved.

Further work that needs to be done includes establishing information on the temporal variability of beaching, an investigation into the relationship between release and beaching, and checking for possible variability of the arrival of the established main sources. The introduction of an increase in sampling frequency and resolution of the ocean field once the debris approaches the Galápagos Islands could improve the employed beaching framework. This can be achieved by using the current output to determine the locations of debris entering the Galápagos EEZ. These locations can then be employed in a spatially smaller simulation with higher resolution in space and time. This will allow for more precise statements on the trajectories of debris from fisheries. Additional research might explore how the estimate of global fractions of fishing gear being lost, provided by Richardson et al. (2021), could be employed in the established framework. One way to do this could be to refine the Lagrangian Analysis, by specifying the properties of marine debris released, for example, drag. Other effects that play a role in the advection of debris can be used to improve particle tracking. One of them is the inclusion of a 3D-velocity field. To allow for the investigation into the prediction of fishing effort, an inquiry into the use of boosted regression trees is recommended. Before starting excursion on this matters I suggest the following literature: (De'Ath, 2007; Kroodsma et al., 2018; Lange et al., 2017; Soykan et al., 2014).

This research has shown many deficiencies in our knowledge of marine debris from fisheries. The most pressing is a mass estimate of global annual debris released. But the study also raised questions other questions: How and when is marine debris discarded? Which big a fraction of this debris is fishing gear, and how much non-gear debris is discarded? To answer these questions observations of fishing boats are necessary. Debris releasing behavior likely varies between fishing boats. Also, a study into the release of debris on boats might be hampered by the presence of the observer and their effect on the release behavior of the crew. To validate the results of this research, observations of pollution on the shores are required. Since regular observations of beached debris at the Galápagos shores, the framework could be tested in a location with more rigorous information on beached debris.

What would I have done differently were I to do this research again. I would spend the first month composing a research proposal to set out a structure and compass. The goal would be to set targets that are achievable and lay out a road to accomplish them. Within this proposal, the main tool would be defined and create the focal point for the months to come. In this case, this would have been OceanParcels. Once the goal is achieved the findings will be denoted, presented, and discussed. Only after satisfying understanding, an investigation into further targets would be started. This would include a profound literature review and the contacting of researchers to figure out if their approach applies to my case study. The contact with other researchers as part of the research I was looking forward to and did not exercise enough. I learned that before launching into an investigation, it is important and appropriate to discuss with colleagues or supervisors if this approach makes sense. With the knowledge achieved by now, a better set-up for coding will be possible, and there are surely ways in which I could have been more efficient. Being more proactive in contacting people to ask questions would have helped. To ease the writing process, I will in the future write biweekly summaries and further split up the writing process. I was certain that I could not do this. But having learned so much about the writing process, I now know that in the future I will be able to separately write each section.

References

Andrades, R., Santos, R. G., Joyeux, J.-C., Chelazzi, D., Cincinelli, A. & Giarrizzo, T. (2018). Marine debris in trindade island, a remote island of the south atlantic. *Marine pollution bulletin*, *137*, 180–184.

Aumont, O., Éthé, C., Tagliabue, A., Bopp, L. & Gehlen, M. (2015). Pisces-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development*, *8*(8), 2465–2513.

Browne, M. A., Chapman, M. G., Thompson, R. C., Amaral Zettler, L. A., Jambeck, J. & Mallos, N. J. (2015). Spatial and temporal patterns of stranded intertidal marine debris: Is there a picture of global change? *Environmental Science & Technology*, *49*(12), 7082–7094.

Butcher, J. C. (2016). *Numerical methods for ordinary differential equations*. John Wiley & Sons.

Castrejón, M. & Charles, A. (2020). Human and climatic drivers affect spatial fishing patterns in a multiple-use marine protected area: The galapagos marine reserve. *PloS one*, *15*(1), e0228094.

Chávez-Castrillón, F., Coltorti, M., Ivaldi, R., Lucena-Sánchez, E. & Sciavicco, G. (2020). Temporal aspects of chlorophyll-a presence prediction around galapagos islands. *International Conference on Technologies and Innovation*, 98–110.

Chen, Y.-C. (2017). A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, *1*(1), 161–187.

Cimino, M. A., Anderson, M., Schramek, T., Merrifield, S. & Terrill, E. J. (2019). Towards a fishing pressure prediction system for a western pacific eez. *Scientific reports*, *9*(1), 1–10.

Cushman-Roisin, B. & Beckers, J.-M. (2011). *Introduction to geophysical fluid dynamics: Physical and numerical aspects*. Academic press.

De'Ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology*, *88*(1), 243–251.

Downey, A. B. (2021). *Think bayes*. " O'Reilly Media, Inc."

Drillet, Y., Chune, S. L., Levier, B. & Drevillon, M. (2019). Smoc: A new global surface current product containing the effect of the ocean general circulation, waves and tides. *Geophysical Research Abstracts*, *21*.

Eriksen, M., Lebreton, L. C., Carson, H. S., Thiel, M., Moore, C. J., Borerro, J. C., Galgani, F., Ryan, P. G. & Reisser, J. (2014). Plastic pollution in the world's oceans: More than 5 trillion plastic pieces weighing over 250,000 tons afloat at sea. *PloS one*, *9*(12), e111913.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X. et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *kdd*, *96*(34), 226–231.

Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *science*, *281*(5374), 237–240.

FlandersMarineInstitute. (2016). *Maritime boundaries geodatabase, version 9* [Available online at https://www.marineregions.org/ https://doi.org/10.14284/317].

Goodman, C. (2009). The regime for flag state responsibility in international fisheries law-effective fact, creative fiction, or further work required? *Austl. & NZ Mar. LJ*, *23*, 157.

Groll, A. & Tutz, G. (2012). Regularization for generalized additive mixed models by likelihood-based boosting. *Methods of Information in Medicine*, *51*(02), 168–177.

Hannachi, A., Unkel, S., Trendafilov, N. & Jolliffe, I. (2009). Independent component analysis of climate data: A new look at eof rotation. *Journal of Climate*, *22*(11), 2797–2812.

Harati-Mokhtari, A., Wall, A., Brooks, P. & Wang, J. (2007). Automatic identification system (ais): Data reliability and human error implications. *The Journal of Navigation*, *60*(3), 373–389.

Hardesty, B. D., Lawson, T., van der Velde, T., Lansdell, M. & Wilcox, C. (2017). Estimating quantities and sources of marine debris at a continental scale. *Frontiers in Ecology and the Environment*, *15*(1), 18–25.

Hazen, E. L., Palacios, D. M., Forney, K. A., Howell, E. A., Becker, E., Hoover, A. L., Irvine, L., DeAngelis, M., Bograd, S. J., Mate, B. R. et al. (2017). Whalewatch: A dynamic management tool for predicting blue whale density in the california current. *Journal of Applied Ecology*, *54*(5), 1415–1428.

Hoyer, S. & Hamman, J. (2017). Xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, *5*(1). https://doi.org/10.5334/jors.148

Jones, J. S., Porter, A., Muñoz-Pérez, J. P., Alarcón-Ruales, D., Galloway, T. S., Godley, B. J., Santillo, D., Vagg, J. & Lewis, C. (2021). Plastic contamination of a galapagos island (ecuador) and the relative risks to native marine species. *Science of The Total Environment*, *789*, 147704.

Jones, M. M. (1995). Fishing debris in the australian marine environment. *Marine Pollution Bulletin*, *30*(1), 25–33.

Jones, P. J. (2013). A governance analysis of the galápagos marine reserve. *Marine Policy*, *41*, 65–71.

Kaandorp, M. L., Dijkstra, H. A. & van Sebille, E. (2020). Closing the mediterranean marine floating plastic mass budget: Inverse modeling of sources and sinks. *Environmental science & technology*, *54*(19), 11980–11989.

Kershaw, P. (2016). *Marine plastic debris and microplastics–global lessons and research to inspire action and guide policy change*. United Nations Environment Programme.

Kroodsma, D. A., Mayorga, J., Hochberg, T., Miller, N. A., Boerder, K., Ferretti, F., Wilson, A., Bergman, B., White, T. D., Block, B. A. et al. (2018). Tracking the global footprint of fisheries. *Science*, *359*(6378), 904–908.

Kuczenski, B., Vargas Poulsen, C., Gilman, E. L., Musyl, M., Geyer, R. & Wilson, J. (2022). Plastic gear loss estimates from remote observation of industrial fishing activity. *Fish and Fisheries*, *23*(1), 22–33.

Lan, K.-W., Shimada, T., Lee, M.-A., Su, N.-J. & Chang, Y. (2017). Using remote-sensing environmental and fishery data to map potential yellowfin tuna habitats in the tropical pacific ocean. *Remote Sensing*, *9*(5), 444.

Lange, M. & van Sebille, E. (2017). Parcels v0. 9: Prototyping a lagrangian ocean analysis framework for the petascale age. *Geoscientific Model Development*, *10*(11), 4175–4186.

Law, K. L. (2017). Plastics in the marine environment. *Annual review of marine science*, *9*, 205–229.

Lebreton, L.-M., Greer, S. & Borrero, J. C. (2012). Numerical modelling of floating debris in the world's oceans. *Marine pollution bulletin*, *64*(3), 653–661.

Lehodey, P., Senina, I. & Murtugudde, R. (2008). A spatial ecosystem and populations dynamics model (seapodym)–modeling of tuna and tuna-like populations. *Progress in Oceanography*, *78*(4), 304–318.

Li, T. & Philander, S. G. H. (1996). On the annual cycle of the eastern equatorial pacific. *Journal of Climate*, 2986–2998.

Liubartseva, S., Coppini, G., Lecci, R. & Clementi, E. (2018). Tracking plastics in the mediterranean: 2d lagrangian model. *Marine pollution bulletin*, *129*(1), 151–162.

Malarky, L. & Lowell, B. (2018). Avoiding detection: Global case studies of possible ais avoidance. *Oceana. url: https://usa. oceana. org/publications/reports/avoiding-detection-global-case-studies-possible-ais-avoidance*.

McCauley, D. J., Woods, P., Sullivan, B., Bergman, B., Jablonicky, C., Roan, A., Hirshfield, M., Boerder, K. & Worm, B. (2016). Ending hide and seek at sea. *Science*, *351*(6278), 1148–1150.

McInnes, L., Healy, J. & Astels, S. (2017). Hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, *2*(11), 205.

McKinney, W. et al. (2011). Pandas: A foundational python library for data analysis and statistics. *Python for high performance and scientific computing*, *14*(9), 1–9.

Messié, M. & Chavez, F. (2011). Global modes of sea surface temperature variability in relation to regional climate indices. *Journal of Climate*, *24*(16), 4314–4331.

Natale, F., Gibin, M., Alessandrini, A., Vespe, M. & Paulrud, A. (2015). Mapping fishing effort through ais data. *PloS one*, *10*(6), e0130746.

O'hara, K. J. et al. (1988). *A citizen's guide to plastics in the ocean: More than a litter problem.* ERIC.

Onink, V., Jongedijk, C. E., Hoffman, M. J., van Sebille, E. & Laufkötter, C. (2021). Global simulations of marine plastic transport show plastic trapping in coastal zones. *Environmental Research Letters*, *16*(6), 064053.

Organisation, I. (2014). International convention for the prevention of pollution from ships (marpol). *Annex IV Prevention of Pollution by Sewage from Ships (entered into force 27 September 2003) http://www. imo. org/About/Conventions/ListOfConventions/Pages/International-Convention-for-the-Prevention-of-Pollution-from-Ships-(MARPOL). aspx (date accessed April 2022)*.

Page, B., McKenzie, J., McIntosh, R., Baylis, A., Morrissey, A., Calvert, N., Haase, T., Berris, M., Dowie, D., Shaughnessy, P. D. et al. (2004). Entanglement of australian sea lions and new zealand fur seals in lost fishing gear and other marine debris before and after government and industry attempts to reduce the problem. *Marine pollution bulletin*, *49*(1-2), 33–42.

Paulino, C. & Escudero, L. (2011). Use of night satellite imagery to monitor squid fishery in peru. *Handbook of satellite remote sensing image interpretation: applications for marine living resources conservation and management. EU PRESPO and IOCCG, Darthmouth*, 143–153.

Perruche, C. (2018). Product user manual for the global ocean biogeochemistry hindcast global_reanalysis_bio_001_029. version 1.

Phillips, J. S., Gupta, A. S., Senina, I., van Sebille, E., Lange, M., Lehodey, P., Hampton, J. & Nicol, S. (2018). An individual-based model of skipjack tuna (katsuwonus pelamis) movement in the tropical pacific ocean. *Progress in Oceanography*, *164*, 63–74.

Pierard, C., Bassotto, D., Meirer, F. & Van Sebille, E. (2021). Attribution of river-sourced floating plastic in the south atlantic ocean using bayesian inference.

Richardson, K., Hardesty, B. D. & Wilcox, C. (2019). Estimates of fishing gear loss rates at a global scale: A literature review and meta-analysis. *Fish and Fisheries*, *20*(6), 1218–1231.

Richardson, K., Haynes, D., Talouli, A. & Donoghue, M. (2017). Marine pollution originating from purse seine and longline fishing vessel operations in the western and central pacific ocean, 2003–2015. *Ambio*, *46*(2), 190–200.

Richardson, K., Wilcox, C., Vince, J. & Hardesty, B. D. (2021). Challenges and misperceptions around global fishing gear loss estimates. *Marine Policy*, *129*, 104522.

Ruiz, I., Rubio, A. et al. (2022). Modelling the distribution of fishing-related floating marine litter within the bay of biscay and its marine protected areas. *Environmental Pollution*, *292*, 118216.

Ryan, P. G., Moore, C. J., Van Franeker, J. A. & Moloney, C. L. (2009). Monitoring the abundance of plastic debris in the marine environment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1526), 1999–2012.

Santos, A. M. P. (2000). Fisheries oceanography using satellite and airborne remote sensing methods: A review. *Fisheries Research*, *49*(1), 1–20.

Scales, K. L., Hazen, E. L., Maxwell, S. M., Dewar, H., Kohin, S., Jacox, M. G., Edwards, C. A., Briscoe, D. K., Crowder, L. B., Lewison, R. L. et al. (2017). Fit to predict? ecoinformatics for predicting the catchability of a pelagic fish in near real time. *Ecological Applications*, *27*(8), 2313–2329.

Schofield, J., Wyles, K. J., Doherty, S., Donnelly, A., Jones, J. & Porter, A. (2020). Object narratives as a methodology for mitigating marine plastic pollution: Multidisciplinary investigations in galápagos. *Antiquity*, *94*(373), 228–244.

Schubert, E., Sander, J., Ester, M., Kriegel, H. P. & Xu, X. (2017). Dbscan revisited, revisited: Why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, *42*(3), 1–21.

Soykan, C. U., Eguchi, T., Kohin, S. & Dewar, H. (2014). Prediction of fishing effort distributions using boosted regression trees. *Ecological Applications*, *24*(1), 71–83.

van Duinen, B., Kaandorp, M. L. & van Sebille, E. (2022). Identifying marine sources of beached plastics through a bayesian framework: Application to southwest netherlands. *Geophysical Research Letters*, e2021GL097214.

van Sebille, E., Delandmeter, P., Schofield, J., Hardesty, B. D., Jones, J. & Donnelly, A. (2019). Basin-scale sources and pathways of microplastic that ends up in the galápagos archipelago. *Ocean Science*, *15*(5), 1341–1349. https://doi.org/10.5194/os-15-1341-2019

van Sebille, E., Aliani, S., Law, K. L., Maximenko, N., Alsina, J. M., Bagaev, A., Bergmann, M., Chapron, B., Chubarenko, I., Cózar, A. et al. (2020). The physical oceanography of the transport of floating marine debris. *Environmental Research Letters*, *15*(2), 023003.

van Sebille, E., Griffies, S. M., Abernathey, R., Adams, T. P., Berloff, P., Biastoch, A., Blanke, B., Chassignet, E. P., Cheng, Y., Cotter, C. J. et al. (2018). Lagrangian ocean analysis: Fundamentals and practices. *Ocean Modelling*, *121*, 49–75.

Virto, L. R. (2018). A preliminary assessment of the indicators for sustainable development goal (sdg) 14 "conserve and sustainably use the oceans, seas and marine resources for sustainable development". *Marine Policy*, *98*, 47–57.

Wheeler, M. C. & Hendon, H. H. (2004). An all-season real-time multivariate mjo index: Development of an index for monitoring and prediction. *Monthly weather review*, *132*(8), 1917–1932.

Wilson, G., Aruliah, D. A., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., Haddock, S. H., Huff, K. D., Mitchell, I. M., Plumbley, M. D. et al. (2014). Best practices for scientific computing. *PLoS Biol*, *12*(1), e1001745.