Optimizing the Path Towards Plastic-Free Oceans

Dick den Hertog
Amsterdam Business School, University of Amsterdam, Amsterdam, The Netherlands
d.denhertog@uva.nl

Jean Pauphilet
Management Science and Operations, London Business School, London, United Kingdom
jpauphilet@london.edu

Yannick Pham, Bruno Sainte-Rose
The Ocean Cleanup, Rotterdam, The Netherlands
{y.pham,bruno.sainte-rose}@theoceancleanup.com

Baizhi Song
Management Science and Operations, London Business School, London, United Kingdom
bsong@london.edu

Increasing ocean plastic pollution is irreversibly harming ecosystems and human economic activities. We partner with a non-profit organization and use optimization to help them clean up oceans from plastic faster. Specifically, we optimize the route of their plastic collection system in the ocean to maximize the quantity of plastic collected over time. We formulate the problem as a longest path problem in a well-structured graph. However, since collection directly impacts future plastic density, the corresponding edge lengths are non-convex and non-decomposable. After analyzing the structural properties of the edge lengths, we propose a search-and-bound method, which leverages a relaxation of the problem solvable via dynamic programming, to efficiently find high-quality solutions with certificates of near optimality (around 6% in practice). On one-year of ocean data, our optimization-based routing approach increases the quantity of plastic collected by over 60% compared with their current routing strategy, hence speeding up the progress towards plastic-free oceans. It also provides a tool to evaluate the impact of system characteristics on the overall efficiency and inform the design of future systems.

Key words: Ocean cleaning; Sustainable operations; Longest path problem; Dynamic programming;

1. Introduction

Oceans are vital to life on earth: they home a vast array of plant and animal species and play a critical role in regulating the climate. In addition, they provide important economic benefits, supporting industries like fishing, aquaculture, tourism, and the extraction of oil and minerals. However, oceans are being threatened by growing and severe plastic pollution. As of 2015, 80% of the 6.3 billion tonnes of plastic waste ever generated ended up in landfills or the natural environment (Geyer et al. 2017). According to the latest estimates, there were around 3 million tonnes of plastic waste floating in the ocean as of 2020 (Kaandorp et al. 2023). Furthermore, the amount of plastic emissions in the ocean increases by 4% every year (Kaandorp et al. 2023), with 0.5–2.7 million tonnes emitted by rivers every year (Meijer et al. 2021, Lebreton et al. 2017, Schmidt et al. 2017). Plastic pollution is posing a threat to the marine ecosystem and the species that rely on it (Gall and Thompson 2015, Wilcox et al. 2015). It also has a detrimental impact on human activities. For example, We refer to Li et al. (2016) for a comprehensive review of marine plastic pollution, its sources and effects. Because of its environmental and economic relevance, the reduction of oceans pollution has been listed as an explicit target in the United Nations' Sustainable Development Goal #14 'Life Below Water'.

The reduction of marine plastic pollution needs to be addressed from two fronts: (i) reducing yearly emissions and (ii) removing plastic already emitted in the oceans. Regarding the first effort, many legislative and non-legislative actions have been taken to ban (or discourage) the use of single-use plastic (e.g., plastic bags or straws), with varying degrees of efficiency (see Schnurr et al. 2018, for a review). Given the importance of land-based pollution and the role of rivers in transporting land-based pollution into the oceans, solutions also include improved in-land plastic waste management, recycling, and plastic interception in rivers (see, e.g., Dijkstra et al. 2021, Winterstetter et al. 2021). On the other hand, the active removal of plastic already emitted in the oceans has received lower attention and may be regarded as less efficient than preventing emissions due to the low average concentration of floating plastic in the oceans.

Fortunately, floating plastic debris get trapped in large circulating currents, called gyres, and tend to accumulate in specific areas in the oceans called 'garbage patches'. The largest of these five patches, the "Great Pacific Garbage Patch" (GPGP), is situated halfway between California and Hawaii. Latest estimates are that nearly 80,000 tonnes of plastic float inside the GPGP, an area of 1.6 million km^2 or three times the size of France (Lebreton et al. 2018). Figure 1 displays a map of the GPGP together with yearly average plastic density estimates. In short, plastic density in the GPGP is about 20 times higher than average.

The Ocean Cleanup is a Dutch NGO whose mission is to clean up oceans from plastic. In addition to interception activities in rivers, they have developed a technology to collect plastic debris in the oceans. They have been trialling their solutions in the GPGP since 2018, and operating their newest system since 2021. Their system consists of a large (600-meter wide and 4-meter deep, at the beginning of our collaboration) U-shaped screen, slowly dragged by two ships, that can capture floating plastics without capturing any marine animals. In this collaboration, we investigate the

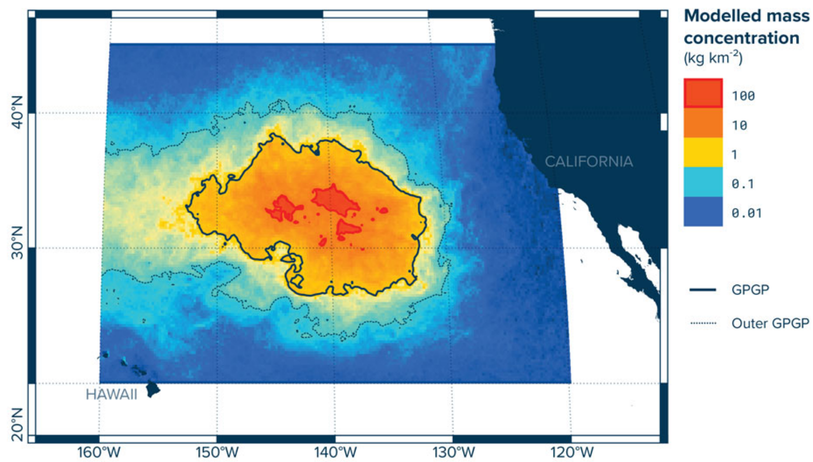


Figure 1 Plastic density map in the Great Pacific Garbage Patch (GPGP). (Source: `https://theoceancleanup.com/great-pacific-garbage-patch/`)

potential for improving the efficiency of their plastic collection system by optimizing its route in the GPGP. In particular, we use data and models about weather conditions and plastic density in the GPGP to construct an optimization-based routing algorithm that directly maximizes the quantity of plastic collected, hence speeding up the progress towards cleaner and healthier oceans.

1.1. Problem description

The Ocean Cleanup's plastic collection system (which we later refer to as the "system") is composed of two ships and a U-shaped screen (or net), as shown in Figure 2. In its original configuration (system 002/B), the screen had a span of 600 meters, but it has been increased to 1.4 km in the latest version of the system (system 003). It acts like an artificial coastline that intercepts floating debris. These floating plastic particles then gradually accumulate in a partially closed contraption at the end of the screen, called the retention zone.

Routing of the system in the GPGP needs to satisfy some navigation requirements. To preserve the physical integrity of the screen, for example, the system can only slowly change course and cannot make any sharp turns. In addition, in order not to catch any fish or other marine life, the system moves at a fixed and low speed of around 1.5 knots (2.78 km/h). Finally, as any sea vessel, it is sensitive to weather and navigation conditions such as waves and wind. For example, when the wave height exceeds 4.5 meters, the system has to head against the waves to protect the screen. Above 6-meter waves, the screen no longer intercepts any plastic.

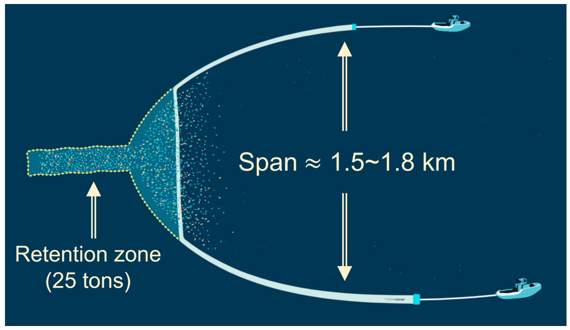


Figure 2 Plastic collection system developed by The Ocean Cleanup. Parameter values correspond to version 03. (Source: `https://theoceancleanup.com/oceans/`)

The retention zone has a limited capacity of 25 metric tons and needs to be emptied regularly. The process of emptying the retention zone is called an extraction and is a complex maneuvre: the screen is taken out of the water by a crane and the collected plastic is discharged on deck, before being sorted and sent on-shore for recycling. In particular, the crane cannot be operated when the wave height exceeds 2.5 meters. Overall, an extraction takes around 24 hours, during which the collection is stopped. Hence, extractions play an important role in the overall collection efficiency and extraction scheduling should be incorporated in our search for a better routing system.

Our primary objective is to maximize the quantity of plastic collected. The Ocean Cleanup has developed a suite of models to estimate the density of plastic in the GPGP (Klink et al. 2022). By using hindcast and forecast models of ocean currents, waves and wind, the dispersal of marine plastics is modelled with a Lagrangian approach. Assimilation methods (Peytavin et al. 2021) and plastic-specific transport models are also investigated (Sainte-Rose et al. 2022). On the sensing front, satellite imaging (Park et al. 2021, 2022) and remote sensing techniques (de Vries et al. 2021) are developed to acquire field data. These models provide a picture of where plastics are located and how they move within this region, hence creating a dynamic view of present and future plastic density (similar to Figure 1, yet evolving over time). Our objective is to integrate these predictions directly into an optimal routing problem, so the system naturally accounts for plastic movements, which is crucial because both the system and the plastics are moving at comparable speeds. A central challenge is to account for the fact that the collection process removes plastic from the oceans and, as such, should directly impact the (estimate of) future plastic density. Hence, plastic density cannot be seen as an exogenous input to our model only, it is also impacted by our routing decisions.

To summarize, our objective is to find a route for the system and to schedule extractions of the retention zone, in order to maximize the total quantity of plastic collected by the system. In particular, we need to account for weather and operational constraints, plastic dynamics, and the direct impact of our decision on future plastic density.

1.2. Contributions and structure

In this work, we develop and validate an optimization approach to jointly optimize the routing and the extractions of the plastic collection system. After reviewing the relevant literature in Section 2, we make the following contributions:

- In Section 3, by discretizing space and time, we model the routing and scheduling decisions as paths in a directed acyclic graph (DAG). Among others, this model can account for relevant operational and weather constraints and provides efficient dynamic programming algorithms for longest-path type of optimization problems.
- Under this lens, the quantity of plastic collected can be seen as lengths of edges in this graph and our problem as a longest path optimization problem. However, due to the direct impact of our routing decisions on future plastic density, our resulting optimization problem is a non-linear and non-decomposable longest path problem. We formally analyze the structure of our path-dependent lengths in Section 4 and bound the estimation error obtained when ignoring the path dependency.
- We propose a search-and-bound strategy to efficiently find a high-quality solution for this class of problems, with certificates of near-optimality (Section 4). Our algorithm leverages a relaxation of the problem to partition and search through the space of trajectories. We also propose a tailored branch-and-bound scheme to solve this class of problems exactly, using our search-and-bound algorithm as the root node analysis (Appendix D). On small instances (2-day planning), our search-and-bound strategy finds the optimal solution, while scaling better with respect to the problem size than exact approaches.
- Finally, we evaluate the benefit of our search-and-bound algorithm on a one-year dataset of ocean weather conditions and plastic density in Section 5. We find that our optimization approach yields at least a 60% improvement in terms of average collection efficiency compared with their current routing strategy. In particular, we observe greater benefits (+100%) during winter months, because weather conditions (and wave height in particular) are limiting the ability to extract, hence exacerbating the benefit of jointly optimizing the route and the extraction schedule. In addition,

our algorithm allows The Ocean Cleanup to explore the non-linear impact of strategic system dimensioning decisions (e.g., span of the system and size of the retention zone) on the resulting efficiency.

2. Literature Review

Our problem can be summarized as a ship joint routing and scheduling problem, where the objective is to steer the system in the GPGP and schedule extractions (i.e., emptying of the retention zone) in order to maximize the quantity of plastic collected. In Section 2.1, we review the optimization literature related to marine operations and ship routing. We then focus on methods for fishing optimization, which is similar to our plastic collection problem. Eventually, we will model our problem as that of a longest path in an appropriately defined graph, so we review the literature on longest path optimization in Section 2.3.

2.1. Optimization for ship routing problems

Following Granado et al. (2021), we divide the literature into weather and tactical routing.

In weather routing, the objective is to find a route that connects a given origin with a given destination and minimizes travel time or fuel consumption, which depend on weather and navigation conditions. The typical planning horizon in weather routing is a few weeks. The great circle passing through these two locations provides the shortest route in terms of travel distance. So, the optimal route is often to be found in the vicinity of the shortest route. Most approaches create a discrete grid of potential locations around the great circle using isochrone lines (James 1957, Hagiwara and Spaans 1987) or a fixed grid (Zoppoli 1972, de Wit 1990). By representing a trajectory as a sequence of locations, the weather routing problem can thus be formulated as a shortest path problem, which can be solved efficiently by Dynamic Programming (DP; Zoppoli 1972, de Wit 1990, Meng and Wang 2011, Ting and Tzeng 2003, Aydin et al. 2017) or Dijkstra's algorithm (Takashima et al. 2009, Skoglund 2012, Sen and Padhy 2015). Additional decision variables, such as engine power in Shao et al. (2012), can be modeled within a shortest path formulation by extending the description of the 'state' of the ship. Heuristic methods have also been used to deal with more complex objectives or constraints, such as simulated annealing (Kosmas and Vlachos 2012), the A* algorithm (Yoon et al. 2018, Langbein et al. 2011), or particle swarm optimization methods (Zheng et al. 2019). Recently, Cheng and Zhang (2018), Chen et al. (2019) used reinforcement learning approach to optimize the route while learning the complex dynamics between waves and speed or fuel consumption. We refer to Zis et al. (2020) for a comprehensive review on ship weather routing.

For our problem, we adopt a similar modeling paradigm by discretizing the location of the system in the GPGP. However, we adopt a more fine-grained discretization of time (3-hour time steps) and space (8 km), and extend the state variable to account for extraction decisions as well, so our resulting graph is of much larger scale, i.e., in the order of 10^6 nodes. In addition, the destination in our problem is not fixed, which can lead to more complex trajectories, such as circling or crossing. In terms of objective, we assume that the fuel efficiency does not depend on the routing decision because of the limited propelling speed so our primary objective is to maximize the amount of plastic collected in a given amount of time. After appropriately defining edge weights, we formulate our problem as a longest path optimization problem and solve it using DP strategies similar to the ones used in weather ship routing.

Tactical ship routing consists in finding the lowest cost route for a ship that needs to visit different locations (e.g., a cargo ship visiting different ports). Since the time horizon is long (several weeks or months) and the ports are fixed isolated location, the problem can be formulated as a Traveling Salesperson Problem (TSP), solved by branch-and-bound (Appelgren 1971, Stalhane et al. 2015), branch-cut-and-price (Battarra et al. 2014), heuristic methods (Malaguti et al. 2018), or DP (Fagerholt and Christiansen 2000). We refer to Christiansen et al. (2004) for a comprehensive review of the literature and its connection to supply chain management.

2.2. Fish routing

Among all maritime activities, fishing is the most comparable to our plastic collection problem because the objective is to capture floating elements in the oceans.

Before solving any route optimization problem, one needs to first predict the density of fish at different locations. However, unlike plastic, fishes are actively moving, which makes their precise location highly unpredictable. Instead, most works describe fish density with coarse granular distributions (see, e.g., Jones et al. 2012, Parra et al. 2017, Coll et al. 2019); see Robinson et al. (2017) for a review of marine-based species distribution models. There is also opportunity to improve these predictions using real-time remote sensing measurements (Iglesias et al. 2007). Unfortunately, estimating the accuracy of these different approaches remains an open challenge. Indeed, 94% of the studies reviewed by Robinson et al. (2017) failed to report the uncertainty of their model.

In the weather routing literature, algorithms that consider wave and wind forecast to design safe and efficient routes have been applied to fishing (e.g., Vettor et al. 2016). In these use cases, the objective of maximizing the quantity of fish collected is captured in the choice of the target destination, and is typically left to the end-user. This implementation bypasses the issue of inaccurate predictions by letting the human user identify (based on quantitative models and intuition) the destination. To the best of our knowledge, no weather ship routing approach uses quantitative fish density predictions directly as an input to optimize the short-term (within the next days) route of fishing ships. Instead, predictions on the presence and movements of fish banks or 'clusters' are mostly used as locations in a tactical ship routing problem. For tuna fishing for example, floating devices are dispersed in the ocean to attract fishes. Groba et al. (2015, 2018, 2020) model the problem of visiting all devices as a dynamic TSP, where locations can drift due to sea current.

In our problem, pieces of plastic move passively according to sea currents, like the floating aggregate devices. However, there are no devices to attract plastic and keep the integrity of plastic clusters. Instead, high-plastic-density clusters are constantly forming and breaking depending on the currents. Accordingly, we use a fluid dynamics model to predict plastic location and movements, and integrate it into an optimization formulation analogous to that of weather ship routing.

2.3. Optimization for longest path

Given weights on the edges of a graph, the length of a path is defined as the sum of the weights of the edges composing the path. The problem of finding the longest path in a graph is shown to be $\mathcal{NP}$-complete as a generalization of the Hamiltonian path problem (Karp 2010). Actually, the longest path problem cannot be approximated in polynomial time unless $\mathcal{P} = \mathcal{NP}$, as proved by Karger et al. (1997) for undirected and Björklund et al. (2004) for directed graph.

In contrast, finding the shortest path in a graph can be solved in polynomial time using algorithms like the greedy-type Dijkstra's algorithm (Dantzig 1960, Dijkstra 2022) or the Bellman–Ford algorithm (Shimbel 1954, Ford Jr 1956, Bellman 1958, Moore 1959). We refer to Pollack and Wiebenson (1960), Schrijver (2012) for comprehensive reviews. Understanding the structural differences between the longest and shortest path problems and their implications for problem complexity has been a vivid research topic (see, e.g., Cormen et al. 2022), unravelling conceptual connections between shortest path algorithms and DP (Sniedovich 2006).

Nonetheless, polynomial time algorithms for longest path problems exist for particular classes of graphs such as trees (Bulterman et al. 2002, Uehara and Uno 2007), block graphs, cactus graphs

(Uehara and Uno 2007), and cocomparability graphs (Ioannidou and Nikolopoulos 2013). The graph we propose in Section 3.1 is a DAG. The longest path problem in a DAG can be solved in linear time by transforming it into the shortest path problem (Pandit 1962, Cormen et al. 2022), or using DP on the topological sort of the DAG (Madraki and Judd 2019). The DAG in our project has a natural topological sort, and we use DP in Section 3.2 to find the longest path.

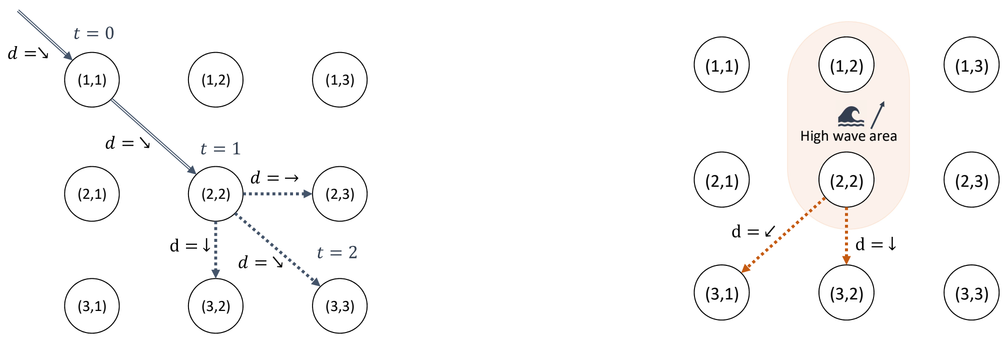
3. Graph-based Routing Model

In this section, we propose a graph-based formulation for our problem. By discretizing time and space, we show in Section 3.1 how the routing decision can be modeled as a path in a sparse directed acyclic graph. Accordingly, longest path optimization problems can be solved efficiently over this graph using DP approaches, as presented in Section 3.2. We conclude this section by discussing how extraction scheduling decisions can be incorporated within this graph-based model (details are deferred to Appendix A.2) and formally identifying the set of tractable optimization problems we can solve in Section 3.3.

3.1. Discretization and graph representation

To describe the system's trajectory and model the key decisions and constraints of our problem, we discretize space onto a finite grid, as represented in Figure 3. Similarly, we divide our planning horizon using a fixed timestep. In our implementation, we use an 8-km step to discretize space and a 3-hour time step. Denoting $\mathcal{L}$ the set of all possible locations and $\mathcal{T} := \{0, 1, \dots, T\}$ the set of time periods, we can represent a system trajectory as a sequence of locations, $\{\ell_t\}_{t \in \mathcal{T}}$ with $\ell_t \in \mathcal{L}$.

However, as explained in Section 1.1, the steering direction also plays an important role in our problem because of operational (e.g., no sharp turns) and weather constraints (e.g., if the wave height exceeds 4.5 meters, need to navigate against the waves). Hence, at a given time t, knowledge of the current location ℓ_t is not sufficient to determine whether the constraints are satisfied and what the accessible next locations are. Accordingly, we describe a trajectory by a sequence $\{(\ell_t, d_t)\}_{t \in \mathcal{T}}$, where ℓ_t is the location at time t and $d_t \in \mathcal{D}$ is the steering direction at time t. Here, $\mathcal{D}$ denotes the (finite) set of allowable steering directions in our grid, $\mathcal{D} := \{\uparrow, \nearrow, \rightarrow, \searrow, \downarrow, \swarrow, \leftarrow, \nwarrow\}$. Note that we introduce redundancy in the information used to describe the trajectory of the system at time t. Essentially, providing ℓ_t and d_t is equivalent to providing ℓ_t and ℓ_{t-1}. Alternatively, given an initial location ℓ_0, the sequence of directions $\{d_t\}_{t \in \mathcal{T}}$ is sufficient to describe the system's trajectory.



(a) Discretization of path (b) High wave region and direction at $t = 1$

Figure 3 Example: routing in a small grid.

However, as we will see in this section, our description using both location and direction allows us to efficiently account for the different constraints in our problem.

Figure 3a shows an example of a discretized trajectory. For simplicity, we associate $\mathcal{L}$ with a set of discrete coordinates in $\mathbb{R}^2$, which are in a one-to-one correspondence with the latitude and longitude of the system. In this example, at $t = 0$, the system is at location $(0,0)$, moving south east ($\searrow$). It keeps the same steering direction at $t = 1$ and reaches $(1,1)$. At $t = 2$, it can reach three different locations depending on whether it continues south east or decides to change course and move $\rightarrow$ or $\downarrow$. Because the propelling speed of our system is limited (in order not to catch any marine life), we use discretization steps for time (3 hours) and space (8 km) that are consistent with this low propelling speed (1–1.5 knots) and assume that the system in one location can only reach the neighbouring locations at the next time period. We could relax this assumption and adopt a finer discretization strategy to account for travel time differences between diagonal and horizontal/vertical moves or allow for different propelling speed depending on the steering direction (e.g., to maintain a constant speed relative to water).

Using terminology from DP, we refer to the triplet $s := (\ell, d, t) \in \mathcal{L} \times \mathcal{D} \in \mathcal{T}$ as the state of the system. For each state $s = (\ell, s, t)$, we can then define its set of *successors*, i.e., the set of admissible next states $s' = (\ell', s', t+1)$ that satisfy all operational and weather constraints, such as:

- **Consistency between locations and directions:** The next location ℓ' needs to correspond to the location reached from ℓ after following the direction d', which we could express algebraically as "$d' = \ell' - \ell$" after appropriately mapping $\mathcal{L}$ and $\mathcal{D}$ to vectors in $\mathbb{R}^2$.
- **No sharp angles:** In our problem, the angle between the steering directions d and d' is at most 45 degrees (or $\pi/4$). For example, for $s = ((2,2), \searrow, 1)$ in Figure 3a, we must have $d' \in \{\rightarrow, \searrow, \downarrow\}$.

- **High-wave regions:** When the wave height exceeds 4.5 meters, the system has to navigate against the waves. For example, in Figure 3a-3b, assume that the location at $t=1$ is in a high-wave region with waves going north-north-east. Then, according to this constraint, the next direction d' can only be $\swarrow$ or $\downarrow$.

Our model can thus account for any constraint defined on the location or direction of the system—we provide a list of the operational constraints of our problem in Appendix A.1. In Appendix A.2, we describe how to extend the state space further, to incorporate the decision of extraction scheduling and impose the associated constraints (in particular, extraction can only be performed when the wave height is below 2.5 meters). All together, these constraints define the successors of a state s, which we concisely denote $\mathrm{succ}(s)$. For the example, in Figure 3, we have $\mathrm{succ}((2,2),\searrow,1)) = \{((3,2),\downarrow,2)\}$. In other words, the system only has one feasible next state given the current state and weather conditions.

With these notations, we can represent admissible trajectories as paths on a graph $\mathcal{G}=(\mathcal{S},\mathcal{E})$. The set of nodes $\mathcal{S}$ can be naturally partitioned by the time period $t\in\mathcal{T}$, i.e., $\mathcal{S}=\cup_{t\in\mathcal{T}}\mathcal{S}_t$, where $\mathcal{S}_t$ is the set of feasible states at time t and is defined recursively. At time $t=0$, if we are given an initial location ℓ_0 only, then the set of all possible initial states is $\mathcal{S}_0=\{(\ell_0,d,0):d\in\mathcal{D}\}$. We then apply the recursion

$$\mathcal{S}_{t+1}=\cup_{s\in\mathcal{S}_t}\,\mathrm{succ}(s).$$

Similarly, the set of edges can be decomposed into $\mathcal{E}=\cup_{t=1}^{T}\mathcal{E}_t$, with $\mathcal{E}_{t+1}=\{(s,s')\in\mathcal{S}_t\times\mathcal{S}_{t+1} : s'\in \mathrm{succ}(s)\}$. In particular, observe that the graph $\mathcal{G}$ is a DAG. Furthermore, it is relatively sparse. Because of the no-sharp-turn constraint, the number of edges satisfies $|\mathcal{E}_t|=\mathcal{O}(|\mathcal{S}_t|)$. Figure 4 shows the graph corresponding to the example of Figure 3. Here, $\mathcal{S}_0=\{((1,1),\searrow,0)\}$. The dashed orange nodes (edges) represent the states (transitions) forbidden by weather-related constraints.

In our implementation, we plan for 7 days with 3-hour time steps, so $T=7\times 8=56$ and the grid of all reachable locations is of size $|\mathcal{L}|=(56+1+56)\times(56+1+56)=12,769$. Hence, for each time t, the number of possible states for time t is bounded as follows: $|\mathcal{S}_t|\leq|\mathcal{S}|\times|\mathcal{D}|\approx 10^5$.

Thanks to the convenient structure of the graph $\mathcal{G}$, given fixed weights on the edges, we can efficiently find the longest path, i.e., the reward-maximizing trajectory. In the next section, we describe an efficient DP approach for solving such longest path problems, assuming all edges are associated with a fixed and known reward. This observation motivates us to define the reward of an

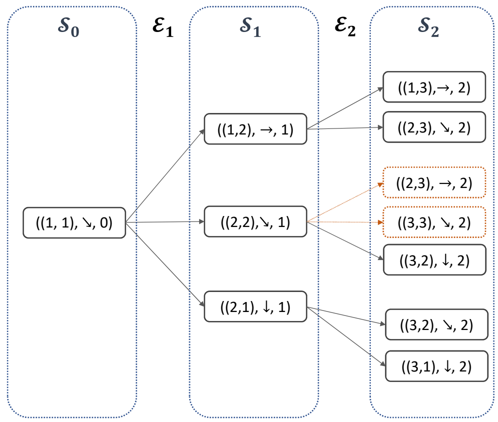


Figure 4 The structure of graph in the example of Figure 3. Edges and nodes in orange dashes are forbidden by the weather constraints.

edge as the quantity of plastic collected when the system passes through that edge. Unfortunately, as we discuss in Section 1.1, properly defining these rewards is non-trivial and introduces non-linearities in our optimization problem due to plastic dynamics and the impact of past decisions and future plastic density.

3.2. Efficient search for longest path

For this section, we assume that a set of weights for each edge of our graph $\mathcal{G}$ is given, w_e for $e \in \mathcal{E}$. Again, intuitively, the rewards should correspond to the quantity of plastic collected when the system moves along edge e at time t, but we defer a formal definition to Section 4. Under this assumption, our plastic collection problem would be equivalent to finding the longest path in $\mathcal{G}$ with edges weighted by $\boldsymbol{w}$. Because $\mathcal{G}$ is a DAG, the longest path can be found efficiently using a DP algorithm.

We describe the DP algorithm in a forward manner, but it could equivalently be solved and described in a backward manner. For any state $s \in \mathcal{S}_{t+1}$, let $V^{t+1}(s)$ denote the length of the longest path connecting s to $\mathcal{S}_0$. Formally,

$$V^{t+1}(s) := \max_{s_0 \in \mathcal{S}_0, \ldots, s_t \in \mathcal{S}_t} \sum_{\tau=0}^{t} w_{s_\tau, s_{\tau+1}} \text{ with } s_{t+1} = s.$$

The key idea in the DP algorithm is that the solution of the optimization problem above can be computed recursively by connecting the longest path between $\mathcal{S}_0$ and s' and the edge (s', s) for some $s' \in \mathcal{S}_t$, i.e.,

$$V^{t+1}(s) = \max_{s' \in \mathcal{S}_t : s \in \mathrm{succ}(s')} \left\{ w_{s',s} + V^t(s') \right\}.$$

The maximization above can be solved by exhaustively searching through $\mathcal{S}_t$. Fortunately, in our graph $\mathcal{G}$, because of the no-sharp-angle constraint, we have $|\{s' \in \mathcal{S}_t : s \in \mathrm{succ}(s')\}| \leq 3$, so this maximization problem can be solved in $\mathcal{O}(1)$ operations. Note that the solution to the maximization problem $V^{t+1}(s)$ is a trajectory or sequence of states from time 0 to time t, which can also be computed recursively when computing the value of $V^{t+1}(s)$. This recursive procedure is the basis of the DP algorithm described in Algorithm 1. Figure 5 shows an example of a 7-day collection route, where the background map represents the plastic density on day 4, when the system is at the triangle location.

Algorithm 1: Dynamic Programming Algorithm for Finding the Longest Path

```
Data: Weighted graph 𝒢 with weight {w_{s,s'}}_{(s,s')∈ℰ};
Initialize (values and optimal paths): V^0(s) = 0, path[s] = {s}, for all s ∈ 𝒮_0;
for t = 1:T do
    for s ∈ 𝒮_t do
        Find the optimal previous state, s* ∈ argmax_{s'∈𝒮_{t-1}:s∈succ(s')} {w_{s',s} + V^{t-1}(s')};
        Update value function: V^t(s) = w^t_{s*,s} + V^{t-1}(s*) ;
        Update optimal path: path[s] ← path(path[s*], s);
    end
end
Find the optimal terminal state s⋆ ∈ argmax_{s∈𝒮_T} V^T(s);
Return: value V^T(s⋆), longest path path[s⋆].
```

Observe the exceptional computational efficiency of the DP algorithm. At each iteration t, the algorithm performs $\mathcal{O}(|\mathcal{S}_t|) = \mathcal{O}(|\mathcal{L}||\mathcal{D}|)$ operations. Hence, the total computational complexity is of the order $\mathcal{O}(T|\mathcal{L}||\mathcal{D}|)$, i.e., linear in T, although there is an exponential number of possible trajectories—$\mathcal{O}(3^T)$ given the no-sharp-turn constraint. In our implementation, the total number of operations required by Algorithm 1 is bounded by $|\mathrm{succ}(s)| \times T|\mathcal{L}||\mathcal{D}| \approx 17 \times 10^6$, while modern computers can execute around $> 10^8$ additions per second. In contrast, evaluating all $3^{56} \approx \times 10^{27}$ trajectories would require billions of years.

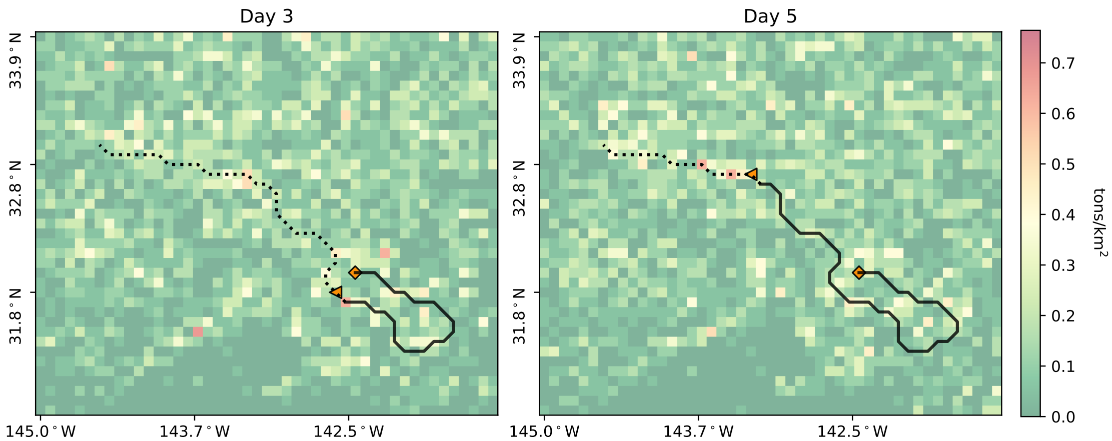


Figure 5 Example of an optimal 7-day route (starting at the diamond), represented on day 3 and day 5 (triangle location).

3.3. Summary: Power of the graph-based modeling

In this section, we propose a graph-based modeling to represent the routing decision as a path in a sparse DAG, $\mathcal{G}$. In Appendix A.2, we show how we can extend the state space $\mathcal{S}$ of our system and the graph $\mathcal{G}$ to also account for extraction scheduling. Hence, we obtain a similar DAG where each path now corresponds to a sequence of routing and extraction scheduling decisions.

The DP algorithm described in Section 3.2 (Algorithm 1) is an efficient approach for solving longest path problems over this graph. In this work, we will be particularly interested in problems where rewards are associated with states instead of edges, i.e., longest path problems of the form

$$\max_{\boldsymbol{x}\in\mathcal{X}} \quad \sum_{t\in\mathcal{T}} \sum_{s\in\mathcal{S}_t} r_s^t x_s^t, \tag{1}$$

where $x_s^t \in \{0,1\}$ indicates whether the system is in state s at time t and $\mathcal{X}$ denotes the set of admissible such binary variables. A formal definition of the feasible set $\mathcal{X}$ is provided in A.3. Of course, problems of the form (1) can be solved by Algorithm 1, as longest path problems with weights $w_{s,s'} = r_{s'}^t$ for $(s,s') \in \mathcal{E}_t$. Conceptually, one can interpret the DP approach as an efficient partitioning of the set $\mathcal{X}$. At the end of Algorithm 1, the set of trajectories $\mathcal{X}$ is partitioned according to the terminal state they reach. $V^T(s)$ corresponds to the value of the longest path problem across all paths terminating at s, and the variable path$[s]$ contains one path achieving this value.

4. Path-Dependent Reward Structure

Since our graph-based formulation presented in Section 3.1 allows us to efficiently find trajectories that maximize a given reward, a natural approach is to cast our plastic collection problem as a longest path problem. Because plastic are freely floating in the oceans, however, plastic particles are constantly moving and the plastic-collecting system directly impacts these dynamics. We present the dynamics of plastic movements in the absence of any collection in Section 4.1, and then derive the resulting dynamics for our objective function in Section 4.2. In particular, the rewards (or edge length) we need to consider are path-dependent, i.e., they depend on the entire past trajectory of the system. We analyze the structure of such path-dependent rewards in Section 4.3, and propose an efficient search-and-bound algorithm to find near-optimal solutions to the resulting non-linear longest path optimization problem in Section 4.4.

4.1. Fluid mechanics model of free-floating plastic dispersal

Plastic particles move passively in the oceans and, as such, their movements can be modeled and predicted using fluid dispersal models. Among others, the engineering team at The Ocean Cleanup models the velocity of plastic particles in the oceans using data on sea currents and waves, and taking into account the Stokes drift (Stokes 1847) and Eddy diffusivity (Taylor 1915) phenomena.

Denoting r_ℓ^t the quantity of plastic present at time $t \in \mathcal{T}$ and at location $\ell \in \mathcal{L}$, these fluid dispersal models provide us with estimates on the quantity of plastic present in the region at times $(\boldsymbol{r}^0, \ldots, \boldsymbol{r}^T)$ as well as structural relationships connecting the vectors. Formally, from the models developed by The Ocean Cleanup, we also obtain matrices $\boldsymbol{Q}^t \in \mathbb{R}_+^{\mathcal{L}\times\mathcal{L}}$ such that

$$\boldsymbol{r}^{t+1} = \boldsymbol{Q}^t \boldsymbol{r}^t. \tag{2}$$

Each entry $Q_{\ell,\ell'}^t$ of the matrix $\boldsymbol{Q}^t$ indicates the fraction of plastic present at location ℓ' at time t that moves to location ℓ at time $t+1$. Of course, if the total quantity of plastic is constant (which is a reasonable assumption given our relatively short planning horizon), then the matrix $\boldsymbol{Q}^t$ should be left stochastic, i.e., $\sum_{\ell\in\mathcal{L}} Q_{\ell',\ell}^t = 1$ for all $\ell \in \mathcal{L}$. In this case, we can interpret $\boldsymbol{Q}^t$ as the transition matrix of a Markov process. Yet, the only property of $\boldsymbol{Q}^t$ that we leverage in our approach is that it has non-negative entries.

Remark 1. In practice, note that the matrices $\boldsymbol{Q}^t$ are large, $125,000 \times 125,000$ in our implementation, so computing matrix-vector products involving $\boldsymbol{Q}^t$'s is computationally challenging, although the vectors $\boldsymbol{r}^t$ can be computed during a preprocessing step. Even the constructino of the matrix $\boldsymbol{Q}^t$ from the particle-level fluid dispersal model can be time consuming. We discuss computational aspects related to the plastic dynamics model in Appendix C.

4.2. Path dependency

Given this information, we now define a relevant objective for our longest path problem. While the plastic density vectors (or 'maps') $\boldsymbol{r}^t$, $t \in \mathcal{T}$, defined in the previous section, describe the plastic dynamics in absence of any collection process, our system actively removes plastic from the ocean and the collected plastic no longer evolves according to (2). In other words, the quantity of plastic collected by our system (and the locations where this plastic has been collected) directly impacts the future spatial distribution of plastic. It is relevant to our optimization problem because our system moves at a speed comparable to that of the plastic. We refer to this phenomenon as *path dependency* and now appropriately define a reward (or length) vector for our optimization problem that takes this phenomenon into account.

Let us denote the location of the system at time t through a one-hot vector $\boldsymbol{x}^t \in \{0,1\}^{\mathcal{L}}$, where $x^t_\ell = 1$ if and only if the system is in ℓ at time t. We denote the quantity of plastic present (or reward) associated with each location at time t $\boldsymbol{r}^t_{|\boldsymbol{x}^{0:t-1}} \in \mathbb{R}^{\mathcal{L}}_+$, where $\boldsymbol{x}^{0:t-1}$ concisely denotes the sequence $\{\boldsymbol{x}^0, \ldots, \boldsymbol{x}^{t-1}\}$ and emphasizes the dependency on the past trajectory. If the system is in location ℓ at time t, it collects a fraction $\alpha \in [0,1]$ of the plastic present. Hence, it collects $\alpha r^t_{\ell|\boldsymbol{x}^{0:t-1}}$ and the remaining $(1-\alpha) r^t_{\ell|\boldsymbol{x}^{0:t-1}}$ continues to float in the ocean, together with the plastic present in other locations, $r^t_{\ell'|\boldsymbol{x}^{0:t-1}}$ for $\ell' \neq \ell$. All together, the spatial density of plastic at time $t+1$, $\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}}$, should depend on $\boldsymbol{r}^t_{|\boldsymbol{x}^{0:t-1}}$ and $\boldsymbol{x}^t$ through the following recursion

$$\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} = \boldsymbol{Q}^t \left(\boldsymbol{r}^t_{|\boldsymbol{x}^{0:t-1}} - \alpha\, \boldsymbol{r}^t_{|\boldsymbol{x}^{0:t-1}} \circ \boldsymbol{x}^t \right). \tag{3}$$

In (3), the symbol $\circ$ denotes the Hadamard or element-wise product between two vectors. Hence, $\boldsymbol{r}^t_{|\boldsymbol{x}^{0:t-1}} - \alpha\, \boldsymbol{r}^t_{|\boldsymbol{x}^{0:t-1}} \circ \boldsymbol{x}^t$ corresponds to the density map where we remove a fraction α of the plastic in the location of the cleaning system.[1] Note that since $\boldsymbol{Q}^t$, $\boldsymbol{x}^t$, and $\boldsymbol{r}^0$ have non-negative entries, one can show by induction that $\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} \geq 0$.

[1] The dynamics in (3) implicitly assume that we can decompose the time interval $[t, t+1)$ into two distinct steps: a first step where some of the plastic present at time t is removed, and a second step where the remaining plastic float according to the dynamics captured by $\boldsymbol{Q}^t$. Of course, this is a simplification of reality where these two steps occur concurrently. Yet, we believe it is an appropriate model of reality, which captures the essence of path dependency.

With these dynamics in mind, the problem of jointly routing the system and scheduling the extractions in order to collect the maximum amount of plastic possible can be formulated as the following longest path optimization problem:

$$\max_{\boldsymbol{x} \in \mathcal{X}} \quad \sum_{t \in \mathcal{T}} \sum_{s \in \mathcal{S}_t} r^t_{s|\boldsymbol{x}^{0:t-1}} x^t_s \quad \text{s.t.} \quad \boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} = \mathcal{Q}^t(\boldsymbol{r}^t_{|\boldsymbol{x}^{0:t-1}}, \boldsymbol{x}^t), \tag{4}$$

which is analogous to the longest path problem (1) except that the rewards are no longer fixed but also depend on the past decisions, $\boldsymbol{x}^{0:t}$.

Remark 2. Note that, with a slight abuse of notations, we use the variable x^t_s in (4) to encode for the state of the system at time t, while plastic dynamics (3) are described using binary variables x^t_ℓ encoding for the location of the system (location being one component of the state only)—and similarly for the associated rewards. However, it should be clear that we can recover the location from the system's state via a simple affine mapping and that the reward dynamics described at a location level in (3) imply similar dynamics for the state rewards. We formally define this mapping in Appendix B.1 and introduce a generic operator $\mathcal{Q}^t$ in the optimization problem (4) to concisely capture the resulting dynamics on the state variables/rewards. In the remainder of this section, for ease of notations, we will implicitly work with location-based $\boldsymbol{x}$ variables when analyzing the structure of the rewards generated by the recursive formula (3) in Section 4.3 but refer to the state-level variables when describing optimization algorithms for solving (4) in Section 4.4. This simplification is valid because the mapping between the two descriptions is monotonous.

Unfortunately, Problem (4) is much more challenging to solve than (1) because the reward vector depends on $\boldsymbol{x}$ itself so the objective is non-linear. In other words, (4) can be seen as a non-linear non-decomposable longest path problem, i.e., where the length of a path can no longer be decomposed into independent edge lengths, because the reward of each state depends on the past trajectory. To better understand the dynamics and complexities of the problem, we first study analytically the path-dependent reward vectors defined by the recursion (3) in Section 4.3 before proposing an efficient numerical algorithm for finding solutions to (4) in Section 4.4.

4.3. Reward decomposition

We now analytically analyze the structure of the path-dependent reward $\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}}$.

To build intuition, we start by the special case where plastic does not move, i.e., when the matrices $\boldsymbol{Q}^t$ are the identity matrices. In this case, in the absence of any plastic collection, the plastic density maps $\boldsymbol{r}^t$ defined by Equation (2) are constant over time, $\boldsymbol{r}^0 = ... = \boldsymbol{r}^T =: \boldsymbol{r}$. Accordingly, we drop the time superscript in our (path-dependent) rewards $\boldsymbol{r}^t$ (and $\boldsymbol{r}^t_{|\boldsymbol{x}^{0:t-1}}$, although the path-dependent reward $\boldsymbol{r}_{|\boldsymbol{x}^{0:t-1}}$ still depends on time t through the past trajectory $\boldsymbol{x}^{0:t-1}$. In this case, we have the following expansion:

Lemma 1. *When the matrices $\boldsymbol{Q}^t$ are all equal to the identity matrix, we have*

$$\boldsymbol{r}_{|\boldsymbol{x}^{0:t}} = \boldsymbol{r} + \sum_{1 \le k \le t} (-\alpha)^k \sum_{0 \le t_1 < \cdots < t_k \le t} \boldsymbol{r} \circ \boldsymbol{x}^{t_1} \circ \cdots \circ \boldsymbol{x}^{t_k}.$$

Proof of Lemma 1 In this case, the plastic dynamics (3) simplifies as

$$\boldsymbol{r}_{|\boldsymbol{x}^{0:t}} \;=\; \boldsymbol{r}_{|\boldsymbol{x}^{0:t-1}} - \alpha\, \boldsymbol{r}_{|\boldsymbol{x}^{0:t-1}} \circ \boldsymbol{x}^t \;=\; \boldsymbol{r}_{|\boldsymbol{x}^{0:t-1}} \circ (\boldsymbol{1} - \alpha \boldsymbol{x}^t) \;=\; \boldsymbol{r} \circ (\boldsymbol{1} - \alpha \boldsymbol{x}^0) \circ \cdots \circ (\boldsymbol{1} - \alpha \boldsymbol{x}^t).$$

The Hadamard product being commutative and associative, we can use the classical polynomial expansion technique to obtain

$$\boldsymbol{r}_{|\boldsymbol{x}^{0:t}} = \boldsymbol{r} \circ \left[\sum_{0 \le k \le t} (-\alpha)^k \sum_{0 \le t_1 < \cdots < t_k \le t} \boldsymbol{x}^{t_1} \circ \cdots \circ \boldsymbol{x}^{t_k} \right] = \sum_{0 \le k \le t} (-\alpha)^k \sum_{0 \le t_1 < \cdots < t_k \le t} \boldsymbol{r} \circ \boldsymbol{x}^{t_1} \circ \cdots \circ \boldsymbol{x}^{t_k}.$$

$\square$

Remark 3. The expansion of the path-dependent reward in Lemma 1 is analogous to the inclusion-exclusion principle. Indeed, given n finite sets $A_1, A_2, \dots, A_n$, the inclusion-exclusion principle states that the cardinality of their union is given by

$$|\cup_{i=1}^n A_i| = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \le i_1 < \cdots < i_k \le n} |A_{i_1} \cap \cdots \cap A_{i_k}|.$$

Actually, the result of Lemma 1 can be directly derived from the formula above by appropriately defining the sets A_i as the plastic particles encountered by the system in each location and time.

Lemma 1 shows that the path-dependent reward $\boldsymbol{r}_{|\boldsymbol{x}^{0:t}}$ can be computed from $\boldsymbol{r}$ by applying successive corrections. The kth-order correction in this expansion involves Hadamard products of the form $\boldsymbol{x}^{t_1} \circ \cdots \circ \boldsymbol{x}^{t_k}$, each of them being different from the $\boldsymbol{0}$ vector if and only if there exists a location ℓ such that $x^{t_1}_\ell = \cdots = x^{t_k} = 1$. In other words, the first-order correction consists in removing a fraction α of the plastic in locations visited at least once by the system; the second-order correction adds a fraction α^2 of the plastic in locations visited as least twice by the system;

and so on. Accordingly, the terms in this expansion decay exponentially in k because of the α^k term and because the number of locations being updated decreases. Indeed, for any $k \geq 0$, we have

$$\sum_{0 \leq t_1 < \cdots < t_k < t_{k+1} \leq t} \boldsymbol{r} \circ \boldsymbol{x}^{t_1} \circ \cdots \circ \boldsymbol{x}^{t_k} \circ \boldsymbol{x}^{t_{k+1}} \leq \sum_{0 \leq t_1 < \cdots < t_k \leq t} \boldsymbol{r} \circ \boldsymbol{x}^{t_1} \circ \cdots \circ \boldsymbol{x}^{t_k}.$$

In the general case, the plastic particles move according to the matrix $\boldsymbol{Q}^t$ so they are not assigned to a fixed location. Still, the intuition of Lemma 1 holds: the path-dependent rewards $\boldsymbol{r}^t_{|\boldsymbol{x}^{0:t-1}}$ can be obtained from the original rewards $\boldsymbol{r}^t$ by removing a fraction α of the plastic particles encountered once, adding a fraction α^2 of the plastic particles encountered twice, ... We derive analytically the first-order expansion of the path-dependent reward in the general case in Proposition 1.

Proposition 1. *The path-dependent reward $\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}}$ can be approximated as*

$$\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} = \boldsymbol{r}^{t+1} - \alpha \sum_{0 \leq t_1 \leq t} (\boldsymbol{Q}^t \times \cdots \times \boldsymbol{Q}^{t_1}) (\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}) + \mathcal{O}(\alpha^2). \tag{5}$$

Proof of Proposition 1 We prove the result by induction. For $t = 0$, $\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} = \boldsymbol{r}^1_{|\boldsymbol{x}^{0:0}}$ and Equation (3) leads to

$$\boldsymbol{r}^1_{|\boldsymbol{x}^{0:0}} \quad = \boldsymbol{Q}^0 \boldsymbol{r}^0 - \alpha \boldsymbol{Q}^0 \left(\boldsymbol{r}^0 \circ \boldsymbol{x}^0\right) \quad = \boldsymbol{r}^1 - \alpha \boldsymbol{Q}^0 \left(\boldsymbol{r}^0 \circ \boldsymbol{x}^0\right).$$

Hence, in this case, $\boldsymbol{r}^1_{|\boldsymbol{x}^{0:0}}$ is exactly equal to the first-order expansion (in α) proposed in (5).

Let us now assume that the expansion (5) holds for some $t \geq 0$. Then, at time $t+1$, we have

$$\boldsymbol{r}^{t+2}_{|\boldsymbol{x}^{0:t+1}} = \boldsymbol{Q}^{t+1} \boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} - \alpha \boldsymbol{Q}^{t+1} \left(\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} \circ \boldsymbol{x}^{t+1}\right) \qquad \text{by (3).}$$

We then expand $\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}}$ to obtain

$$\begin{aligned} \boldsymbol{Q}^{t+1} \boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} &= \boldsymbol{Q}^{t+1} \left(\boldsymbol{r}^{t+1} - \alpha \sum_{0 \leq t_1 \leq t} (\boldsymbol{Q}^t \times \cdots \times \boldsymbol{Q}^{t_1}) (\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}) + \mathcal{O}(\alpha^2) \right) \\ &= \boldsymbol{r}^{t+2} - \alpha \sum_{0 \leq t_1 \leq t} \boldsymbol{Q}^{t+1} (\boldsymbol{Q}^t \times \cdots \times \boldsymbol{Q}^{t_1}) (\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}) + \mathcal{O}(\alpha^2), \end{aligned}$$

and

$$\alpha \boldsymbol{Q}^{t+1} \left(\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} \circ \tilde{\boldsymbol{x}}^{t+1}\right) = \alpha \boldsymbol{Q}^{t+1} \left(\boldsymbol{r}^{t+1} + \mathcal{O}(\alpha)\right) \circ \boldsymbol{x}^{t+1} \quad = \alpha \boldsymbol{Q}^{t+1} \left(\boldsymbol{r}^{t+1} \circ \boldsymbol{x}^{t+1}\right) + \mathcal{O}(\alpha^2).$$

Regrouping the two pieces together leads to

$$\boldsymbol{r}^{t+2}_{|\boldsymbol{x}^{0:t+1}} = \boldsymbol{r}^{t+2} - \alpha \sum_{0 \leq t_1 \leq t+1} (\boldsymbol{Q}^{t+1} \times \boldsymbol{Q}^t \cdots \times \boldsymbol{Q}^{t_1}) (\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}) + \mathcal{O}(\alpha^2),$$

i.e., the expansion (5) for $t+1$. □

Proposition 1 exhibits the same structure as in the static-plastic case except that now the cumulative product of transition matrices $\boldsymbol{Q}^t \times \cdots \times \boldsymbol{Q}^{t_1}$ accounts for the movement of plastics encountered at time t_1, between time t_1 and $t+1$. We also provide the second-order expansion in Appendix B.3.

Finally, from the analogy with the inclusion-exclusion principle, one can expect that truncating the expansion at a fixed order k with k even (resp. odd) leads to an upper (resp. lower) bound on the path-dependent reward. Proposition 2 shows that zero-th and first-order expansion provide valid upper and lower bound respectively on the path-dependent reward, which can be a computationally efficient way to estimate these rewards given the size of the matrices $\boldsymbol{Q}^t$'s.

PROPOSITION 2. *The path-dependent reward $\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}}$ satisfies the following bounds:*

$$\boldsymbol{r}^{t+1} - \alpha \sum_{0 \le t_1 \le t} \left(\boldsymbol{Q}^t \times \cdots \times \boldsymbol{Q}^{t_1}\right)\left(\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}\right) \quad \le \quad \boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} \quad \le \quad \boldsymbol{r}^{t+1}. \tag{6}$$

The proof of Proposition 2 (by induction) is deferred to Appendix B.2.

4.4. Adjusted dynamic programming algorithm

In this section, we propose an efficient algorithm for solving our longest path problem with path-dependent rewards (4).

Let us concisely denote $\boldsymbol{r}$ the plastic density maps obtained by applying the fluid advection equations (2) without any active collection, and $\boldsymbol{r}_{|\boldsymbol{x}}$ its path-dependent version. Our objective is to find the longest path according to the path-dependent rewards $\boldsymbol{r}_{|\boldsymbol{x}}$, (4). If, instead, we use $\boldsymbol{r}$ as rewards, we obtain a problem of the form (1), which we call the *relaxed* problem. Both problems optimize over the same feasible space, $\boldsymbol{x} \in \mathcal{X}$, but differ in their objective function, $\langle \boldsymbol{r}, \boldsymbol{x} \rangle$ and $\langle \boldsymbol{r}_{|\boldsymbol{x}}, \boldsymbol{x} \rangle$ respectively. For concision, we use $\langle \cdot, \cdot \rangle$ to denote the inner product between two vectors. Let us denote $\boldsymbol{x}^\star(\boldsymbol{r})$ and $\boldsymbol{x}^\star(\boldsymbol{r}_{|\boldsymbol{x}})$ their respective solutions.

In fact, $\boldsymbol{x}^\star(\boldsymbol{r})$ is the solution returned by Algorithm 1. Because $\boldsymbol{r}_{|\boldsymbol{x}} \le \boldsymbol{r}$ (Proposition 2) and because flows are non-negative,

$$\langle \boldsymbol{r}_{|\boldsymbol{x}^\star(\boldsymbol{r}_{|\boldsymbol{x}})}, \boldsymbol{x}^\star(\boldsymbol{r}_{|\boldsymbol{x}}) \rangle \le \langle \boldsymbol{r}, \boldsymbol{x}^\star(\boldsymbol{r}_{|\boldsymbol{x}}) \rangle \le \langle \boldsymbol{r}, \boldsymbol{x}^\star(\boldsymbol{r}) \rangle =: UB,$$

where the last inequality follows from the fact that $\boldsymbol{x}^\star(\boldsymbol{r}_{|\boldsymbol{x}}) \in \mathcal{X}$ is feasible for (1) and the optimality of $\boldsymbol{x}^\star(\boldsymbol{r})$. In other words, the value of the relaxed optimization problem (1) provides a valid upper bound on the value of (4). This result is intuitive: ignoring the effect of our collection on future plastic collection leads to an optimistic estimate of the plastic we can actually collect. In addition, $\boldsymbol{x}^\star(\boldsymbol{r}) \in \mathcal{X}$ is feasible for (4) so

$$\langle \boldsymbol{r}_{|\boldsymbol{x}^\star(\boldsymbol{r})}, \boldsymbol{x}^\star(\boldsymbol{r}) \rangle \leq \langle \boldsymbol{r}_{|\boldsymbol{x}^\star(\boldsymbol{r}_{|\boldsymbol{x}})}, \boldsymbol{x}^\star(\boldsymbol{r}_{|\boldsymbol{x}}) \rangle.$$

Note that the reward vectors in the two sides of this inequality are different because they correspond to different paths $\boldsymbol{x}^\star(\boldsymbol{r})$ and $\boldsymbol{x}^\star(\boldsymbol{r}_{|\boldsymbol{x}})$. Hence, $\boldsymbol{x}^\star(\boldsymbol{r})$ also provides a valid lower bound on the value of (4). Altogether, Algorithm 1 can be used to return a feasible solution to Problem (4) alongside an optimality gap, which, according to Proposition 2, scales quadratically with T and linearly with α.

Furthermore, any other feasible solution $\boldsymbol{x} \in \mathcal{X}$ provides a valid, and potentially better, solution. In Algorithm 2, we propose to search for improved solutions by leveraging not only the solution provided by Algorithm 1 but its partitioning of the feasible space by terminal states. Formally, in its final stage, Algorithm 1 partitions the space of trajectories $\mathcal{X}$ according to the state they reach at time T. For a state s, Algorithm 1 returns the length (according to $\boldsymbol{r}$) of the longest path reaching s, $V^T(s)$, as well as one path of that length, $\boldsymbol{x}^s$. Accordingly, we use $V^T(s)$ to select a few (K) high-potential terminal states s. For each of these states, we consider its associated trajectory $\boldsymbol{x}^s$ and compute its path-dependent rewards $\boldsymbol{r}_{|\boldsymbol{x}^s}$, which is the most expensive step in Algorithm 2 ($\star$). In Appendix D.1, we describe and compare different strategies for selecting K high-potential terminal states. In our final implementation, we geographically cluster the space of terminal states into K regions and consider one candidate trajectory per cluster, as represented in Figure 6.

Algorithm 2, which we call the adjusted DP algorithm, can be interpreted as a relaxation-induced search (Danna et al. 2005) or the root node analysis in a branch-and-bound algorithm. We use the value of the relaxed problem (1) as an upper bound on the final value and efficiently search for high-quality feasible solutions to obtain a lower bound, hence the term ‘search-and-bound’. In Appendix D.3, we describe a tailored branch-and-bound algorithm that can be used to further tighten the quality of the upper bound and eventually solve (4) to provable optimality. However, the branch-and-bound trees it generates are highly imbalanced so we expect a very slow convergence in practice.

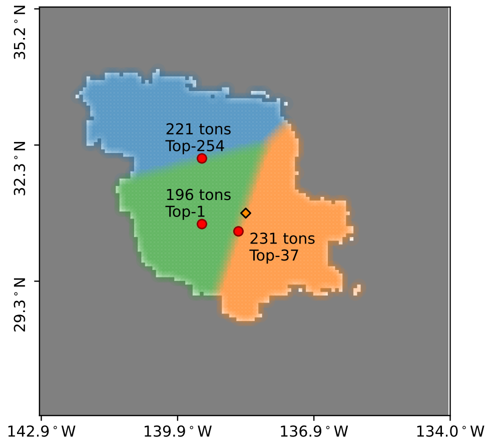


Figure 6 Illustration of our cluster-based search strategy on one 7-day planning instance (Jan 15, 2002). We first exclude locations that are provably suboptimal (grey region), cluster the remaining area into K clusters (here, $K = 3$), and then evaluate one trajectory per cluster (red circle dots). The starting location is indicated by a diamond at the center of the map.

Algorithm 2: Search-and-Bound Algorithm for Problem (4)

Data: Weighted graph $\mathcal{G}$ with dynamic plastic density estimates $\boldsymbol{r}^0$, $\{\boldsymbol{Q}^t\}_{t\in\mathcal{T}}$;

Initialize: Compute $\boldsymbol{r}^t$ for all $t \in \mathcal{T}$ according to (2);

Stage 1: Run Algorithm 1, obtain values $V^T(s)$ and solutions path$[s]$ for $s \in \mathcal{S}_T$;

Stage 2: Search-and-bound;

Initialize upper bound $UB = \max_{s\in\mathcal{S}_T} V^T(s)$;

Initialize lower bound $LB = 0$, best solution $\hat{\boldsymbol{x}} = \boldsymbol{0}$;

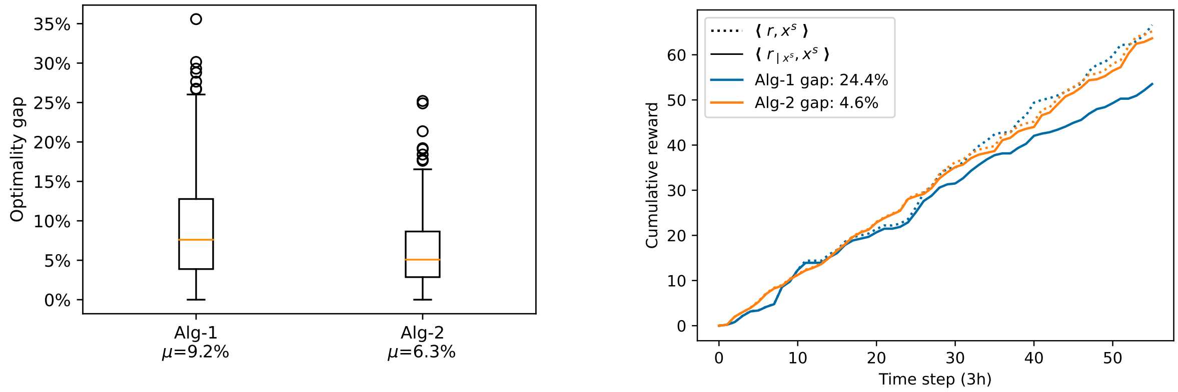
for *some specific states* $s \in \mathcal{S}_T$ **do**
  Get the solution $\boldsymbol{x}^s \leftarrow$ path$[s]$;
  ($\star$) Compute the path-adjusted reward $\boldsymbol{r}_{|\boldsymbol{x}^s}$;
  if $\langle \boldsymbol{r}_{|\boldsymbol{x}^s}, \boldsymbol{x}^s\rangle > LB$ **then**
    Update lower bound $LB = \langle \boldsymbol{r}_{|\boldsymbol{x}^s}, \boldsymbol{x}^s\rangle$;
    Update best solution $\hat{\boldsymbol{x}} = \boldsymbol{x}^s$;
  end
end

return the solution $\hat{\boldsymbol{x}}$ and optimality gap $(UB - LB)/UB$.

4.5. Numerical validation

The adjusted DP Algorithm 2 improves upon taking the solution provided by Algorithm 1 directly by searching for a better solution in (some) other terminal states $s \in \mathcal{S}_T$.

To assess the benefit from conducting this additional search, we use some of the weather and plastic ocean data described in Section 5.1. and consider 250 instances of the 7-day routing problem.



(a) Optimality gaps from 250 simulations (b) Cumulative rewards (single simulation)

Figure 7 Simulation results comparing the performance of Algorithm 1 and Algorithm 2 (with $\alpha = 0.2$).

We generate the instances by considering 50 different starting time (the beginning of each week, except for the first and last week of the year) and 5 different initial locations. Figure 7a represents the distribution (box plot) of the optimality gaps achieved by Algorithm 1 and Algorithm 2. We observe that Algorithm 1 achieves an average optimality gap of 9.2%, compared with 6.3% for Algorithm 2—a 2.9 percentage points or 32% improvement. Note that, since both approaches consider the same upper bound, this improvement is only due to an improved lower bound, i.e., Algorithm 2 finding a higher-quality solution. In addition, evidence on small-scale experiments (see Appendix D.2) suggests that the solution returnd by Algorithm 2 is often closer to optimality than what is suggested by the optimality gap due to the upper bound not being tight.

To illustrate the benefit of our algorithm, we focus on one of these 250 instances and represent on Figure 7b the quantity of plastic collected over the 7 days (56 time periods) by the solution of Algorithm 1 (blue lines) and Algorithm 2 (orange lines), both when the density vectors are upper bounded by $\langle \boldsymbol{r}, \boldsymbol{x} \rangle$ (path-independent, dashed lines) and adjusted for path-dependency (solid lines). By comparing the dashed lines, we observe that the solution returned by Algorithm 2 is indeed sub-optimal for the path-independent problem. However, when adjusting for path dependency, the reward associated with solution of Algorithm 1 is significantly lower than expected (especially for time periods $t \geq 27$), much more than for the solution of Algorithm 2. Eventually, the solution provided by Algorithm 2 performs significantly better—here, the optimality gap reduces from 24.4% to 4.6%.

5. Numerical Experiment

In this section, we evaluate our method on one-year weather and plastic density data. After presenting our experimental setting in Section 5.1, we compare the performance of different implementations of our algorithm in Section 5.2. In Section 5.3, we delve deeper into the differences in plastic collection efficiency across seasons. Finally, we use our algorithm to investigate the non-linear relationships between some strategy system design decisions (such as total span or size of the retention zone) on the overall efficiency in Section 5.4.

5.1. Experimental setting

We work with one year of weather and plastic density data (year 2002 in our data set). The weather data provides the height and direction of the waves and the wind. The plastic density data is provided as trajectories of a particle-based fluid dispersal model, as described in Appendix C. The level of spatial (resp. time) granularity of the data is 8 km (resp. 3 hours), in line with our discretization. Given a span of 1.8 km, we consider a plastic collection rate $\alpha = 20\%$. The capacity of the retention zone is fixed to 25 metric tons. We assume that an extraction takes one day to complete (8 time periods). The wave height needs to be below 2.5 meters during the first 6 hours for the crane to operate and needs to be below 3.5 meters during at least 12 arbitrary hours for workers to sort the plastic.

We divide the year into 13 non-overlapping 28-day periods[2], which we later refer to as 'simulations'. For each simulation, we assume the system starts in the center of the GPGP, whose coordinates are (31.92°N, 142.4°W). Given the propelling speed, the system is able to reach any point on the boundary of the GPGP from its initial location within 31 days.

We compare the performance of our algorithms with that of The Ocean Cleanup's current heuristic. For each direction d, we let $\mathcal{L}_t(\ell, d)$ denote the set of locations that are reachable in t steps from ℓ when looking in the direction of d (here, we associate each direction d with a cone corresponding to all steering angle within $\pm 22.5°$ of d). We associate each direction with a value:

$$\sum_{t \in [T]} \frac{1}{|\mathcal{L}_t(\ell, d)|} \sum_{\ell' \in \mathcal{L}_t(\ell, d)} \frac{r_{\ell'}^t}{\|\ell' - \ell\|_2^2},$$

[2] Since each 4-week simulation requires 5 weeks of data (because some of our optimization algorithms are forward-looking and take into account information of the next 7 days), we would need 53 weeks to conduct 13 non-overlapping 28-day simulations. Instead, we start the last experiment one week earlier (on day 329 instead of 336).

and pick the best feasible direction according to this criterion. In other words, the benchmark goes in the direction leading to the highest distance-weighted reward. The benchmark does not include any rule for extraction scheduling so we start an extraction as soon as the system reaches capacity. If the current weather does not allow it, we replace rewards with wave heights in the above formula and steer towards the lowest distance-weighted wave height to start an extraction.

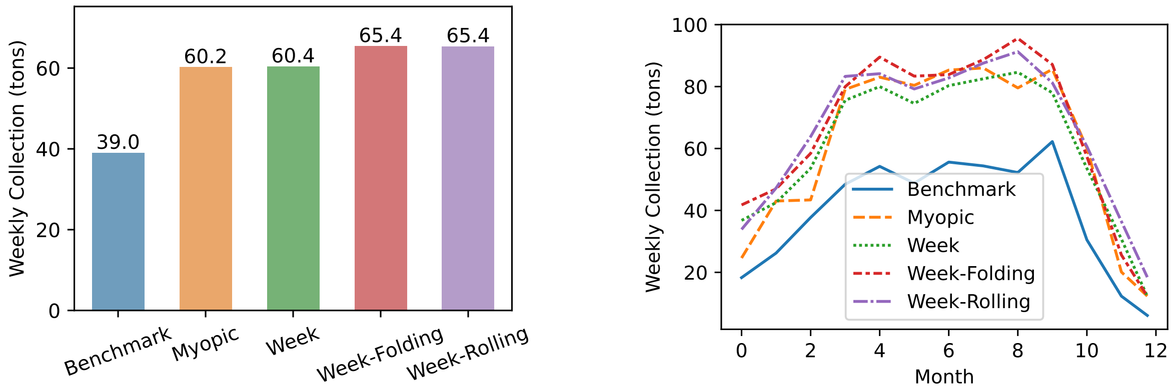
We evaluate the performance of four different implementations of Algorithm 2, with different optimization and implementation horizons: First, we consider solving our longest path problem (4) for $T = 8$ time periods only (one day), by solving and implementing the resulting solution (routing and extraction scheduling) every day. We refer to this implementation as **Myopic**. To be more forward looking, we consider using $T = 56$ time periods instead (one week). At the beginning of each week, we run Algorithm 2, obtain a solution, and implement it for the following 7 days (**Week**). In these two variants, the planning horizon (used in the definition of the optimization problem) and the implementation horizon are the same. However, they do not have to be. For example, one can optimize over the next 7 days ($T = 56$) but only implement the first 8 time periods of the solution. In particular, we can use a rolling horizon by solving each day a 7-day longest path problem (4) and implementing its solution for the first day (**Week-Rolling**). Alternatively, to navigate the system during a week, we can use a folding (or shrinking) horizon by finding the longest path 'until the end of the week' (i.e., over 7 days in the beginning, 6 days after the first day,...) and implementing its solution for the first day (**Week-Folding**).

Note that before resolving an optimization problem—which happens every day for the Myopic, Week-Rolling and Week-Folding implementations, and every week for Week—we update our plastic density estimates based on the past-day trajectory and collection of the system. Hence, resolving can also be seen as a way to mitigate path-dependency issues described in Section 4. We decided not to consider optimization models over horizons longer than 7 days for two reasons: unavailability of reliable weather forecast for more-than-7-day-ahead predictions and computational considerations.

5.2. Overall improvement in plastic collection rate

Figure 8a represents the weekly quantity of plastic collected by each method, averaged over our 13 four-week simulations.

Based on these results, we make the following observations: First, Figure 8a illustrates the edge of optimization, with all methods significantly improving over the benchmark—as identified formally



(a) Results averaged over the 13 simulations (b) Results per simulation

Figure 8 Weekly quantity of plastic collected, for the benchmark and each of the optimization-based approaches

by paired t-test with $\leq 10^{-5}$ p-values in Table E.1. Among all methods, Week-Rolling collects the most plastic (65.4 tons/week), which is around 1.67 times more than the benchmark (39.0 tons/week). Among all optimization-based approaches, Myopic and Week perform the worst and comparably. This suggests that the benefit of being forward-looking of the Week implementation is counter balanced by the path dependency issue (to which Week is more sensitive because it re-optimizes every week only). Accordingly, methods that re-optimize every day *and* consider a longer planning horizon, namely Week-Folding and Week-Rolling, have a clear edge.

While the performances reported in Figure 8a are averaged over the 13 simulations, Figure 8b reports the performance for each of the 13 simulations, where each simulation is identified by its start date (in month). Figure 8b confirms the main findings from Figure 8a but also exhibits an interesting behavior with respect to month: We observe that the quantity of plastic collected (by any method) is higher in summer months (April-August) than in winter months (November-February) and we observe that the relative benefit from our optimization methods (compared with the benchmark) is higher during these winter months. We investigate the mechanisms driving this pattern in the coming section.

5.3. Heterogeneity across seasons and the impact of extraction scheduling

Figure 8b raises the question of the impact of season (or time of the year) on the plastic collection efficiency. We should emphasize from the start that the behavior we observe is not driven by differences in the overall plastic density during the year. Values of plastic density continuously increase over the year (roughly by 10% in our 2002 data, by around 2% nowadays) but do not exhibit this inverted U-shape (see Figure E.3a in Appendix E.2).

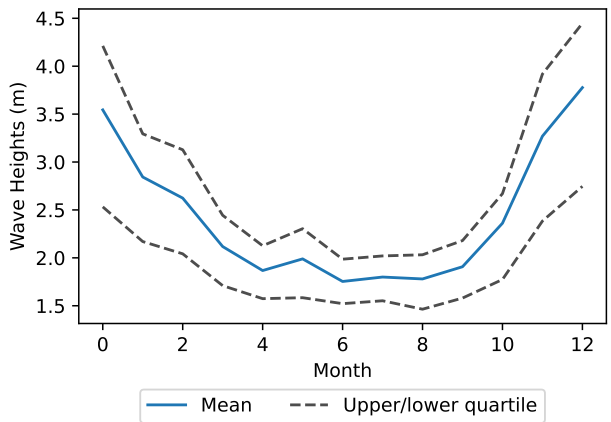
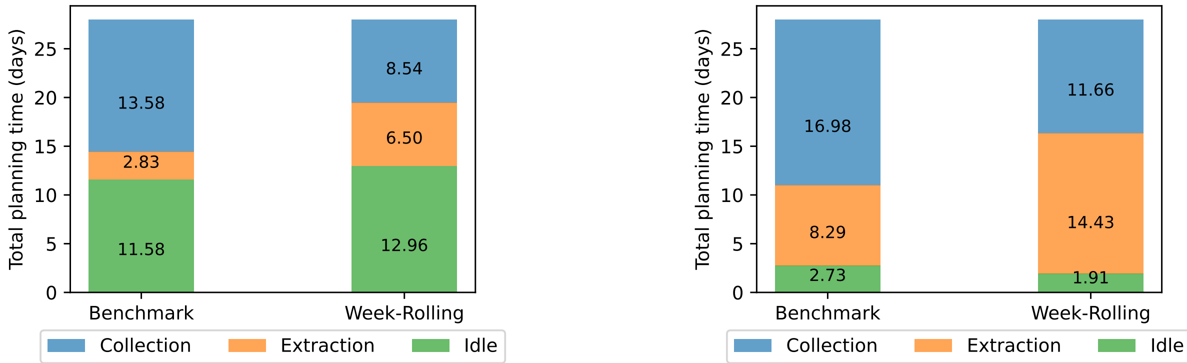


Figure 9 Average wave height in the GPGP across the year 2002

If plastic density (i.e., the objective of our optimization problem) cannot explain this behavior, it is natural to consider the impact of weather (which drives most of the operational constraints) on the heterogeneous performance across seasons. Figure 9 represents the average wave height in the GPGP (together with upper and lower quartiles in dashed lines) for each month. We observe that wave height follows the same pattern as the quantity of plastic collected, with lower waves experienced in the middle of the year (April-August) and higher waves during November-February. High waves affect the collection process in two ways. First, the system cannot operate when the wave height exceeds 6 meters. Hence, the average 'collectable' plastic density is much lower in winter than in summer (see Figure E.3b in Appendix E.2). Furthermore, extractions require the waves to be below 2.5 meters for the first 6 hours, and below 3.5 meters for 12 hours. Hence, weather and its impact on the feasibility of extractions also contributes to the behavior observed in Figure 8b.

To confirm this intuition, we quantify the time spent by the system while waiting to extract. At any point in time, the system can be in one of three phases: it can be actively collecting plastic, it can be undergoing an extraction, or it can be idle (i.e., unable to collect plastic because it has reached capacity but unable to start an extraction either because of weather). For each 4-week simulation, we compute the number of days the system spent in collecting, extracting (which is equivalent to the number of extractions performed), and staying idle. For each simulation, the above three numbers should add up to 28 days. We averaged these numbers per season (winter/summer), where we define 'winter' as the first three and last three simulations and 'summer' as the remaining 7 ones. Figure 10 reports these metrics for the benchmark and Week-Rolling methods in both winter months (Figure 10a) and summer months (Figure 10b).



(a) Winter

(b) Summer

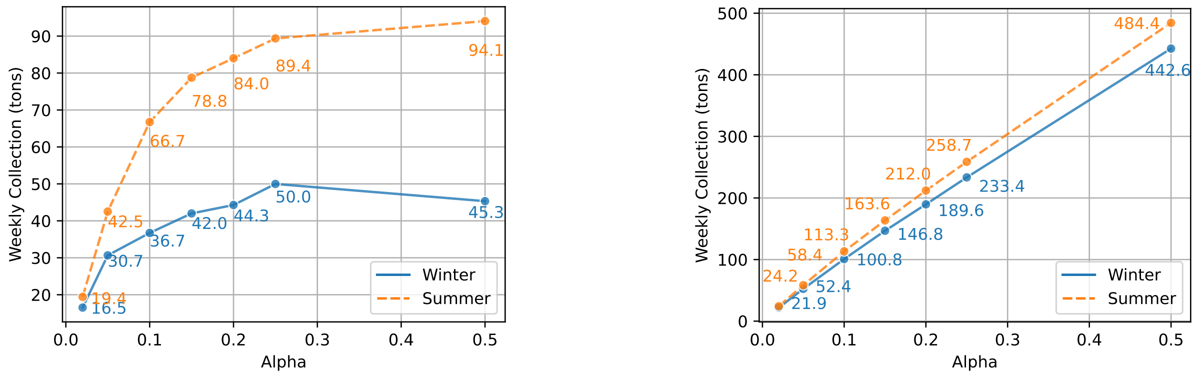
Figure 10 Number of days spent on collection, extraction and idle per 4-week simulation, for the benchmark and Week-Rolling methods. Results are aggregated over the winter (left panel) and summer (right panel) months.

First, we observe that, in both winter and summer, Week-Rolling spends less time collecting than the benchmark. Given that Week-Rolling collects more plastic (67% more on average), this indicates that our approach is more efficient: it collects more in less time. Comparing Figure 10a and Figure 10b, we observe that idle time is significantly higher in winter, confirming the fact that weather conditions limit the ability to extract (hence, to collect further) during winter. Surprisingly, Week-Rolling does not reduce total idle time in winter (around 12–13 days out of 28 for both method). However, Week-Rolling performs twice as many extractions, around 6.5 times on average compared with 2.8 times for the benchmark (which aligns with the increase in quantity of plastic collected), so Week-Rolling experiences a much lower idle time per extraction than the benchmark.

The above observations highlight the importance of jointly finding a collection route and a schedule for the extractions for overall efficiency. On this matter, our optimization approach that can explicitly account for weather-related constraints experiences greater benefits in the winter, when the ability to extract constitutes the main bottleneck.

5.4. Designing a new system: is bigger better?

In Section 5.2, we show that with optimization, we can improve the collection speed from 40 tons/week to 60+ tons/week, using their current system. In this section, we use our optimization model to help answer strategic dimensioning decision for the next-generation system. The Ocean Cleanup was first considering increasing the span of the system from $0.6 \sim 0.8$ km ($\alpha \approx 0.1$) to $1.6 \sim 1.8$ km ($\alpha \approx 0.2$), *without increasing the size of the retention zone (25 tons)*. Indeed, the size



(a) System capacity 25 tons (b) System infinite capacity (no extraction)

Figure 11 Weekly quantity of plastic collected by DP (Rolling) under different α. Results are aggregated over the winter (blue solid line) and summer (orange dashed line) months, and with (left panel) or without (right panel) extraction.

of the retention zone is partially constrained in practice by the size of the ship used to store and sort the plastic collected.

Figure 11 represents the weekly quantity of plastic collected in winter (solid blue lines) and summer (dash orange lines), for increasing values of α, in the case of a 25-ton capacity (left panel, Figure 11a) and an infinite capacity (right panel, Figure 11b). Without capacity constraints from the retention zone, one expects the total quantity of plastic collected to depend linearly in the span of the system α, as displayed in Figure 11b. However, with a finite capacity (Figure 11a), we observe (i) an overall lower quantity of plastic collected (which is due to the need to extract and the fact that we stop collection during extraction), and (ii) a strong concave dependency of the plastic collected on α. Indeed, by doubling the span size from $\alpha = 0.1$ to $\alpha = 0.2$, the weekly collection increases by 26% in summer (from 66.7 to 84.0 tons/week) and by 20% only in winter (from 36.7 to 44.3 tons/week). Moreover, increasing the span beyond $\alpha = 0.25$ (i.e., 2 km span) provides barely any improvement.

Intuitively, this different behavior across seasons is due to the fact that a larger span requires more frequent extractions, which are very sensitive to weather conditions. The comparison between Figure 11a and Figure 11b highlights the impact of having a finite-capacity retention zone on the overall performance. In the future, the difficulty to extract could largely erode the benefit of having a larger system. This leads us to the next question: how to design a new system with better extraction?

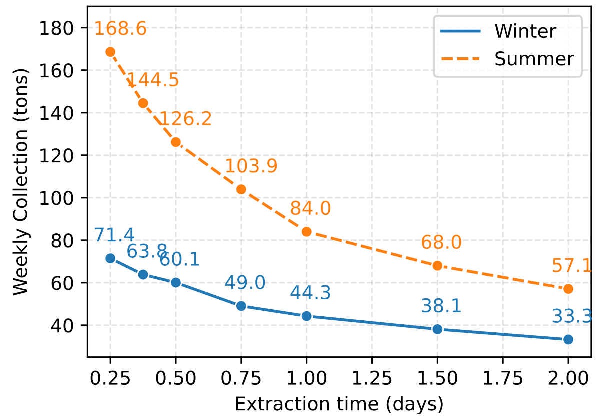


Figure 12 Weekly collections of DP (Rolling) with different unit extraction times. Results are aggregated over the winter (blue solid line) and summer (orange dashed line) months.

There are several ways to improve the current extraction process. One solution could be to reduce the unit extraction time, namely, the time spent per extraction. In practice, this could be achieved via more efficient extraction operations. For example, one could first empty the plastic from the retention zone on the deck (taking approximatively 6 hours), put the system back into the water, and sort the plastic while resuming the plastic collection Another solution could be to increase the total capacity of the retention zone, which we do not discuss in this paper.

Figure 12 represents the weekly quantity of plastic collected in winter (solid blue lines) and summer (dash orange lines), for increasing values of unit extraction time, in the case of a 25-ton capacity. By reducing the time per extraction from 1 day (current practice) to 0.25 days (or 6 hours), the weekly collection increases by 101% in summer (84.0 to 168.6 tons/week) and by 61% in winter (44.3 to 71.4 tons/week). Observe, in comparison, that the potential improvement by further increasing the span size beyond $\alpha = 0.2$ (Figure 11a) is less than 12%. Reducing the unit extraction time demonstrates a greater potential for impact, both in winter and summer.

We emphasize that the above improvement solely comes from a shorter extraction time, not from a lower impact of weather constraints since we kept, in our implementation, the same weather constraints for extraction (described in Appendix A.1) irrespective of the extraction time. In practice, shorter extraction time might also translate into less stringent weather constraints, which could in turn provide additional benefits.

6. Conclusion

Our oceans are being threatened by the millions of tons of plastic that have been emitted over the recent decades. To limit future harm to marine ecosystems and activities, we need to clean up oceans from plastic, as quickly as possible. To this end, we develop a graph-based model and formulate the problem of routing a plastic-collecting system in the GPGP to maximize the quantity of plastic encountered as a longest path problem. However, due to the plastic dynamics and the direct impact of collection on these dynamics, our resulting longest path problem (4) is non-convex and non-separable over edges. To deal with these computational difficulties, we propose to relax the reward dynamics and solve large-scale instances of this relaxation in linear time using a DP algorithm. Then, we obtain near-optimal solutions to our original problem, together with certificates of near optimality, by building an efficient search algorithm based on the DP algorithm (and not only on its optimal solution).

On one-year weather and plastic density data, we observe that our optimization approaches increase the quantity of plastic collected by 67% compared with the status quo, thus accelerating the path to plastic-free oceans. We also leverage our optimization algorithms to explore the non-linear relationships between system characteristics and system performance. For example, because of difficulties to extract (i.e., empty the capacity of the system) in winter, we find that increasing the span of the system beyond 1.8 km will have barely any impact on collection efficiency in winter.

For our current application, the main concern and area for future research is to account for uncertainty in weather predictions and plastic dynamics. In particular, we are currently investigating whether collection of real-world data by drones or satellites could help quantify uncertainty and lead to robust versions of our longest path problem. More broadly, we are excited to study whether whether the class of longest path problems we identify in (4) could find other applications, as a model for operations with nature dynamics.

References

Appelgren LH (1971) Integer programming methods for a vessel scheduling problem. *Transportation Science* 5(1):64–78.

Aydin N, Lee H, Mansouri SA (2017) Speed optimization and bunkering in liner shipping in the presence of uncertain service times and time windows at ports. *European Journal of Operational Research* 259(1):143–154.

Battarra M, Pessoa AA, Subramanian A, Uchoa E (2014) Exact algorithms for the traveling salesman problem with draft limits. *European Journal of Operational Research* 235(1):115–128.

Bellman R (1958) On a routing problem. *Quarterly of Applied Mathematics* 16(1):87–90.

Björklund A, Husfeldt T, Khanna S (2004) Approximating longest directed paths and cycles. *International Colloquium on Automata, Languages and Programming* 3142:222–233.

Bulterman RW, van der Sommen FW, Zwaan G, Verhoeff T, van Gasteren AJM, Feijen WHJ (2002) On computing a longest path in a tree. *Information Processing Letters* 81(2):93–96.

Chen C, Chen XQ, Ma F, Xiao-Jun Zeng, Zeng XJ, Wang J (2019) A knowledge-free path planning approach for smart ships based on reinforcement learning. *Ocean Engineering* 189:106299.

Cheng Y, Zhang W (2018) Concise deep reinforcement learning obstacle avoidance for underactuated unmanned marine vessels. *Neurocomputing* 272:63–73.

Christiansen M, Fagerholt K, Ronen D (2004) Ship routing and scheduling: Status and perspectives. *Transportation Science* 38(1):1–18.

Coll M, Pennino MG, Steenbeek J, Solé J, Bellido JM (2019) Predicting marine species distributions: Complementarity of food-web and Bayesian hierarchical modelling approaches. *Ecological Modelling* 405:86–101.

Cormen TH, Leiserson CE, Rivest RL, Stein C (2022) *Introduction to Algorithms* (MIT press).

Danna E, Rothberg E, Pape CL (2005) Exploring relaxation induced neighborhoods to improve mip solutions. *Mathematical Programming* 102:71–90.

Dantzig GB (1960) On the shortest route through a network. *Management Science* 6(2):187–190.

de Vries R, Egger M, Mani T, Lebreton L (2021) Quantifying floating plastic debris at sea using vessel-based optical data and artificial intelligence. *Remote Sensing* 13(17):3401.

de Wit C (1990) Proposal for low cost ocean weather routeing. *The Journal of Navigation* 43(3):428–439.

Dijkstra EW (2022) A hidenote on two problems in connexion with graphs. *Edsger Wybe Dijkstra: His Life, Work, and Legacy*, 287–290.

Dijkstra H, van Beukering P, Brouwer R (2021) In the business of dirty oceans: Overview of startups and entrepreneurs managing marine plastic. *Marine Pollution Bulletin* 162:111880.

Fagerholt K, Christiansen M (2000) A travelling salesman problem with allocation, time window and precedence constraints — an application to ship scheduling. *International Transactions in Operational Research* 7(3):231–244.

Ford Jr LR (1956) Network flow theory. Technical report, Rand Corp Santa Monica Ca.

Gall SC, Thompson RC (2015) The impact of debris on marine life. *Marine Pollution Bulletin* 92(1-2):170–179.

Geyer R, Jambeck JR, Law KL (2017) Production, use, and fate of all plastics ever made. *Science Advances* 3(7):e1700782.

Granado I, Hernando L, Galparsoro I, Gabiña G, Groba C, Prellezo R, Fernandes JA (2021) Towards a framework for fishing route optimization decision support systems: Review of the state-of-the-art and challenges. *Journal of Cleaner Production* 320:128661.

Groba C, Sartal A, Bergantiño G (2020) Optimization of tuna fishing logistic routes through information sharing policies: A game theory-based approach. *Marine Policy* 113:103795.

Groba C, Sartal A, Vázquez XH (2015) Solving the dynamic traveling salesman problem using a genetic algorithm with trajectory prediction: An application to fish aggregating devices. *Computers & Operations Research* 56:22–32.

Groba C, Sartal A, Vázquez XH (2018) Integrating forecasting in metaheuristic methods to solve dynamic routing problems: Evidence from the logistic processes of tuna vessels. *Engineering Applications of Artificial Intelligence* 76:55–66.

Hagiwara H, Spaans JA (1987) Practical weather routing of sail-assisted motor vessels. *The Journal of Navigation* 40(1):96–119.

Iglesias A, Dafonte C, Arcay B, Cotos JM (2007) Integration of remote sensing techniques and connectionist models for decision support in fishing catches. *Environmental Modelling and Software* 22(6):862–870.

Ioannidou K, Nikolopoulos SD (2013) The longest path problem is polynomial on cocomparability graphs. *Algorithmica* 65(1):177–205.

James RW (1957) *Application of Wave Forecasts to Marine Navigation* (New York University).

Jones MC, Dye SR, Pinnegar JK, Warren R, Cheung WWL (2012) Modelling commercial fish distributions: Prediction and assessment using different approaches. *Ecological Modelling* 225:133–145.

Kaandorp MLA, Lobelle D, Kehl C, Dijkstra HA, van Sebille E (2023) Global mass of buoyant marine plastics dominated by large long-lived debris. *Nature Geoscience* ISSN 1752-0908, URL `http://dx.doi.org/10.1038/s41561-023-01216-0`.

Karger D, Motwani R, Ramkumar GDS (1997) On approximating the longest path in a graph. *Algorithmica* 18(1):82–98.

Karp RM (2010) *Reducibility Among Combinatorial Problems* (Springer).

Klink D, Peytavin A, Lebreton L (2022) Size dependent transport of floating plastics modeled in the global ocean. *Frontiers in Marine Science* 9:903134.

Kosmas OT, Vlachos D (2012) Simulated annealing for optimal ship routing. *Computers & Operations Research* 39(3):576–581.

Langbein J, Stelzer R, Frühwirth T (2011) A rule-based approach to long-term routing for autonomous sailboats. *Robotic Sailing: Proceedings of the 4th International Robotic Sailing Conference*, 195–204 (Springer).

Lebreton L, Slat B, Ferrari FF, Sainte-Rose B, Aitken J, Marthouse R, S Hajbane, Hajbane S, Cunsolo S, Schwarz A, Levivier A, Noble K, Debeljak P, Maral H, Schoeneich-Argent R, Brambini R, Reisser J (2018) Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic. *Scientific Reports* 8(1):4666–4666.

Lebreton L, Van Der Zwet J, Damsteeg JW, Slat B, Andrady AL, Reisser J (2017) River plastic emissions to the world's oceans. *Nature Communications* 8(1):15611–15611.

Li WC, Tse H, Fok L (2016) Plastic waste in the marine environment: A review of sources, occurrence and effects. *Science of the Total Environment* 566:333–349.

Madraki G, Judd RP (2019) Recalculating the length of the longest path in perturbed directed acyclic graph. *IFAC-PapersOnLine* 52(13):1560–1565.

Malaguti E, Martello S, Santini A (2018) The traveling salesman problem with pickups, deliveries, and draft limits. *Omega* 74:50–58.

Meijer LJ, Van Emmerik T, Van Der Ent R, Schmidt C, Lebreton L (2021) More than 1000 rivers account for 80% of global riverine plastic emissions into the ocean. *Science Advances* 7(18):eaaz5803.

Meng Q, Wang T (2011) A scenario-based dynamic programming model for multi-period liner ship fleet planning. *Transportation Research Part E: Logistics and Transportation Review* 47(4):401–413.

Moore EF (1959) The shortest path through a maze. *Proceedings of International Symposium on the Theory of Switching, 1959*, 285–292.

Pandit SN (1962) Some observations on the routing problem. *Operations Research* 10(5):726–727.

Park YJ, Garaba S, Sainte-Rose B, Han HJ (2022) Satellite remote sensing of marine litter floating in open ocean and coastal waters. *Living Planet Symposium 2022* (The European Space Agency (ESA)).

Park YJ, Garaba SP, Sainte-Rose B (2021) Detecting the great pacific garbage patch floating plastic litter using worldview-3 satellite imagery. *Optics Express* 29(22):35288–35298.

Parra H, Pham CK, Menezes GM, Rosa A, Tempera F, Morato T (2017) Predictive modeling of deep-sea fish distribution in the Azores. *Deep Sea Research Part II: Topical Studies in Oceanography* 145:49–60.

Peytavin A, Sainte-Rose B, Forget G, Campin JM (2021) Ocean plastic assimilator v0. 2: assimilation of plastic concentration data into lagrangian dispersion models. *Geoscientific Model Development* 14(7):4769–4780.

Pollack M, Wiebenson W (1960) Solutions of the shortest-route problem-a review. *Operations Research* 8(2):224–230.

Robinson NM, Nelson WA, Costello MJ, Sutherland JE, Lundquist CJ (2017) A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. *Frontiers in Marine Science* 4:421.

Sainte-Rose B, Rakotonirina AD, van den Bremer T, Pham Y (2022) Numerical simulation of the wave-induced drift of floating marine plastic debris modeled as discrete particles. *International Conference on Offshore Mechanics and Arctic Engineering*, volume 85925, V007T08A008 (American Society of Mechanical Engineers).

Schmidt C, Krauth T, Wagner S (2017) Export of plastic debris by rivers into the sea. *Environmental Science & Technology* 51(21):12246–12253.

Schnurr RE, Alboiu V, Chaudhary M, Corbett RA, Quanz ME, Sankar K, Srain HS, Thavarajah V, Xanthos D, Walker TR (2018) Reducing marine pollution from single-use plastics (sups): A review. *Marine Pollution Bulletin* 137:157–171.

Schrijver A (2012) On the history of the shortest path problem. *Documenta Mathematica* 17(1):155–167.

Sen D, Padhy CP (2015) An approach for development of a ship routing algorithm for application in the North Indian Ocean region. *Applied Ocean Research* 50:173–191.

Shao W, Zhou P, Thong SK (2012) Development of a novel forward dynamic programming method for weather routing. *Journal of Marine Science and Technology* 17(2):239–251.

Shimbel A (1954) Structure in communication nets. *Proceedings of the Symposium on Information Networks*, 119–203 (Polytechnic Institute of Brooklyn).

Skoglund L (2012) A new method for robust route optimization in ensemble weather forecasts. Master thesis, KTH, School of Engineering Sciences (SCI).

Sniedovich M (2006) Dijkstra's algorithm revisited: the dynamic programming connexion. *Control and Cybernetics* 35(3):599–620.

Stalhane M, Hvattum LM, Skaar V (2015) Optimization of routing and scheduling of vessels to perform maintenance at offshore wind farms. *Energy Procedia* 80:92–99.

Stokes GG (1847) On the theory of oscillatory waves. *Transactions of the Cambridge Philosophical Society* 8:441–455.

Takashima K, Mezaoui B, Shoji R (2009) On the fuel saving operation for coastal merchant ships using weather routing. *International Journal on Marine Navigation and Safety of Sea Transportation* 3(4):401–406.

Taylor GI (1915) I. Eddy motion in the atmosphere. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 215(523-537):1–26.

Ting SC, Tzeng GH (2003) Ship scheduling and cost analysis for route planning in liner shipping. *Maritime Economics & Logistics* 5(4):378–392.

Uehara R, Uno Y (2007) On computing longest paths in small graph classes. *International Journal of Foundations of Computer Science* 18(05):911–930.

Vettor R, Tadros M, Ventura M, Guedes Soares C (2016) Route planning of a fishing vessel in coastal waters with fuel consumption restraint. *Maritime Technology and Engineering* 3:167–173.

Wilcox C, Van Sebille E, Hardesty BD (2015) Threat of plastic pollution to seabirds is global, pervasive, and increasing. *Proceedings of the National Academy of Sciences* 112(38):11899–11904.

Winterstetter A, Grodent M, Kini V, Ragaert K, Vrancken KC (2021) A review of technological solutions to prevent or reduce marine plastic litter in developing countries. *Sustainability* 13(9):4894.

Yoon H, Nguyen V, Nguyen T (2018) Development of solution for safe ship considering seakeeping performance. *TransNav, International Journal on Marine Navigation and Safety od Sea Transportation* 12(3):517–525.

Zheng J, Zhang H, Yin L, Liang Y, Wang B, Li Z, Song X, Zhang Y (2019) A voyage with minimal fuel consumption for cruise ships. *Journal of Cleaner Production* 215:144–153.

Zis T, Psaraftis HN, Ding L (2020) Ship weather routing: A taxonomy and survey. *Ocean Engineering* 213:107697.

Zoppoli R (1972) Minimum-time routing as an N-stage decision process. *Journal of Applied Meteorology (1962-1982)* 11(3):429–435.

This appendix will be published as a separate Electronic Companion.

Appendix A: Details on the Graph-Based Optimization Formulation

A.1. List of all operational and weather constraints

In Table A.1, we summarize all the operational and weather constraints The Ocean Cleanup encountered in their practice.

Constraints	Description
Speed	The propelling speed cannot exceed 1.5 knots.
Angle	The system can turn at most $45°$ within 3 hours.
Wave (direction)	If wave height > 4.5m, head against the direction of the wave.
Wave (screen)	If wave height > 6m, the screen cannot work, no plastic is collected.
Wave (extraction 1)	For the first 6 hours of each extraction, wave height ≤ 2.5m.
Wave (extraction 2)	For the remaining 18 hours of extraction, at least 12 hours with height ≤ 3.5m.
Wind (extraction)	If wind speeds > 25 knots, the crane cannot work.

Table A.1 Operational and weather constraints for The Ocean Cleanup's system 002/B

A.2. Extended graph representation for modeling extraction scheduling decisions

In Section 3.1, we presented a graph-based modeling to describe the trajectory of the system as a path in an appropriate graph. In this section, we expand our description of the state of the system to incorporate the decision of extraction scheduling, while maintaining our efficient longest path interpretation. As described in Section 1.1, the plastic collected accumulates in a retention zone with fixed capacity and that needs to be emptied regularly. In addition, this emptying process, called extraction, requires the weather condition to be favorable (in particular, wave height need to be below 2.5 meters). Accordingly, we need to enrich the state space to include the type of activity the system is performing (collection or extraction) as well as an indication of the load of the retention zone.

The system can perform two types of activity: it can either be collecting plastic or extracting plastic from the retention zone. Actually, because the system cannot extract when weather conditions are unfavorable, the system can also be idle, i.e., not actively collecting because the retention zone is full but not extracting either. Accordingly, we define five possible activities, denoted $\mathcal{V} = \{v_1^c, v_\emptyset^i, v_1^i, v_\emptyset^e\}$, where the subscript indicates the propelling speed of the system ($\emptyset$ for not moving and 1 for moving at normal speed) and the superscript indicates the activity (*c*ollection, *i*dle, and *e*xtraction). Here, we assume that the system does not move during an extraction (because of operations). However, when idle, it can either stay at the same location and wait for the weather to improve or move towards an area with nicer weather. We denote ΔT_e the time required for extraction (in time periods). By construction, we will ensure that the status $v^e_\emptyset$ corresponds to the extraction being finished, since complexity of the extraction process does not provide the flexibility to perform partial extractions. In addition, the time ΔT_e includes extraction and time needed to re-direct the system so the no-sharp-turn constraint does not apply between the steering directions before and after extraction.

In addition to v, we introduce a state dimension to capture the current load of the retention zone. More precisely, we divide the capacity of the retention zone into C increments, corresponding to the typical quantity of plastic collected during one time period. Hence, we can introduce a variable $c \in \mathcal{C} := \{0, 1, \dots, C\}$ capturing load of the retention zone.

With these additional variables, we can describe the state of the system as a 5-tuple $s = (\ell, d, v, c, t) \in \mathcal{L} \times \mathcal{D} \times \mathcal{V} \times \mathcal{C} \times \mathcal{T}$ and we can represent the set of all possible joint trajectories and extraction schedules as a path in a graph, whose nodes and edges are constructed recursively thanks to a successor operator. For this state s, an admissible next state $s' = (\ell', d', v', c', t')$ should satisfy the following constraints:

- If the retention zone is empty and the system is currently extracting (and is done emptying), then the system resumes collection and can go in any direction. Formally, if $c = 0$ and $v = v^e_\emptyset$, we have $s' = (\ell', d', v^c_1, 1, t+1)$ where (ℓ', d') satisfies all constraints defining $\mathrm{succ}((\ell, d, t))$ except for the no-sharp-turn constraint.

- If the retention zone is not full and the system is currently collecting, then the system can either continue collecting (and obey the same constraints as those described in Section 3.1) or proactively start an extraction (if feasible). Formally, if $c < C$ and $v = v^c_1$, we have either $s' = (\ell', d', v^c_1, c+1, t+1)$ with $(\ell', d', t+1) \in \mathrm{succ}((\ell, d, t))$ or $s' = (\ell, d, v^e_0, 0, t + \Delta T_e)$ if possible to empty at location ℓ.

- If the retention zone is currently full, then the next state can either be idle or extraction. Formally, if $c = C$, then s' must fall in one of the three cases: $(\ell, d, v^i_\emptyset, C, t+1)$, $(\ell', d', v^i_1, C, t+1)$ with $(\ell', d', t+1) \in \mathrm{succ}((\ell, d, t))$, or $(\ell, d, v^e_\emptyset, 0, t + \Delta T_e)$ if extraction is possible at location ℓ.

As displayed in Figure A.1, compared with the simple model presented in Section 3.1, this extended state space and successor operator allows the system to "jump" some time periods: when deciding to extract, we directly connect the state of the system at the beginning of the extraction with the state of the system after extraction, i.e., ΔT_e time periods later. Nonetheless, the resulting graph

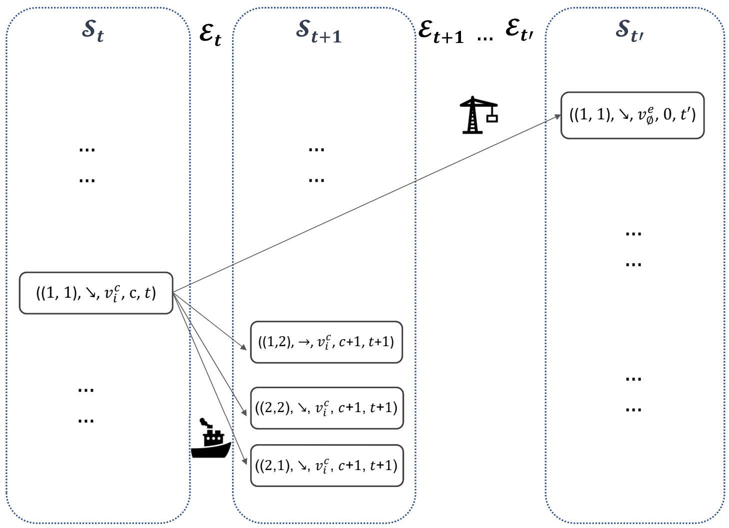


Figure A.1 Successor of a state in a graph that includes both collection and extraction

$\mathcal{G}$ conserves the good properties we leverage in our algorithm: First, $\mathcal{G}$ is a DAG, because all edges are of the form $(s, s') \in \mathcal{S}_t \times \mathcal{S}_{t'}$ with $t' > t$, hence ensuring the correctness of the DP algorithm for longest path problems. Second, for a given state s, its number of successors (or predecessors) is uniformly bounded by a constant.

A.3. Longest path and network flow optimization problems

Mathematically, the longest path problem (1) can be formulated as a mixed-integer optimization problem by introducing binary variables $x_s^t \in \{0,1\}$ indicating whether the system is in state s at time t. The resulting optimization writes as follows:

$$\max_{\boldsymbol{x} \in \mathcal{X}} \quad \sum_{t \in \mathcal{T}} \sum_{s \in \mathcal{S}_t} r_s^t x_s^t,$$

with

$$\mathcal{X} := \left\{ \boldsymbol{x} \in \{0,1\}^{\mathcal{S}} \,\middle|\, \begin{array}{ll} \sum_{s \in \mathcal{S}_t} x_s^t = 1, & \forall t \in \mathcal{T}, \\ \sum_{s' \in \mathrm{succ}(s)} x_{s'}^{t+1} \geq x_s^t, & \forall t \in \mathcal{T}, s \in \mathcal{S}_t \end{array} \right\}.$$

Despite its compactness, we should emphasize that this optimization formulation is not the most computationally efficient because of the integrality constraints. Alternatively, the optimization problem above can be seen as a special case of a network flow optimization problem for which ideal formulations are well known.

Formally, let us now view any sequence of states $\{s_t\}_{t \in \mathcal{T}}$ as binary flow variables $\boldsymbol{f} \in \{0,1\}^{\mathcal{E}}$ where $f_e = 1$ if and only if $e = (s_t, s_{t+1})$ for some $t \in \mathcal{T}$. In other words, $\boldsymbol{f}$ indicates the transitions that the system goes through, while $\boldsymbol{x}$ is encoding the system states. Naturally, one can recover the flow variables from the state variables, and vice versa, via the relations

$$\begin{aligned} f_{s,s'} &= x_s^t x_{s'}^{t+1}, & \forall (s,s') \in \mathcal{E}_t, \\ x_s^t &= \sum_{s' \in \mathrm{succ}(s)} f_{s,s'}, & \forall s \in \mathcal{S}_t. \end{aligned}$$

With these notations, the DP algorithm, Algorithm 1, can solve any optimization problem of the form

$$\max_{\boldsymbol{f} \in \mathcal{F}} \quad \sum_{e \in \mathcal{E}} w_e f_e. \tag{7}$$

where $\mathcal{F}$ is the set of admissible binary flows in $\mathcal{G}$:

$$\mathcal{F} := \left\{ \boldsymbol{f} \in \{0,1\}^{\mathcal{E}} \,\middle|\, \begin{array}{l} \sum_{(s,s') \in \mathcal{E}_0} f_{s,s'} = 1, \\ \sum_{s' : s \in \mathrm{succ}(s')} f_{s',s} = \sum_{s' \in \mathrm{succ}(s)} f_{s,s'} \ \forall t \in \mathcal{T}, s \in \mathcal{S}_t \end{array} \right\}.$$

The first constraint is the initial condition and ensures that the path starts at some state $s \in \mathcal{S}_0$. The second set of constraints corresponds to the common flow conservation constraints, ensuring that, at each node s, inflow equals outflow. Together, they imply that $f_e \leq 1$ for any $e \in \mathcal{E}$. So, the binary constraints can be replaced by an integrality constraint, $\boldsymbol{f} \in \mathbb{Z}_+^{\mathcal{E}}$. Furthermore, it is well known that the polytope of unconstrained flows in a graph is integral, i.e., its extreme points are integral flows. Hence, without loss of optimality, we can assume that the optimization problem (1) occurs over the convex set

$$\mathcal{F} = \left\{ \boldsymbol{f} \in \mathbb{R}_+^{\mathcal{E}} \,\middle|\, \begin{array}{l} \sum_{(s,s') \in \mathcal{E}_0} f_{s,s'} = 1, \\ \sum_{s' : s \in \mathrm{succ}(s')} f_{s',s} = \sum_{s' \in \mathrm{succ}(s)} f_{s,s'} \ \forall t \in \mathcal{T}, s \in \mathcal{S}_t \end{array} \right\}.$$

Appendix B: Analytical analysis of the path-dependent reward

In this section, we supplement the analysis of path-dependent rewards from Section 4.1, by eliciting the dynamics operator $\mathcal{Q}^t$ in (4), the proof of Proposition 2, and deriving the second-order expansion (in α) of the path-dependent reward $\boldsymbol{r}_{|\boldsymbol{x}}$.

B.1. Path dependency for state-based rewards

In our optimization formulations, we use binary variables that are defined over states, $\tilde{\boldsymbol{x}} \in \{0,1\}^{\cup_t \mathcal{S}_t}$, while plastic and reward dynamics are naturally expressed using location indicators $\boldsymbol{x}^t \in \{0,1\}^{\mathcal{L}}$—for clarity, in this section, we will systematically use $\tilde{\cdot}$ to indicate variables defined at a state level. The latter can be recovered from the former via the linear relationship

$$x_\ell^t = \sum_{s' = (\ell', d', t) \in \mathcal{S}_t : \ell' = \ell} \tilde{x}_{s'}^t$$

which we write in matrix form $\boldsymbol{x}^t = \boldsymbol{W}^t \tilde{\boldsymbol{x}}^t$ for some matrix $\boldsymbol{W}^t \in \{0,1\}^{\mathcal{L} \times \mathcal{S}_t}$. With these notations, the reward dynamics (3) can be expressed as a function of $\tilde{\boldsymbol{x}}$ as

$$\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} = \boldsymbol{Q}^t \left[\boldsymbol{r}^t_{|\boldsymbol{x}^{0:t-1}} - \alpha\, \boldsymbol{r}^t_{|\boldsymbol{x}^{0:t-1}} \circ (\boldsymbol{W}^t \tilde{\boldsymbol{x}}^t) \right].$$

We now need to map rewards defined for each locations into rewards defined for each state. For example, if we assign to each state the reward associated with its location, we have $\tilde{\boldsymbol{r}} = (\boldsymbol{W}^t)^\top \boldsymbol{r}$. In practice, we consider a slightly different mapping where the reward for a state equals the reward of its location if the state is collecting, and 0 if the state is idle or extraction. In any case, we have a linear mapping of the form $\tilde{\boldsymbol{r}} = \boldsymbol{V}^t \boldsymbol{r}$ with $\boldsymbol{V}^t \in \{0,1\}^{\mathcal{S}_t \times \mathcal{L}}$ and $\boldsymbol{V}^t \leq (\boldsymbol{W}^{t+1})^\top$. Similarly, we can obtain the reverse mapping $\boldsymbol{r} = \boldsymbol{U}^t \tilde{\boldsymbol{r}}$ with $\boldsymbol{U} \in \{0,1\}^{\mathcal{L} \times \mathcal{S}_t}$, i.e., for each location ℓ, $\boldsymbol{U}^t$ maps ℓ with one state $s \in \mathcal{S}_t$ whose reward is the reward at location ℓ (e.g., a collecting state currently at location ℓ).

Consequently, we are interested in maximizing

$$\sum_{t \in \mathcal{T}} \sum_{s \in \mathcal{S}_t} \tilde{r}^t_{s|\tilde{x}^{0:t-1}} \tilde{x}^t_s,$$

where the path-dependent rewards (defined at a state-level) $\tilde{\boldsymbol{r}}^{t+1}_{|\tilde{\boldsymbol{x}}^{0:t}}$ evolve according to the following dynamics:

$$\tilde{\boldsymbol{r}}^{t+1}_{|\tilde{\boldsymbol{x}}^{0:t}} = \boldsymbol{V}^{t+1} \underbrace{\boldsymbol{Q}^t \left[\boldsymbol{U}^t \tilde{\boldsymbol{r}}^t_{|\tilde{\boldsymbol{x}}^{0:t-1}} - \alpha \left(\boldsymbol{U}^t \tilde{\boldsymbol{r}}^t_{|\tilde{\boldsymbol{x}}^{0:t-1}} \right) \circ (\boldsymbol{W}^t \tilde{\boldsymbol{x}}^t) \right]}_{\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}}} =: \mathcal{Q}^t(\tilde{\boldsymbol{r}}^t_{|\tilde{\boldsymbol{x}}^{0:t-1}}, \tilde{\boldsymbol{x}}^t).$$

Finally, observe that the matrices $\boldsymbol{V}^t$ have non-negative entries. Accordingly, it preserves vector ordering, i.e., $\boldsymbol{r}^1 \leq \boldsymbol{r}^2 \implies \boldsymbol{V}^t \boldsymbol{r}^1 \leq \boldsymbol{V}^t \boldsymbol{r}^2$. Hence, upper and lower bounds on path-dependent rewards in Proposition 2 directly translate to rewards defined for each state.

B.2. Proof of Proposition 2

Proof of Proposition 2 We prove the result by induction. For $t = 0$, $\boldsymbol{r}^1_{|\boldsymbol{x}^{0:0}} = \boldsymbol{r}^1 - \alpha \boldsymbol{Q}^0 (\boldsymbol{r}^0 \circ \boldsymbol{x}^0)$. Hence, in this case, $\boldsymbol{r}^1_{|\boldsymbol{x}^{0:0}}$ is exactly equal to the lower bound in (6). Since $\boldsymbol{Q}^0$, $\boldsymbol{r}^0$, and $\boldsymbol{x}^0$ have non-negative entries, $\boldsymbol{Q}^0 (\boldsymbol{r}^0 \circ \boldsymbol{x}^0) \geq 0$ and the upper bound in (6) holds as well.

Let us now assume that the bounds (6) hold for some $t \geq 0$ and let us show that they also hold for $t+1$.

For the upper bound,

$$\boldsymbol{r}^{t+2}_{|\boldsymbol{x}^{0:t+1}} = \boldsymbol{Q}^{t+1} \boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} - \alpha \boldsymbol{Q}^{t+1} \left(\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} \circ \boldsymbol{x}^{t+1} \right) \leq \boldsymbol{Q}^{t+1} \boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} \leq \boldsymbol{Q}^{t+1} \boldsymbol{r}^{t+1} = \boldsymbol{r}^{t+2}.$$

For the lower bound

$$\begin{aligned}
\boldsymbol{r}^{t+2}_{|\boldsymbol{x}^{0:t+1}} &= \boldsymbol{Q}^{t+1}\left[\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} \circ \left(\mathbf{1} - \alpha \boldsymbol{x}^{t+1}\right)\right] \\
&\geq \boldsymbol{Q}^{t+1}\left[\left(\boldsymbol{r}^{t+1} - \alpha \sum_{0\leq t_1 \leq t} \left(\boldsymbol{Q}^{t} \times \cdots \times \boldsymbol{Q}^{t_1}\right)\left(\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}\right)\right) \circ \left(\mathbf{1} - \alpha \boldsymbol{x}^{t+1}\right)\right] \\
&= \boldsymbol{Q}^{t+1}\boldsymbol{r}^{t+1} - \alpha \boldsymbol{Q}^{t+1} \sum_{0\leq t_1 \leq t} \left(\boldsymbol{Q}^{t} \times \cdots \times \boldsymbol{Q}^{t_1}\right)\left(\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}\right) - \alpha \boldsymbol{Q}^{t+1}\left(\boldsymbol{r}^{t+1} \circ \boldsymbol{x}^{t+1}\right) + \alpha^2 \underbrace{\cdots}_{\geq 0} \\
&\geq \boldsymbol{r}^{t+2} - \alpha \sum_{0 \leq t_1 \leq t+1} \boldsymbol{Q}^{t+1} \times \left(\boldsymbol{Q}^{t} \times \cdots \times \boldsymbol{Q}^{t_1}\right)\left(\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}\right),
\end{aligned}$$

which concludes the proof. $\square$

B.3. Second-order expansion of the path-dependent reward

Proposition 1 provides the first-order expansion (in α) of $\boldsymbol{r}_{|\boldsymbol{x}}$ in the general case. Compared with the case of static plastic (Lemma 1), the first-order term involves cumulative products of $\boldsymbol{Q}^t$'s to account for plastic movements over time. In this section, for any $s \geq t$, we denote $\boldsymbol{Q}^{s:t} := \boldsymbol{Q}^s \times \boldsymbol{Q}^{s-1} \times \cdots \times \boldsymbol{Q}^t$. Hence, $\boldsymbol{Q}^t$ captures the plastic dynamics between time t and $t+1$, while $\boldsymbol{Q}^{s:t}$ captures the plastic dynamics from t to $s+1$. With these notations, we have:

PROPOSITION 3. *The second-order expansion of the path-dependent reward $\boldsymbol{x}$ is given by:*

$$\begin{aligned}
\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} =& \boldsymbol{r}^{t+1} \\
&- \alpha \sum_{0 \leq t_1 \leq t} \boldsymbol{Q}^{t:t_1}\left(\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}\right) \\
&+ \alpha^2 \sum_{0 \leq t_2 < t_1 \leq t} \boldsymbol{Q}^{t:t_1}\left\{\left[\boldsymbol{Q}^{t_1-1:t_2}\left(\boldsymbol{r}^{t_2} \circ \boldsymbol{x}^{t_2}\right)\right] \circ \boldsymbol{x}^{t_1}\right\} \\
&+ \mathcal{O}(\alpha^3).
\end{aligned}$$

Proof of Proposition 3 When $t=0$, we already shown that the first-order expansion is exact and the second-order correction is 0, so the result holds.

Let us assume that the above expansion holds for $\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}}$, for some $t \geq 0$, then we have:

$$\boldsymbol{r}^{t+2}_{|x^{0:t+1}} = \boldsymbol{Q}^{t+1}\boldsymbol{r}^{t+1}_{|x^{0:t}} - \alpha \boldsymbol{Q}^{t+1}\left(\boldsymbol{r}^{t+1}_{|x^{0:t}} \circ \boldsymbol{x}^{t+1}\right).$$

We proceed for each term separately. The expansion of $\boldsymbol{Q}^{t+1}\boldsymbol{r}^{t+1}_{|x^{0:t}}$ is simply that of $\boldsymbol{r}^{t+1}_{|x^{0:t}}$ multiplied by $\boldsymbol{Q}^{t+1}$, i.e.,

$$\begin{aligned}
\boldsymbol{Q}^{t+1}\tilde{\boldsymbol{r}}^{t+1}_{|x^{0:t}} = \quad & \boldsymbol{Q}^{t+1}\boldsymbol{r}^{t+1} - \alpha \boldsymbol{Q}^{t+1} \textstyle\sum_{0 \leq t_1 \leq t} \boldsymbol{Q}^{t:t_1}\left(\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}\right) \\
& + \alpha^2 \boldsymbol{Q}^{t+1} \textstyle\sum_{0 \leq t_2 < t_1 \leq t} \boldsymbol{Q}^{t:t_1}\left\{\left[\boldsymbol{Q}^{t_1-1:t_2}\left(\boldsymbol{r}^{t_2} \circ \boldsymbol{x}^{t_2}\right)\right] \circ \boldsymbol{x}^{t_1}\right\} + \mathcal{O}(\alpha^3) \\
= \quad & \boldsymbol{r}^{t+2} - \alpha \textstyle\sum_{0 \leq t_1 < t+1} \boldsymbol{Q}^{(t+1):t_1}\left(\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}\right) \\
& + \alpha^2 \textstyle\sum_{0 \leq t_2 < t_1 < t+1} \boldsymbol{Q}^{(t+1):t_1}\left\{\left[\boldsymbol{Q}^{t_1-1:t_2}\left(\boldsymbol{r}^{t_2} \circ \boldsymbol{x}^{t_2}\right)\right] \circ \boldsymbol{x}^{t_1}\right\} + \mathcal{O}(\alpha^3).
\end{aligned}$$

The second-order expansion (in α) of $\alpha \boldsymbol{Q}^{t+1}\left(\boldsymbol{r}^{t+1}_{|x0:t} \circ \boldsymbol{x}^{t+1}\right)$ can be obtained by using the first-order expansion of $\boldsymbol{r}^{t+1}_{|x0:t}$. We have

$$\begin{aligned}\alpha \boldsymbol{Q}^{t+1}\left(\boldsymbol{r}^{t+1}_{|x0:t} \circ \boldsymbol{x}^{t+1}\right) &= \alpha \boldsymbol{Q}^{t+1}\left(\left[\boldsymbol{r}^{t+1} - \alpha \sum_{0\le t_1\le t} \boldsymbol{Q}^{t:t_1}\left(\boldsymbol{r}^{t_1}\circ \boldsymbol{x}^{t_1}\right) + \mathcal{O}(\alpha^2)\right] \circ \boldsymbol{x}^{t+1}\right) \\ &= \alpha \boldsymbol{Q}^{t+1}\left(\boldsymbol{r}^{t+1}\circ \boldsymbol{x}^{t+1}\right) - \alpha^2 \sum_{0\le t_1\le t} \boldsymbol{Q}^{t+1}\left\{\left[\boldsymbol{Q}^{t:t_1}\left(\boldsymbol{r}^{t_1}\circ \boldsymbol{x}^{t_1}\right)\right]\circ \boldsymbol{x}^{t+1}\right\} + \mathcal{O}(\alpha^3).\end{aligned}$$

Combining the two expansions yield the desired result. □

We see that the second-order term follows the same counting intuition as in the static case, with matrices $\boldsymbol{Q}^t$ accounting for plastic particle movements. At time t_1, to count the plastic particles already encountered at an earlier time t_2, one needs to account for their potential movements between these two time steps. $\boldsymbol{Q}^{t_1-1:t2}$ map plastic locations at time t_1 to locations at time t_2, hence the Hadamard product between $\boldsymbol{Q}^{t_1-1:t2}(\boldsymbol{r}^{t_2}\circ \boldsymbol{x}^{t_2})$ and $\boldsymbol{x}^{t_1}$.

Appendix C: Efficient Computations of Plastic Dynamics

In Section 4 we introduce the dynamics of plastic movement and how they are affected by the history collection. In this section, we describe how these predictions of plastic dynamics are stored in practice, and how the inputs to our optimization model can be efficiently computed.

C.1. Data set description

Plastic density data is provided by describing the trajectory of individual plastic 'particles', where a particle is a discretization unit for the fluid mechanical model. This description is referred to as a Lagrangian description. Our one-year plastic data contains the trajectory of $N = 1,212,613$ particles, discretized over T equally-spaced intervals. For example, the original model used a 1-day discretization timestep ($T = 365$), which we extended to a 3-hour discretization timestep via interpolation $T = 365 \times 8$). Eventually, the latitude and longitude of all particles are stored in two $N \times T$ arrays, which we denote as matrices $\boldsymbol{Lo}$ and $\boldsymbol{La}$.

For any location ℓ (defined as an 8×8 km region) and any time t, we can efficiently identify the set of particles present in this location at that time by filtering the rows of $\boldsymbol{Lo}$ and $\boldsymbol{La}$, i.e., we can compute

$$\mathcal{P}^t_\ell := \{n \in [N] \ : \ (La_{n,t}, Lo_{n,t}) \in \ell\},$$

where $(La_{n,t}, Lo_{n,t}) \in \ell$ means that particle n is in location ℓ. Observe that these sets $\{\mathcal{P}^t_\ell\}_{\ell\in\mathcal{L}}$, for a fixed t, can all be computed simultaneously in $\mathcal{O}(N)$ operations, by doing a single pass through all particles $n \in [N]$.

C.2. Computation of raw plastic density map

Based on the data set described above, our first task is to compute the raw plastic density map for the whole year, without considering any history collection. To do so, we can simply count the number of particles in each location at each time, i.e.,

$$r_\ell^t = \left|\mathcal{P}_\ell^t\right| \times \text{weight factor},$$

where the weight factor simply converts a number of particle into a mass of plastic. As discussed above, for a given time t, the entire map $\boldsymbol{r}^t$ can be computed in $\mathcal{O}(N)$ operations. So, we obtain the T maps $\{\boldsymbol{r}^t\}_{t\in[T]}$ in $\mathcal{O}(NT)$ time.

Alternatively, one could have computed $\boldsymbol{r}^0$, in $\mathcal{O}(N)$ time, and the transition matrices $\{\boldsymbol{Q}^t\}_{t\in[T]}$. This is, however, significantly more time consuming.

Indeed, we can compute the element $Q_{\ell,\ell'}^t$ by computing the ratio of particles that move from ℓ to ℓ' between t and $t+1$:

$$Q_{\ell',\ell}^t = \frac{|\mathcal{P}_\ell^t \cap \mathcal{P}_{\ell'}^{t+1}|}{|\mathcal{P}_\ell^t|}.$$

Computing a column of $\boldsymbol{Q}^t$, $\{Q_{\ell',\ell}^t\}_{\ell'\in\mathcal{L}}$ takes $\mathcal{O}(N)$ operations so computing the full matrix takes $\mathcal{O}(N|\mathcal{L}|)$. Accordingly, computing the density maps using the transition probability matrices $\{\boldsymbol{Q}^t\}_{t\in[T]}$ requires $\mathcal{O}(N|\mathcal{L}|T)$ iterations, where $|\mathcal{L}| \approx 10^5$.

C.3. Computation of path-dependent plastic density map

Compared to the raw map, it is more difficult to compute the path-dependent density map. Intuitively, we need to consider the effect of historical collection on the multi-period future. Mathematically, the reward decomposition described in Proposition 1 and Proposition 3 requires multiple computations between $\boldsymbol{Q}^{t:t_1} = \boldsymbol{Q}^t \times \cdots \times \boldsymbol{Q}^{t_1}$, $\boldsymbol{x}$, and α. To solve that, we propose two ways:

1. First, we can pre-compute $\boldsymbol{r}^0$ and $\{\boldsymbol{Q}^t\}_{t\in[T]}$ and then recursively compute $\boldsymbol{r}^t_{|\boldsymbol{x}^{0:t}}$ according to (3). This approach requires a number of iteration in the order of $\mathcal{O}(N|\mathcal{L}|T + |\mathcal{L}|^2 T)$.

2. Alternatively, we can use the expansion presented in Propositions 1–3 and compute each term sequentially. The zero-order term is simply the original density maps, without any collection, and can be computed in $\mathcal{O}(NT)$ time. Then, for the first-order term, we can compute the term $\boldsymbol{Q}^{t:t_1}(\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1})$ directly, without pre-computing the matrices $\boldsymbol{Q}^t$ individually. Indeed, $\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}$ is a vector with only one non-zero element, equal to $|\mathcal{P}_{\ell_1}^{t_1}|$, where ℓ_1 denotes the unique location such

that $x_\ell^{t_1} > 0$. Correspondingly, $\boldsymbol{Q}^{t:t_1}(\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1})$ represents the vector of locations at time t of plastic particles that were in ℓ_1 at time t_1, i.e.,

$$(\boldsymbol{Q}^{t:t_1}(\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}))_\ell = |\mathcal{P}_{\ell_1}^{t_1} \cap \mathcal{P}_\ell^t\}| \times \text{weight factor}.$$

Hence, each correction term $(\boldsymbol{Q}^{t:t_1}(\boldsymbol{r}^{t_1} \circ \boldsymbol{x}^{t_1}))$ can be computed in $O(N)$ operations. Altogether, we can compute the first-order approximation for the path-adjusted rewards in $O(NT)$. We can proceed analogously for the higher-order correction terms, as presented in Section 4.3. However, it becomes prohibitive for $k > 1$. On the positive side, truncating our expansion at an order $k < T$ is guaranteed to provide either a lower or an upper bound on the actual path-dependent reward.

3. Finally, we propose a faster approximation which relies on the efficient particle-level description of plastic dynamics and assigning dynamic weights to the particles. Formally, we assign each particle with a weight $w_n^t(\boldsymbol{x})$ (we will later drop the dependency on $\boldsymbol{x}$ for concision) such that the weight of the particle is divided by α each time it is seen by the system. Mathematically, we compute

$$\begin{aligned} w_n^0(\boldsymbol{x}) &= 1, \quad \forall n \in [N], \\ w_n^{t+1}(\boldsymbol{x}) &= \begin{cases} \alpha w_n^t(\boldsymbol{x}), & \text{if } n \in \mathcal{P}_{\ell_t}^t \text{ with } \ell_t \text{ s.t. } x_{\ell_t}^t > 0, \\ w_n^t(\boldsymbol{x}), & \text{otherwise.} \end{cases} \end{aligned}$$

With these notations, we can count the particles in a given location ℓ by weighting them according to their weight, i.e.,

$$|\mathcal{P}_\ell^t|_{\boldsymbol{w}} := \sum_{n \in [N] \,:\, (La_{n,t}, Lo_{n,t}) \in \ell} w_n,$$

and approximate the path-dependent plastic density at time t as

$$r_{\ell|\boldsymbol{x}^{0:t}}^{t+1} \approx |\mathcal{P}_\ell^{t+1}|_{\boldsymbol{w}^{t+1}(\boldsymbol{x})} \times \text{weight factor}.$$

The above computations can be efficiently executed in vectorized manner, therefore reducing the computational burden to compute path-dependent reward down to $\mathcal{O}(NT)$.

Appendix D: Solving the Path-Dependent Longest Path Problem

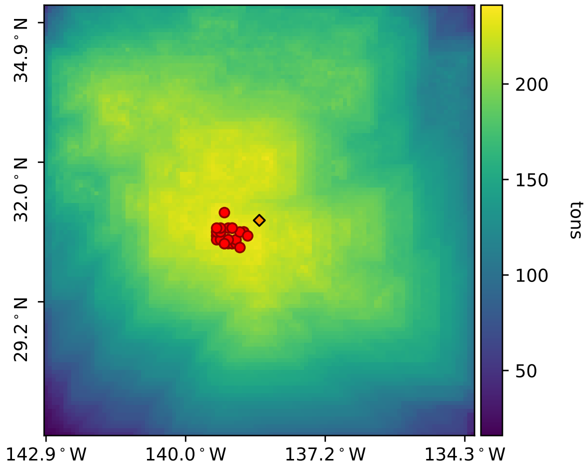
In this section, we provide further details on how to solve the path-dependent longest path problem (4). First, we describe and compare different strategies for searching for a high-quality solution in Algorithm 2. Then, we describe a valid branching strategy that could theoretically solve the problem to provable optimality, although its slow convergence makes it practically irrelevant.

D.1. Comparison of search-and-bound strategies

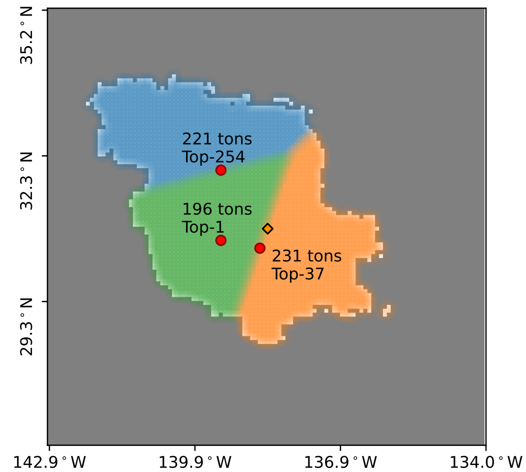
Recall that in the search-and-bound strategy (Algorithm 2), we search through the terminal states of the DP algorithm (Algorithm 1) to find high-quality trajectories. For a given terminal state s, the terminal value provided by Algorithm 1, $V^T(s)$, can be used as a proxy of the quality of the best trajectory that terminates in state s. However, computing the actual path-dependent reward, i.e., the quantity we are interested in, is time-consuming. Hence, we are interested in identifying a few terminal states $s \in \mathcal{S}_T$ and only evaluating the objective function at these states to make the search more efficient.

A first strategy is to search through $s \in \mathcal{S}_T$ by decreasing order of $V^T(s)$. To control for computational complexity, we can, for example, search through the top-K states only, with K an hyper-parameter controlling the number of path-dependent reward calculations (step ($\star$) in Algorithm 2). This strategy, however, may compare trajectories that are very similar. For illustration purposes, Figure D.1a displays the map of $V^T(s)$ values on one problem instance (the same instance as that of Figure 7b). The red dots indicate the 30 best locations, where, for each location, we consider the highest value of $V^T(s)$, across all terminal states associated with this location. We observe that all top-30 locations are geographically concentrated. Hence, these terminal locations are likely associated with very similar trajectories, leading to very similar path-dependent rewards. In addition, we should keep in mind that one location is associated with multiple states s. So, when searching through the top-K states s, one can compare trajectories that are even more similar (they can end at the same location). All together, while this strategy can improve upon the solution of the DP algorithm (without path-dependency), the improvement will likely be marginal for small values of K.

Alternatively, we propose to force a search through states/locations that are geographically dispersed, as displayed in Figure D.1b. First, we analyze the solution returned by Algorithm 1, compute its path-dependent reward, and obtain a valid lower bound on (4), LB. With this lower bound, we can safely exclude all states s such that $V^T(s) < LB$ (corresponding to the grey area in Figure D.1b). Then, we apply a K-means algorithm to partition the remaining states (the 'active' states) into K disjoint subsets, $\{s : V^T(s) \geq LB\} = \cup_{k=1}^K \mathcal{S}_T^{(k)}$. For the clustering algorithm, we represent each state by a 2-dimensional vector corresponding to its associated latitude and longitude. Finally, for each region $\mathcal{S}_T^{(k)}$, we inspect only the state with the highest terminal value, i.e., in $\arg\max_{s \in \mathcal{S}_T^{(k)}} V^T(s)$. In our example, this process, which we formally describe in Algorithm 2



(a) Heatmap of the $V^T(s)$ values; The states corresponding to the 30 highest $V^T(s)$ are representated by red dots.



(b) Cluster-K search ($K = 3$) with the 3 selected states represented by red dots.

Figure D.1 Comparing top-K (left panel) and cluster-K (right panel) search strategies on one 7-day planning instance (Jan 15, 2002). The starting location is indicated by a diamond (30.76°N, 138.44°W).

– Stage 2, inspects the 3 red locations displayed in Figure D.1b. Compared to the top-K approach illustrated in Figure D.1a, this procedure searches through locations that are more dispersed. We represent the path-dependent reward for each inspected location in Figure D.1b. Compared to the trajectory with the highest $V^T(s)$, the best trajectory we found (230 tons) only ranked 37th according to its $V^T(s)$ value but increased the weekly collection by 18% (from 231 to 196 tons).

We now compare both approaches numerically by applying them on $5 \times 50 = 250$ instances of 7-day routing (the same instances as those used in Section 4.5, 5 different starting locations, 50 weeks). The upper bound is the same for both approaches and obtained from the DP relaxation. We compute the optimality gap (which is the same as comparing the quality of the solution found) obtained by each approach, the top-K search and the cluster-K search, for different values of K. We report the box plot of the optimality gap for each approach in Figure D.2.

As expected, searching through multiple terminal states improves upon the solution of the DP relaxation. We observe that the Cluster-K search strategy is more efficient than the top-K one. Indeed, for the same search budget K, cluster-K achieves lower optimality gaps. For example, Top-12 achieves a 7.4% gap on average, while Cluster-12 achieves 6.3% on average —a 1.1-percentage point (or 17%) improvement. Cluster-3 (7.1% average gap) evaluates 4 times less trajectories than Top-12, yet achieves comparable performance. In addition, we observe that the standard deviation of the gap is smaller for cluster-K (4.6%–5.6%) than for top-K (6.2%–6.5%), indicating that the

Algorithm 2 – Stage 2: Cluster-K search for Algorithm 2

Stage 2: Search-and-bound;

Find $s^\star \in \arg\max_{s\in\mathcal{S}_T} V^T(s)$ and $\boldsymbol{x}^\star \leftarrow \text{path}[s^\star]$;

Initialize $UB = V^T(s^\star)$;

Compute $\boldsymbol{r}_{|\boldsymbol{x}^\star}$ and initialize $LB = \langle \boldsymbol{r}_{|\boldsymbol{x}^\star}, \boldsymbol{x}^\star \rangle$;

Apply K-means clustering and construct the partition $\{s : V^T(s) \geq LB\} = \cup_{k=1}^{K} \mathcal{S}_T^{(k)}$;

for $k = 1, \ldots, k$ **do**

  Get $s^{(k)} \in \arg\max_{s\in\mathcal{S}_T^{(k)}} V^T(s)$;

  Define $\boldsymbol{x}^{(k)} \leftarrow \text{path}[s^{(k)}]$;

  Compute the path-adjusted reward $\boldsymbol{r}_{|\boldsymbol{x}^{(k)}}$;

  if $\langle \boldsymbol{r}_{|\boldsymbol{x}^{(k)}}, \boldsymbol{x}^{(k)} \rangle > LB$ **then**

    Update lower bound $LB = \langle \boldsymbol{r}_{|\boldsymbol{x}^{(k)}}, \boldsymbol{x}^{(k)} \rangle$;

    Update best solution $\hat{\boldsymbol{x}} = \boldsymbol{x}^{(k)}$;

  end

end

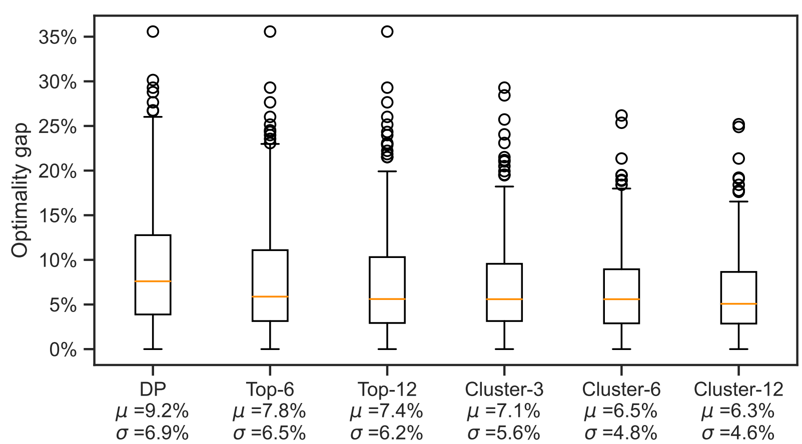


Figure D.2 Distribution of the optimality gap achieved by different search-and-bound approaches: Algorithm 1, top-K search, and cluster-K search. Results are obtained 250 instances of our 7-day routing problem. We also report the mean (μ) and standard deviation (σ) of each distribution.

performances of cluster-K are also more reliable. Therefore, we use the Cluster-K search with $K = 12$ in our implementation.

D.2. Evaluation of the 'true' performance of Algorithm 2 on small instances

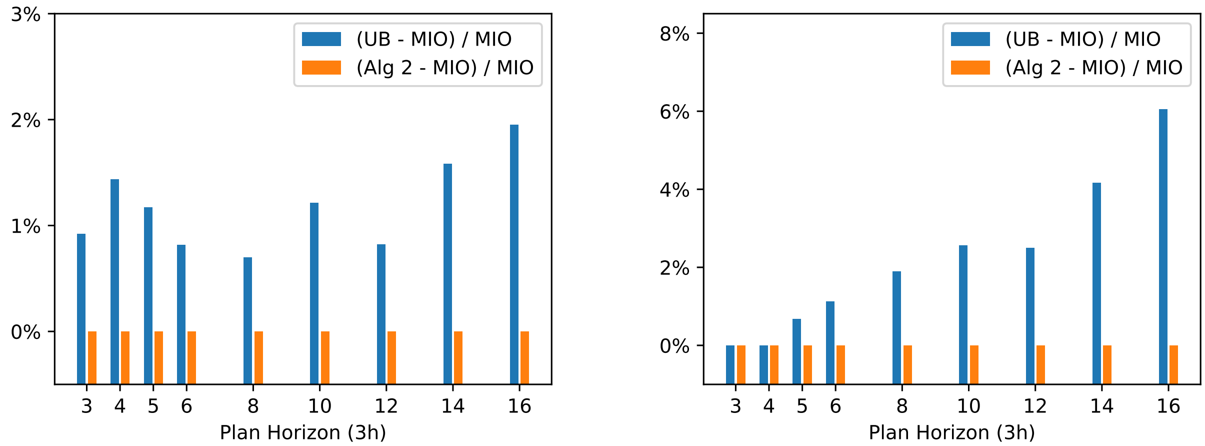
The optimality gap returned by Algorithm 2 is a conservative estimate of the sub-optimality of the returned solution. Indeed, it captures both the distance of the returned solution to the optimal solution (in terms of objective value) and the relaxation gap (i.e., the difference between the upper-bound provided by the path-independent relaxation and the true objective value). Accordingly, in this section, we investigate numerically whether the optimality gaps returned by Algorithm 2 are primarily driven by the quality of the solution or by the quality of the relaxation (i.e., the certificate of optimality). To do so, we apply Algorithm 2 on small instances where the optimal solution can be computed exactly.

The path-dependent longest path problem (4) can be formulated as a mixed-integer optimization (MIO) problem

$$\max_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\eta},\boldsymbol{\rho}\geq 0,} \quad \sum_{t\in\mathcal{T}}\sum_{s\in\mathcal{S}_t}\eta_s^t \quad \text{s.t.} \quad \begin{aligned}&\boldsymbol{\rho}^{t+1} = \boldsymbol{Q}^t\left(\boldsymbol{\rho}^t - \alpha\,\boldsymbol{\eta}^t\right),\\ &\boldsymbol{\eta}^t \leq \boldsymbol{\rho}^t,\\ &\boldsymbol{\eta}^t \leq M\boldsymbol{x}^t,\end{aligned}$$

where the additional variable $\boldsymbol{\rho}^t$ encodes for the vector of path-dependent rewards $\boldsymbol{r}^t_{|\boldsymbol{x}^{0:t}}$ and $\boldsymbol{\eta}^t$ for the component-wise product $\boldsymbol{r}^t_{|\boldsymbol{x}^{0:t}} \circ \boldsymbol{x}^t = \boldsymbol{\rho}^t \circ \boldsymbol{x}^t$. However, this formulation is notably inefficient to solve for two reasons. First, unlike the DP Algorithm 1, it requires the computation of the matrices $\boldsymbol{Q}^t$ for all time periods t, which is time-consuming (see Appendix C for a discussion). Second, it involves a large number of binary variables (the location-level variables x_ℓ^t and the state-level ones $\tilde{x}_s^t$). Even when a network flow formulation is used, integrality constraints cannot be relaxed due to the presence of the additional decision variables $\boldsymbol{\rho}$ and $\boldsymbol{\eta}$.

Nonetheless, we can solve this MIO formulation for small values of planning horizon (T) and compare the optimal objective value with (i) the upper bound obtained from Algorithm 1 (represented as UB) and (ii) the best solution found by Algorithm 2 (represented as Alg 2). We report these results in Figure D.3, for six instances (generated by taking 6 initial time period, uniformly spread across the year). Across all instances and for T ranging from 3 to 16 (2 days), we observe that across Algorithm 2 recovers the optimal solution ((Alg 2 - MIO) / MIO = 0%). However, the upper bound from the path-independent relaxation is systematically 1% higher than the true objective value. In other words, Algorithm 2 returns a positive optimality gap around 1%, although the solution returned is actually optimal. These results on smaller values of T suggest that the value



(a) Result averaged over six simulations (starting at day 43, 99, 155, . . . , 323)

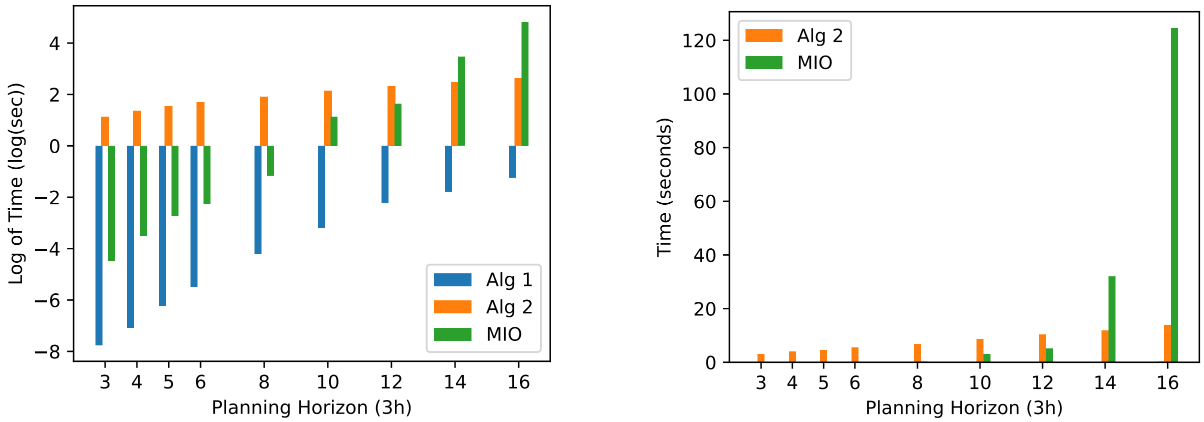
(b) Result of one simulation (starting at day 155)

Figure D.3 Gap of upper bound (UB) and our best solution (by Alg 2) compared to the exact solution (by MIO) under different planning horizons. All simulations start at (30.76°N, 138.44°W).

of optimality gap returned by Algorithm 2 is mostly an indicator of the tightness of the relaxation, rather than an indicator of the sub-optimality. While the average optimality gap is small, there is discrepancy across instances. We report the results for one instance with particularly large gap in Figure D.3b. In this case, we observe that the quality of the upper-bound degrades as the planning horizon increases (as suggested by the bounds provided in Proposition 2).

Unfortunately, the MIO approach becomes computationally prohibitive when T increases. We report the average computational time of Algorithm 1, Algorithm 2, and the MIO approach in Figure D.4. We observe that the time needed to solve the MIO formulation increases exponentially as the planning horizon increases and is already 2 orders of magnitude larger than Algorithm 2 for $T = 16$ (2 days). Moreover, in addition to the raw computational time needed to solve the MIO formulation, we need to compute the matrices $\boldsymbol{Q}^t$. For $T = 16$, there are 16 such matrices of size $39^2 \times 39^2$ and it takes 4.3 to 5.8 hours to compute them. We did not include this processing time in the computational time of MIO reported in Figure D.4.

In summary, on instances with up to $T = 16$ time steps, we observe that Algorithm 2 finds the optimal solution in most of the instances, while reducing computational time by two orders of magnitude compared with an MIO approach. However, it does return a positive optimality gap due to fact that the relaxation is not tight.



(a) Log of average computational time.

(b) Average computational time.

Figure D.4 Computational times under different planning horizons, averaged over 6 simulations (starting at day 43, 99, 155, . . . , 323).

D.3. Exact branch-and-bound scheme

In our algorithmic strategy, we solve a relaxation of (4) where we use the raw density maps $\boldsymbol{r}^t$ (path-*in*dependent) as edge lengths. Since for any admissible path $\boldsymbol{x}$, $\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} \leq \boldsymbol{r}^{t+1}$, this relaxation provides a valid upper bound on the objective of (4). In this section, we propose a branching strategy to refine this upper approximation and converge towards the optimal solution.

Since the path-dependent rewards, $\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}}$, depend on the specific locations visited by the system, a branching rule must fix the location of the system (at a particular time) in order to improve the reward approximation. Let us consider a subset of trajectories $\tilde{\mathcal{X}} \subseteq \mathcal{X}$ such that trajectories within $\tilde{\mathcal{X}}$ must pass through k locations at k particular time points. Formally, we consider an integer k, times $t_1, \dots, t_k \in \mathcal{T}$, and locations $\ell_1, \dots, \ell_k \in \mathcal{L}$ and assume that $\tilde{\mathcal{X}} = \{\boldsymbol{x} \in \mathcal{X} : x^{t_i}_{\ell_i} = 1, \forall i = 1, \dots, k\}$. We construct an upper bound on the path-dependent reward, $\tilde{\boldsymbol{r}}$, as follows:

$$\tilde{\boldsymbol{r}}^0 = \boldsymbol{r}^0,$$
$$\tilde{\boldsymbol{r}}^{t+1} = \begin{cases} \boldsymbol{Q}^t \left(\tilde{\boldsymbol{r}}^t - \alpha \tilde{\boldsymbol{r}}^t \circ \boldsymbol{x}^t\right) & \text{if } t \in \{t_1, \dots, t_k\}, \\ \boldsymbol{Q}^t \tilde{\boldsymbol{r}}^t & \text{otherwise.} \end{cases} \tag{8}$$

With this definition, we have

$$\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} \leq \tilde{\boldsymbol{r}}^{t+1}, \forall t \geq 0, \quad \forall \boldsymbol{x} \in \tilde{\mathcal{X}}. \tag{9}$$

We delay a formal proof of this property to the end of this section. Instead, let us now describe how we can use this construction to design a tailored branch-and-bound algorithm for our problem.

We consider the subproblem of finding the longest (path-dependent) path among all trajectories in $\tilde{\mathcal{X}} \subseteq \mathcal{X}$:

$$\max_{\boldsymbol{x}\in\tilde{\mathcal{X}}} \quad \sum_{t\in\mathcal{T}} \sum_{s\in\mathcal{S}_t} r^t_{s|\boldsymbol{x}^{0:t-1}} x^t_s \quad \text{s.t.} \quad \boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} = \mathcal{Q}^t(\boldsymbol{r}^t_{|\boldsymbol{x}^{0:t-1}}, \boldsymbol{x}^t). \tag{10}$$

Given the valid upper bound on the path-dependent rewards over $\tilde{\mathcal{X}}$, $\tilde{\boldsymbol{r}}$, defined by (8), we can apply Algorithm 2 on this restricted problem and obtain an upper and lower bound on (10). To refine this approximation, we pick a time $t_0 \in \mathcal{T}$ and a location $\ell_0 \in \mathcal{L}$ and partition $\tilde{\mathcal{X}}$ into

$$\begin{tikzcd}[column sep=small] & \tilde{\mathcal{X}} \arrow[dl, dash] \arrow[dr, dash] & \\ \tilde{\mathcal{X}}_0 & & \tilde{\mathcal{X}} \setminus \tilde{\mathcal{X}}_0 \end{tikzcd}$$

where $\tilde{\mathcal{X}}_0 := \{\boldsymbol{x} \in \tilde{\mathcal{X}} \,:\, x^{t_0}_{\ell_0} = 1\}$. In other words, $\tilde{\mathcal{X}}_0$ fixes a new time/location for the admissible trajectories. Hence, we can construct a tighter upper approximation of the path-dependent rewards over $\tilde{\mathcal{X}}_0$ (the left child node) by applying (8). The subproblem $\tilde{\mathcal{X}}_0$ benefits from both a tighter approximation and a reduced search space, so we should expect to effectively reduce the optimality gap on this child node. For the right child node, however, the benefit only comes from reducing the search space from $\tilde{\mathcal{X}}$ to $\tilde{\mathcal{X}} \setminus \tilde{\mathcal{X}}_0$, which is not very restrictive because this set still contains trajectories that can be close to ℓ_0 at time t_0 (or trajectories that visit ℓ_0 at time $t_0 \pm 1$). Hence, the improvement in the upper bound should be marginal (as confirmed in preliminary numerical experiments we conducted). Because of this imbalance, we expect this branching scheme to experience very slow convergence towards the optimal solution and do not view it as practically relevant for our application. Nonetheless, theoretically, this branching strategy eventually enumerates all possible trajectories so is guaranteed to converge to the optimal solution after a finite yet exponential number of iterations.

Proof of (9) We prove the result by induction on k.

For $k = 0$ (no location is fixed), we have $\tilde{\mathcal{X}} = \mathcal{X}$. Furthermore, $\tilde{\boldsymbol{r}}^0 = \boldsymbol{r}^0$, and $\tilde{\boldsymbol{r}}^{t+1}$ and $\boldsymbol{r}^{t+1}$ follow the same recursive formula. So, for any $t \geq 0$, $\tilde{\boldsymbol{r}}^{t+1} = \boldsymbol{r}^{t+1}$. We conclude by noting that $\boldsymbol{r}^{t+1} \geq \boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}}$ for any $\boldsymbol{x} \in \mathcal{X}$ (Proposition 2).

Fix $k \geq 0$. We assume that the result holds for k, namely, we assume that for any set $\tilde{\mathcal{X}}$ of the form $\{\boldsymbol{x} \in \mathcal{X} : x^{t_i}_{\ell_i} = 1, \forall i = 1, \dots, k\}$, the reward vector constructed according to the recursive formula (8) satisfies the property (9). Let us consider a set of the form $\tilde{\mathcal{X}} = \{\boldsymbol{x} \in \mathcal{X} \,:\, x^{t_i}_{\ell_i} = 1, \forall i = 1, \dots, k+1\}$ and assume, without loss of generality, that $t_1 < \cdots < t_k < t_{k+1}$. Define $\bar{\mathcal{X}} := \{\boldsymbol{x} \in \mathcal{X} \,:\, x^{t_i}_{\ell_i} = 1, \forall i =$

$1,\dots,k\}$, the set obtained by imposing the first k (out of $k+1$) locations imposed by $\tilde{\mathcal{X}}$, and denote $\bar{\boldsymbol{r}}$ the associated reward vector defined by (8).

By the induction hypothesis applied to the set $\bar{\mathcal{X}}$, we have

$$\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} \leq \bar{\boldsymbol{r}}^{t+1}, \forall t \geq 0, \quad \forall \boldsymbol{x} \in \bar{\mathcal{X}}.$$

Furthermore, the definition of $\tilde{\boldsymbol{r}}^{t+1}$ only differs from that of $\bar{\boldsymbol{r}}^{t+1}$ after time t_{k+1}. In other words, for any $t \geq 0$, we have

$$\tilde{\boldsymbol{r}}^{t+1} = \begin{cases} \bar{\boldsymbol{r}}^{t+1} & \text{if } t < t_{k+1}, \\ \boldsymbol{Q}^t\left[\bar{\boldsymbol{r}}^t \circ (1-\alpha \boldsymbol{x}^t)\right] & \text{if } t = t_{k+1}, \\ \boldsymbol{Q}^t \tilde{\boldsymbol{r}}^t & \text{if } t > t_{k+1}. \end{cases}$$

We now analyze each case separately.

For $t < t_{k+1}$*:* Since $\tilde{\mathcal{X}} \subseteq \bar{\mathcal{X}}$, we have

$$\tilde{\boldsymbol{r}}^{t+1} = \bar{\boldsymbol{r}}^{t+1} \geq \boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}}, \quad \forall \boldsymbol{x} \in \tilde{\mathcal{X}} \subseteq \bar{\mathcal{X}}.$$

For $t = t_{k+1}$*:* Taking $t = t_{k+1} - 1$ in the inequality above, we have just proved that $\tilde{\boldsymbol{r}}^{t_{k+1}} \geq \boldsymbol{r}^{t_{k+1}}_{|\boldsymbol{x}^{0:t_{k+1}-1}}$. Multiplying both sides of these inequalities (component-wise) by $1 - \alpha \boldsymbol{x}^{t_{k+1}} \geq \boldsymbol{0}$ and multiplying by the non-negative matrix $\boldsymbol{Q}^{t_{k+1}}$ yields

$$\tilde{\boldsymbol{r}}^{t_{k+1}+1} := \boldsymbol{Q}^{t_{k+1}}\left[\tilde{\boldsymbol{r}}^{t_{k+1}} \circ (1-\alpha \boldsymbol{x}^{t_{k+1}})\right] \geq \boldsymbol{Q}^{t_{k+1}}\left[\boldsymbol{r}^{t_{k+1}}_{|\boldsymbol{x}^{0:t_{k+1}-1}} \circ (1-\alpha \boldsymbol{x}^{t_{k+1}})\right],$$

where the right-hand side is precisely the definition of $\boldsymbol{r}^{t_{k+1}+1}_{|\boldsymbol{x}^{0:t_{k+1}}}$ for trajectories $\boldsymbol{x} \in \tilde{\mathcal{X}}$.

For $t > t_{k+1}$*:* We have

$$\tilde{\boldsymbol{r}}^{t+1} \geq \boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}}, \forall \boldsymbol{x} \in \tilde{\mathcal{X}}, \qquad \text{for } t = t_{k+1}.$$

If the inequality holds for some $t \geq t_{k+1}$, then multiplying by $\boldsymbol{Q}^{t+1} \geq \boldsymbol{0}$ leads to

$$\tilde{\boldsymbol{r}}^{t+2} := \boldsymbol{Q}^{t+1}\tilde{\boldsymbol{r}}^{t+1} \geq \boldsymbol{Q}^{t+1}\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} \geq \boldsymbol{Q}^{t+1}\boldsymbol{r}^{t+1}_{|\boldsymbol{x}^{0:t}} \circ (1-\alpha \boldsymbol{x}^{t+1}) = \boldsymbol{r}^{t+2}_{|\boldsymbol{x}^{0:t+1}},$$

for any $\boldsymbol{x} \in \tilde{\mathcal{X}}$, where the last inequality holds because $(1 - \alpha \boldsymbol{x}^{t+1}) \leq \boldsymbol{1}$. Thus, we can prove the desired properties for all $t = t_{k+1}, \dots, T$ by induction. $\square$

Appendix E: Numerical Experiment Supplement Result

In this section, we provide supplementary results to the numerical experiments presented in Section 5.

	Δmean	Std. error	t-stat	p-value	95% Confidence interval	
					Lower	Upper
Myopic vs. Benchmark	21.199	3.032	6.993	1.45×10^{-5}	14.594	27.804
Week vs. Benchmark	21.381	1.940	11.022	1.24×10^{-7}	17.154	25.608
Week-Folding vs. Benchmark	26.422	2.750	9.608	5.51×10^{-7}	20.430	32.413
Week-Rolling vs. Benchmark	26.380	2.154	12.246	3.86×10^{-8}	21.686	31.073

Table E.1 Paired t-test to evaluate the benefit (in terms of weekly plastic collected) of each optimization method compared with the benchmark

E.1. Overall improvement in plastic collection rate

Figure 8a reports the average weekly quantity of plastic collected by the system, for different routing/extraction scheduling algorithms. We can visually infer that optimization methods provide a significant improvement over the benchmark. To formalize this statement, we conduct a paired t-test to compare the 4-week reward of each method to the benchmark, over the 13 simulations and report the results of these comparisons in Table E.1. Note, unfortunately, that we cannot perform paired t-tests at the week level because the starting location at the beginning of each week is different for each method (while the starting location at the beginning of each simulation is the same).

While the results in the main paper focus mostly on the resulting collection efficiency, we now also report results on computational tractability. Figure E.1 computational time for each method. The DP method with rolling windows takes more time than others. In our DP methods, the size of state $\mathcal{S}_T$ increases exponentially with the total planning time frame T. The rolling method requires planning for 7 days repeatedly, leading to a longer computational time. In practice, the route decision is updated daily, and the time budget for model planning is around 12 hours, much larger than the 24 minutes we need.

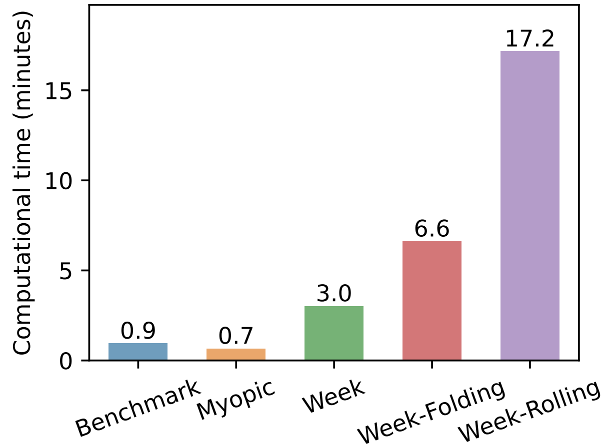
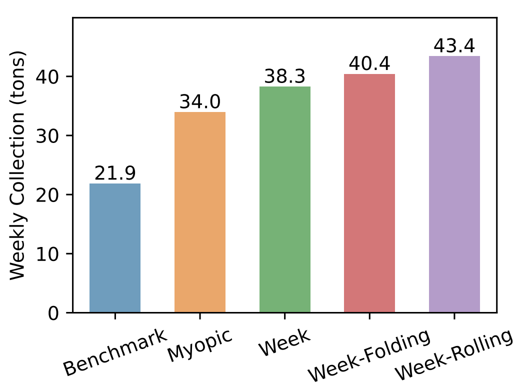


Figure E.1 Average computational time of benchmark and DP methods (1-week planning)

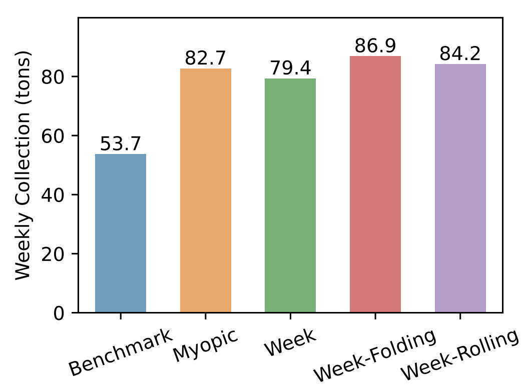
E.2. Heterogeneity across seasons and the impact of extraction scheduling

In Figure 8b, we observe a significant impact of the season on collection efficiency, for all methods. Figure E.2 reports the performance of each method, on the winter and summer months separately. Here, we group days 1–84 (first three simulations) and day 280–365 (last three simulations) into 'Winter' and the remaining 7 simulations as 'Summer'. We see that both our methods and the benchmark collect significantly less plastic during the winter months (around a factor 2 decrease). In addition, the value of optimization is stronger during the winter. For example, Week-Rolling collects 1.5× more plastic than the benchmark on average, which breaks down into a 1.35× improvement and a 2× improvement in the winter.

As discussed in Section 5.3, the heterogeneity in efficiency observed cannot be explained by differences in plastic densities. Indeed, as shown in Figure E.3a, real plastic density in the GPGP

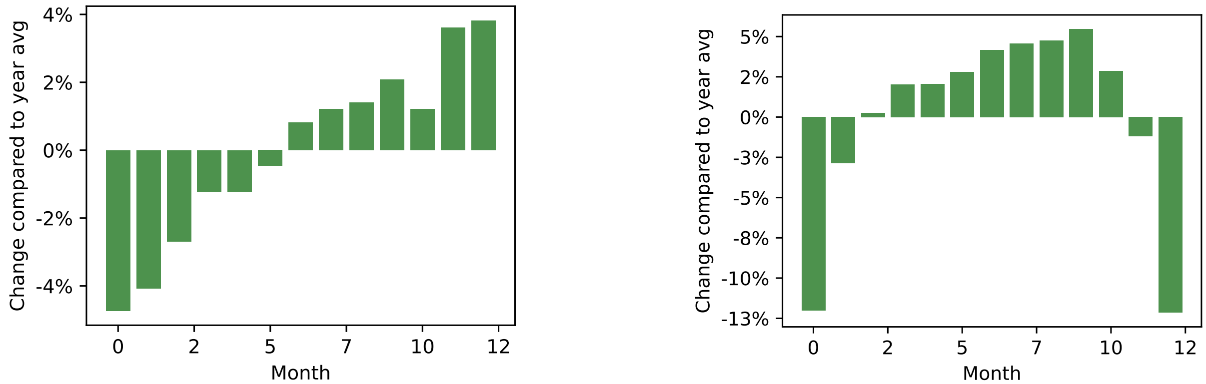


(a) Winter



(b) Summer

Figure E.2 Average weekly reward group by season

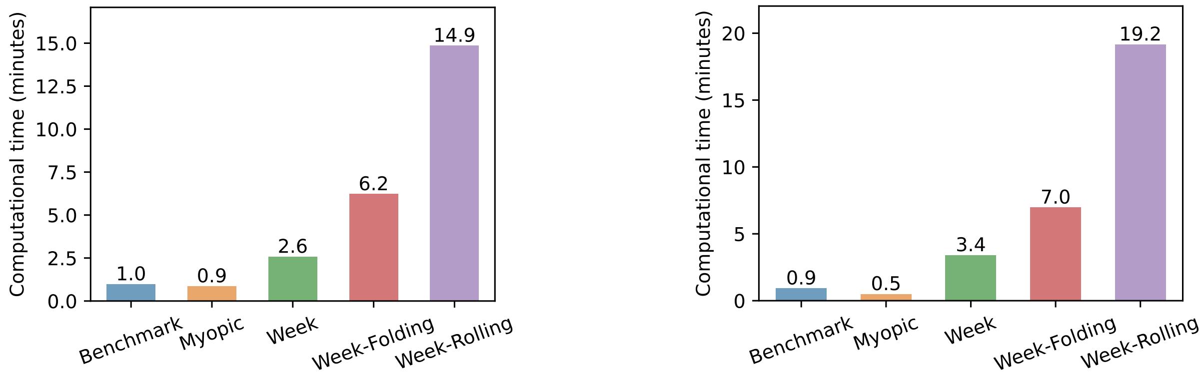


(a) Real plastic density (b) Collectable plastic density

Figure E.3 Average real and collectable plastic density in the GPGP over the year 2002. The collectable plastic density in the right panel is calculated by assigning zero density to any location with wave heights above 6 meters. Values are reported in relative terms compared to yearly average.

globally increases over time, due to new plastic being emitted in the oceans. In our data (year 2002), we observe that the average plastic density in the GPGP increased by 10%. However, we do not observe an inverted U-shaped behavior as in Figure 8b, suggesting that weather (and wave height in particular) is the main driver of performance here. To consider the effect of high wave on the collection process, we calculate the average collectable plastic in the GPGP in Figure E.3b, where we treat the plastic density of a high wave location (wave height $\geq$ 6 meters as zero. It can be seen that the collectable plastic density has such inverted U-shaped, but the difference between the highest collectable density month only exceed 20% more density than the lowest collectable density month. In contrast, in Figure 8b, the maximum gap between the highest collection month and lowest collection month is much higher (more than 300% for Week-Rolling and more than 500% for the Benchmark approach). Therefore, the change in collectable plastic density cannot fully explain the seasonal heterogeneity in performance efficiency.

In Figure E.4, we report the computational time of each method, grouped by season instead of averaged over the year. Overall, we observe that bad weather conditions limit the number of feasible states, hence tend to lead to shorter computational times than for instances occurring the summer months.



(a) Winter (b) Summer

Figure E.4 Computational time of different methods by season

References