

# **Identifying the Most Efficient Plastic Clean-up Locations in the Galapagos Archipelago**

**Quinten Bohte**

Supervisors:  
Erik van Sebille & Steffie Ypma

Faculty of Natural Sciences  
Utrecht University  
Netherlands  
09-07-2021

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Method</b>	<b>4</b>
2.1	Model . . . . .	4
2.2	Beaching Parametrization . . . . .	4
2.3	Simulations . . . . .	6
2.4	Transition Matrices and Networks . . . . .	6
2.5	Lagrangian Network Analysis . . . . .	9
2.5.1	Betweenness Centrality . . . . .	9
2.5.2	Eigenvector Centrality . . . . .	10
2.6	Markov Chain Analysis . . . . .	11
2.6.1	Distribution . . . . .	11
2.6.2	Clustering . . . . .	11
2.6.3	Sink Implementation and Plastic Loss . . . . .	13
<b>3</b>	<b>Results</b>	<b>13</b>
3.1	Particle Pathways . . . . .	13
3.2	Transition Matrices . . . . .	16
3.3	Clean-up Locations . . . . .	16
3.3.1	Accumulation zones . . . . .	16
3.3.2	Crossing Hotspots . . . . .	17
3.4	Plastic Loss into the Ocean . . . . .	20
3.5	Use of the constructed methods for a plastic release from San Cristobal . . . . .	21
<b>4</b>	<b>Discussion</b>	<b>24</b>
4.1	Beaching and Resuspension Sensitivity . . . . .	24
4.2	Trapped particles . . . . .	26
4.3	Comparison Methods . . . . .	28
4.4	Clean-up Strategies . . . . .	29
<b>5</b>	<b>Conclusion</b>	<b>30</b>
	<b>Appendices</b>	<b>31</b>
<b>A</b>	<b>Particle Trajectories</b>	<b>31</b>

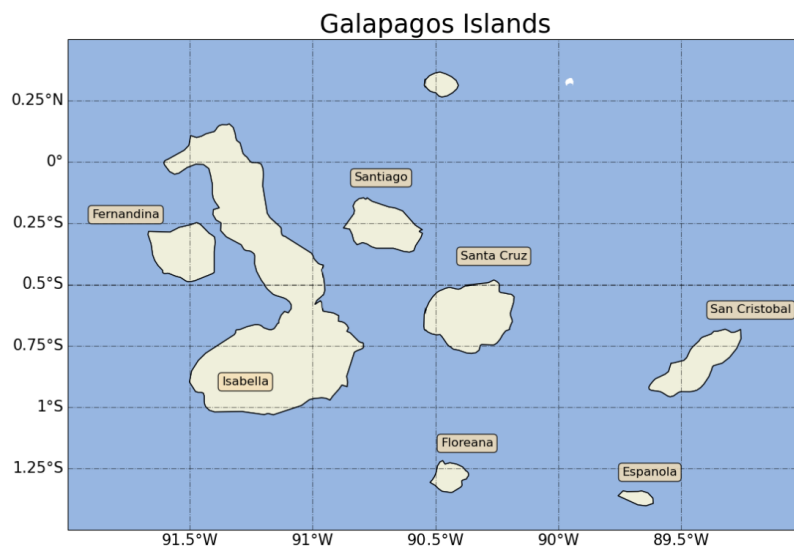
# 1 Introduction

Human activities have a major impact on the decrease of biodiversity [14] and such a worrying situation has arisen, that we have created the conditions for a serious mass extinction [24]. In marine ecosystems, one of the causes of the decrease of biodiversity is the pollution by plastic debris [14]. The properties of plastic, lightweight, durable and cheap, make them very suitable for the manufacture of many products. These properties in particular make them also very harmful for marine environments [5]. Plastics are highly resistant to natural decay, but will also break up into smaller pieces. Due to the buoyancy of many plastics, they become widespread in the sea and on the beaches [5]. The production of plastic is rapidly increasing in the last decades [10] and so does the amount of plastic that ends up in marine environments, especially in coastal zones [12][3]. The ecosystems in these coastal zones are particularly sensitive for these plastics [19]. Plastic debris affects marine species primarily by mechanical threats. Entanglement in plastic debris, like old fishing gear, and the ingestion of plastic debris are a huge threat for marine species and seabirds [34][14].

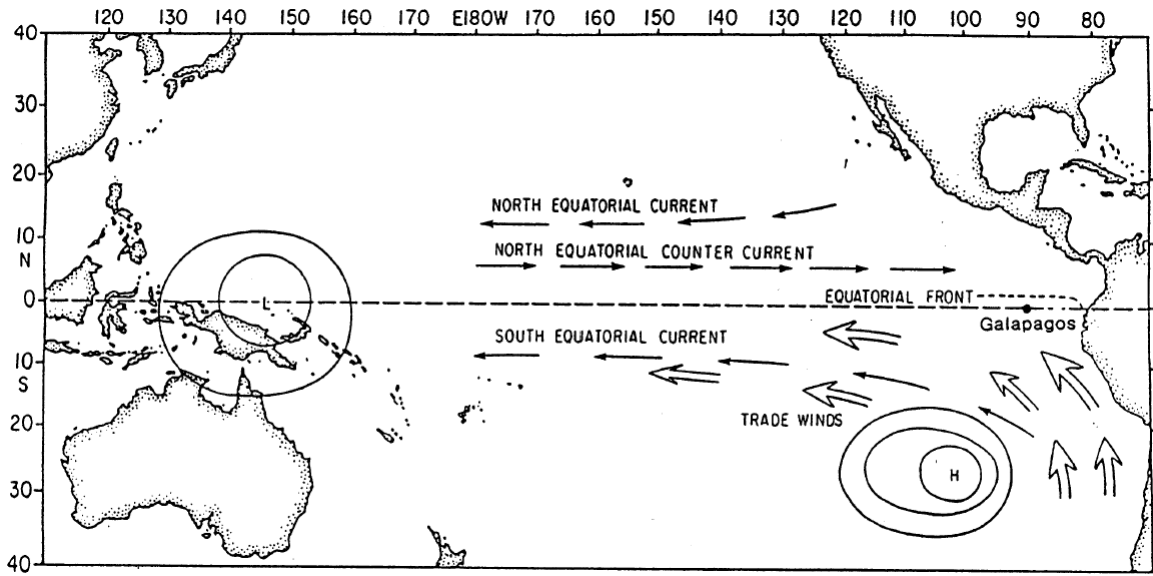
The Galapagos Islands have a unique marine ecosystem that is severely threatened by plastic debris [16]. They are home to more than 3200 marine species [16]. A risk scoring analysis has identified 32 species which have ended up in a vulnerable situation and have an urgent need for mitigation of the effects of plastics [16].

Plastics are not only harmful for the coastal ecosystems of these islands and their beaches but also decrease their economic value. The economic value of beaches can drop 97 % due to plastic debris [7]. Therefore, plastic pollution in the Galapagos Islands can have huge ecological and subsequent socioeconomic consequences, since tourism covers 80% of the local economy [16].

The Galapagos Islands belong to the Republic of Ecuador and are located in the Eastern Pacific Ocean, 1000 kilometers from the west-coast of South America (figure 2). They consist of nine islands with areas from 15 km<sup>2</sup> to almost 4600 km<sup>2</sup> (figure 1). The Eastern Pacific Ocean is characterized by a high pressure system from which the southeast trade winds flow towards a low pressure system around Indonesia (figure 2). These winds are driving the Humboldt (Peru) Current and the South Equatorial Current (SEC), ranging from 5° North to 10° South. The Peru Current moves along the coast of South America and merges into the westward flowing SEC. The SEC causes a general westward flow within the Galapagos archipelago. The Galapagos Islands are located south of the Inter Tropical Convergence Zone (ITCZ) and are consequently located in the south east regime of the trade winds. Due to the high pressure system in the east of the Pacific (figure 2) and the location of the ITCZ, the Galapagos Islands are subject to an average southerly wind [9].



**Figure 1:** Topography map of the Galapagos Islands.



**Figure 2:** Map of the Pacific Ocean showing the low and high pressure systems forcing the Humboldt current and the south equatorial current. The Galapagos Islands are crossed by these currents [9].

Modelling studies suggest that most plastic in the Galapagos Islands originates from the coast of Ecuador and Peru and nearby fishing regions [32]. The plastic originating from Ecuador and Peru is transported to the Galapagos Islands by the Humboldt current and the SEC [16].

The Galapagos Islands can be considered as a network in which plastic is transported. When plastic enters the waters of the Galapagos Islands a process of beaching and resuspension starts [22], causing plastic being transported between different beaches of the islands. The beaches in the network where most plastic eventually ends up, are called *accumulation zones*. The beaches where most plastic crosses along their paths through the network, are in this research referred to as *crossing hotspots*. Due to the limited resources of the Ecuadorian Government, it is important to remove the plastic from the network as efficient as possible. The accumulation zones as well as the crossing hotspots are considered as the most efficient clean-up locations.

The aim of this research is to identify the accumulation zones and crossing hotspots, by analysing the connectivity of plastic transport between the islands. The intended objective is to improve clean-up strategies in the Galapagos Islands, by giving recommendations about the most efficient clean-up locations.

From literature [4] [2] [1] [20] [15] [31] [26] [25], theoretical concepts were found on the basis of which one method has been constructed for identifying the accumulation zone and two methods have been constructed for identifying the crossing hotspots. The construction of these methods will be explained in detail in the methodology section (section 2). The Lagrangian networks combined with network science tools will be used to identify the connectivity of plastic transport between the island in order to identify the accumulation zones and crossing hotspots. For the crossing hotspots an additional method, based on Markov Chains is constructed. With these Markov Chain models the propagation of a plastic input through the network can be determined within a specific time window. By implementing sinks in these Markov Chain models, the crossing hotspots can be identified. The results of the identified crossing hotspots by both methods will be compared to identify differences and similarities and whether these can be explained based on the conceptual differences between the two methods.

## 2 Method

This chapter will explain how the methods are constructed to identify the most efficient clean-up location in the Galapagos islands. The first section explains how the simulations of the particle trajectories between the islands were executed (section 2.1 and 2.3). These particle trajectories were converted into transition matrices (section 2.4), used for the construction of the Lagrangian Networks (section 2.4) and the Markov Chain Models (section 2.4). Subsequently, the Lagrangian network analysis (section 2.5) and the Markov Chain analysis (section 2.6.1 and 2.6.3) were conducted in order to identify the most effective clean-up locations.

### 2.1 Model

We used Ocean Parcels [18] to simulate trajectories of virtual plastic particles. This research is limited to only floating plastic, so the virtual plastic particles were advected using ocean surface flow field data. The flow field data used in this research consist of the MITgcm LLC4320 simulation as well as surface Stokes drift estimates from Wavewatch III Hindcast dataset. Floating particles can be subjected to wind, where the effect of the wind depends on the size of the object. The windage and Stokes drift seem to be similar on global scale [22]. Therefore, an extra windage term was not included in our model. The MIT general circulation model (MITgcm) lat-lon-polar cap 4320 (llc4320) simulations are the output of the MIT general circulation model, which solves the Navier Stokes equations using a Boussinesq approximation. The components of the flow are staggered in space in the form of an Arakawa C grid.

This research used a LLC simulation. These LLC simulations are a set of simulations developed with the MITgcm with increasing resolutions  $1/12^\circ$ ,  $1/24^\circ$  (2-5 km) and  $1/48^\circ$  (0.75-2.3 km). The LLC4320 simulation provides the highest resolution ( $1/48^\circ$ ). One of the largest improvements of the LLC simulations compared to previous simulations is the extension of the gridded domain to the arctic. There are three types of gridding the earth. The first type cubed-sphere grid (CS) has been used for MITgcm simulations. However, some shortcomings have been noticed: when the horizontal resolution increases, the loss of orthogonality near the cube corners increases, and some of the grid vertices are located in the ocean and not on land. These shortcomings led to the development of the second type, which is the Lon-Lat-Cap (LLC) curvilinear grid. The grid vertices are located on land and the grid heterogeneities remains acceptable with high resolutions ( $1/48^\circ$ )[13]. Between  $70^\circ\text{S}$  and  $57^\circ\text{N}$  the grid reverts to the third type, the lon-lat grid.

The model is forced at the surface with 6-hourly atmospheric fields from the European center for Medium-Range Weather Forecasts atmospheric operational model analysis which are converted to surface fluxes using the bulk formulae. Bathymetry is from Smith and Sandwell 1997 Version 14.1 and IBCAO Version 2.23. Additionally, it also includes atmospheric pressure and 16 tidal constituents, which are applied as additional atmospheric forcing. The MITgcm was run with a time step of 25 seconds and with 90 vertical levels, ranging from a thickness of 1 m at the surface to 480 m near the bottom, with a maximum model depth of 7 km. The processes resolved by the MITgcm are eddies, fronts, coastal upwelling, internal and surface tides and waves and barotropic variability. The simulation ran from September 10, 2011 to November 15, 2012 and saved all the model variables at hourly intervals [28][33]. The validation of the model is based on tides and internal gravity waves. It accurately simulates tidal flow fields and internal gravity waves continuum [33]. Additionally, the MITgcm captures the upper ocean stratification and variability very well [6].

The surface Stokes drift plays an important role in the shoreward transport of particles [23] and so for the grounding of particles, however, it is not included in the MITgcm LLC 4320 simulations. Therefore, the surface Stokes drift estimates from the WaveWatch III hindcast dataset were added to the model, having a spatial resolution of  $1/2^\circ$  and a temporal resolution of 3 hours [22].

### 2.2 Beaching Parametrization

An important factor in our research is the beaching of particles. Hence, before we ran the particle simulations as described in section 2.3 and analysed the connectivity between the islands, we parametrized the beaching process of particles. From [22][11][27] we know that the amount of beached plastic on coastlines is influenced by many different processes, such as wind direction and speed, coast angle, aspect and morphology, local runoff and the degree of human usage. However, these factors do not have high predictive power in statistical models [11][27]. Therefore, it is very unclear how these processes should be parametrized. Due to this it was decided to parametrize

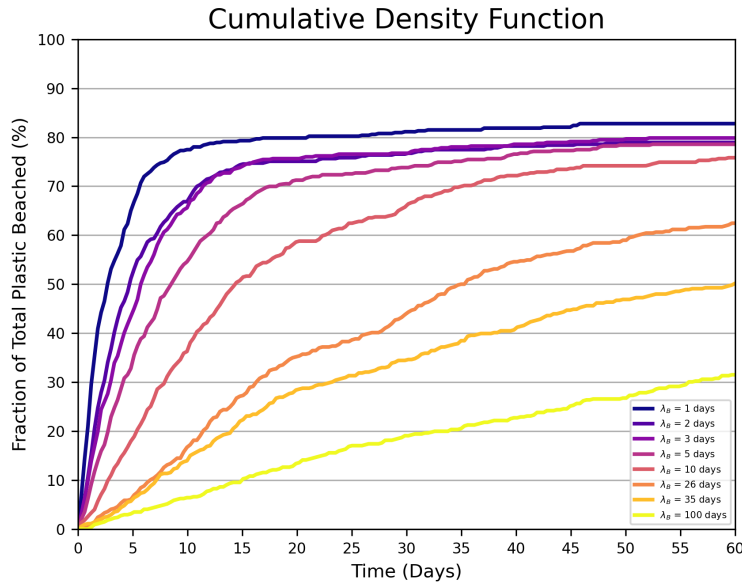
the beaching process with a simple beaching probability as done in [22]. This beaching probability,  $p_B$ , considers beaching as stochastic process in the coastal zone and is calculated for every time step as follows:

$$p_B = \begin{cases} \text{if } d \leq D, p_B = 1 - \exp(-dt/\lambda_B) \\ \text{if } d > D, p_B = 0 \end{cases} \quad (1)$$

Where  $d$  is the distance of the particle to the coast,  $D$  is the distance in which beaching can occur, the beaching zone,  $dt$  is the integration time step and  $\lambda_B$  is the characteristic beaching timescale. As in [22] the width of the beaching zone was chosen in such a way that it included all the coastal cells, resulting in a beaching zone of 2km. To decide whether a particle beaches or not we generated a random number between zero and one. When  $p_B$  was larger than this randomly created number, a particle was marked as beached.

From equation 1 it becomes clear that the probability of beaching is determined by the beaching timescale  $\lambda_B$ , which is the number of days that a particle must spend in the beaching zone in order to have a 63.2 % probability of beaching [22]. Different values for  $\lambda_B$  have been determined by different studies. [17] finds  $\lambda_B = 76$  days by analysing drifter trajectories and a value of 26 days for plastic debris using inverse modelling. Based on the previous papers, [22] comes with a range of plausible beaching timescales, 1-100 days. However, the results of [22] shows that global beaching patterns are fairly robust against the value of the beaching probability and consequently the value of  $\lambda_B$ . As in [22], we acknowledge that the value of  $\lambda_B$  is still a major uncertainty in our analysis.

In order to pick a beaching timescale a sensitivity analysis was conducted. We assessed the sensitivity of the beaching patterns in the Archipelago and the fraction of beached particles with respect to the characteristic beaching timescale. The values used for the analysis are within the plausible range of 1-100 days as in [22],  $\lambda_B \in [1, 2, 5, 10, 26, 35, 100]$ . Figure 3 displays the sensitivity of the fraction of beached plastic to  $\lambda_B$ . As expected, with increasing values for  $\lambda_B$  less particles will beach. The beaching patterns seem to be quite robust against the value of  $\lambda_B$ . These results are discussed in more detail in the discussion section.



**Figure 3:** This cumulative density function displays the sensitivity of the fraction of beached particles to the value of the characteristic beaching timescale  $\lambda_B$ . We observe that with increasing values of  $\lambda_B$  the fraction of beached particles within 60 days decreases.

Combining the results of the sensitivity analysis and previous research we decided to continue our research with  $\lambda_B = 5$  days. We picked a value which is slightly lower than in [22], since our beaching zone is slightly smaller. When the exact same coastal conditions are considered, a particle will beach typically faster when it is close to shore then when it is further away from shore. Justifying the smaller value for  $\lambda_B$ , with a smaller beaching zone.

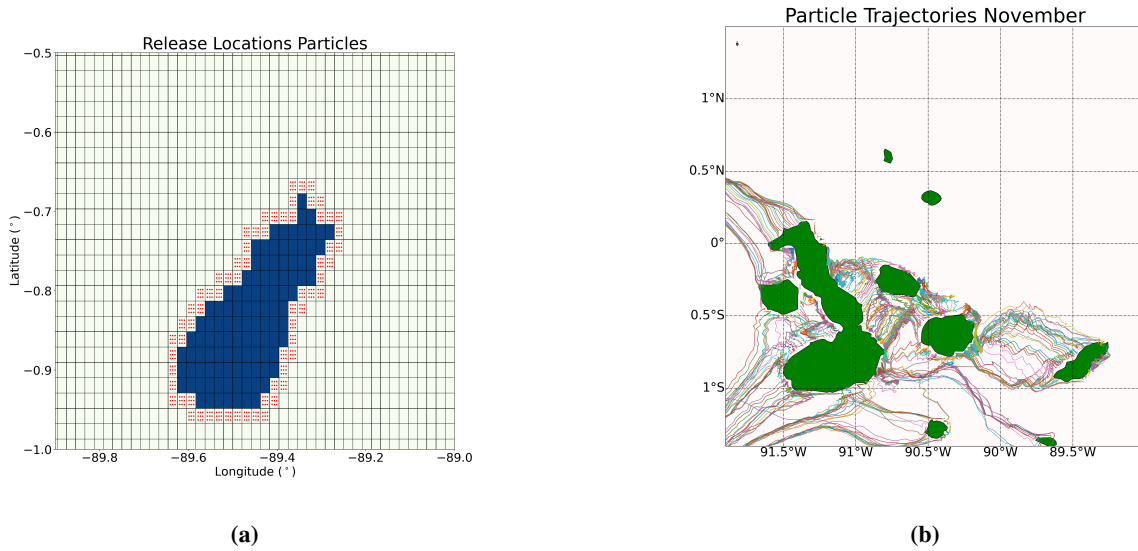
Additionally, with  $\lambda_B = 5$  days a good quantity of the plastic has beached. This is important for our analysis, since we want to have a good picture of the connectivity between the islands. When only a very small quantity of the particles beach the connectivity is not represented very well.

### 2.3 Simulations

With the beaching parametrization, explained in the previous section, we can run the simulations and track the beaching of particles in the Galapagos Islands. Flow field data from November till December were used for the simulations. In this research two types of simulations were conducted. The first type,  $S_1$ , was used to construct transition matrices for the Markov Chain analysis and requires a relative short simulation length, referred to as *interval* length from this point. Based on the results of the sensitivity analysis in the previous section we made our intervals 15 days. Within 15 days almost 80 percent of all the plastic that will beach had beached (figure 3) and the interval is still relatively small. In  $S_1$  particles are only released at  $t = 0$ .

The second type,  $S_2$  is used for the Lagrangian flow network analysis, which requires a longer interval. From figure 3 we know that within 30 days almost all particles have beached, so simulations were run for 30 days. In order to get an idea of the global connectivity within a month, we released particles every five days.

Both simulations had a time step (dt) of 15 minutes and the particle locations were saved every 15 minutes. The modelling domain runs from  $-1.5^\circ$  till  $1.5^\circ$  latitude and from  $-92^\circ$  till  $-88^\circ$  longitude, resulting in grid cells with a size of  $0.02^\circ \times 0.02^\circ$  degrees. The intervals of  $S_1$  and  $S_2$  were defined as  $l \in [\text{Nov}_1, \text{Nov}_2, \dots, \text{Apr}_1, \text{Apr}_2]$  and  $k \in [\text{Nov}, \dots, \text{Apr}]$ , respectively. The release locations of the particles are the same for  $S_1$  and  $S_2$ . In every land-adjacent ocean grid cells nine particles were released (Figure 4a). The modelling contains 545 land-adjacent ocean grid cells, which results in the release of a total of 4905 particles ( $9 \times 545$ ) in total for the  $S_1$  simulations. During the  $S_2$  simulations, 29430 particles were released in total, 4905 particles every five days for a month ( $4905 \times 6$ ). Figure 4b shows the trajectories of one of the intervals.



**Figure 4:** Particles are released at the coastal grid cells of the islands of the Galapagos. (a) The release locations of the particles on the land-adjacent coastal cells of the island San Cristobal. (b) Particle trajectories for the month November 2012.

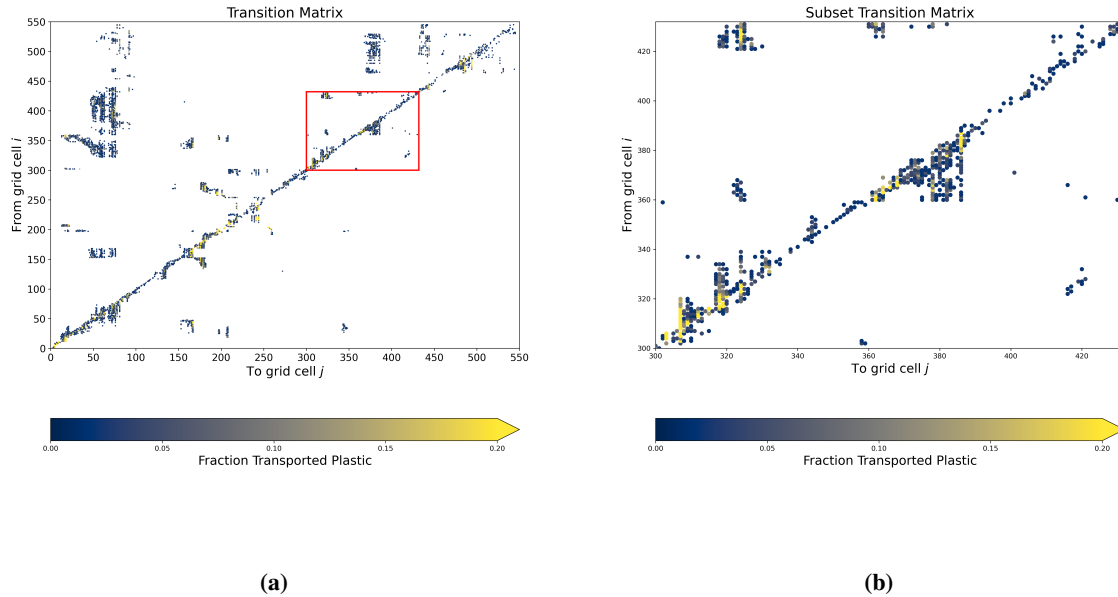
### 2.4 Transition Matrices and Networks

The particle trajectories from  $S_1$  and  $S_2$  were converted into transition matrices [26][31][1]. These transition matrices capture all the information about the plastic transport between the coastal grids. Both analyses, the *Lagrangian flow*

network analysis and Markov Chain analysis, were based on these transition matrices. How these analyses were exactly executed will be discussed in more detail in section 2.5 and 2.6.

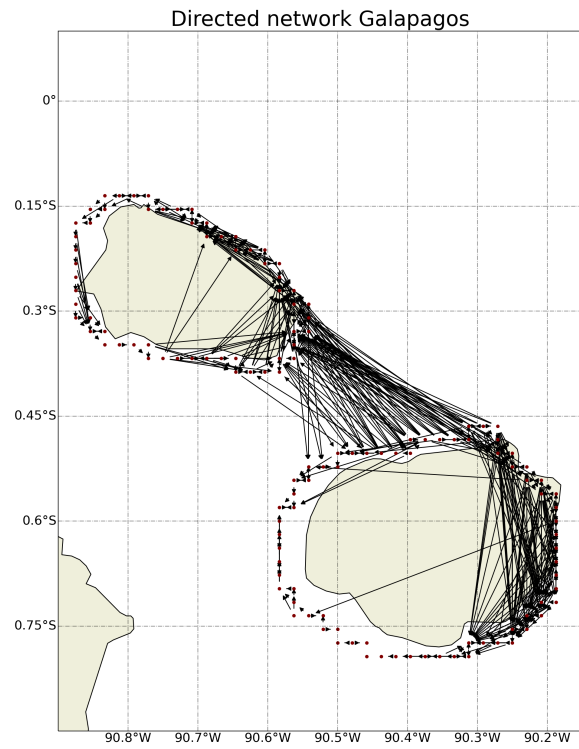
The transition matrices have the shape,  $\mathbf{A}(r, b)$ , with  $b$  and  $r$  representing the release grids and the beach grids, respectively. Note that one ocean grid is included in the transition matrix, since a fraction of the particles ends up in the ocean. The number of times a particle released, at time  $t$ , from grid  $r$  and beached at grid  $b$  within  $t + \Delta t$ , with  $\Delta t$  as the length of the simulation, were added to  $\mathbf{A}_{b,r}$ . Consequently, the rows of the transition matrices were normalized, resulting in entries representing transport probabilities or transported plastic fractions. These matrices now represent the fractional distribution of the plastic at the end of each simulation. and can be interpreted as a probability distribution of all the plastic particles at the end of the simulation. Transition matrices constructed for  $S_1^k$  and  $S_2^l$  will be denoted as  $\mathbf{N}_{r,b}^k$  and  $\mathbf{M}_{r,b}^l$ , respectively. With the superscript indicating the simulation interval. Figure 5a shows the transition matrix of the month of November and figure 5b shows a subset of the transition matrix, representing the connections between the islands Santa Cruz and Santiago.

At this point these matrices were now suitable to construct Markov Chain models, to determine plastic distribution in our system. For the Lagrangian network flow analysis we first converted  $\mathbf{M}_{r,b}^l$  into directed graphs as done in [26]. These networks represent the connectivity within the system. The graphs are denoted as  $G^k = (V, E)$  and consist of a set of nodes  $V$  and a set of edges  $E$ , which are ordered pairs of elements in  $V$  and where an edge  $i, j \in E$  represent a connection between nodes  $i$  and  $j$ , which are coastal grid cells. This means that the edges are represented by the non-zero entries of  $\mathbf{M}_{r,b}^l$ . For every graph,  $V = \{1, \dots, 545\}$  and  $E$  will vary for every interval, since the connectivity between the islands is different for every interval due to changing currents. The weights of edges are denoted as  $w(e_{r,b}) = \mathbf{M}_{r,b}^l$ , representing the fraction of plastic transport from node  $r$  to node  $b$ , i.e an edge  $(i,j)$  with  $w(e_{i,j}) = 0.7$ , means that 70% of the plastic from node  $i$  is transported to node  $j$ . Since  $\mathbf{M}_{r,b}^l$  is not necessarily the same as  $\mathbf{M}_{b,r}^l$  our network is an directed network. Figure 6a shows the directed network constructed from a subset of the transition matrix shown in figure 5b. The subset is indicated with a red box. The large amount of connections between all nodes causes this figure to become very chaotic. In order to visualize this network more clearly, the nodes are rearranged in figure 6b. This clearly shows the directed network constructed from the subset of the total network.

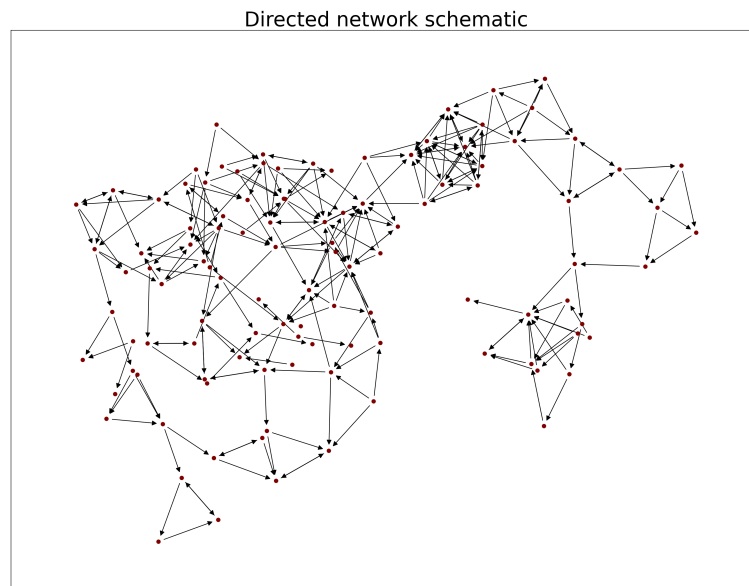


**Figure 5:** Figure a shows the transition matrix of the month November. Figure b shows the subset of the transition matrix enclosed by the red lines in figure a. In figure b it is more clear how the transported plastic fraction variate between nodes.





(a)



(b)

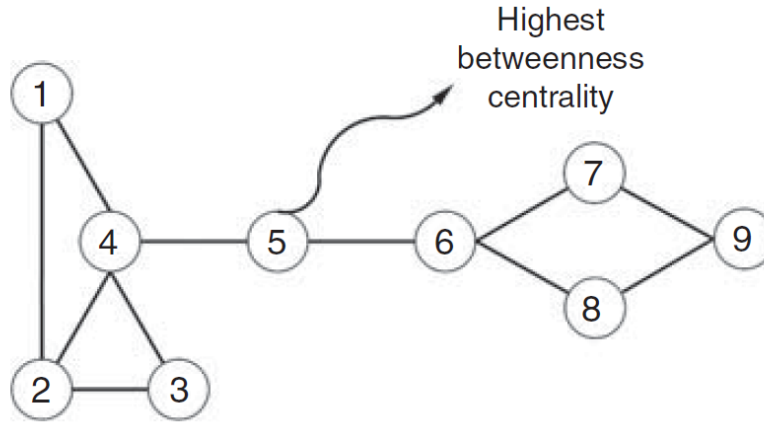
**Figure 6:** The sub network of the islands Santiago and Santa Cruz is drawn in figure a. Where the red dots represent the coastal grids and the arrows represent transport between them. Due to the arrangement of the coastal grids the arrows overlap very much. When we rearrange the nodes, it is easier to observe how the nodes are connected and that it is a directed network (figure b).

## 2.5 Lagrangian Network Analysis

The connectivity of the plastic transport between the islands is now represented by the constructed networks. With these networks and the Python module Networkx, we were now able to analyse the networks for all intervals. As mentioned in the introduction, the aim of this research is to identify the accumulation zones and the crossing hotspots within the network. With the use of centrality measures of the nodes, these locations can be identified. A *centrality measure* of a node says something about its relative importance within the network. Many centralities measures are known from network theory [15][25]. The centrality measures used in this research are the betweenness centrality and the eigenvector centrality. The betweenness centrality indicates the nodes where most plastic crosses (the crossing hotspots) and eigenvector centrality indicates the nodes where most plastic accumulates. The following sections will explain how these methods were constructed.

### 2.5.1 Betweenness Centrality

The *betweenness centrality* of a particular node indicates the number of times this particular node is crossed along the shortest path between any pairs of nodes[20]. Nodes with high betweenness centralities are important for the flow of information through the network[15]. A common analogy of such a node, is a bus stop (node) in a public transport network in which the highest betweenness centrality has the most traffic[15]. Figure 7 visualizes a simple network that contains nine nodes. Node five has the highest betweenness centrality, since most shortest paths are crossing this node. We consider that plastic is transported from the left side of the network (node 1-4) to the right side of the network (node 6-9) and vice versa. All the plastic has to cross node five in order to reach the the other side of the network. Therefore, node five is the most efficient clean-up location in this network. In our networks the nodes with the highest betweenness centralities represent the nodes where most plastic crosses and thus the most efficient clean-up locations, as mentioned in the introduction. The betweenness centrality of a node  $v$  is calculated with expression 2.



**Figure 7:** An example of a network with one node with a very high betweenness centrality. When plastic is transported through the this network, node five would be the most efficient location to remove plastic from the system [4]

$$g(v) = \sum_{r \neq v \neq b} \frac{\sigma_{r,b}(v)}{\sigma_{r,b}} \quad (2)$$

Where,  $\sigma_{r,b}$  represent all shortest paths from release nodes  $r$  to beach nodes  $b$  and  $\sigma_{r,b}(v)$  all shortest paths crossing node  $v$ . This can easily be calculated using the Networkx module in python, which only needs a graph/network as input. This betweenness centrality was calculated using the shortest paths between nodes, however, a more interesting measure would be when we use the most likely path, since plastic does not necessarily takes the shortest path. In [21] a good approach is given for finding the most likely path between two nodes. First, all possible paths  $P_{r,b}$  between release node  $r \in V$  and beach node  $b \in V$  are defined as:

$$P_{r,b} = \{ \mathbf{p} = (p_r, p_1, p_2, \dots, p_b) : p_i \in V, \forall i \in \{1, \dots, n-1\}, p_r = r, p_b = b, p_{i-1} \neq p_i \}$$

Using the transition matrices,  $\mathbf{M}_{r,b}^l$ , the probability of a path is given by:

$$P(p) = \prod_{i=0}^{n-1} M_{p_i, p_{i+1}} \quad (3)$$

When all paths  $p \in P_{r,b} = \{p_0, p_1, p_2, \dots, p_n\}$  are considered, the most likely path  $\hat{\mathbf{p}}$ , is the path with the highest probability. This path is found by the following expression:

$$\hat{\mathbf{p}} = \arg \max_{\mathbf{p} \in P_{r,b}} \{ P(\mathbf{p}) \} = \arg \max_{\mathbf{p} \in P_{r,b}} \left\{ \prod_{i=0}^{n-1} T_{p_i, p_{i+1}} \right\} \quad (4)$$

For solving the optimization of equation 4, first the *log* of  $P(\mathbf{p})$  is taken:

$$\log P(\mathbf{p}) = \sum_{i=0}^{n-1} \log T_{p_i, p_{i+1}} \quad (5)$$

Then the following statement is used:

$$\begin{aligned} \hat{\mathbf{p}} &= \arg \max_{\mathbf{p} \in P_{r,b}} \{ \log P(\mathbf{p}) \} \\ &= \arg \min_{\mathbf{p} \in P_{r,b}} \{ -\log P(\mathbf{p}) \} \\ &= \arg \min_{\mathbf{p} \in P_{r,b}} \left\{ \prod_{i=0}^{n-1} T_{p_i, p_{i+1}} \right\} \end{aligned}$$

This expression can now be solved using the Dijkstra's Algorithm for finding the shortest path. The weights of the edges are now denoted as  $w(e_{r,b}) = -\log(\mathbf{M}_{r,b}^k)$ . The Dijkstra's algorithm will find the path  $P = \{p_1, p_2, \dots, p_n\}$ , such that this path minimizes:

$$\sum_{i=1}^{n-1} w(e_{p_i, p_{i+1}}) \quad (6)$$

This solves the optimization problem of equation 4. The betweenness centrality is now calculated, using the most likely paths instead of shortest paths, as expressed below.

$$g(v) = \sum_{r \neq v \neq b} \frac{\hat{\mathbf{p}}_{r,b}(v)}{\hat{\mathbf{p}}_{r,b}} \quad (7)$$

## 2.5.2 Eigenvector Centrality

The *eigenvector centrality* measure defines a relative score for a node based on its connections. The *centrality* of a node is the number of connections it has with other nodes. The eigenvector centrality is a not only a measure of the centrality of the node it self, but also takes into account the centralities of the indirect ones of any length, so nodes being away more than only one edge [20]. Connections with high centrality nodes contribute more to the score of the node than connections with low centrality nodes[15]. If a node is pointed to by many nodes, also having

high eigenvector centralities, then that node will have a high eigenvector centrality [20]. Hence, we can consider nodes with high eigenvector centralities as accumulation zones in the system. The locations of nodes with high eigenvector centralities may be the locations in the Galapagos Islands where most plastic accumulates. With the Networkx module the eigenvector centralities for all nodes and all intervals were calculated. The weights ( $w$ ) of the edges in our network are also taken into account when computing the eigenvector centrality.

## 2.6 Markov Chain Analysis

This section will explain how the transition matrices, constructed in section 2.4, are used in the Markov Chain models. With these Markov Chain models the distribution of a plastic input can be determined (section 2.6.1). By implementing sinks in the Markov Chain models the regions where most plastic crosses, referred to as crossing hotspots, can be identified (section 2.6.3).

### 2.6.1 Distribution

The transition matrices,  $\mathbf{N}^l$ , define the propagation of the plastic through the system as a Markov Chain. They tell us how a plastic input is distributed over the system within the time window of the Markov Chain. More specifically, the fraction from the total input released at grid cell  $r$  at time  $t$  that is transported to grid cell  $b$  at  $t + \Delta t$ , with  $\Delta t = 15$  days. The distribution of a plastic input, over multiple time windows, can be determined by vector-matrix multiplication as is done in [31][1]. This vector-matrix multiplication is conducted as follows, a plastic input at time  $t$  is given as the vector  $\mathbf{v}_t = [v_1, \dots, v_{546}]$ . The entry indices and the entry values of  $\mathbf{v}_t$  represent the release grids and the plastic fraction released, respectively. The distribution of this input vector from  $t$  to  $t + \Delta t$  is,  $\mathbf{v}_{t+\Delta t} = \mathbf{v}_t \times \mathbf{N}^l$ . The non-zero entries of  $\mathbf{v}_{t+\Delta t}$  represent the fraction of the plastic input that ended in that grid after 15 days. The expression below gives an example how the plastic distribution of  $\mathbf{v}_t$  in November is determined.

$$\mathbf{v}_{\text{end}} = \mathbf{v}_t \times \mathbf{N}^{\text{nov}_1} \times \mathbf{N}^{\text{nov}_2} \quad (8)$$

The initial condition of the input vector can vary. However, in our research we used an evenly spread plastic input over all coast grids. It has the same fraction of plastic for every entry, a value of  $1/546$ . With this input vector we obtained a regional picture of the plastic distribution over time.

From figure 5a it becomes clear that the constructed transition matrices are quite sparse. After a couple of vector-matrix multiplications the sparsity increases rapidly. In order to reduce the sparsity, a clustering algorithm is used to find clusters in the transition matrices. We will first explain what clustering algorithm was used and how it works. Thereafter, we will explain what these clusters exactly represent and finally, we will explain the approach to reduce the sparsity using these clusters.

### 2.6.2 Clustering

The detection of regions with high densities of beached plastics was the goal of the clustering. Therefore, we used the density based clustering algorithm DBSCAN. DBSCAN is able to find arbitrary shaped clusters and define isolated points as noise. It accepts two parameters, the radius,  $\epsilon$ , and the number of minimal points (MinPts), which have to lie within  $\epsilon$ . In order to explain the concept of DBSCAN some terms are explained as follows [8].

**Definition 1** (*neighbor*). A point  $q$  within range  $\epsilon$  of point  $p$ .

**Definition 2** (*Eps-neighborhood*). The Eps-neighborhood of a point,  $p$  in a database,  $D$  is defined by  $\{q \in D | \text{dist}(p, q) \leq \epsilon\}$

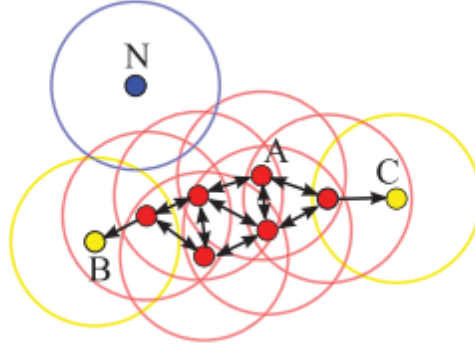
**Definition 3** (*Core object*). A core object refers to a point of which the neighborhood with a given radius (Eps) contains at least a minimum number (MinPts) of other points.

**Definition 4** (*Border object*). A point which lies within an Eps-neighborhood, but is not a core object.

**Definition 5** (*Noise*). A point which does not lie within any Eps-neighborhood of any core object.

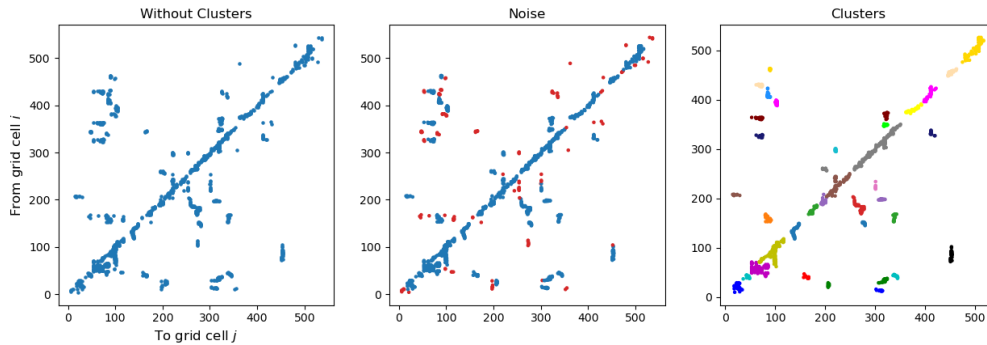
The DBSCAN algorithm starts with the first point in  $D$ . It defines the *Eps-neighborhood* of this point. When the total points of the *Eps-neighborhood* are larger than the MinPts, all points and  $p$  are assigned to a new cluster. This process

is repeated for all core points in the new cluster and continues until no core points are found, but only border points. All points found by the latter process are now assigned to the new cluster. The process is now repeated until all points in  $D$  are assigned to a cluster or labeled as noise. The figure below gives a visualization of the described process.



**Figure 8:** Illustration of the DBSCAN model. DBSCAN starts at point A and executes the explained process. At some point it reaches point B and C, the border points. All red and yellow points belong to one cluster. Point N does not lie within any Eps-Neighborhood so is assigned as noise. [30]

Figure 9 gives a visualization of the clustering of one of the transition matrices. The left plot shows the transition matrix before clustering. The middle plot shows the points which are labeled as noise (red) by the DBSCAN algorithm and the right plot shows the clusters found by the DBSCAN algorithm. As the aim of our clustering is to find regions with high densities, the points labeled as noise should be isolated and points in dense regions should not be labeled as noise. From the middle plot it can be observed that DBSCAN assigns mainly isolated points as noise. Therefore, it works as it supposed to.



**Figure 9:** Visualization of the clustering procedure of one of the transition matrices. The left plot shows the initial transition matrix, the middle plot shows the points labeled as noise in red and the right plot shows all clusters mapped to different colors.

From section 2.4 we know that every entry in the transition matrix represents a release grid and beach grid. Therefore, every cluster found in the transition matrix represents a release cluster and a beach cluster. The release clusters are denoted as  $C_r^l$  and the beach clusters as  $C_b^l$ , with  $l$  indicating the interval.  $C_r^l$  and  $C_b^l$  are both arrays with the entry indices corresponding to the grid number and the entry values corresponding to the cluster which the grid is assigned to. With this in mind we can explain how these clusters are used in the plastic distribution calculation with the transition matrices.

As mentioned before, the evolution of the input vector  $\mathbf{v}_t$  is given by,  $\mathbf{v}_{t+\Delta t} = \mathbf{v}_t \times \mathbf{N}^l$ . With  $C_r^{l+1}$  we can check to what clusters all the non-zero entries of  $\mathbf{v}_{t+\Delta t}$  belong to. The plastic fraction of the grids belonging to a cluster will be spread over all grids of that cluster. This whole procedure will be executed between every vector-matrix

multiplication and consequently making the output vectors less sparse. We are now able to determine the distribution of a plastic input through the system.

### 2.6.3 Sink Implementation and Plastic Loss

As explained in the previous section, now it is possible to compute the distribution of a plastic input over the coastlines in the Galapagos Islands. Plastic can be removed from the system by the implementation of plastic sinks (locations where all plastic is removed from the system). Most plastic will be removed from the system when sinks are located in the regions where most plastic crosses, crossing hotspots. Therefore, indicating the most effective sink locations will indicate the regions where most plastic crosses along their pathways through the system, crossing hotspots, considered as optimal clean-up locations (introduction).

We consider a sink as a location where all plastic that arrives will be removed. Sinks can be implemented relatively easy by setting all entries of a row of  $\mathbf{N}^l$  to zero,  $\sum_{N}^{j=1} \mathbf{N}^l = 0$ . All the plastic that arrives at this grid will now be removed from the system and not be further distributed. Sinks are implemented in transition matrices in a similar way in [1]. The efficiency of a sink can be determined by calculating the removed plastic after a full vector-matrix multiplication. Consider the multiplication of November, equation 8. When we set one of the rows to zero in  $\mathbf{N}^{nov1}$  and  $\mathbf{N}^{nov2}$ , execute this multiplication and calculate  $\sum_N^i \mathbf{v}_{end}$ , we know how much plastic is left in the system after the multiplication. This process is iteratively executed until every row is set to zero once and after every iteration we save  $\sum_N^i \mathbf{v}_{end}$ . The difference between  $\sum_N^i \mathbf{v}_{end}$  when setting a row to zero compared to  $\sum_N^i \mathbf{v}_{end}$  without setting any row to zero, represent the sink efficiency of the grid corresponding to that row.

As mentioned in section 2.4, every transition matrix also contains a row representing plastic lost in the ocean. One can imagine that after every multiplication more plastic will end up in the ocean instead of on an island. Plastic that ends up in the ocean may be considered as lost since it is very difficult to remove it from the ocean. The plastic loss after a vector-matrix multiplication into the ocean is simply the value of the entry representing the ocean,  $\mathbf{v}_{end}^{i=ocean}$ . These values may be interesting in order to find out how quickly we lose plastic into the ocean and how quickly we should remove plastic from the Galapagos Islands.

## 3 Results

This chapter presents and discusses the results of our research. Section 3.1 presents the particle pathways and discusses the mean pathways across the investigated period as well as the variation that exists between the months. The transition matrices constructed from these particle pathways will be presented in section 3.2, which are subsequently used for the Lagrangian Network and Markov Chain analyses. The optimal beach clean-up locations, the accumulation zones and crossing hotspots, identified by these analyses will be discussed in section 3.3. In subsection 3.3.1 the results of the accumulation zones will be discussed and in subsection 3.3.2 the results of the crossing hotspots will be discussed. As mentioned in the introduction and explained in the method section, the crossing hotspots are identified using the Lagrangian method (betweenness centrality) and the Markov Chain method (sink implementation). The identified crossing hotspots by these two different methods will also be compared in section 3.3.2. Thereafter, the constructed methods will be used to identify the most efficient clean-up locations when plastic is released solely from San Cristobal. The results from this analysis are discussed in section 3.5.

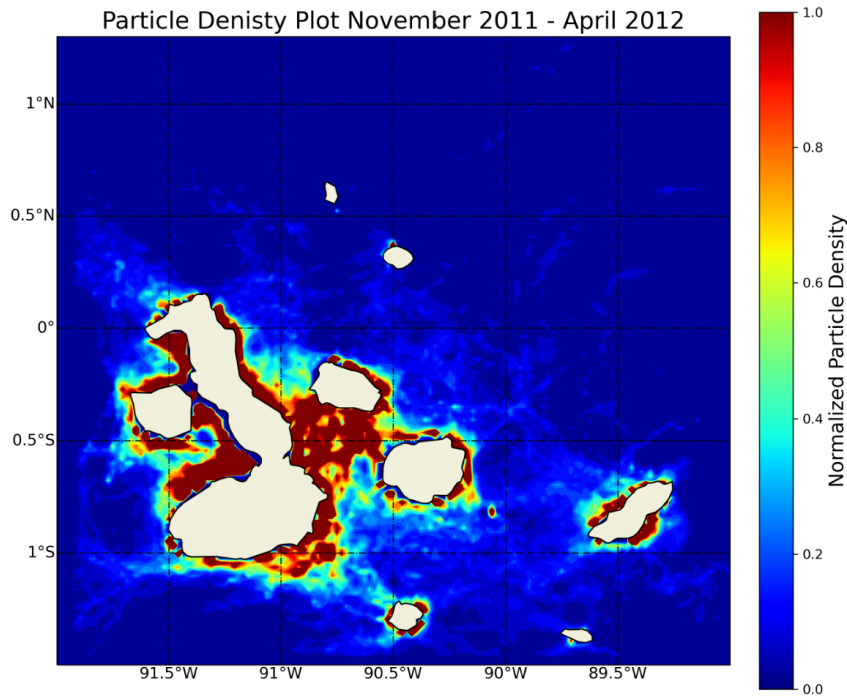
### 3.1 Particle Pathways

This section displays the particle pathways, which are the output of the simulations (section 2.3) and the basis of all the executed analyses in our research. In order to analyse the particle pathways two figures were constructed. First, a particle density plot, visualizing the particle densities over the whole investigated period for the whole modelling domain (figure 10). This plot makes it easy to observe the mean pathways of the particles over all months. The second figure displays the particle trajectories for every month separately (figure 11). From these plots, the variety in the particle pathways between the months, can be observed. Additionally, in the appendix three figures with the consecutive 10 days for every month are shown. With these figures the temporal variation of the trajectories within a month is more clearly displayed.

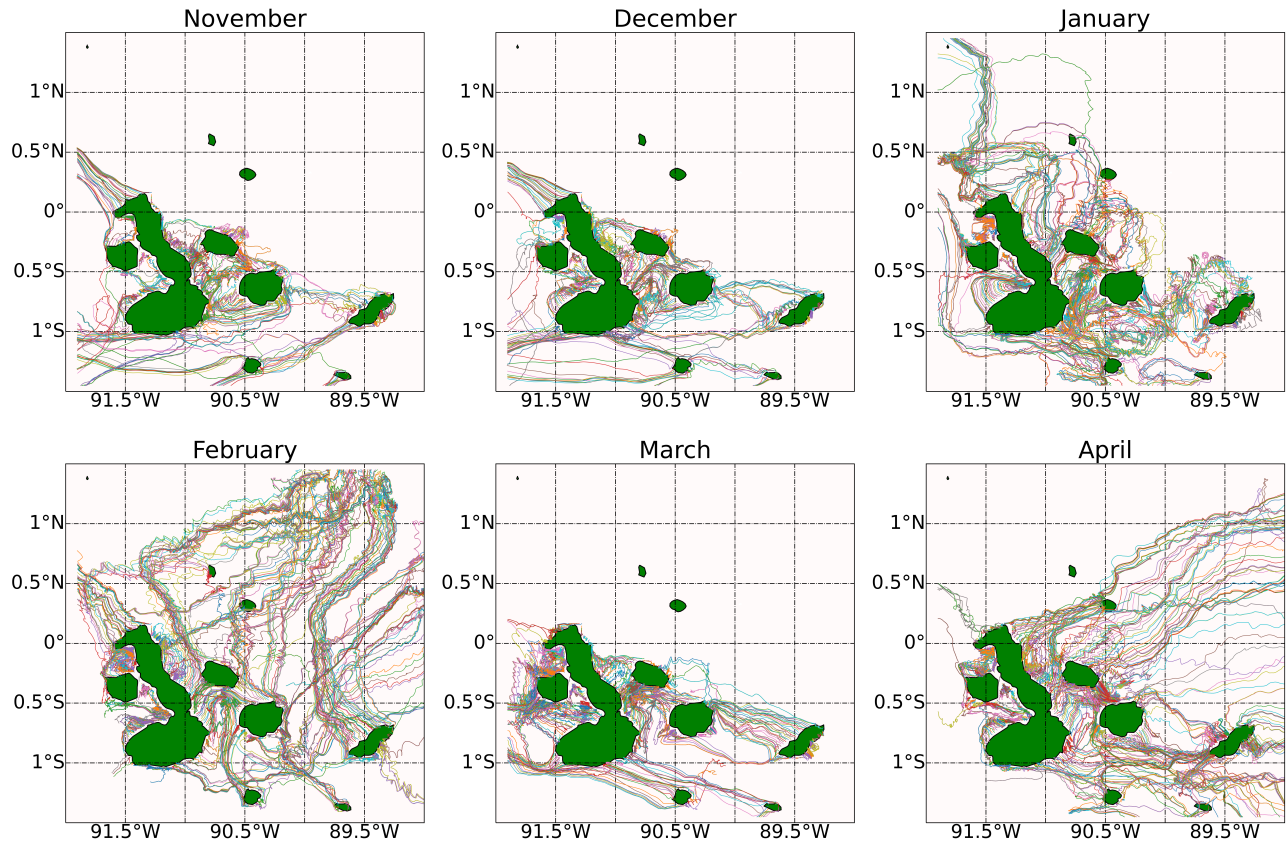
From figure 10 it can be observed that the mean pathways are generally directed to the West, however, some variation is observed towards the North, North-East. With the mean pathways directed to the West, higher particle densities

would be expected on the west side of the islands when releasing particles from all coastal cells. Additionally, particles released on the east side of islands would 'stick' to the coast instead of moving away from it. When looking at the values around San Cristobal, low particle densities on the East and higher particle densities on the West can be observed, indicating particle pathways to the West. Very high particle densities are found on the west side of Santa Cruz and Santiago, while on the east side very low densities are found. Particles released from Floreana also seem to mainly move towards Isabella and not to Espanola Island. From all the particles released at the east coast of Isabella only very small fractions appear to move away from the coast. All things mentioned above, indicate that the mean direction of the particle pathways in the Galapagos Islands between November and April is towards the West. As expected, since the Galapagos Islands experience currents with an average direction to the West, causing particles moving to the West as well.

Though the mean pathways are directed to the West, figure 11 shows some variation in particle pathways between the months. In February and April the direction of the particle pathways changes and they are more directed to the North, North-East. This may be due to the weakening of the North Equatorial Countercurrent (NECC) in these months. The NECC normally bounds the South Equatorial Current (SEC) in the North. However, when the NECC weakens during the months of February, March and April it is not present until 120° West [9]. This may cause the SEC to move more northwards and as a consequence the particle pathways are directed to the North. All transition matrices constructed from these particle pathways are presented in the following section.



**Figure 10:** Particle density plot of the particle pathways in the whole investigated period. The locations of the particles are binned for all grid cells in our modelling domain and thereafter normalized.

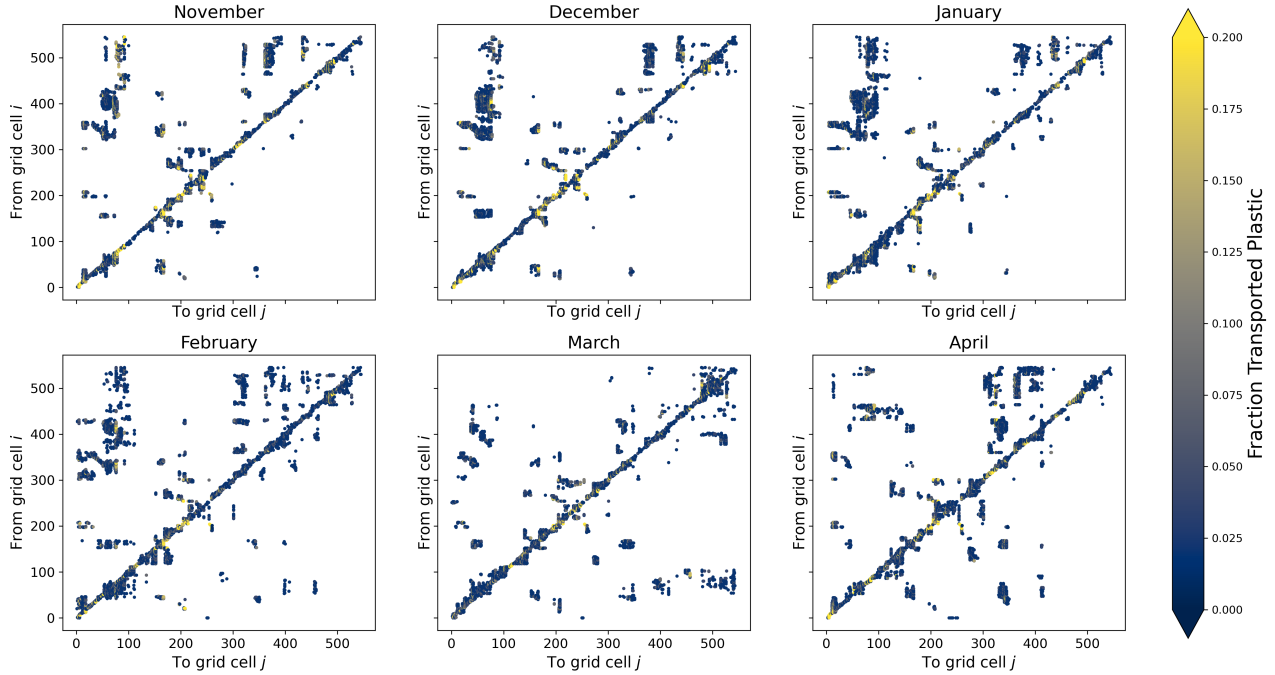


**Figure 11:** The particle pathways for every month in the investigated period. Each figure represents the particle pathways for every month in the investigated period.



## 3.2 Transition Matrices

As explained in section 2.4 the particle pathways are used to construct a set of transition matrices. The transition matrices from the particle trajectories of the  $S_2$  simulations are shown in figure 12. As expected, the transition matrices that represent the months with more varying currents (February, March and April), show more scattered patterns. These transition matrices are the basis of the Lagrangian Network and the Markov Chain analyses. These analyses are used to identify the optimal clean-up locations, which will further be discussed in the next section.



**Figure 12:** The transition matrices of the  $S_2$  simulation of each month in the investigated period. The color scheme represents the fraction of the transported plastic.

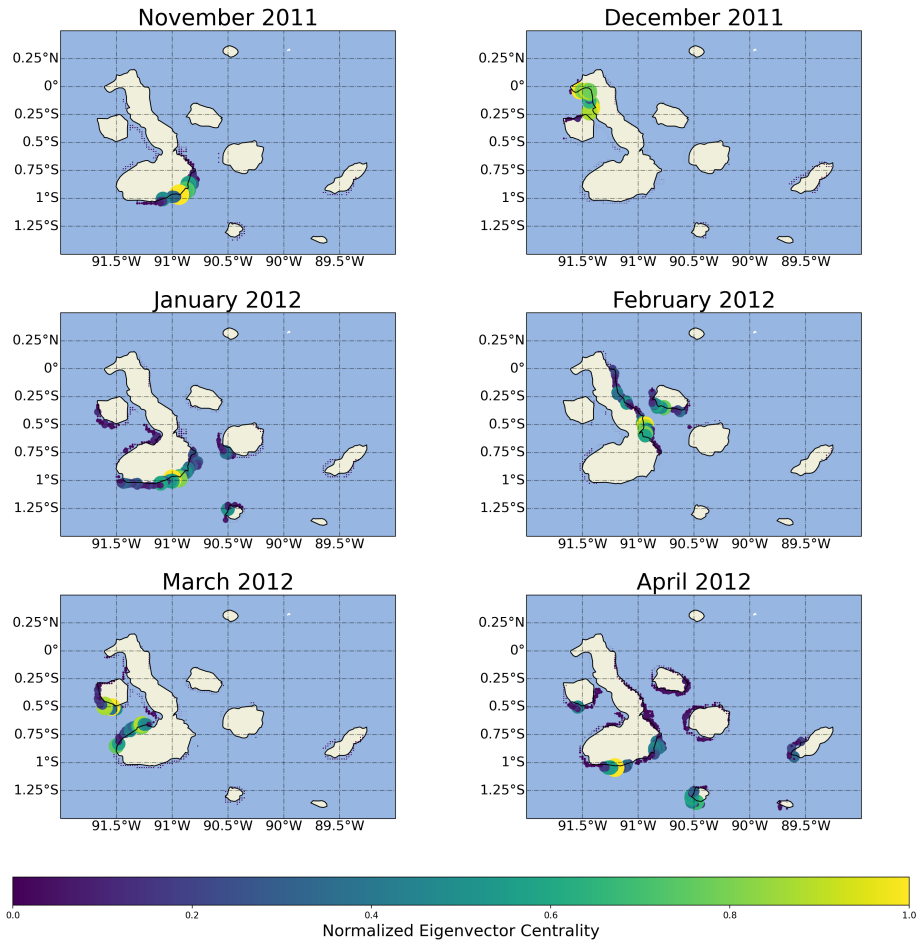
## 3.3 Clean-up Locations

This section will discuss the clean-up locations indicated by the Lagrangian Network analysis and the Markov Chain analysis. As mentioned before, the accumulation zones and the regions where most plastic crosses along their pathways through the system, are considered to be the most efficient clean-up locations. First, the indicated accumulation zones are discussed in section 3.3.1. Thereafter, the regions where most plastic crosses, referred to as crossing hotspots, indicated with both the Lagrangian Network and the Markov Chain analyses, are shown in section 3.3.2. Subsequently, the results of both methods, indicating the crossing hotspots, are compared in order to identify the differences and similarities between the results of both methods. It will also be discussed whether these differences and similarities can be explained by conceptual differences between the methods.

### 3.3.1 Accumulation zones

With the constructed networks from the transition matrices (section 2.4) we are able to identify the accumulation zones in the Galapagos Archipelago, by computing the eigenvector centralities for every node in the network (section 2.5.2). The values of the computed eigenvector centralities are mapped to the color and size of the scatters on the map of the Galapagos (figure 13), from which we can observe some temporal and spatial variation. The relatively high values on the south of Isabella during November, January and April seem likely when considering the trajectories in these months, generally directed to the west. However, in January and April some higher eigenvector values are also

found on the west side of Floreana, Santa Cruz and San Cristobal. This may be explained by the temporal changing current direction to the North-East during these Months, causing particles to accumulate on the west side of these islands. In February, the plastic accumulation shifts more to the North of the Galapagos Islands. This seems likely, since the currents in February are mainly directed to the North. Particles released on the southern islands, Floreana and Espanola, and on south of Isabella will travel northward, in the direction of Santa Cruz and Santiago, explaining higher eigenvector centralities in these regions. In addition, some particles will travel along the coastline of Isabella to the North and beach during their path along this coast, resulting in higher eigenvector centralities on the north part of the east coast of Isabella. In December and March high eigenvector centralities can be observed on the west side of Isabella, which seem very unlikely, considering the isolated nature of this region from the other island and the main direction of the currents during the investigated months. These regions, will be discussed in more detail in section 4.2.



**Figure 13:** The normalized values of the eigenvector centralities mapped to the color and size of the scatters on the map of the Galapagos Islands. Each figure represents these values for every month.

### 3.3.2 Crossing Hotspots

As mentioned in the introduction, two methods are constructed to identify the crossing hotspots (section 2.5.1 and 2.6.1). The results of the betweenness centralities and sink efficiencies are shown in figure 14 and figure 15, respectively. We mapped the values of the betweenness centrality and sink efficiency to the color and size of the scatters. The locations where most plastic crosses indicated by both methods seem to vary in time and sometimes, the higher values are quite scattered over the Islands. However, the Lagrangian network method shows some relative high values on Santiago and Santa Cruz throughout most months. Also around 0.5°S on the east side Isabella high values are found in most months. From the results of the Markov Chain analysis, shown in figure 15, some regions

with consistent higher values can be observed, but not as evident as in the Lagrangian network analysis. The region around 0.5°S on the east side of Isabella shows some consistent higher values and in some months we observe peak values on Santiago and Santa Cruz. However, the network analysis definitely shows more evident regions with consistent higher values. We expect that the high values on the east side of Isabella are a result of trapped particles in these regions, discussed in section 4.2.

For this research, the differences between both methods are very interesting. Spatial and temporal differences between both methods were found. To quantify how much the results of both methods correspond, the Pearson correlation between the two methods is calculated, see table 1. The Pearson correlation has a value of 0.2 and a p-value lower than 0.05 for all months except for April. April shows a non-significant Pearson correlation of 0.07. We can conclude that the correlation found between the two methods is almost negligible.

**Table 1:** The Pearson correlation between the indicated crossing hotspots by the two methods. The correlations are, although significant, almost negligible

Month	Pearson Correlation	p-value
November	0.19	<0.05
December	0.17	<0.05
January	0.21	<0.05
February	0.19	<0.05
March	0.20	<0.05
April	0.07	0.12

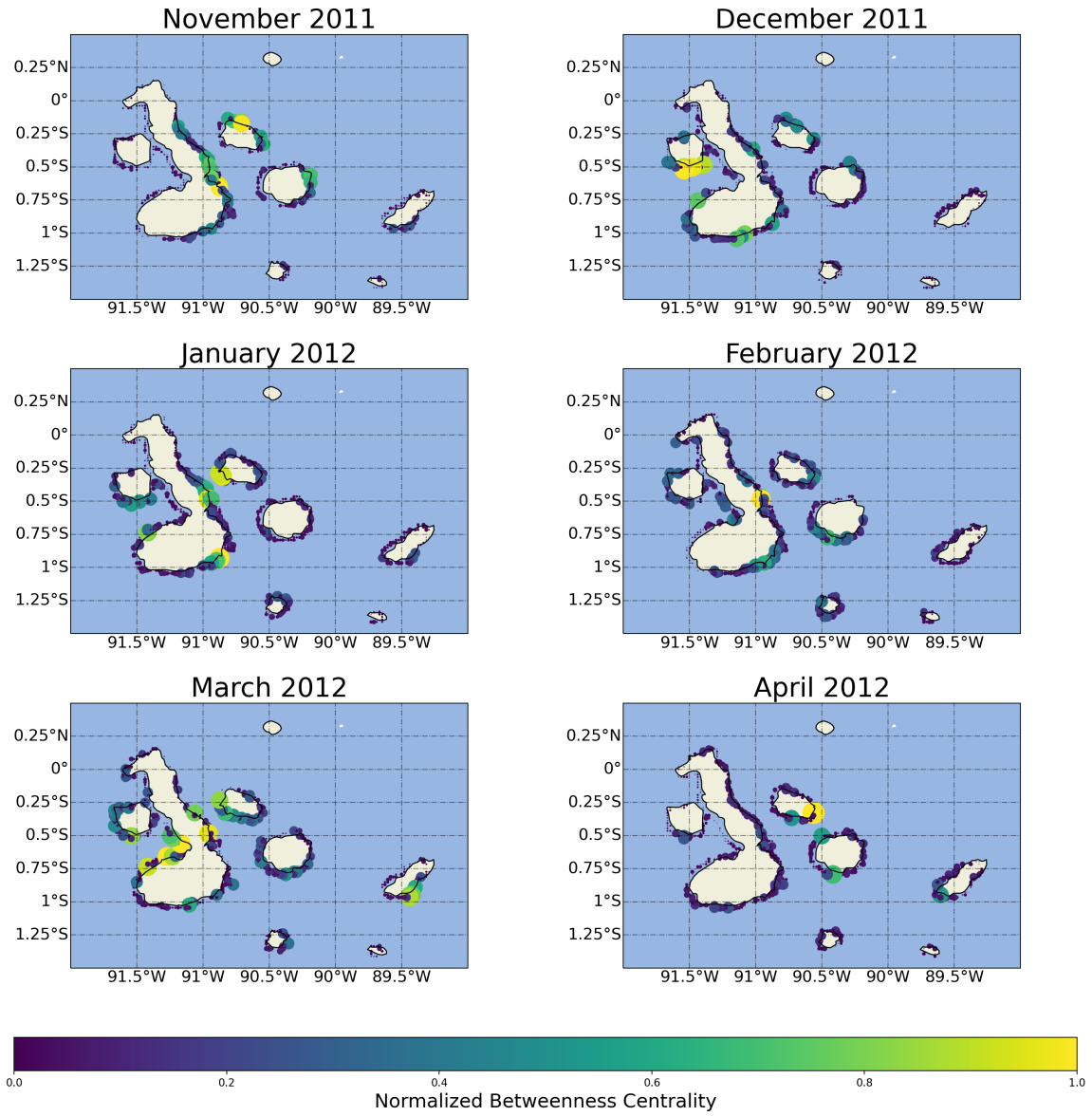
It is not incomprehensible that the two methods show different results, since they are based on different simulations and different concepts, which both will influence the information captured by the methods. We expect that this information difference is one of the primary reasons causing these different results.

First the amount of information captured by both methods is different due to the different simulations they are based on. The Lagrangian method contains more information than the Markov Chain method, due to the repeated five day release, in contrast to the single release every 15 days for the Markov Chain analysis. Additionally, the way the temporal dimension is captured in the information differs, due to the way both methods process the information of the trajectories, explained in the next paragraph.

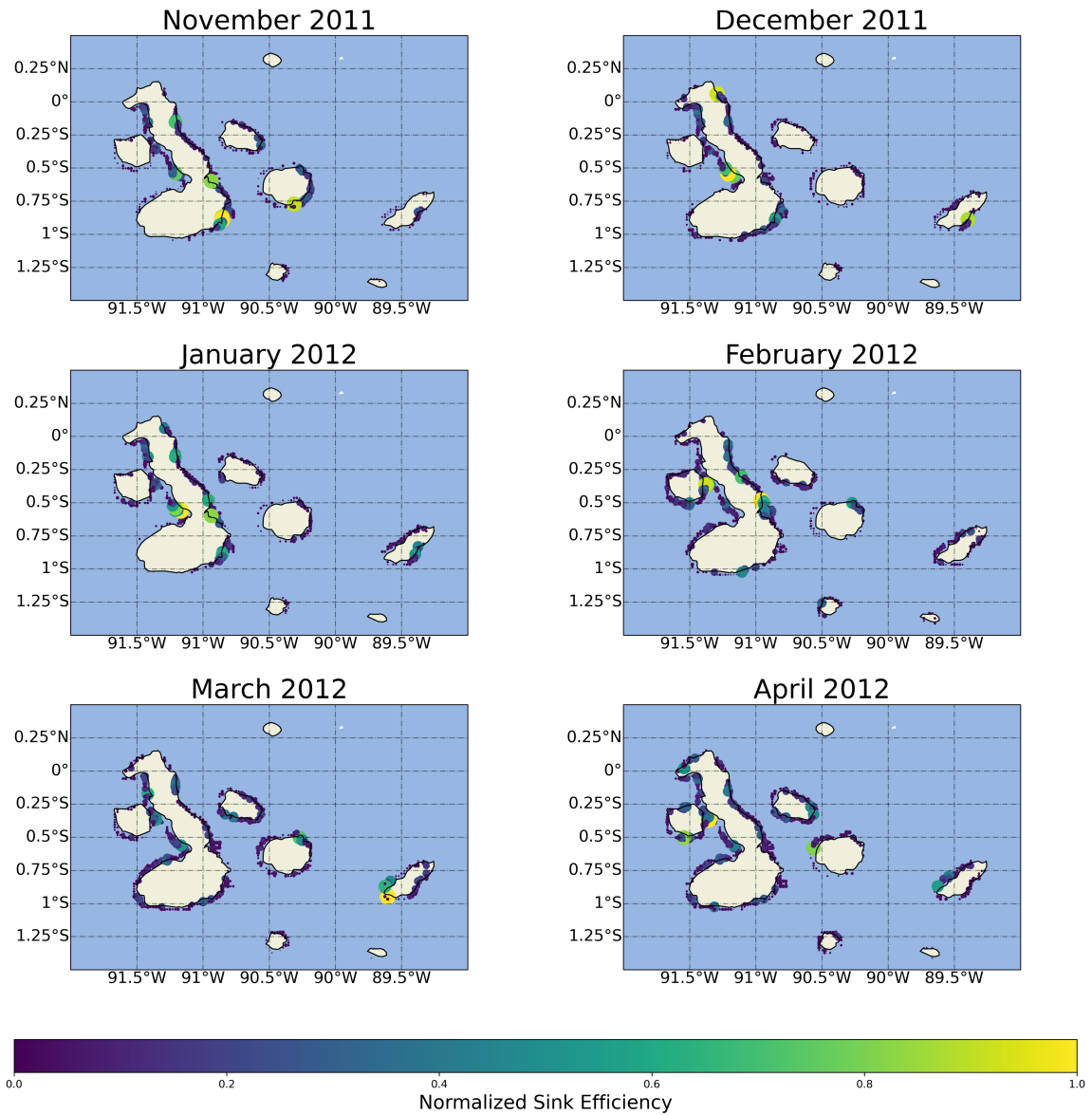
The trajectories for the Lagrangian network analysis contain information about the temporal variation within a month, since particles are released throughout the month, i.e the trajectory of a particle released on day 20, contains information about the connectivity from day 20 until it beaches, in contrast to a particle released on day one, containing information from day one until it beaches. When the trajectory information for the Lagrangian analysis is converted into transition matrices and consequently into networks, we lose the information about the temporal variation of the connectivity within the month. The constructed networks give a more average picture of the connectivity within a month. On the other hand, the Markov Chain models contains very precise information about the initial and final locations of particles released at day one of the interval. Consequently, the Markov Chain method captures the temporal variation of the connectivity for these particles very accurate. However, no information about connectivity at intermediate times is captured by the Markov Chains.

With this in mind we can explain what caused the differences in the results. The betweenness centrality is based on networks that contain average information, but a large amount of information, about the transport of particles between coastal grids within a particular month. The indicated regions by the betweenness centrality tell us where on average within a particular month most plastic crosses. The sink efficiencies are calculated by constructing Markov Chain models and will tell us with relatively high accuracy what regions are effective sinks when releasing particles at day one of the interval. This accuracy will decrease for particles released further away from this initial day, due to the missing information about connectivity at intermediate times. This explains why both methods show different results, since the crossing hotspots on average within a month differ from the crossing hotspots for plastic released at a specific point in time. We expect that Markov Chain models outperform the Lagrangian Network method when the exact moment of release is known. However, we expect that the Lagrangian Network method performs better when the moment of release is not known, since this method contains more information about the connectivity and gives a

more average indication. What method is more suitable in what situation and has more perspective, is discussed in more detail in section 4.3.



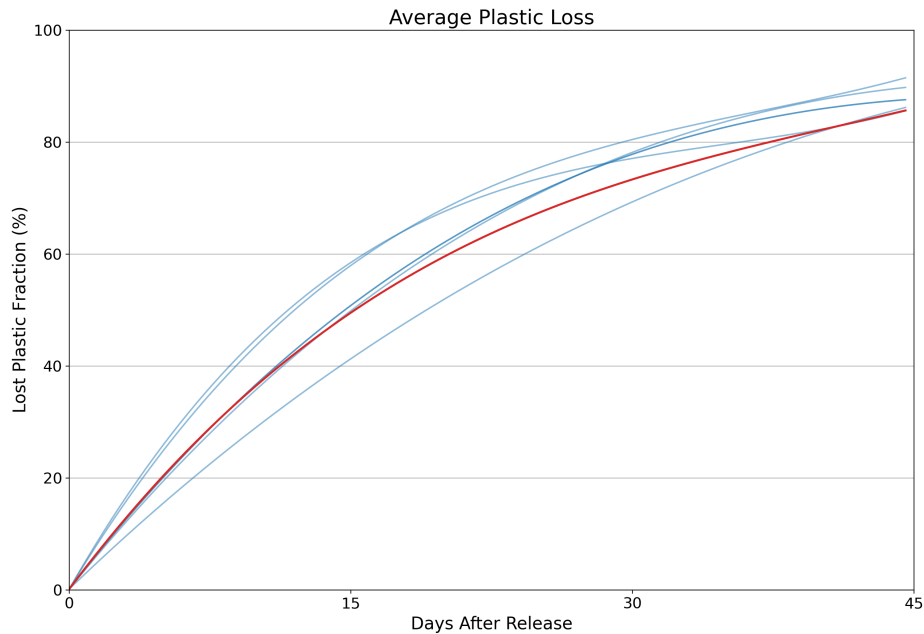
**Figure 14:** The Betweenness Centralities are plotted on the map of the Galapagos. All values are normalized with the highest values having a value of one and the lowest a value of zero.



**Figure 15:** The sink efficiencies are plotted on the map of the Galapagos Islands. The values are normalized with the most efficient sink having a value of one and less efficient sink a value of zero.

### 3.4 Plastic Loss into the Ocean

From the matrix-vector multiplication executed in the Markov Chain analysis, the plastic loss into the ocean can be determined (section 2.6). Figure 16 shows the evolution of lost plastic over 45 days, after it is released into the system. The blue lines represent the plastic loss for all different months and the red line represents the average plastic loss over all months. From figure 16 it becomes evident that plastic is lost in the ocean quite rapidly. As mentioned before, it is easier to remove plastic from the beaches than from the ocean. Therefore, when we develop clean-up strategies, this rapid plastic loss into the ocean should be considered.

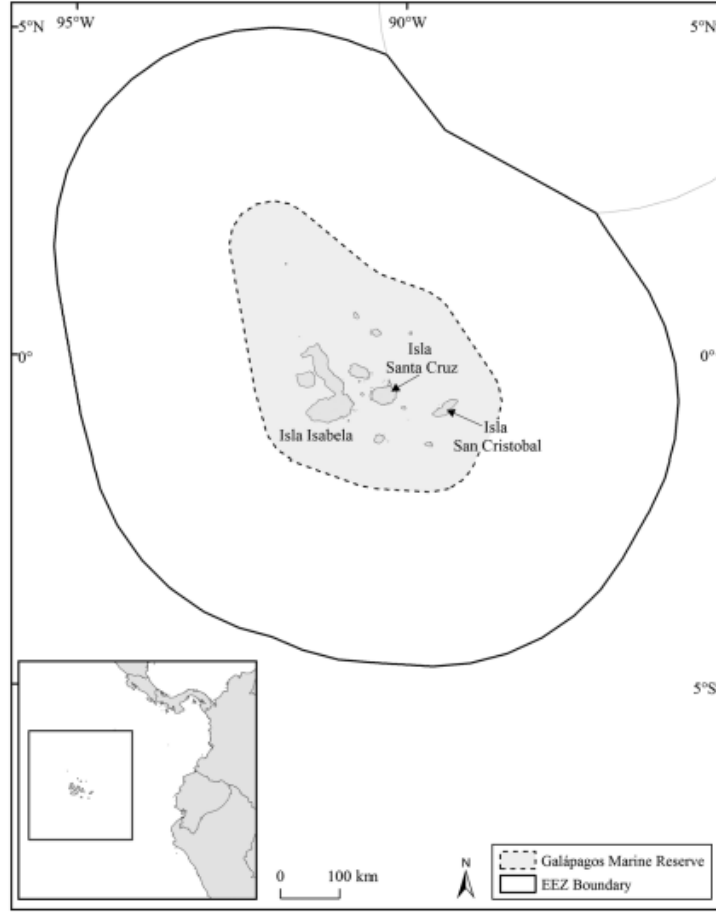


**Figure 16:** The average fraction of lost plastic over 45 days after it is released into the system. The blue lines represent the plastic loss for each individual month and the red line represents the average plastic loss.

### 3.5 Use of the constructed methods for a plastic release from San Cristobal

The results shown and discussed in section 3.3 are based on an evenly spread plastic input. This is very useful in order to determine the global connectivity of the Galapagos Islands and efficient clean-up locations. However, for the development of clean-up strategies it may be insightful to indicate efficient clean-up locations for non-evenly spread plastic inputs, since in reality plastic is not entering the system from all coast grids. Therefore, we decided to apply the constructed methods on a case where the plastic input is limited to a smaller region.

First, the location particles are released from, has to be determined. Using Ocean Parcels, an approximation is made about the number of particles, originating outside the Galapagos, that arrive at each coastal grid cell. This is conducted by releasing particles daily at the Exclusive Economic Zone (EEZ) of the Galapagos (figure 17) for four years, using the SMOG data set. The particles are advected for three months until they beach on one of the islands in the Galapagos or have left the EEZ. The results are shown in figure 18, displaying the number of particles arrived at each coastal grid cell. Note that this is still an indication and not the reality. Figure 18 shows that most particles originating outside the Galapagos, beach on the coast of San Cristobal. Consequently, this would be the region where most plastic enters the system. Therefore, we decided to release particles solely from San Cristobal for this analysis.



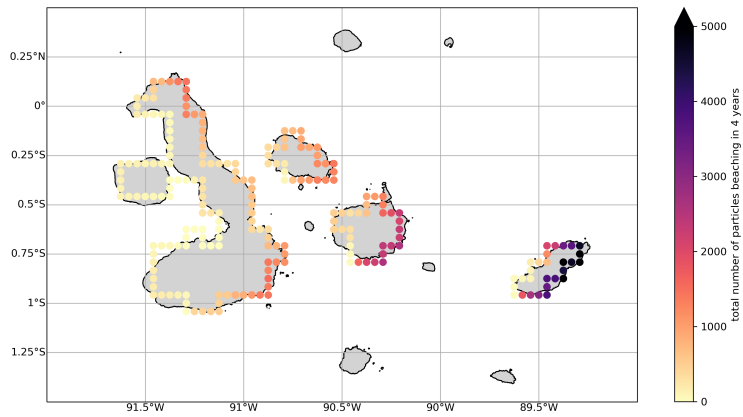
**Figure 17:** The Exclusive Economic Zone (EEZ) of the Galapagos Islands indicated with the solid black line. The particles are released from the border of the EEZ. [29]

We decided to use the Lagrangian Network method to indicate the crossing hotspots since we do not know the exact moment the plastic enters the system from San Cristobal. Therefore we expect that this method is more suitable, as argued in the last paragraph of section 3.3.2. When applying the Lagrangian Network method on this plastic input from San Cristobal, the betweenness centrality is calculated in a slightly different way. In section 2.5.1 the betweenness centrality is determined by computing the most likely path,  $\hat{P}$ , between every pair of nodes. In the case of a selected plastic release,  $\hat{P}$  is only computed between the nodes where plastic is released and all other nodes in the network. The betweenness centrality is now computed for every node, as done in section 2.5.1, using only these paths, see expression 9.

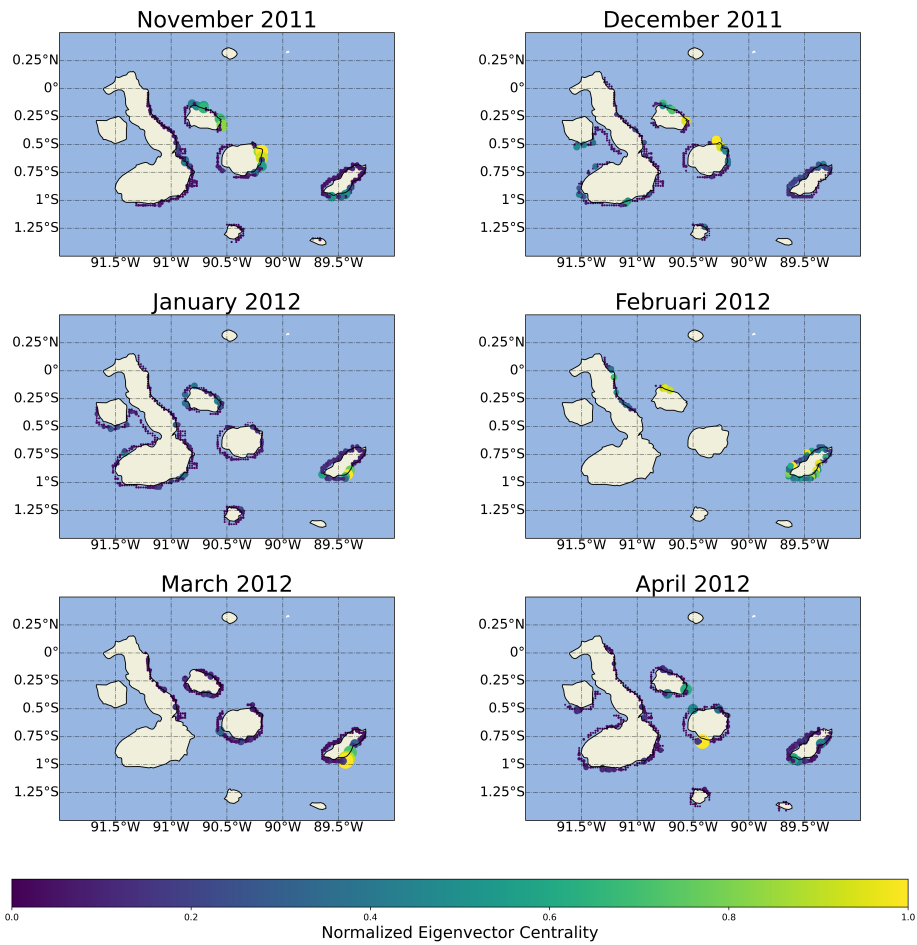
$$g(v) = \sum_{r.sel \neq v \neq b} \frac{\hat{p}_{r.sel,b}(v)}{\hat{p}_{r.sel,b}} \quad (9)$$

In this expression the subscript  $r.sel$  represent all the nodes of the selected plastic release, in this case the nodes from San Cristobal. The betweenness centralities computed with expression 9 indicate the crossing hotspots when plastic is released solely from San Cristobal.

The results are shown in figure 19. The highest betweenness centralities are found on Santiago and Santa Cruz. We would expect that the coastal grids of these Islands would be crossed most often by particles released in the West, considering the mean direction of the currents in the Galapagos. Relative high values are also found on San Cristobal, the release location of the particles, which is caused by particles traveling only small distances between beaching.



**Figure 18:** Approximation of the number of beached particle.



**Figure 19:** The Betweenness Centralities plotted on the Galapagos Map for a selected release on San Cristobal. The values of the betweenness centralities are normalized.



## 4 Discussion

In this research it was investigated what the optimal clean-up locations are in the Galapagos Islands, by indicating the accumulation zones and crossing hotspots. One method was constructed to identify the accumulations zones, based on Lagrangian networks (section 2.5.2) and two methods were constructed to identify the crossing hotspots, of which one was based on Lagrangian Networks and the other on Markov Chains (section 2.5.1 and 2.6).

The accumulation zones indicated by the constructed method were mostly found on the southern regions of the island Isabella. When considering the general particle pathways directed to the West, it was expected that most particles would accumulate in the western regions of the Archipelago. This is reflected in the results. In periods where the particle pathways changed direction to the North, North-East, we expected the accumulation zones to move northwards as well. This also reflected in the results. Therefore, we expect that the method used for indicating the accumulation zones performs quite well. However, validation of the results with drifter data in the Galapagos Islands would be very useful for the improvement of our method.

Both methods used for the identification of the crossing hotspots, indicated crossing hotspots on the islands Santa Cruz and Santiago, located in the middle of the archipelago. Considering the general pathways, it was expected that the beaches in the middle of the Galapagos Islands are crossed most often. However, a part of the indicated crossing hotspots by both methods are quite scattered over the islands, as mentioned in section 3.3.2. Therefore, it is difficult to give an indication about the degree of accuracy of the results. In order to validate the accuracy of the results, real data about plastic transport between the islands are needed. However, they are yet not available at the moment. With the calculated Pearson correlation between the results of both methods, indicating the crossing hotspots, it is shown that, though significant, the correlations are quite small. As mentioned and explained in section 3.3.2, we assume that this is due to the difference in information captured by both methods. In section 4.3 it is discussed what method we expect to perform better and has more potential for further research.

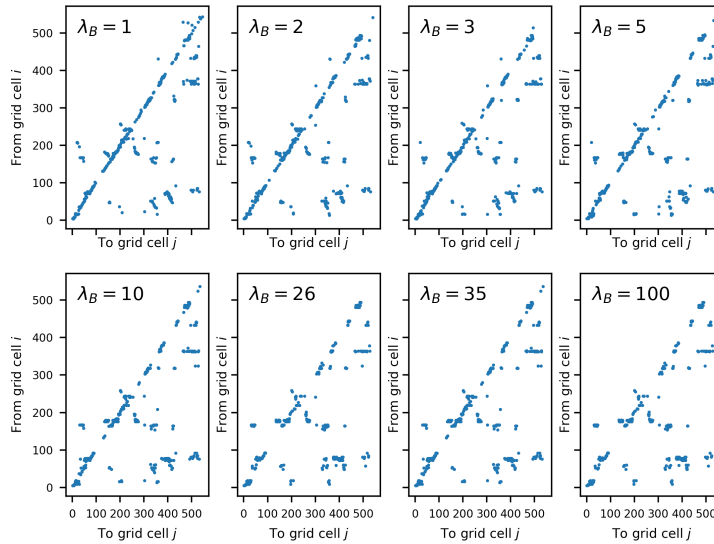
The constructed methods still contain some uncertainties concerning the beaching and resuspension of particles, discussed in section 4.1. Therefore, drifter data in the Galapagos Islands would be beneficial for the improvement of these parametrizations. All constructed methods did show some very unlikely results on the west side of Isabella. We expect this to be caused by trapped particles in these regions. This is discussed in section 4.2. In the last section, section 4.4 we will combine all results of our research in order to give some recommendation about clean-up strategies in the Galapagos Archipelago.

### 4.1 Beaching and Resuspension Sensitivity

The beaching and resuspension processes are important processes for the transport of plastic between the islands. As mentioned in section 2.2 the beaching process is still a large uncertainty in the used methods. In this section another sensitivity analysis was conducted to assess the sensitivity of the particle distribution with respect to the beaching timescale. Using this sensitivity analysis as well as the analysis in section 2.2, it will be discussed how much we expect the beaching parametrization to influence our results. The second part of this section discusses how the resuspension of particles could be implemented in our methods and how much we expect that this would influence our results.

The beaching of particles is parametrized with a beaching probability, calculated with a characteristic beaching timescale,  $\lambda_B$ . In section 2.2 the sensitivity of the fraction of beached particles to the value of the beaching timescale was shown. The sensitivity of the distribution of the particles with respect to  $\lambda_B$  is also very important to take into account. In order to investigate this sensitivity, simulations were run for a range of beaching timescale,  $\lambda_B \in [1, 2, 3, 5, 10, 26, 35, 100]$ . Subsequently, a transition matrix was constructed for each simulation, since the scatter pattern of a transition matrix represent the plastic distribution. Figure 20 shows the transition matrices for all beaching timescales. This figure shows that the density of the diagonal line in the transition matrix decreases with increasing  $\lambda_B$ , indicating less particles beaching close to their release location. This is what we would expect when the characteristic beaching timescale decreases. However, the scatter patterns do not seem to change much with increasing beaching timescales. This indicates that the beaching timescale does influence the amount of particles that beaches close to their release location. However, it does not affect the distribution of particles much. Based on both sensitivity analyses the beaching timescale was determined, which was used in our research. However, this value could still deviate a lot from the true value. We expect that the value of the beaching timescale would not influence the results enormously, since the distribution patterns of the particles are not changing significantly with

respect to the beaching timescale.



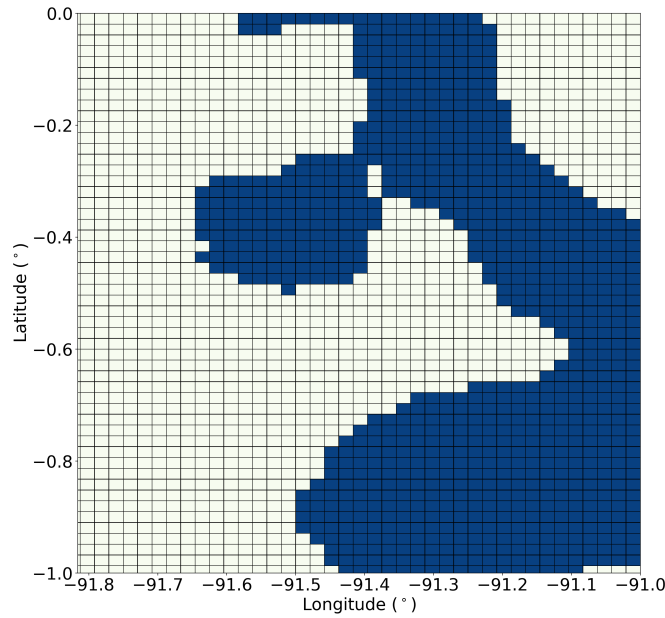
**Figure 20:** The sensitivity of the plastic distribution with respect to the beaching timescale. The density of the diagonal line, representing the particles beaching close to the release location, decreases with increasing beaching timescales. Less plastic beaches close to their release location with increasing beaching timescales. The distribution patterns are not influenced very much by the beaching timescale, as shown by the scatter patterns.

The resuspension of particles has not been included in the constructed methods. In case of the Lagrangian networks, the potential paths of the particles are based on the edge weights ( $w$ ) between the nodes, representing a transport probability. A resuspension parametrization would have affected this transport probability. When not including a resuspension parametrization it is assumed that the whole fraction of particles at a node resuspend and is transported to the next node. However, this is not the case in reality. The same applies for the Markov Chain models, where no resuspension is included. We assume that all particles will resuspend at the moment of the next multiplication, which in reality is not the case. In order to get a more realistic picture of the potential pathways of particles within the system we should implement a resuspension probability. It could be implemented in a similar way as the beaching of particles, by implementing a resuspension probability as done in [22]. For the Lagrangian network this re-suspension probability could be implemented in order to determine when a particle is re-released. Instead of releasing particles every five days, particles are re-released when a particular probability is reached, based on this re-suspension probability. In the case of the Markov Chain analysis, the re-suspension probability could be implemented in between the matrix multiplications. This resuspension would determine whether particles from a particular coastal grid would be re-released in the next multiplication. As mentioned before, this resuspension probability is very uncertain, so when it would be implemented it would add uncertainty to the results. The results of our analysis should definitely be affected by including a resuspension probability. We expect that including a resuspension parametrization will largely influence the rate at which plastic propagates through the system. Consequently, this will influence the paths of the particles, since the moment particles enter the system changes. However, since the Lagrangian network approach gives a more average picture of the connectivity, we expect that it will not have a significant impact on the indicated regions of interest.

As we argued in previous paragraphs, the values of the beaching and resuspension probabilities would not influence the results very much, however, it is still very important to obtain more realistic and certain value for the beaching and resuspension probabilities. Analysing data of beached and resuspend drifters would be a good approach to obtain a more realistic parametrization for beaching and resuspension. However, these data are currently not available and so we're not able to do this analysis. This would be very useful and interesting information in order to improve the performance of our methods.

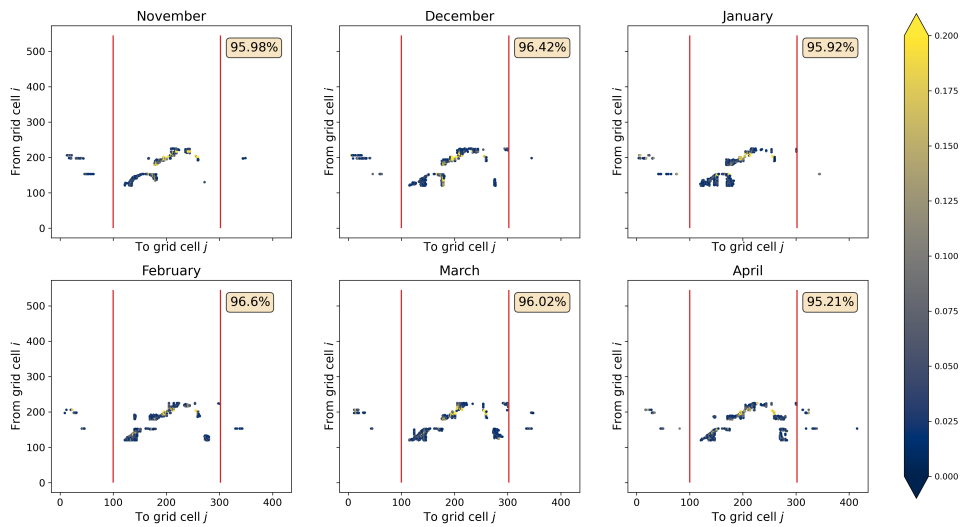
## 4.2 Trapped particles

As mentioned in section 3.3, the centralities on the west side of Isabella are considered separately. Relative high values were found in the two bays between the west side of Isabella and Fernandina. By looking at the particle pathways (figure 11) it can be observed that particles get stuck in the bays between Isabella and Fernandina. We expect the trapped particles to be trapped for two reasons. The first reason is the shape of the bays, which make it difficult for the particles released in these bays to leave the bays. The second reason is the grid resolution of our modelling domain. Figure 21, shows the grids of our modelling domain. Here it can be seen that Isabella and Fernandina are connected, which is not the case in reality. Due to this connection, particles are not able to move between Fernandina and Isabella, which makes these regions even more isolated. Consequently, it is even more difficult for particles to get out of these bays.



**Figure 21:** The grid resolution of the modelling domain of this research. The islands Fernandina and Isabella are connected, while this is not the case in reality.

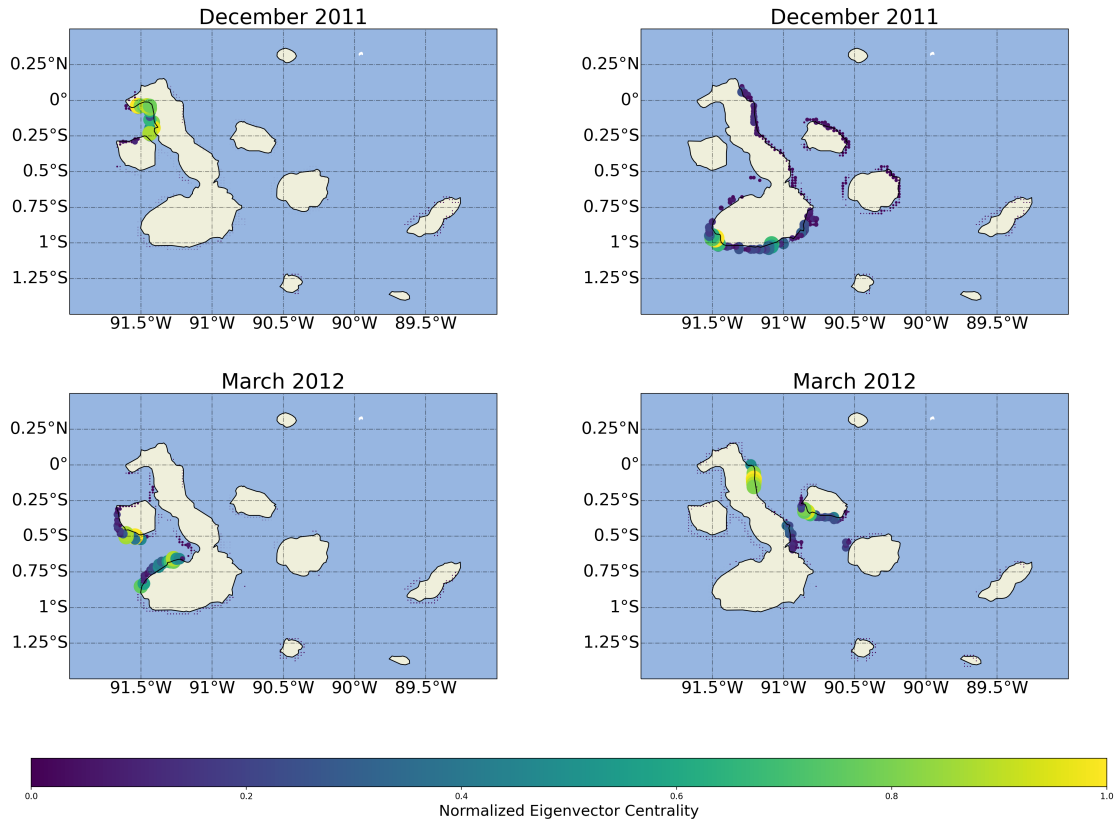
One can imagine that when a particle is trapped in a bay, the probability that it beaches somewhere in that particular bay becomes extremely high. From the transition matrices we can confirm that particles truly get trapped in these regions. In figure 22 all transition matrices are shown with all rows set to zero, except the rows representing the release grids on the west side of Isabella. The two vertical red lines indicate the same region, but for the beaching grids. All the scatters within between the two red lines represent particles that are both released and beached in these isolated regions. In the top right corner of the figure, the percentage of particles trapped in these regions is indicated. In every month more than 95 % of the particles released in these regions also beached in these regions. In network theory, this is called a sub network,  $G'$ . This consists of a set of nodes  $S$  and a set edges consisting of all edges in  $E$  with both endpoints in  $S$ .



**Figure 22:** Trapped particles visualized in the transition matrices. The area between the two red lines indicates the regions on the west side of the largest island. This clearly shows that almost all particles released in these regions also beach in those regions.

Consequently, large amounts of plastic will be transported between the coastal grids within these bays, resulting in high values in the transition matrices and so high centrality measures. However, the high centralities measures found in these regions do not truly represent influential nodes in the system, since they are completely isolated from the rest of the system.

Section 3.3.1 shows very high eigenvector centralities during December and March in the isolated regions discussed above. It seems very unlikely that most particles released in the whole system will end up here. Since the eigenvector centrality is a relative measure of importance, the extremely high values may mask the values in other areas. By calculating the eigenvector centralities without releasing particles in the isolated regions we can find the locations of the accumulation zones in the system without any trapped particles. Figure 23 shows how the eigenvector values change without particles released in the isolated regions. In December, the accumulation zones of the system moved to the south of Isabella and in March the accumulation zones are more located in the middle and top of Isabella. From this it becomes evident that the very high values were indeed caused by the trapped particles, and without any trapped particles the locations of the accumulation zones move to much more likely regions, considering the particle trajectories in these months. The eigenvector centralities are also recalculated for the other months. However, no large changes of results were noticed. Apparently, the results in the other months are not influenced heavily by the trapped particles.



**Figure 23:** The Eigenvector centrality mapped for the months December and March with trapped particles (left) and without trapped particles (right). This clearly shows that the regions with high eigenvector centralities move to much more likely regions.

The betweenness centralities found in the isolated regions are only based on the connections within this sub-network. It is very unlikely that particles originating outside the Galapagos will beach in these regions considering the main currents in the Galapagos Islands. With this in mind we can argue that the betweenness centralities found in these regions are not very relevant for our analysis and that we should focus on the values found on the remaining regions of the Galapagos Archipelago.

### 4.3 Comparison Methods

As discussed in section 3.3.2 the results of both methods for indicating the crossing hotspots differ, due to the different information captured by both methods. The Lagrangian network method gives a more average picture of the locations of the crossing hotspots, while the Markov Chain methods gives a more precise indication for plastic entering the system at a particular point in time. The accuracy of the Markov Chain will decrease when particles are released further away from the start of the time window. Unfortunately, no data are available to validate the results of both methods, so we cannot conclude which method performs better.

Considering the nature of the Markov Chain we expect that when we know the moment that plastic enters the system, the Markov Chain approach would be more accurate in tracking the propagation through the system and consequently finding crossing hotspots. Unfortunately, it is very uncertain when plastic will resuspend into the system. Therefore, we expect that, due to the fact that the Lagrangian networks contain more information about the connectivity throughout the month and give a more average indication, it will outperform the the Markov Chain approach in most cases.

The computation time of both methods also differ. Iteratively implementing sinks into the system is computational much heavier than computing the betweenness centrality. This is not a huge issue in our research, since the number of coast grids in our domain is relatively small, so the absolute computation time is acceptable. However, when in further research the resolution increases and consequently the number of coastal grids increases, the sink implementation will become computational unfeasible. This methods would also become unfeasible when investigating the connectivity of a much larger Archipelago, e.g. Indonesia or the Philippines. Therefore, we assume that the Lagrangian network approach has a greater potential.

There is plenty of room for improvement in the constructed methods. As mentioned in the section 4.1, the implementation of a more accurate beaching and resuspension parametrization would be very interesting for both methods. The Markov Chain approach may be improved by using transition matrices with shorter intervals, capturing more information about the connectivity in the intermediate times. On the other hand less particles will beach and so information is lost about the connectivity. Finding a better balance between these two could improve the performance of the Markov Chain method. The Lagrangian network approach could be improved by using simulations releasing particles every day. However, this would decrease the computational feasibility.

#### **4.4 Clean-up Strategies**

The intended objective of this research was to improve clean-up strategies in the Galapagos. By combining the indicated accumulation zones, crossing hotspots and the knowledge about the rate at which plastic is lost into the ocean, current clean-up strategies could be improved. As argued in section 4.3 we expect that the Lagrangian network approach would on average perform better than the Markov Chain approach, so we'll consider these results in this section.

The results from section 3.3.2 and section 3.5 indicate that the crossing hotspots are often located on Santiago and Santa Cruz and therefore these are efficient areas to remove the plastic from. However, the exact locations of the crossing hotspots on Santiago and Santa Cruz differ throughout the months. These locations may be crossed by plastic most often. However, not all plastic in the system will cross these regions. The most effective location to remove this plastic from, are the indicated accumulation zones, since these are the regions where most plastic in the system eventually accumulates. From section 3.3.1 and section 4.2 we know that the accumulation zones in the system are most often located on the south of Isabella. Additionally, we know that plastic gets lost into the ocean relatively quickly, therefore it is important to remove plastic from the system relatively fast after it has entered the system. Otherwise large fractions will get lost into the ocean and it will be difficult to remove them.

Based on the results of our analysis we would recommend to remove plastic on beaches located on Santiago and Santa Cruz and on beaches on the south of Isabella. This would be, based on our analyses, the best strategy to remove plastic from the Galapagos Archipelago.

## 5 Conclusion

The aim of this research was to identify the accumulation zones and crossing hotspots in the Galapagos Islands, in order to improve clean-up strategies. One method was constructed to identify the accumulations zones, based on Lagrangian networks (section 2.5.2), and two methods were constructed to identify the crossing hotspots, of which one was based on Lagrangian Networks, and the other on Markov Chains (section 2.5.1 and 2.6).

The accumulation zones indicated by the Lagrangian method are located on the southern regions of Isabella in most months. This is what was expected when considering the mean particle pathways in the investigated period. The two methods indicating the crossing hotspots both indicate crossing hotspots on the Santa Cruz and Santiago. However, they also show quite some differences in results. The differences in the results are probably caused by the difference in information captured by both methods. The Lagrangian network method contains more information. However, the time dependency of the information within a month is lost, so it can be considered as a more average indication of crossing hotspots. The Markov Chain approach contains less information since it is based on a single release. However, the time dependency within a month is conserved. This probably results in accurate indications of crossing hotspots when the moment of release is well known. Considering the fact that the exact moment particles entered the system and resuspend from beaches is quite uncertain, we expect the Lagrangian network method performs better in most situations. Based on the results of our analyses, the most efficient approach to remove plastic from the Galapagos system is to remove plastic from beaches of Santa Cruz and Santiago and from beaches located on the south of the island Isabella.

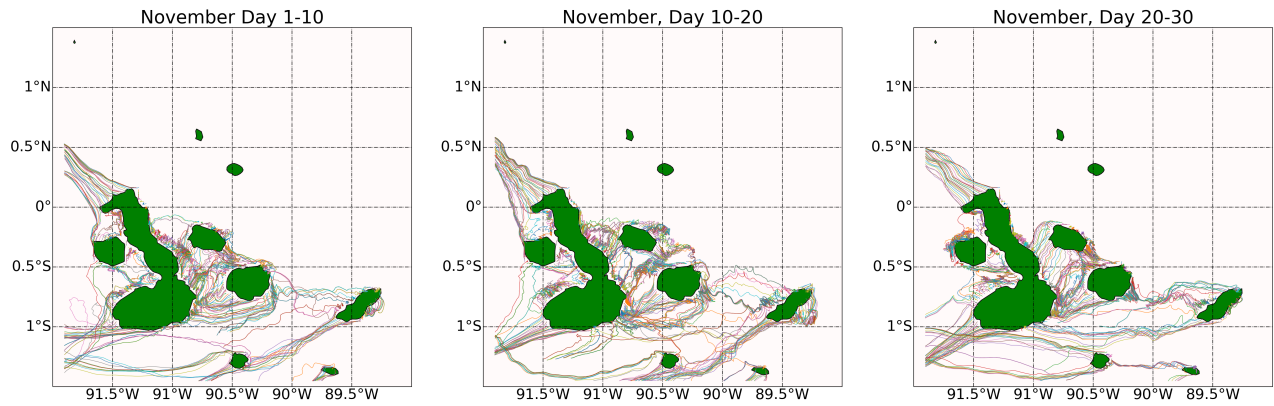
The constructed methods still contain some uncertainties. One of the large uncertainties in the constructed methods is the beaching process of particles, since the true value of the characteristic beaching timescale is not known very well. However, it is expected that the characteristic beaching timescale would not influence the results significantly, since the distribution does not depend much on it. That the resuspension of particles is not parametrized in our methods, leads to another uncertainty in the results. Instead of the parametrization, several assumptions about the resuspension were made. In the methods based on the Lagrangian networks, it is assumed that all particles resuspend when transported between nodes. In the method based on the Markov Chain model, it is assumed that between all matrix multiplication all particles resuspend. By parametrizing the resuspension of particles and implementing this in the model, we expect our methods would improve and come closer to reality. However, we expect that the resuspension of particles would not heavily influence our results, as argued in section 4.1.

The results shown in section 3.3 indicate some accumulation zones and crossing hotspots in the isolated regions on the west of Isabella, which seem unlikely. In section 4.2 it is shown that the accumulation zones and crossing hotspots in these regions were caused by the particles released in these isolated regions which got trapped. When executing the analyses again without releasing particles in these regions, the accumulation zones and crossing hotspots were indeed found in more likely regions.

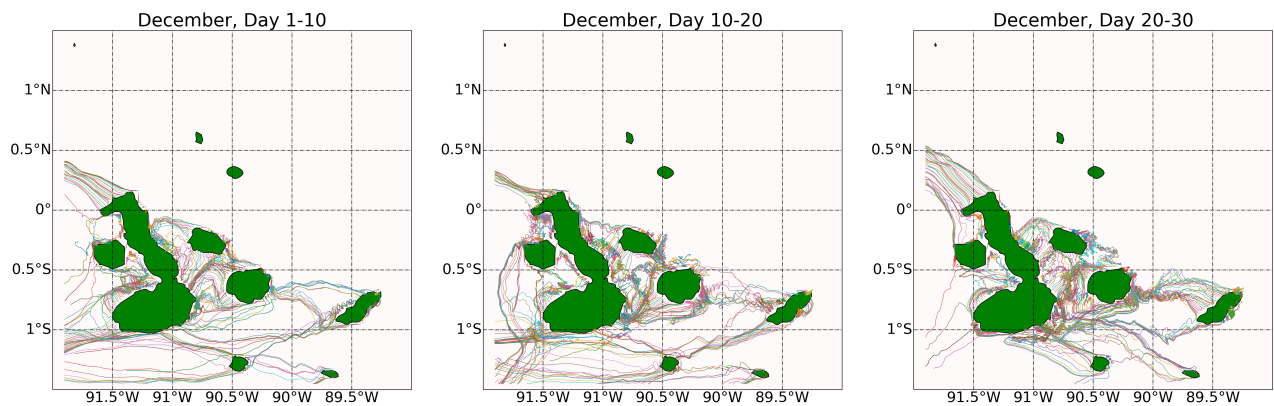
In future research it is recommended to parametrize the resuspension of particles in order the constructed methods. The future availability of drifter data in the Galapagos Island, which are not yet available at the moment, would make it possible to validate the results of the constructed methods, and improve the beaching and resuspension parametrizations.

# Appendices

## A Particle Trajectories

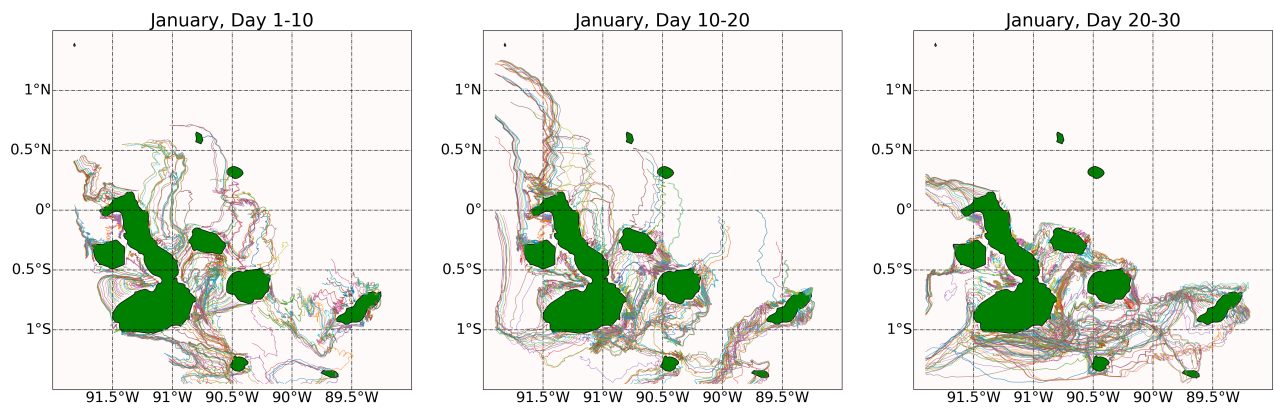


**Figure 24:** Particle trajectories with 10 day intervals

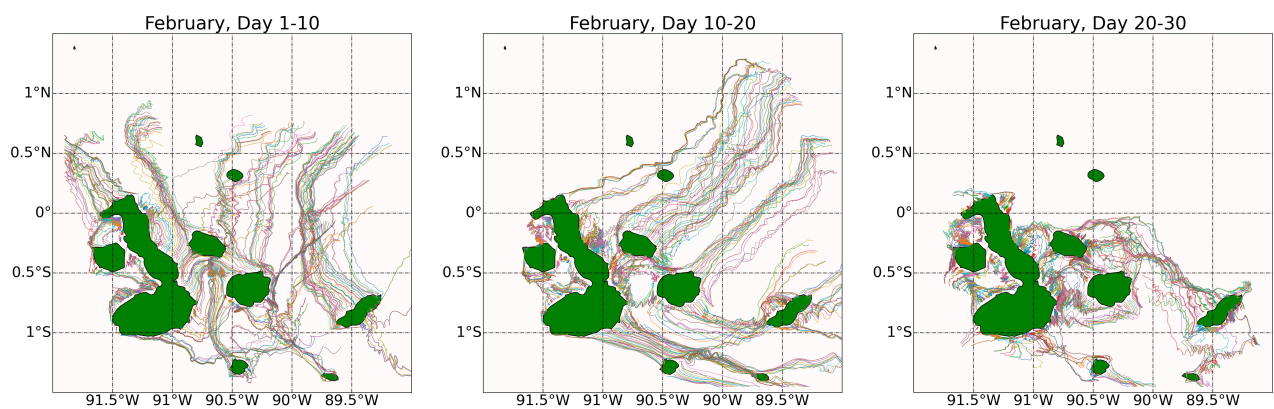


**Figure 25:** Particle trajectories with 10 day intervals

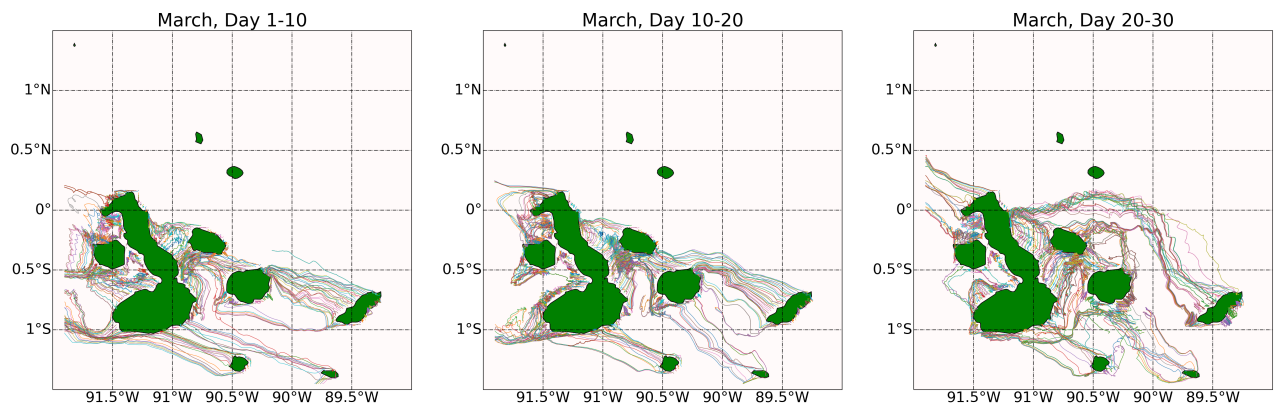




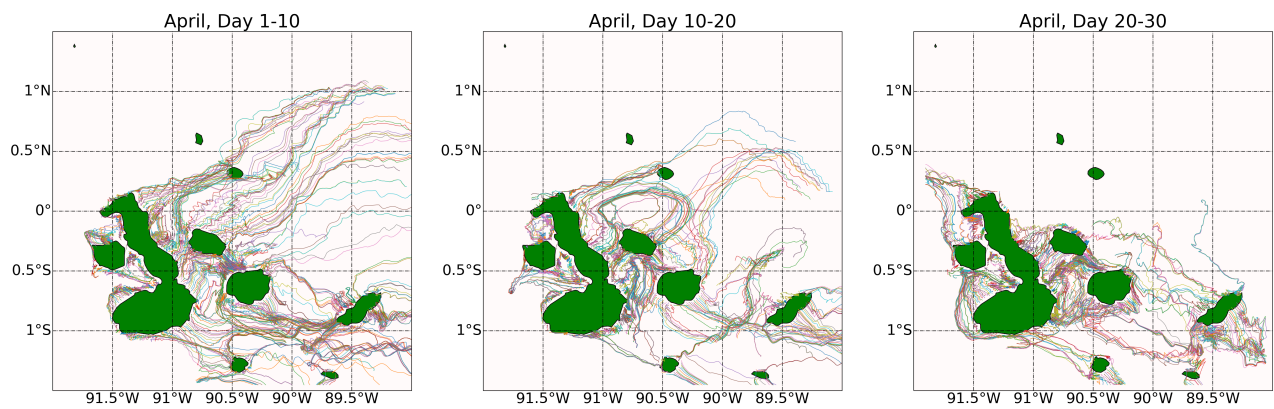
**Figure 26:** Particle trajectories with 10 day intervals



**Figure 27:** Particle trajectories with 10 day intervals



**Figure 28:** Particle trajectories with 10 day intervals



**Figure 29:** Particle trajectories with 10 day intervals

## References

- [1] Modeling marine surface microplastic transport to assess optimal removal locations. 2015.
- [2] Explicit and implicit network connectivity: Analytical formulation and application to transport processes. 2020.
- [3] M. Anthony Browne, M. Gee Chapman, C. T. Richard, and L. A. Amaral Zettler. Spatial and temporal patterns of stranded intertidal marine debris: Is there a picture of global change? 2015.
- [4] S. Ashkan Zarghami and I. Gunawan. A domain-specific measure of centrality for water distribution networks. 2018.
- [5] P. A.T. Sources, quantities and distribution of persistent plastics in the marine environment. 1987.
- [6] C. B. Rocha, S. T.Gille, T. K. Chareskin, and D. Menemenlis. Seasonality of submesoscale dynamics in the kuroshio extension. 2016.
- [7] A. Ballance, P. Ryan, and J. Turpie. How much is a clean beach worth? the impact of litter on beach users in the cape peninsula, south africa. 2000.
- [8] D. Birant and A. Kut. St-dbscan: An algorithm for clustering spatial-temporal data. Data Knowledge Engineering, 2006.
- [9] F. P. Chavez and R. C. Brusca. Galapagos marine invertebrates. 1991.
- [10] Y. Chen, A. Awasthi Kumar, F. Wei, Q. Tan, and J. Li. Single-use plastics: Production, usage, disposal, and adverse impacts. 2021.
- [11] B. Denise Hardesty, T. Lawson, T. van der Velde, M. Landsdell, and C. Wilcox. Estimating quantities and sources of marine debris at a continental scale. 2017.
- [12] M. D.Robards, J. F.Piatt, and K. D.Wohl. Increasing frequency of plastic particles ingested by seabirds in the subarctic north pacific. 1995.
- [13] G. Forget, J.-M. Campin, P. Heimbach, C. Hill, R. M. Ponte, and C. Wunsch. Ecco version 4 : an integrated framework for 1 non-linear inverse modeling and global ocean state estimation. Geosci. Model Dev, 2015.
- [14] J. G.B Derraik. The pollution of the marine environment by plastic debris: a review. 2002.
- [15] L. Iguar and S. Segui. Introduction to Data Science. Undergraduate Topics in Computer Science. Springer, 2017.
- [16] J. Jones, A. Porter, and J. Munoz-Perze. Plastic contamination of the galapagos marine food web and the relative risks to native species. Research Square, 2020.
- [17] L. A. M. Kaandorp, A. H. Dijkstra, and E. Seville. Closing the mediterranean marine floating plastic mass budget: Inverse modeling of sources and sinks. Environmental Science and Technology, 2020.
- [18] M. Lange and E. Seville. Parcels v0.9: prototyping a lagrangian ocean analysis framework for the petascale age. Geoscientific Model Development, 10(11):4175–4186, 2017.
- [19] R. Li, L. Yu, M. Chai, H. Wu, and X. Zhu. The distribution, characteristics and ecological risks of microplastics in the mangroves of southern china. 2020.
- [20] J. McKay Fletcher and T. Wennekers. From structure to activity: Using centrality measures to predict neuronal activity. International Journal of Neural Systems, 2017.
- [21] M. O’Mally and A. M. Sykulski. Estimating the travel time and the most likely path from lagrangian drifters. Journal of Atmospheric and Oceanic Technology, 2021.
- [22] V. Onink, C. Jongedijk, M. Hoffman, E. Seville, and C. Laufkötter. Global simulations of marine plastic transport show plastic trapping in coastal zones. Environmental Research, 2021.
- [23] V. Onink, D. Wichmann, P. Delandmeter, and E. Seville. The role of ekman currents, geostrophy, and stokes drift in the accumulation of floating microplastic. 2019.

- [24] T. Pievani. *The sixth mass extinction: Anthropocene and the human impact on biodiversity*. 2013.
- [25] C. A. Pinheiro. *Social network analysis in telecommunications*. John Wiley Sons, 2011.
- [26] D. Reijnders, E. J. van Leeuwen, and E. van Sebille. *Ocean surface connectivity in the arctic: Capabilities and caveats of community detection in lagrangian flow networks*. 2020.
- [27] P. G. Ryan, V. Perold, A. Osborne, and C. L. Moloney. *Consistent patterns of debris on south african beaches indicate that industrial pellets and other mesoplastic items mostly derive from local sources*. 2018.
- [28] C. A. Savage. *Spectral decomposition of internal gravity wave sea surface height in global models*. *JGR Oceans*, 2017.
- [29] L. Schiller, J. Jose Alava, J. Grove, G. Recl, and D. Pauly. *The demise of darwin’s fishes: evidence of fishing down and illegal shark finning in the galápagos islands*. 2014.
- [30] E. Schubert, S. Jörg, M. Ester, H. P. Kriegel, and X. Xu. *Dbscan revisited, revisited: Why and how you should (still) use dbscan*. *ACM Transaction on Database Systems*, 2019.
- [31] v. E. Sebille. *Adrift.org.au — a free, quick and easy tool to quantitatively study planktonic surface drift in the global ocean*. *Journal of Experimental Marine Biology and Ecology*, 2014.
- [32] v. E. Sebille, P. Delandmeter, J. Schofield, B. D. Hardesty, and A. Donnelly. *Basin-scale sources and pathways of microplastic that ends up in the galápagos archipelago*. *Ocean Science*, 2019.
- [33] A. Sinha, D. Balwada, N. Tarshisa, and R. Abernathey. *Modulation of lateral transport by submesoscale flows and inertia-gravitywaves*. *Journal of Advances in Modelling Earth Systems*, 2019.
- [34] D. W.Laist. *Overview of the biological effects of lost and discarded plastic debris in the marine environment*. 1987.