Location Optimization of Catching Systems for Plastic Waste Removal from Waterways

A Plastic Waste Flow Capturing Location Model

MSc Thesis Applied Mathematics

Anne-Fleur Dijkhorst

Delft University of Technology

TU Delft

Noria Sustainable Innovators

Location Optimization of Catching Systems for Plastic Waste Removal from Waterways

a Plastic Waste Flow Capturing Location Model

by

Anne-Fleur Dijkhorst

Performed at

Noria Sustainable Innovators

to obtain the degree of Master of Science in Applied Mathematics at Delft University of Technology,

to be defended publicly on June 4th, 2024 at 14:30.

Thesis committee:

Daily Supervisor TU Delft:	Dr. J.T. van Essen
Daily Supervisor Noria:	J. van Wijk, MSc
Committee member:	Dr. J.L.A. Dubbeldam
Project Duration:	September, 2023 - June, 2024
Faculty:	Faculty of Electrical Engineering, Mathematics and Computer Science, Delft
Student number:	4713370

TU Delft Delft University of Technology

NORIA SUSTAINABLE INNOVATORS

Contents

PREFACE

This thesis is written to complete my MSc in Applied Mathematics at Delft University of Technology. The thesis aims to address the challenge of optimizing the placement of catching systems to prevent plastic waste from reaching our oceans. The research gave me the opportunity to dive into the problem of plastic waste pollution in waterways and apply my mathematical knowledge to an important practical problem. I learned a lot from designing the model and developing it from scratch, while working at a sustainable start-up company.

This project would not have been possible without the guidance of my daily supervisors, Theresia van Essen and Jur van Wijk. I am very grateful for the time they invested in our meetings and the advice they gave me, not just on the project's content but also on the process, my personal goals, and career orientation along the way. Furthermore, I would like to thank the entire team at Noria for the warm atmosphere at the company that made my time spent on the thesis a very pleasant period. I always enjoyed being at the office and the shared dedication to the environment and the mission of the company energized me a lot.

Finally, I would like to thank my friends and family for their support throughout this thesis and my entire time studying at Delft University of Technology. From reading parts of my work, to listening to my struggles with the project, to sitting together while working, to sharing cozy dinners or fun activities to relax, all of these have been an immense help to me. Thank you so much!

I hope you enjoy reading this thesis, and I hope this work contributes in some small way to the ongoing efforts to protect our environment.

A.F.X.H. (Anne-Fleur) Dijkhorst
Delft, May 2024

ABSTRACT

Plastic waste transported to oceans through canals and rivers becomes increasingly challenging to retrieve and harmful to ecosystems. Catching systems designed by Noria Sustainable Innovators can be used to capture the plastics as close to their source as possible. Deciding the best locations to place these systems is a difficult task, which is why a model for location optimization of catching systems for plastic waste removal from waterways is designed in this thesis: the Plastic Waste Flow Capturing Location Model (PW-FCLM).

In this model, the plastic waste flow through a network of waterways is represented as a Markov chain, using environmental data as inputs to estimate the initial probabilities and transition probabilities of plastic waste in the network. The PW-FCLM extends an existing Markov Decision Process-based Flow Capturing Location Model by incorporating various types of catching systems, considering sensitive areas, and specifying orientations for the systems. The equivalence of the linearized version of the extended model is demonstrated, a proof of NP-hardness of the problem is given and a greedy heuristic is presented as an alternative solution method.

A sensitivity analysis on the different types of input parameters is performed, and the runtimes for different problem sizes and solution methods are tested for case studies of Delft and Groningen in the Netherlands. The model is most sensitive to changes in the distance between the nodes in the network and to the probability of getting stuck due to water vegetation. For budgets up to $B = 2$ and problem sizes up to $n = 375$ nodes, the exact optimal solution can be found efficiently without a commercial solver license. For larger problem sizes or a higher budget, the heuristic appears to be a more appropriate solution method.

For future research, it is recommended to further study the influence of the distance between the nodes on the optimal solution and to investigate the plastic flow representation in the Markov chain further. Exploring the model's application to larger areas, such as provinces or countries, would be beneficial. The real-life effectiveness of placing catching systems at the optimal locations suggested by the model, depends on the accuracy of the input parameters. In this thesis, the values of the input parameters are primarily based on estimations of experts. It would be beneficial for the user of the model to further validate the values of the input parameters through experiments.

1

INTRODUCTION

Plastic waste pollution is an emerging issue (Borrelle et al., 2020). Once plastic waste leaves land and enters water, it becomes increasingly difficult to retrieve it, causing it to harm surrounding ecosystems, to degrade into microplastics and sequentially often to enter food chains (Wagner et al., 2014). It has been shown that a significant amount of the plastics is transported to the ocean via canals and rivers (Meijer et al., 2021). This provides an opportunity of catching the waste closer to the source in these canals and rivers.

Noria Sustainable Innovators tackles this problem by performing research on the location and flow of plastic waste in local waterways and removing the plastic waste using catching systems that they developed. However, to put these catching systems to their best use, it is important to identify which locations are the best positions to place them. The goal is to remove as much plastic waste from the water as possible; that means trying to find the locations where most plastics can be caught by the systems. Therefore, the goal of this thesis is to design an optimization model to facilitate decision making of the locations of plastic waste catching systems, in order to catch as much plastic waste as possible.

A model in which locations are chosen to capture moving flow along its way was presented in literature: the Flow Capturing Location Model (FCLM) (Berman et al., 1992). There are two versions of this model; a path-based (Berman et al., 1992) and a Markov Decision Process (MDP-)based (Berman et al., 1995c) approach. There is a lack of availability of data on exact flow paths of plastic waste. However, it is feasible to make well-founded estimates of the probability that plastic waste starts at a location, transitions to locations in different directions and ends at a location, based on geographic data. Therefore, the MDP-based FCLM is used as a foundation and is adjusted to design a new model in this thesis that takes all aspects of plastic waste capturing into account: the Plastic Waste Flow Capturing Location Model (PW-FCLM).

Previous research on waste removal from water and on the mathematics of location optimization models is presented in Chapter 2. The technical background of previous work at Noria and mathematical background of the FCLM is shown in Chapter 3.

A problem description and a model formulation for the PW-FCLM are shown in Chapter 4. A linearization of the PW-FCLM, an analysis of the complexity of the model and a greedy heuristic are shown in Chapter 5. The processing of geographic data into input parameters is explained in Chapter 6. The computational results of a sensitivity analysis of the input parameters is shown in Chapter 7, together with results on the runtime and performance of open source and commercial solvers compared to the greedy heuristic. A discussion of the results and recommendations for further improvements of the model are given in Chapter 8. The final conclusions of the thesis are presented in Chapter 9.

2

LITERATURE RESEARCH

In this chapter, an overview of previous research from literature is presented. This includes research on optimization of waste removal from water in Section 2.1, but also research on mathematical models for location optimization in other settings that are interesting in the scope of this thesis in Section 2.2.

2.1 WASTE REMOVAL FROM WATER

Several efforts have been made in different circumstances for location optimization of waste removal from water. The first was done by Sherman and Sebille (2016). The movement of marine microplastics was modeled and local optimization methods were used to determine favorable locations for placing a passive catching system using floating barriers (referred to as a "sink"). Ocean drift trajectories were mapped onto a $1^\circ \times 1^\circ$ grid with seasonal transition matrices and the amount of waste released in each grid cell along the coast was modeled to be proportional to the amount of mismanaged waste in the country and the population density. As a second scenario, the net primary production measured by satellite observations was considered as a measure of the size of ecosystems. The model was run in the first scenario to maximize removal of microplastics mass, and in the second scenario to minimize overlap between microplastics and the ecosystems, thereby minimizing total impact on ecosystems.

The initial solution of the placement of 29 sinks used as imput to their model was based on coordinates with the maximum cumulative plastic mass flux after a decade. Then, the locations were moved randomly from the initial solution to a new location in a $30^\circ \times 30^\circ$ neighborhood and then to the optimal cell within a $5^\circ \times 5^\circ$ neighborhood of this new location. The results of this method showed that according to the model, it is more beneficial to place the sinks closer to the shore than placing them in the middle of the Great Pacific Garbage Patch, which is one of the largest accumulation zones of marine plastics (Lebreton et al., 2018). This indicates that it is worth investigating in this thesis, whether it is more beneficial to catch the plastic waste close to the source, or closer to accumulating hotspots.

More recently, network theory has been used to identify optimal locations for plastic debris cleanup (Ypma et al., 2022). A transition matrix is constructed from Lagrangian simulation of plastic flow between several positions along the shore of the Galapagos Marine Reserve islands. The matrix is used to generate a network with transition probabilities on the edges and several network centrality measures of this network are considered to find which centrality measures can be used to optimize for maximum waste removal. Overall, it seems that a positive *Source-Sink Index*, which shows if a node is a net sink (instead of a net source) of plastics, provides the highest benefit in terms of the amount of plastics removed. The Retention Rate (percentage of particles for which the sink node is the same as the source node), PR_{in} centrality (the importance of the node in comparison to neighbors on incoming edges) and the betweenness centrality (whether nodes lie on many shortest paths between other nodes) are also shown to be good indicators of effective cleanup location nodes. A significant difference between the case study and this thesis is that shore/beach cleanup with land-based nodes is considered instead of cleanup in the water.

Additionally, other types of ocean cleanup aside from plastic waste cleanup have been researched and show useful techniques for plastic waste cleanup. Oil spill cleanup response in the Gulf of Mexico has been optimized using a Mixed Integer Linear Program (MILP) model that minimizes costs while meeting a minimum volume of oil cleanup goal (Grubesic et al., 2017). In this research, day to day drift of the spilled oil is modeled and Navy Coastal Ocean Model American Seas (AmSEAS) ocean data was used as oil transport data. The optimal response using vessels that move from different coastal harbors to the spill sites is calculated for each day. The model was solved to optimality using Python with the Gurobi solver. The use of vessels and a daily varying response is different from removing waste using a stationary catching system that remains in the same location for a period of time. However, the setup of the MILP was used as an inspiration to create a MILP formulation of the problem in this thesis as presented in Section 4.2.

A more dynamical optimization model for oil spill response operations was presented by You and Leyffer (2011), where three different types of cleanup methods (mechanical cleanup, in-situ burning and chemical dispersants) are considered in a cost-minimizing objective function. The changing area of oil slick was modeled by numerically solving differential equations which are integrated into a mixed integer non-linear optimization model. An approximate MILP is formed to create an initial solution that is used to find an approximate solution to the non-linear non-convex problem.

2.2 SIMILAR MATHEMATICAL PROBLEMS

Restricting to existing ocean and river cleanup research would be a limitation to this research. Essentially, the problem is about capturing flows (of pieces of plastic waste) in discrete locations in a (water) network. There are several traditional location allocation models where distance to locations is minimized and coverage of

the locations is maximized. However, generally these models concern node-based demand. In literature, this problem is called the "Maximal Covering Location Problem/Model" (MCLP/MCLM) (Church and ReVelle, 1974). In this model, demand at a node is considered covered when the node lies within a certain distance from a located facility. A well known application of this model is the ambulance location problem, where demand by patients is covered when the ambulance is located within a certain distance/time from the patient (Brotcorne et al., 2003).

However, demand in networks is not always located at nodes. In the case of plastic waste in waterways, the plastic is moving along the arcs of the network. This is mathematically equivalent to the "Flow Capturing Location Problem/Model" (FCLP/FCLM) that has been widely researched in the past 30 years for purposes where demand is not expressed at the nodes of a network, but rather intercepted on pre-planned paths. This is the case with for example fuel stations, banking machines, billboards or security checkpoints (Hodgson et al., 1996).

Different versions of this FCLM together with solution methods and applications from literature are summarized in this section. Deterministic versions of the FCLM are summarized in Section 2.2.1 and stochastic versions in Section 2.2.2. Mathematical details that are important for the development of the model in this thesis are further elaborated on in Section 3.2.

2.2.1 DETERMINISTIC FLOW CAPTURING LOCATION MODEL

The FCLM was first introduced independently by Hodgson (1990) and Berman et al. (1992). The basic idea is that the flow on a network $G(V, E)$ with vertices V and edges E is defined along paths from the set P of nonzero flow paths on the network arcs. The flow f_p on a path $p \in P$ is considered captured when there is at least one node on the path p where a facility is located. There is a fixed number of facilities m that are located on the nodes of the network and the objective function is to maximize the total amount of captured flow. When there is more than one facility on a path p, the flow on this path f_p is only counted as captured once. The MILP formulation of this model is equivalent to the Maximal Coverage Location Model and both are NP-hard (nondeterministic polynomial time) problems (Berman et al., 1992).

A greedy heuristic is proposed by Berman et al. (1992) and a worst case bound is proven. The greedy algorithm consists of sequentially placing a facility at the node which intercepts the maximum amount of flow, after which the intercepted flow is removed. A new facility is placed at the node which intercepts the maximum residual flow, until m facilities are located. This method does not guarantee finding an optimal solution, but works very intuitive and fast. A branch-and-bound algorithm that finds the optimal solution is also proposed, where upper bounds of partial solutions are calculated to find the maximum amount of flow that can possibly be caught after setting one variable x_j of locating a facility at node j to zero or one. These upper bounds show if the partial solution could be part of the optimal solution to the problem. A solution from the greedy algorithm can be used as input to the branch-and-bound algorithm.

Computational results on random test cases for networks of size $|V| \leq 100$ of both the greedy algorithm and the branch-and-bound algorithm are presented in the paper.

The difference between the FCLM and the MCLM lies in the constraint where demand is considered covered/captured. In the FCLM, demand of a path is considered captured when there is at least one facility located on a node *on the path* of the flow (demand). In contrast, demand is considered covered when there is at least one facility located on a node *in the neighbourhood* of a certain distance from the demand. Considering paths instead of neighbourhoods ensures that flow is not "*cannibalized*". Cannibalization is a term that Hodgson used when flow on a path is counted as captured multiple times by different facilities which happens when flow through a location instead of flow on a path is considered. The naive approach of counting all flow that passes a facility in a location as captured, could lead to effectively capturing less flow than expected, since the flow might have been captured by another facility on its path before reaching this facility (Hodgson, 1990).

Computational results and the effect of "cannibalization" on a real-life transportation network in Edmonton, Canada are shown by Hodgson et al. (1996). The flow on paths of the network was determined using origin-destination pairs with shortest paths between them. The network consists of 703 nodes and 23350 flows and the problem is solved exactly for $m \in \{1, ..., 15\}$ with a computation time of < 6 days and using a vertex substitution heuristic (VSH) for the larger cases with $m \in \{16, ..., 50\}$ as a benchmark. The greedy algorithm presented previously by both Hodgson (1990) and Berman et al. (1992) was then compared to the best found solutions (exact or VSH). On this real-life network, the solution to the greedy algorithm is never more than 0.9 percent worse in the amount of captured flow than the best found solution (exact or VSH). The performance of the greedy algorithm also always outperforms its worst case performance bound shown by Berman et al. (1992). The cannibalization effect on the network is between 10-30 percent for $m > 3$, which means that this percentage of flow is not caught compared to the best found solution because flow is counted as captured more than once in the naive solution. These results show that the first proposals of the FCLM are promising in their performance for capturing demand on the arcs of a network.

An improvement of the aforementioned greedy heuristic was proposed by Boccia et al. (2009). In each iteration of this improved heuristic, not only the leftover interceptable flow is considered but also the amount of flow already intercepted by facilities that were located in previous iterations. When locating a new facility makes a facility that was located during a previous iteration redundant, the previous facility is removed during an extra step in the algorithm for the improved heuristic. This improved heuristic was also used by Kuby et al. (2009) to locate hydrogen fuel stations in Florida.

VARIATIONS FLOW CAPTURING LOCATION MODEL

Several variations of the FCLM were proposed for applications in different areas. For example for the application of refueling, the range of a vehicle is taken into account

(Kuby et al., 2009)(Wu and Sioshansi, 2017) and deviations from the pre-planned path by customers are allowed (Berman et al., 1995a). For services where multiple exposures to facilities on each path are beneficial, multi-counting is introduced instead of only counting the first facility on a path as capturing. In Averbakh and Berman (1996), the FCLM is adapted with multi-counting under the assumption that the objective value is a nondecreasing concave function in the number of facilities on a path. A greedy heuristic that maximizes the increase of this nondecreasing concave objective function in each step is proposed and a worst case upper bound is shown.

The FCLM can generally be formulated with two different possible objective functions. One option is maximizing the amount of flow that is captured under certain constraints including the budget, which sets a limit on the number of facilities that can be located. The other option is to find the minimum number of facilities that is necessary to capture (a fraction of) the total flow. Boccia et al. (2009) refer to these two options as M1 and M2. They also present the possibility to write the model in a "gain-oriented" way (M3). A coefficient that represents the obtainable gain for each node on a path is included in the model. The value of this coefficient decreases along the length of a path. Intuitively, this can be used when it is beneficial to capture the flow as early on the path as possible, for example with drunk drivers. A greedy heuristic, an ascent heuristic and a local search were presented for M3. The different models and heuristics were computationally tested on random networks. The heuristics all perform well and within a small range of each other.

2.2.2 STOCHASTIC FLOW CAPTURING LOCATION MODEL

The FCLM was first proposed in a deterministic setting, meaning that knowledge of all paths with non-zero flows is necessary as input data. However, in most real-life situations, these paths and their flows are not known exactly and are subject to stochasticity. Therefore, several different approaches to include stochasticity in the model have been proposed in literature. This includes using Markov Decision Processes and simulations.

STOCHASTIC APPROACH USING MARKOV DECISION PROCESSES

For traffic applications, origin-destination pairs are often used as input data with the assumption of shortest paths between origin and destination. This assumption does not represent reality fully for traffic applications, for instance, because drivers are more likely to take the fastest route which is not always the shortest route. The shortest path assumption is even less appropriate for the problem of plastic waste propagation through water, since environmental factors such as wind, water flow and vegetation causes the plastics to move in paths that are likely not the shortest possible between origin and destination. An alternative representation of the data is possible using transition probabilities. This leads to a transition matrix between nodes of the network, making the process equivalent to a Markov Decision Process (MDP).

This MDP approach of the FCLM was proposed by Berman et al. (1995b) and

is appropriate when exact numbers of flows on paths are unknown, but reliable knowledge about the probability that demand flow moves from node i to node j is known. Artificial nodes are added to the network for nodes that are a destination for some flow in the network. Input data for the MDP model include the transition matrix between all nodes and an initial distribution vector. The objective function which maximizes the captured flow is now no longer a linear function, since a matrix inversion of the transition matrix is necessary to calculate the objective value. However, using average-reward MDP theory in (Berman et al., 1995b) and total-reward MDP theory in (Berman et al., 1997), this non-linear model was rewritten to an Integer Linear Programming (ILP) formulation. The equivalence between the non-linear model and the derived ILP was also proven without MDP theory in (Berman et al., 1995c).

A greedy heuristic (similar to the previously mentioned heuristics) with a worst case analysis was presented for the MDP version of the FCLM (Berman et al., 1995c). The nodes with the largest improvement of the MDP objective function were sequentially added to the set of facilities. Nodes that become redundant in later iterations of the heuristic are removed in a step that checks their profit margin. The worst case bound derived in the paper is $f_{\text{greedy}}/f_{\text{opt}} \geq 1 - e^{-1}$. Computational results show that the greedy heuristic finds solution within less than 1% of the objective value of the optimal solution for randomly generated test cases with the number of nodes $n \leq 100$ and the number of facilities $m \leq 10$. For large n and m, the results for the greedy heuristic can be obtained much faster than the exact solution can be obtained.

The same authors presented an extension of their research in 1997 where multi-counting of flows captured is taken into account (Berman et al., 1997). They distinguish between two types of multi-counting. In type I models, it is assumed that a flow visiting multiple facilities on a path increases the objective value, but the value of the visits decays with the number of visits on the path. This type of model is appropriate for marketing purposes such as billboards. In type II models, only the first visit by a customer to a facility adds value to the objective function, but there is a probability that a flow is not intercepted by a facility on its path, for example when there is "lack of visibility" or "obscureness" of facilities.

STOCHASTIC APPROACH USING SIMULATIONS

Another way of taking the probabilistic nature of the problem into account is by performing simulations of the demand on possible paths in the network. This has been done in previous research using several different approaches.

Wu and Sioshansi (2017) developed a stochastic version of the FCLM that they refer to as the SFCLM. The SFCLM uses two stages and they apply the model to electric vehicle refueling. In the first stage, a number of charging locations is placed by maximizing the objective function using the expected number of vehicles that can be captured. This expected value cannot be directly computed, and therefore, a sample-average approximation (SAA) of the expected captured flow is used. In the second stage, the actual number of captured vehicles is calculated using FCLM

subproblems of the samples from the SAA-method. This second stage is used to include uncertainty in the flow instead of only taking the expected value into account. For each subproblem, optimality cuts are derived from the dual and these cuts are iteratively added to the first stage problem. The first stage problem is solved again with the added optimality cuts and it is checked whether the solution of the expected value objective function is an overestimation of true second-stage objective function values. If this is the case, then more cuts are added, otherwise the algorithm terminates. The authors refer to these two stages as an L-shaped SAA problem, which is solved using a branch-and-cut solution method. They compared their stochastic model to a deterministic model where the EV flows are set equal to their expected values and show that the stochastic model outperforms the deterministic model when the maximum number of facilities that can be located is small.

A very similar two stage stochastic model that is solved by an L-shaped method is suggested by Tan and Lin (2014). Again, optimal locations are decided in the first stage based on the expected value of captured flow, and in the second stage uncertainty is included using different scenarios.

Yang et al. (2008) used the Hurwicz rule to incorporate stochastic flow in the FCLM. The Hurwicz rule is used to balance between flow scenarios under extremely optimistic and extremely pessimistic values of the stochastic flow. The FCLM with the Hurwicz rule is then solved using a "hybrid intelligent algorithm" that combines stochastic simulation, greedy search and a genetic algorithm. The first chromosomes for the initial population of the genetic algorithm are generated by performing the greedy heuristic as proposed in Berman et al. (1992) on randomly sampled simulations of the flow on the network, together with randomly located facilities as chromosomes. The chromosomes are used to calculate the objective function of the FCLM with the Hurwicz rule, which is then used to calculate the fitness of the chromosomes. Then, a fixed number of new generations in the genetic algorithm is generated by using selection (based on the fitness), crossover and mutation of the population. This hybrid intelligent algorithm is applied to a random graph on a grid with flows generated by a normal distribution.

3

TECHNICAL BACKGROUND

In this chapter, technical background from previous projects at Noria is presented, together with the mathematical details of the Flow Capturing Location Model (FCLM). Based on the literature research presented in Chapter 2, the FCLM seems very suitable to model the location optimization of plastic waste catching systems. Therefore, the mathematical model formulation of the path-based and MDP-based FCLM are introduced.

3.1 PREVIOUS WORK AT NORIA

Since 2018, the company "Noria Sustainable Innovators" (Noria) has been working on catching plastic waste from local waterways. Projects in different regions with several clients have been carried out using their "3R-method": Research, Remove, Reduce. For the *"Research"* phase of their services, they have developed ways to measure the location of plastic debris using GPS tracking and AI-assisted camera monitoring. Furthermore, they developed a model in QGIS that predicts locations where plastics will likely accumulate: the hotspot prediction model. Using the insights from this research, they *"Remove"* plastic waste using several catching systems that Noria developed. The caught plastic is then analysed to find out more about the origin of the plastics to try to *"Reduce"* plastic waste coming from this source in the future. In this section, the model and the different catching systems that Noria developed are described.

3.1.1 HOTSPOT PREDICTION MODEL

As part of the *"Research"* phase, a hotspot prediction model was developed to predict locations where plastic accumulates because it gets stuck. The model is based on a manual GIS-method that consists of three steps. The first step is to consider all potential sources of plastic waste. Examples of sources are restaurants, recreational sites, public wastebins and marketplaces. Then, it is calculated whether plastics from these sources could get stuck based on wind direction, water flow direction, geometry of the water and shore type. For some areas, Noria has observed that routes traveled by plastics are mostly influenced by the wind. This information

was gathered by interviewing local experts and testing with objects marked by GPS sensors. Water flow caused by tides or pumping stations seemed to have little impact in these areas. Therefore, water flow is often omitted in the model. Finally, locations where plastics could likely get stuck are considered. This is based on shore type, vegetation and sharp corners.

Noria converted these three manual steps into an automatic QGIS model with two steps. In step 1, the model calculates all potential hotspot locations where plastics could likely get stuck based on water geometry and wind directions. In step 2, all extra factors in the surrounding that could result in these locations to become a hotspot are considered. The number of surrounding factors that influence each location is used to calculate a percentage that shows how likely this location is a hotspot. A visualization of the hotspot prediction model is shown in Figure 3.1.

The model has a different outcome for each wind direction. Therefore, the model is run using 8 wind directions to see all potential hotspots in an area under different circumstances. Using this method in the area of Delft, the Netherlands, 50-60% of the predicted hotspots were true positives. During the validation of the model, the average wind direction of the past 12 hours was used as input. To find out how long it takes for the plastics to travel a certain distance, further analysis of the trajectory of the plastics for each wind direction and wind strength should be considered, possibly using objects marked by GPS-sensors.

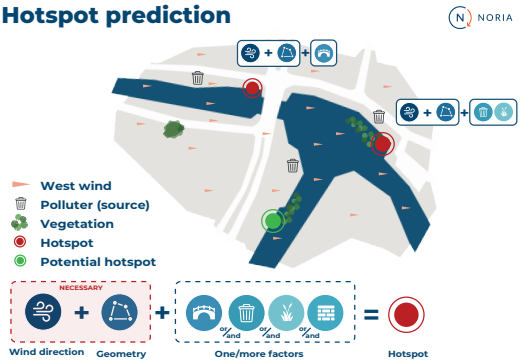


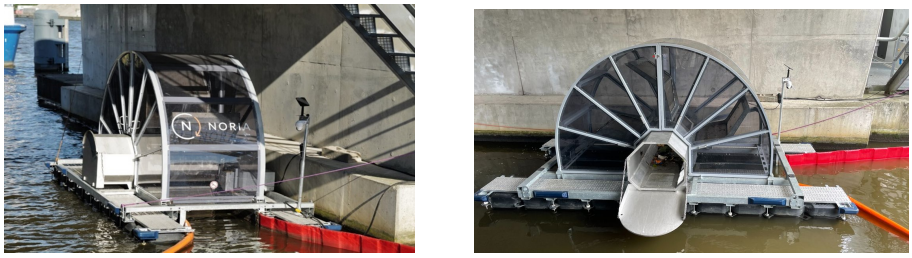
Figure 3.1: Visualization of the QGIS hotspot prediction model designed by Noria.

3.1.2 CATCHING SYSTEMS

There are two types of catching systems that Noria uses to catch floating plastic waste from the water: active and passive catching systems. The active catching system uses electric power to catch plastic with a moving element in the system. The passive catching system does not need any power and uses no movement to catch plastics.

The active catching system "CirCleaner" was developed by Noria and is shown

in Figure 3.2. It has the shape of a wheel with 5 metal shovels that scoop up the plastics during the rotation and lift them into the container in the hollow axis of the wheel. The container has a 0.65m^3 capacity and is emptied manually from the side. Emptying the CirCleaner happens once every 2 weeks at a location with an average amount of plastic flow. The wheel and container are covered with a cap, such that plastics cannot be blown out of the container by strong winds. The CirCleaner has a diameter of 3.5 meters and rotates against the direction of the incoming plastic flow. The rotation uses gear drive that can be powered using an outlet from shore or using solar energy. The CirCleaner rotates slowly with approximately 0.2 rotations per minute, such that the water flow is interrupted as little as possible and plastic is not being pushed out of the system.



(a) (b)

Figure 3.2: Front and side view of Noria's CirCleaner.

A passive system "CanalCleaner" was also developed recently. The system is a floating cage covered by a wooden raft with an opening on one side and is shown in Figure 3.3a. Plastics enter the opening of the CanalCleaner when the water flow direction is towards the opening of the system. When plastics enter the opening of the passive system, they enter a cage with a capacity of 3.7m^2 for plastics floating on the water surface. The height of the cage above the water surface is approximately 30cm. Plastic waste in the water generally does not stick out high above the water surface. The plastic is not lifted from the water by the CanalCleaner, but is floating in the water inside the cage under the wooden raft. To ensure that the plastics do not float back out of the system when the flow direction of the water changes, trapping arms are built inside the cage to prevent movement back towards the opening of the catching system. The trapping arms are built such that plastics can easily flow into the system but get caught in the arms if they are flowing out of the system.

Floating lines are attached to both active and passive systems to guide plastics to the opening of the system from a larger width than the width of the system. This allows for blocking the flow of plastics past a certain location as much as possible. Red and orange floating lines can be seen in Figure 3.2 and Figure 3.3b and yellow floating lines in Figure 3.3a. The length of the floating lines can be adjusted for each location and are installed such that they guide as much plastic as possible to the opening of the system without blocking ship routes.

The CirCleaner can store more plastic waste than the CanalCleaner, since it lifts the plastics from the water into a separate compartment and is not limited by the

(a) (b)

Figure 3.3: Front view of Noria's CanalCleaner. In Delft on the left image and in Amsterdam after Kingsday on the right image.

surface where plastics float in the water in the CanalCleaner. Therefore, the passive system should be manually emptied more often than the active system if the plastic flux is the same. In practice, the CirCleaner is chosen when a location has a very high amount of plastic flow, because the storage capacity of the CirCleaner can be increased more easily. Furthermore, there is a small probability that plastics that were initially caught by the passive system flow back out of the system, since they are not lifted from the water. Therefore the probability that a piece of plastic is caught by the CanalCleaner is slightly lower than the probability that a piece of plastic is caught by the CirCleaner. Finally, when there is a strong influence of tides on the water flow, the passive system is less suitable. However, the passive system is cheaper to purchase than the active system. The exact values of variables of the two types of catching systems that are used in the final model are reported in Chapter 6.

3.2 MATHEMATICAL BACKGROUND FCLM

In this section, we discuss two versions of the FCLM: the FCLM on a network with flows on paths and the Markov Decision Process (MDP) approach to the FCLM with transition probabilities for the flow. These two different models are referred to as the "path-based FCLM" and "MDP-based FCLM" from now on. In Chapter 4, the two models are adjusted to fit the problem of this thesis.

3.2.1 PATH-BASED FCLM

In the path-based FCLM on a network $G(V,E)$, the set P of all nonzero flow paths and the amount of flow f_p on each path $p \in P$ is known. The set of nodes on the path $p \in P$ is denoted as N_p. The set of nodes $I \subseteq V$ on the network represents possible locations for the facilities. A binary decision variable x_i is introduced for

each location $i \in I$:

$$x_i = \begin{cases} 1 & \text{if a facility is placed at location } i \in I, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

A second binary decision variable y_p is introduced for each path $p \in P$:

$$y_p = \begin{cases} 1 & \text{if at least one facility is located on path } p \in P, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

The path-based FCLM formulation as presented in Berman et al. (1992) is shown in Model (3.3). The model maximizes the amount of flow that is captured under Constraint (3.3b), which ensures that a maximum number of m facilities is placed. Constraint (3.3c) ensures that a path is only considered as captured ($y_p = 1$) if there is at least one catching system placed on the path.

$$\text{maximize} \quad \sum_{p \in P} f_p y_p \tag{3.3a}$$

$$\text{subject to} \quad \sum_{i \in I} x_i \leq m, \tag{3.3b}$$

$$\sum_{i \in N_p} x_i \geq y_p, \qquad \forall p \in P, \tag{3.3c}$$

$$x_i \in \{0,1\}, \qquad \forall i \in I, \tag{3.3d}$$

$$y_p \in \{0,1\}, \qquad \forall p \in P. \tag{3.3e}$$

Table 3.1: Notation of Model (3.3).

Notation	Description
Set	
P	set of non-zero flow paths on the network
I	set of locations where a catching system could be located
N_p	set of locations that are on the path $p \in P$
Decision variables	
x_i	1 if a catching system is placed at location $i \in I$, 0 otherwise
y_p	1 if at least one catching system is placed on path $p \in P$, 0 otherwise
Input parameters	
f_p	amount of plastic flow on path $p \in P$
m	number of facilities

3.2.2 MDP-BASED FCLM

The Markov Decision Process (MDP) approach of the FCLM is based on Berman et al. (1995c). In this case, a transition matrix is used as input to the model instead of the set of paths P and their flows f_p.

Let $V = \{1, ..., n\}$ be the set of nodes on the network where the flow travels. The decision variable x_i is introduced, which indicates whether a facility is placed at location $i \in V$ ($x_i = 1$) or not ($x_i = 0$). This is the same definition as the decision variable in Equation (3.1), which was used in the path-based FCLM shown in Model (3.3). We refer to the n-dimensional binary vector $\boldsymbol{x}$ as the "location vector".

The flow paths are now represented by probabilities to start in a location, transition probabilities to move between locations $i \in V$ and $j \in V$ and probabilities to end the path in a location. Let $\boldsymbol{b}$ be the initial distribution factor, such that b_i gives the fraction of the total amount of flow that starts at node $i \in V$. A common sink node "0" is added to the model such that flow that arrives at the final node of its path can be directed to this sink node. An artificial sink node "$n+1$" is also added to direct the intercepted flow to this node. These two extra nodes thus represent a state where flow has ended its path, either by reaching the end of the path, or by being captured. The nodes 0 and $n+1$ are the absorbing states of this Markov chain. This means that once the system enters these states, the system will return to that state infinitely often with probability 1. The other nodes $i \in V = \{1, ..., n\}$ are transient states, which means that once the system enters that state, the system will return to that state only finitely often with probability 1 (Kemeny and Snell, 1960).

The initial probabilities of starting in the sink node or the artificial node are equal to zero, $b_0 = 0, b_{n+1} = 0$. Let q_{ij} represent the probability that flow moves from node $i \in V$ to node $j \in V \cup \{0\}$ when there are no facilities in the network. The probability q_{i0} corresponds to the probability that flow ends in node $i \in V$. Then, we know that $\sum_{j=0}^{n} q_{ij} = 1, \forall i \in V$.

The transition probabilities t_{ij} between nodes $i, j \in V \cup \{0, n+1\}$ are introduced when there are facilities in the network. These transition probabilities are based on the locations of the placed facilities $\boldsymbol{x}$, the effectivity of the facilities and on the transition probabilities q_{ij} when there are no facilities in the network.

First of all, we set $t_{00} = 1$ and $t_{0i} = 0, \forall i \in V \cup \{n+1\}$, since flow that reaches the destination state 0 cannot transition back into the network or to the intercepted state $n+1$. For node $n+1$ that represents the intercepted state, we set $t_{n+1,n+1} = 1$ and $t_{n+1,i} = 0, \forall i \in V \cup \{0\}$, since intercepted flow also cannot transition back into the network. This defines the first and last row of the $(n+2) \times (n+2)$ Markov transition matrix $T^+(\boldsymbol{x})$ given in Equation (3.5) that describes the Markov chain. The remaining entries of this transition matrix are dependent on the choice of locations where catching systems are placed, denoted by $\boldsymbol{x}$.

The model uses a factor β_i that indicates which fraction of the flow through node $i \in V$ is captured when there is a facility in location $i \in V$, thus representing the capturing effectivity of the facility. Intuitively, the transition probabilities $t_{ij}(\boldsymbol{x})$ to move from node $i \in V$ to node $j \in V \cup \{0\}$ given the location vector of the placed facilities are equal to q_{ij} when no facility is located in node $i \in V$ ($x_i = 0$). However, if it is decided that there is a facility placed in node $i \in V$ ($x_i = 1$), then $t_{ij}(\boldsymbol{x})$ is not equal to q_{ij}. In this case, the flow is either intercepted with probability β_i, and thus, transitions from node $i \in V$ to node $n+1$, or it is not intercepted with probability $1-\beta_i$ and moves further in the network according to transition probabilities q_{ij}.

Similar to the path-based FCLM, facility locations can only be placed on the nodes of the network. In the path-based FCLM, this assumption is without loss of generality, however, in the MDP approach, this is not the case. An example is shown by Berman et al. (1997). The example is shown in Figure 3.4. We have a network with n nodes, for which $b_1 = 1, b_i = 0$ for $i = 2,3,...,n$ and $q_{12} = 1, q_{2i} = \frac{1}{n-2}$ for $i = 3,4,...,n$. Suppose that we are allowed to place 3 catching systems ($m = 3$), and $\beta_i = \beta < 1$ for all $i \in V$, which means that a facility does not catch 100% of the flow that passes the facility. Then, the optimal locations for facilities are nodes $i = 1$ and $i = 2$ and a facility on the edge $(1,2)$ as long as $n \geq 4$. It is assumed in this thesis that facilities can only be located at the nodes of the network, similar to previous research on the MDP-based FCLM.

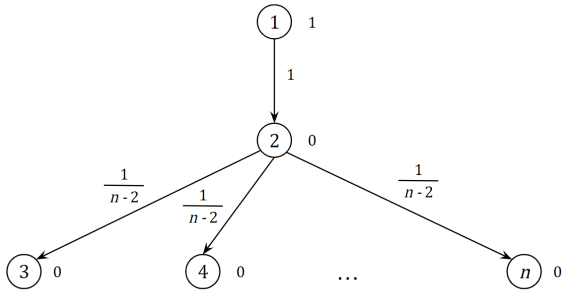


Figure 3.4: Example where optimal locations are not on nodes but also on edges (Berman et al., 1997). The initial probability b_i of each node $i \in V$ is shown on the right of each node. The transition probability of each edge q_{ij} is shown next to each edge (i, j).

The location vector $\boldsymbol{x}$ and the factors β_i define the rest of the entries of the $(n+2) \times (n+2)$ Markov transition matrix $T^+(\boldsymbol{x})$ as follows:

$$\begin{aligned} t_{ij}(\boldsymbol{x}) &= (1 - \beta_i x_i) q_{ij} \quad && \forall i \in V, j \in V \cup \{0\}, \\ t_{i,n+1}(\boldsymbol{x}) &= \beta_i x_i && \forall i \in V. \end{aligned} \tag{3.4}$$

The transition matrix can be decomposed as follows:

$$T^+(\boldsymbol{x}) = \begin{bmatrix} 1, & 0,\ldots,0 & 0 \\ \boldsymbol{t}_0 & T(\boldsymbol{x}) & \boldsymbol{t}_{n+1}(\boldsymbol{x}) \\ 0, & 0,\ldots,0 & 1 \end{bmatrix}, \tag{3.5}$$

where $\boldsymbol{t}_0$ is a column vector containing the transition probabilities to go from a state $i \in V$ to state 0, and $\boldsymbol{t}_{n+1}$ is a column vector containing the transition probabilities to go from a state $i \in V$ to the captured state $n+1$. These two column vectors show the transition probabilities from transient states to the two absorbing states of the Markov chain. The entries of the matrix $T(\boldsymbol{x})$ are the transition probabilities within the set of transient states $i \in V$.

According to Markov chain techniques, the probability of reaching node $j = n+1$ (i.e. the flow is captured) is the hitting probability of reaching the state $j = n+1$

from the initial distribution. The hitting probability h_{ij} is defined as the probability of ever reaching state j starting from initial state i and can be calculated with the following equation:

$$h_{ij} = \begin{cases} 1 & \text{if } i = j, \\ \sum_{k \in V} t_{ik}(\boldsymbol{x}) h_{kj} & \text{otherwise.} \end{cases} \tag{3.6}$$

Using Equation (3.6), we can derive the following matrix equation to calculate the hitting probability $\boldsymbol{h} = [h_{1,n+1}, h_{2,n+1}, \ldots, h_{n,n+1}]^T$ from any initial node $i \in V$ to the recurrent captured state $n+1$:

$$\tilde{\boldsymbol{h}} = \begin{bmatrix} h_{1,n+1} \\ h_{2,n+1} \\ \vdots \\ h_{n,n+1} \\ h_{n+1,n+1} \end{bmatrix} = \begin{bmatrix} T(\boldsymbol{x}) & \boldsymbol{t}_{n+1}(\boldsymbol{x}) \\ 0, \ldots, 0 & 1 \end{bmatrix} \begin{bmatrix} h_{1,n+1} \\ h_{2,n+1} \\ \vdots \\ h_{n,n+1} \\ h_{n+1,n+1} \end{bmatrix} \tag{3.7}$$

Using the fact that $h_{n+1,n+1} = 1$, we obtain the following equation for $\boldsymbol{h}$:

$$\boldsymbol{h} = T(\boldsymbol{x})\boldsymbol{h} + \boldsymbol{t}_{n+1}(\boldsymbol{x}). \tag{3.8}$$

The solution to this equation is $\boldsymbol{h} = [I - T(\boldsymbol{x})]^{-1} \boldsymbol{t}_{n+1}(\boldsymbol{x})$. By taking the inner product with the initial distribution vector $\boldsymbol{b}$, we find the proportion of the total flow in the Markov chain that ends up in state $n+1$ when starting from this initial distribution: $\boldsymbol{b}^T [I - T(\boldsymbol{x})]^{-1} \boldsymbol{t}_{n+1}(\boldsymbol{x})$.

The objective function of the MDP-based FCLM is to maximize this quantity under the same constraints as in Section 3.2.1. The objective function of the model proposed by Berman et al. (1995c) also takes revenue and cost of facilities into account. The input parameter r is the expected value of revenue when all the flow in the network is captured. The input parameter c_i gives the costs that are associated with placing a facility in location $i \in V$. This value includes set-up costs and fixed operating costs of a facility at each location. This leads to the mixed integer Non-Linear Program **(NLP1)** in Model (3.9):

(NLP1)

$$\text{maximize} \quad r\boldsymbol{b}^T [I - T(\boldsymbol{x})]^{-1} \boldsymbol{t}_{n+1}(\boldsymbol{x}) - \sum_{i \in V} c_i x_i \tag{3.9a}$$

$$\text{subject to} \quad \sum_{i \in V} x_i \leq m, \tag{3.9b}$$

$$x_i \in \{0, 1\}, \forall i \in V. \tag{3.9c}$$

Table 3.2: Notation of Model (3.9).

Notation	Description
Set	
V	set of locations where a catching system could be located
Decision variables	
x_i	1 if a catching system is placed at location $i \in V$, 0 otherwise
Input parameters	
r	expected value of revenue when all the flow in the network is captured
$\mathbf{b}$	initial distribution vector of plastics
$T(\boldsymbol{x})$	transition matrix between states $i \in V, j \in V \cup \{0\}$
$\boldsymbol{t}_{n+1}(\boldsymbol{x})$	transition vector between states $i \in V, j = n+1$
c_i	all costs associated with placing a facility at location $i \in V$
m	maximum number of facilities

The following linear Mixed Integer Program **(MIP1)** can be derived from nonlinear Model (3.9) using total-reward Markov Decision Processes (Berman et al., 1997). The extra decision variables v_{i1} and v_{i2} are introduced for all $i \in V$ in the formulation of this equivalent linear model.

(MIP1)

$$\text{maximize} \quad r \sum_{i \in V} \beta_i v_{i2} - \sum_{i \in V} c_i x_i \tag{3.10a}$$

$$\text{subject to} \quad v_{i1} + v_{i2} - \sum_{j \in V} q_{ji} v_{j1} - \sum_{j \in V} (1 - \beta_j) q_{ji} v_{j2} = b_i, \qquad \forall i \in V, \tag{3.10b}$$

$$\sum_{i \in V} x_i \leq m, \tag{3.10c}$$

$$v_{i1} \leq M_{i1}(1 - x_i), \qquad \forall i \in V, \tag{3.10d}$$

$$v_{i2} \leq M_{i2} x_i, \qquad \forall i \in V, \tag{3.10e}$$

$$v_{i1}, v_{i2} \geq 0, x_i \in \{0, 1\}, \qquad \forall i \in V. \tag{3.10f}$$

Constraints (3.10d), and (3.10e) are big-M constraints that ensure that $v_{i1} = 0$ when $x_i = 1$ and $v_{i2} = 0$ when $x_i = 0$. The values M_{i1} and M_{i2} are included to ensure that the constraints are as tight as possible to make it easier to solve MIP1. First, we show the equivalence between **NLP1** and **MIP1**. Then, the interpretation of the extra decision variables v_{i1} and v_{i2} and the values of M_{i1} and M_{i2} are shown.

Let $\boldsymbol{v}_1 = [v_{11}, ..., v_{n1}]$ and $\boldsymbol{v}_2 = [v_{12}, ..., v_{n2}]$, then Theorem 1 shows the equivalence between Model (3.9) and Model (3.10) (Berman et al., 1995c):

Theorem 1.

*(a) If $(\boldsymbol{x}^*, \boldsymbol{v}_1^*, \boldsymbol{v}_2^*)$ is an optimal solution to **MIP1** then $\boldsymbol{x}^*$ is an optimal solution to **NLP1**.*

*(b) If $\boldsymbol{x}^*$ is an optimal solution to **NLP1**, then there exist vectors $\boldsymbol{v}_1^*, \boldsymbol{v}_2^*$ such that $(\boldsymbol{x}^*, \boldsymbol{v}_1^*, \boldsymbol{v}_2^*)$ is an optimal solution to **MIP1**.*

*(c) The optimal values of **MIP1** and **NLP1** are identical.*

Proof. This is proven in two parts.

1. Every feasible solution $\boldsymbol{x}$ of problem **NLP1** has a unique feasible solution $(\boldsymbol{x}, \boldsymbol{v}_1, \boldsymbol{v}_2)$ to **MIP1**.
2. The objective functions of **NLP1** and **MIP1** are equivalent.

Part 1: Suppose we have a location matrix $\boldsymbol{x}$. This uniquely defines the entries of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ because of Constraints (3.10b), (3.10d), and (3.10e). For n of the $2n$ entries of v_{i1} and v_{i2}, we know that they are zero because of Constraints (3.10d) and (3.10e):

$$\begin{aligned} v_{i1} = 0, \forall i \in V \text{ when } x_i = 1, &\qquad \text{because of Constraint (3.10d),} \\ v_{i2} = 0, \forall i \in V \text{ when } x_i = 0, &\qquad \text{because of Constraint (3.10e).} \end{aligned}$$

We can write a new auxiliary vector $\boldsymbol{z}$ filled with the remaining n entries of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ that have not been set to zero so far:

$$z_i(\boldsymbol{x}) = \begin{cases} v_{i1} & \text{if } x_i = 0, \\ v_{i2} & \text{if } x_i = 1. \end{cases}$$

The vector $\boldsymbol{z}$ can be used to write the n constraints from (3.10b) as the following equation:

$$\boldsymbol{z}^T [I - T(\boldsymbol{x})] = \boldsymbol{b}^T.$$

The matrix $[I - T(\boldsymbol{x})]^{-1}$ is also referred to as the fundamental matrix (Kemeny and Snell, 1960). Since $T(\boldsymbol{x})$ is the transition matrix between the transient states of our absorbing Markov chain, we know that $T(\boldsymbol{x})^n$ tends to the zero matrix as $n \to \infty$. Therefore, we know that the inverse $[I - T(\boldsymbol{x})]^{-1}$ is well defined and

$$\boldsymbol{z}^T = \boldsymbol{b}^T [I - T(\boldsymbol{x})]^{-1} \tag{3.11}$$

defines the remaining n nonzero values of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ uniquely.

Part 2: Using the definition of $\boldsymbol{t}_{n+1}(\boldsymbol{x})$ and (3.11), we can see that the objective functions are equivalent as shown in the following equation:

$$\begin{aligned} \boldsymbol{b}^T [I - T(\boldsymbol{x})]^{-1} \boldsymbol{t}_{n+1}(\boldsymbol{x}) &= \boldsymbol{b}^T [I - T(\boldsymbol{x})]^{-1} [\beta_1 x_1, \ldots, \beta_n x_n]^T \\ &\overset{(3.11)}{=} \boldsymbol{z}^T [\beta_1 x_1, \ldots, \beta_n x_n]^T = \sum_{i \in V} \beta_i x_i z_i = \sum_{i \in V} \beta_i v_{i2}(\boldsymbol{x}) \end{aligned}$$

The last step of the equation follows from the fact that the entries z_i where $x_i \neq 0$ are equal to v_{i2}.

Statements (a), (b), and (c) of the theorem follow directly from part 1 and 2 of the proof. □

From Equation (3.11), we can see that

$$v_{i1} = \begin{cases} \left(\boldsymbol{b}^T [I - T(\boldsymbol{x})]^{-1}\right)_i & \text{if } x_i = 0, \\ 0 & \text{if } x_i = 1, \end{cases} \tag{3.12}$$

and

$$v_{i2} = \begin{cases} 0 & \text{if } x_i = 0 \\ \left(\boldsymbol{b}^T [I - T(\boldsymbol{x})]^{-1}\right)_i & \text{if } x_i = 1. \end{cases} \tag{3.13}$$

As explained by Berman et al. (1995c), v_{i1} is the expected number of visits to node $i \in V$ prior to interception by some facility or reaching the destination node 0. This can be seen from the fact that according to Markov chain theory, the fundamental matrix $[I - T(\boldsymbol{x})]^{-1}$ gives the expected number of visits to transient state $j \in V$ from transient state $i \in V$ before hitting an absorbing state (state 0 or $n+1$).

v_{i2} is the expected number of visits to a facility node $i \in V$. The probability of being intercepted in node $i \in V$ is zero when there is no facility in the node ($x_i = 0$). When there is a facility in node $i \in V$, we can see that v_{i2} is equal to the expected number of visits to the facility node. We can also see this from the equivalence between the objective functions of **NLP1** and **MIP1**: $\boldsymbol{b}^T [I - T(\boldsymbol{x})]^{-1} \boldsymbol{t}_{n+1}(\boldsymbol{x}) = \sum_{i \in V} \beta_i v_{i2}$. The left hand side is equal to the proportion of flow that is caught, calculated by the hitting probability of state $n+1$ from initial distribution $\boldsymbol{b}$. Since β_i is the proportion of flow in a node $i \in V$ that is intercepted when there is a facility, we know that v_{i2} should thus be equal to the expected number of visits of flow through this node to make sure that the right hand side of this equation is equal to the left hand side.

Given these interpretations of the variables, it becomes apparent that M_{i1} is the maximum expected number of visits to node $i \in V$ (under any location vector $\boldsymbol{x}$), and M_{i2} is the maximum expected number of visits to a facilitiy node $i \in V$. The values of M_{i1} and M_{i2} can be computed directly. The maximum expected number of visits to a node $i \in V$ is reached when there are no facilities in the network. In that case, $\boldsymbol{x} = \boldsymbol{0}$ and $T(\boldsymbol{x}) = T$ where the entries T_{ij} equal q_{ij} for $i, j \in V$, since there are no facilities in the network. Therefore,

$$M_{i1} = \left(\boldsymbol{b}^T [I - T]^{-1}\right)_i. \tag{3.14}$$

The maximum number of visits to a facility node $i \in V$ that can be intercepted is reached when there is a facility in location $i \in V$ ($x_i = 1$), but there are no other facilities in the network ($x_j = 0, \forall j \in V \setminus \{i\}$). In that case, using $B_i = \text{diag}[1, ..., 1, (1 - \beta_i), 1, ..., 1] \in \mathbb{R}^{n \times n}$, we have $T(\boldsymbol{x}) = B_i T$ and

$$M_{i2} = \left(\boldsymbol{b}^T [I - B_i T]^{-1}\right)_i. \tag{3.15}$$

4

PROBLEM FORMULATION

In this chapter, the problem description and model formulation of the problem of locating plastic waste catching systems in water as introduced in Chapter 1 are further specified. Previous research described in Chapter 2 shows that two mathematical models seem most appropriate to model the problem; a path-based and an MDP-based approach of the Flow Capturing Location Model.

The main difference between the two models is the representation of the flow that is captured. For the path-based FCLM, the flow is represented by a set of nonzero flow paths P and the amount of flow f_p on each path $p \in P$. For the MDP-based FCLM, the flow is represented by a transition matrix T that contains transition probabilities between the nodes in the network. Together with an initial distribution $\boldsymbol{b}$, this transition matrix is a probabilistic representation of the flow in the area. The flow that is captured in the model in this thesis is plastic waste flow in local waterways. Theoretically, pieces of plastic waste travel on distinct paths that could be represented by flow paths, however, practically it is impossible to track all pieces of plastic waste in a certain area from the start to the end of their paths. It is much more realistic to calculate and estimate the probability that a piece of plastic waste enters the water, moves from one place to another and gets stuck at a location. This is why the MDP-based FCLM is chosen and adjusted in this chapter to model the Plastic Waste Flow Capturing Location Problem (PW-FCLP).

4.1 PROBLEM DESCRIPTION

The objective of the PW-FCLP is to catch as much plastic waste as possible in a water network, which is done by intercepting plastic waste in the water with a catching system. The water network is represented by a graph $G(V, E)$. The nodes V are the set of locations where catching systems could possibly be placed. The model should then decide whether a catching system is placed at location $i \in V$ or not. Let b_i be the probability that flow starts in location $i \in V$ and let q_{ij} be the transition probability that plastic waste travels between locations $i, j \in V$, where $q_{i0} = 1 - \sum_{j \in V} q_{ij}$ is the probability that plastic gets stuck at a node $i \in V$. The proportion of flow passing each location can then be calculated using Markov chain theory. When we put a limit m on the number of catching locations, the

optimization problem is to find the locations where to place catching systems such that the caught flow is maximized.

In Chapter 2, it was mentioned that different objective functions are used in previous research. As mentioned in Boccia et al. (2009), it is also possible to minimize the number of catching systems such that at least a predefined fraction of the total flow is captured. Since Noria works on projects with clients who generally have a maximum budget for the catching systems, it is more suitable to maximize the amount of flow that is captured under constraints that limit the number of catching systems that can be placed.

This is the simplest form of the problem, however, there are several other factors that should be incorporated in the model to make it more realistic.

First of all, there is a set K of different types of catching systems that could be placed in each location. These are the types of catching systems presented in Section 3.1.2. At most one system of one type can be placed at each location. Some locations in local waterways might not be suitable for all types of catching systems because of the geometry of the water or ship routes. These locations should be considered in the model such that the solution does not contain a catching system of a certain type that cannot be placed in its suggested location.

Secondly, placing a catching system of type $k \in K$ at location $i \in V$, does not mean that all plastic flow passing this location is caught. Intuitively, one can imagine that a certain catching system can block only 80% of the width of a canal at a certain location, because boats should still be able to pass the system. Furthermore, the effectivity of the floating lines should be taken into account, since plastic could be going under the lines to pass the system when there are waves. The catching systems designed by Noria can only catch floating plastics, which is why submerged plastics are not considered in the model. For each location and each catching system, a parameter β_{ik} can be defined as an input for the model such that β_{ik} gives the probability that a piece of plastic floating by location $i \in V$ is caught by a catching system of type $k \in K$ when placed at location $i \in V$. We refer to this factor β_{ik} as the "catching probability" of the catching system of type $k \in K$ in location $i \in V$. The catching probability also takes the probability that plastics escape the catching system into account. It is assumed in this research that the catching systems can always be emptied before they are full, therefore, the capacity of the catching systems is not a limiting factor.

Since waves that influence the effectivity of the floating lines are actually related to passing ships or stormy weather, we could argue that this variable should be stochastic in our model. However, it was decided to take the average probability that a passing piece of plastic is caught and to leave out the variability of this probability in our model. The catching probability can be used to calculate the proportion of plastics that is caught when encountering a catching system. In this way, we make sure that a piece of plastic is counted as caught at most once, by one catching system.

Noria's technology is generally used by organizations that have a fixed budget for plastic waste cleanup. Therefore, costs c_{ik} of each type of catching system $k \in K$ for

each location $i \in V$ and a budget B should be taken into account in the model, such that the sum of the costs of all placed catching systems is less than or equal to the budget of the client. In case there is no budget, $B = \infty$, it is useful to include a weight w in the model that introduces a tradeoff between catching more plastic flow and paying fewer costs. If there is a finite budget B, then this weight could still be introduced with a small value of w to make sure that the model does not decide to place an extra catching system if the gain in the amount of plastics that is caught by this extra system is small compared to the costs of the extra system.

Furthermore, it is extra beneficial when the plastics are caught as early on their paths as possible to minimize their degradation into microplastics and minimize their impact on animals in the ecosystem. There might also be sensitive areas in the water network that need to be extra protected against plastic waste. For example, it could be important to catch plastic before it moves from local rivers and canals into the sea or ocean. It could also be important to prevent plastics from moving between areas with different responsible municipalities or water boards. An impact factor α_i can be included in the model that is high for sensitive areas and low for less sensitive areas. Uncaught plastic flow in the network at each location $i \in V$ is multiplied by this factor α_i and is subtracted from the objective function. In this way, the objective function is higher when more plastics are caught in sensitive areas. If the impact factor is larger than zero but has the same value for all locations $i \in V$, this ensures that the plastics are caught as early as possible on their paths.

Finally, the probabilities q_{ij} of plastics transitioning between nodes $i, j \in V$ and the probabilities b_i of starting or q_{i0} of getting stuck in a node $i \in V$ are not exactly known. There is little data available on the flow of plastics in most areas. Noria uses GPS tracking of plastics, AI camera's for counting plastics and the aforementioned Static Hotspot Prediction Model with publicly available environmental data to investigate the flow and accumulation of plastics in several areas. Translating this data into the input probabilities that represent the plastic flow is a challenge, but very important for the real life performance of the model. The focus of this thesis is to formulate and solve a mathematical model to find the optimal solution. Therefore, data gathering and processing is of secondary importance for the scope of this thesis. A description of the processing of the data into the input parameters $b_i, q_{ij}, \alpha_i, \beta_{ik}, w, c_{ik}$, and B is given in Chapter 6.

Furthermore, the influence of uncertainty of these impact parameters due to variations over time and dependence on environmental factors on the outcome of the model is investigated in a sensitivity analysis, explained in Section 7.1.

4.2 MODEL FORMULATION

In this section, the aforementioned aspects of the problem are translated into a mathematical model with decision variables, constraints and an objective function. The Flow Capturing Location Model as presented in Section 2.2 and Section 3.2 is used to design a new version of the model based on Markov Decision Processes, which uses transition probabilities for the plastics between locations in the water network as input. The new model is called the Plastic Waste Flow Capturing Location Model (PW-FCLM).

4.2.1 ORIENTATION OF THE CATCHING SYSTEM

In Section 3.1.2, it is shown that the catching systems designed by Noria all have an opening on one side where the plastics can enter the system. This means that when plastic flows arrive in a location $i \in V$ coming from different angles, we cannot assume that all this flow is in the correct direction to be captured by the catching system. To ensure that this is addressed in the model, we add extra nodes and use directed edges instead of undirected edges. This means that we double the nodes that are on straight water segments of the network with directed edges between them. Half of the nodes represent catching systems oriented with their opening in one direction, the other half has its opening in the other direction. It is unlikely that catching systems are placed at intersections where the flow is coming from three or more directions, since it is usually not possible to block a large part of the waterway with floating lines at intersections because boats need to be able to turn there. Therefore, it is not necessary to add extra nodes at intersections.

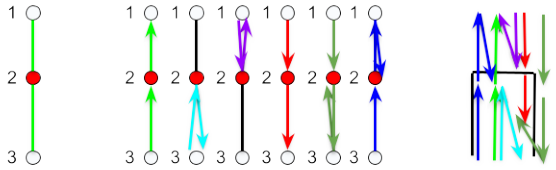


Figure 4.1: Visualization of the importance of the orientation of the catching system.

The problem is visualised in Figure 4.1. Red nodes represent locations where a catching system is placed. On the left of the figure, we see that when a catching system is placed on a flow path (green), it seems that the flow is captured. In most versions of the FCLM this is not a problem, because in applications such as refueling stations it is usually quite easy for a car to change direction to enter the station, no matter its orientation. In the middle of the figure, we see six different possible paths for the plastic waste flow on this same segment. On the right of the figure, we see that the red and purple paths would actually not be caught by the system if the opening of the system is oriented as shown.

Note that it is possible for plastics to turn around, for example when the wind changes. When a piece of plastic turns around after reaching node $i \in V$, it is important to take into account that this plastic is caught when the orientation of the catching system is correct for the incoming direction of the plastic (such as shown in the cyan blue path). This is because we saw in Section 3.1.2 that both passive and active catching systems are designed to trap the plastics, even when the direction of the plastic flow changes after they are caught. When plastic reaches node $i \in V$ from the closed side of the catching system and then turns around, it should not be considered caught by the model (such as shown in the purple path). However, when

the plastic passes the system from the "wrong direction" but then turns around to enter the opening of the system, it should be considered caught (such as shown in the dark green path). Plastics traveling the dark blue path are also considered caught, since they only turn around after being caught by the system.

This problem is solved by adjusting the network by doubling the nodes and adding the directed edges as shown in Figure 4.2 on the left. The orientation of the catching system corresponding to each node is shown on each side. In the two visualizations on the right of Figure 4.2, we can see that this works because the light green, cyan blue, dark green paths and dark blue paths go through the red node with the catching system and are considered caught. The red and purple paths do not go through a facility node and are not considered caught. This is shown for both possible orientations of the catching systems.

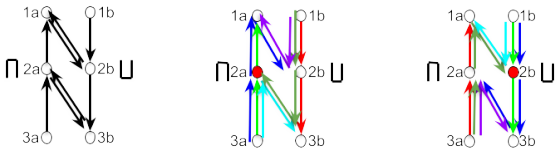


Figure 4.2: Visualization adjusting the network to take the orientation of the catching system into account.

4.2.2 MDP-BASED FCLM

In the problem description in Section 4.1, more parameters are introduced than in the basic MDP-based FCLM shown in Model (3.10) presented by Berman et al. (1995c). These parameters are introduced to take all aspects of the location optimization problem of plastic waste catching systems into account. This includes the possibility of several types of catching systems $k \in K$, impact factors for sensitive areas α_i, the effectivity of the catching systems β_{ik}, a weight w to trade off the amount of plastics caught with the costs, costs of catching systems c_{ik} and total budget B. These aspects lead to the following adjustments to Model (3.10) presented by Berman et al. (1995c).

Let $V = \{1, ..., n\}$ be the set of nodes on the network where the plastic waste travels and the locations where a catching system could be placed. Since the goal is to catch as much plastic waste from the water network as possible by placing catching systems, the model should decide in which locations $i \in V$ a catching system of type $k \in K$ needs to be placed. Therefore, a binary decision variable with a double index is introduced:

$$x_{ik} = \begin{cases} 1 & \text{if a catching system of type } k \in K \text{ is placed at location } i \in V, \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

The constraint:

$$\sum_{k \in K} x_{ik} \leq 1, \qquad \forall i \in V, \tag{4.2}$$

is added to ensure that only 1 catching system of 1 type is allowed to be placed in each location.

The difference with the original model by Berman et al. (1995c) is the extra index $k \in K$ which is added to the decision variable to include the possibility of placing different types of catching systems. To account for locations $i \in V$ where a catching system of type $k \in K$ cannot be placed, we add the following constraints:

$$x_{ik} = 0, \qquad \forall k \in K \backslash K_i, i \in V, \tag{4.3}$$

where $K_i \subseteq K$ is the set of catching systems that can be placed in location $i \in V$. In Section 4.2.1, it was mentioned that catching systems are not placed at intersections due to the orientation of the catching systems. Therefore, $K_i = \emptyset$ for the nodes $i \in V$ that are intersections.

Let $\boldsymbol{b}$ be the initial distribution factor, such that b_i gives the fraction of the total amount of plastic waste that starts at node $i \in V$. Similar to the model proposed by Berman et al. (1995c), a common sink node "0" is added to the model, such that plastics that do not travel further in the network can be directed to this sink node. An artificial sink node "$n+1$" is also added to direct the captured flow of plastics to this node. Intuitively, in the case of location optimization of plastic waste catching systems, this means that plastics that are stuck somewhere in the water (e.g. in a corner, vegetation or a boathouse) reach state 0. Plastics that are caught by a catching system reach state $n+1$. The nodes 0 and $n+1$ are the absorbing states of the Markov chain, and the nodes $i \in V$ are the transient states.

The initial probability of plastic waste starting in the sink node b_0 or the artificial node b_{n+1} is equal to zero. Let q_{ij} represent the probability that plastic waste travels from node $i \in V$ to node $j \in V \cup \{0\}$ when there are no catching systems in the network. The probability q_{i0} corresponds to the probability that plastic waste gets stuck in node $i \in V$. We know that $\sum_{j=0}^{n} q_{ij} = 1, \forall i \in V$.

The transition probabilities q_{ij} of plastics moving from location $i \in V$ to $j \in V$ when there are no catching systems are determined from data using wind and geometry of the water. The probability q_{i0} that plastic waste gets stuck in node $i \in V$ is based on surrounding factors such as corners, bridges, vegetation or boats. The details of how these parameters are determined are given in Chapter 6.

The transition probabilities t_{ij} between nodes $i, j \in V \cup \{0, n+1\}$ are introduced when there are catching systems in the network. These transition probabilities are based on the locations of the placed catching systems $\boldsymbol{x}$, the effectivity of the catching systems β_{ik} and on the transition probabilities q_{ij} when there are no catching systems in the network.

As described in Section 3.2.2, we set $t_{00} = 1$ and $t_{0i} = 0$, for all $i \in V \cup \{n+1\}$. For node $n+1$, which represents the captured state, we set $t_{n+1,n+1} = 1$ and $t_{n+1,i} = 0$, for all $i \in V \cup \{0\}$, since we assume that plastics that are caught by the catching system cannot escape from these systems. If it is physically possible for a piece of plastic to escape from a catching system, this probability is included in the effectivity of the catching system β_{ik}.

This defines the first and last row and the first column of the $(n+2)\times(n+2)$ Markov transition matrix $T^+(\boldsymbol{x})$ that describes the Markov chain. The remaining entries of this transition matrix are dependent on the choice of locations where catching systems are placed. This is indicated by the $n\times|K|$-dimensional binary "location matrix" $\boldsymbol{x}$. The $(n+2)\times(n+2)$ Markov transition matrix $T^+(\boldsymbol{x})$ is defined by the location matrix $\boldsymbol{x}$ similarly to Equations (3.4), however, we also sum over the possible types of catching systems $k\in K$ for each entry $i,j\in V$:

$$\begin{aligned} t_{ij}(\boldsymbol{x}) &= \left(1-\textstyle\sum_{k\in K}\beta_{ik}x_{ik}\right)q_{ij} && \forall\, i\in V, j\in V\cup\{0\},\\ t_{i,n+1}(\boldsymbol{x}) &= \textstyle\sum_{k\in K}\beta_{ik}x_{ik} && \forall\, i\in V. \end{aligned} \tag{4.4}$$

Note that Constraints (4.2) ensure that only one catching system could be placed in each location, such that at most one $\beta_{ik}x_{ik}$ term should remain in these sums.

The transition matrix is decomposed as follows:

$$T^+(\boldsymbol{x}) = \begin{bmatrix} 1, & 0,\ldots,0 & 0 \\ \boldsymbol{t}_0 & T(\boldsymbol{x}) & \boldsymbol{t}_{n+1}(\boldsymbol{x}) \\ 0, & 0,\ldots,0 & 1 \end{bmatrix}. \tag{4.5}$$

Again, the column vectors $\boldsymbol{t}_0$ and $\boldsymbol{t}_{n+1}$ represent the probability that plastic waste flow gets stuck or is captured, which are transition probabilities from transient states to the two absorbing states of the Markov chain. The entries of the matrix $T(\boldsymbol{x})$ are the transition probabilities within the set of transient states $i\in V$, which means that plastic is still moving around in the network.

The hitting probability $\boldsymbol{h}=[h_{1,n+1},h_{2,n+1},\ldots,h_{n,n+1}]^T$ of reaching state $n+1$ from any initial state $i\in V$ is given by $\boldsymbol{h}=[I-T(\boldsymbol{x})]^{-1}\boldsymbol{t}_{n+1}(\boldsymbol{x})$ as shown in Section 3.2.2. By taking the inner product with the initial distribution vector $\boldsymbol{b}$, we find the proportion of the total flow in the Markov chain that ends up in state $n+1$ when starting from this initial distribution: $\boldsymbol{b}^T[I-T(\boldsymbol{x})]^{-1}\boldsymbol{t}_{n+1}(\boldsymbol{x})$. This is the proportion of the total flow of plastics that is caught by placing the catching systems according to $\boldsymbol{x}$, which is maximized in the objective function of the model.

To make sure that the total costs do not exceed the total budget B for the area, the following constraint is included in the model:

$$\sum_{i\in V}\sum_{k\in K} c_{ik}x_{ik}\leq B. \tag{4.6}$$

This constraint replaces Constraint (3.9b), which limits the number of placed facilities in Model (3.9). The new constraint in (4.6) takes different costs for different types of catching systems into account.

Furthermore, a term $-w\sum_{i\in V}\sum_{k\in K}c_{ik}x_{ik}$ can be added to the objective function with the weight w that introduces a tradeoff between catching more plastic flow and paying fewer costs. This replaces the use of the variables r and c_i for revenue and costs of facilities in the objective function of the model proposed in Berman et al. (1995c).

Finally, the impact factor α_i is included in the model by multiplying this by the uncaught flow in the network and subtracting this term from the objective function.

From Markov Chain theory, it follows that $\left(\boldsymbol{b}^T[I-T(\boldsymbol{x})]^{-1}\right)_i$ is the expected number of visits to a state $i \in V$ (Kemeny and Snell, 1960). Therefore, we know that for a location $i \in V$ with no catching systems ($\sum_{k\in K} x_{ik} = 0$), the uncaught remaining flow is equal to $\left(\boldsymbol{b}^T[I-T(\boldsymbol{x})]^{-1}\right)_i$. For a location $i \in V$ with a catching system of type $k \in K$ ($x_{ik} = 1$), the uncaught remaining flow is equal to $(1-\beta_{ik})\left(\boldsymbol{b}^T[I-T(\boldsymbol{x})]^{-1}\right)_i$. This leads to a term $-\sum_{i\in V}\alpha_i(1-\sum_{k\in K}\beta_{ik}x_{ik})\left(\boldsymbol{b}^T[I-T(\boldsymbol{x})]^{-1}\right)_i$ in the objective function to take the sensitive areas into account.

The mixed integer Non-Linear Program **(NLP2)** containing all our extra variables and constraints is shown in Model (4.7) below.

(NLP2)

$$\text{maximize} \quad \boldsymbol{b}^T[I-T(\boldsymbol{x})]^{-1}\boldsymbol{t}_{n+1}(\boldsymbol{x}) - w\sum_{i\in V}\sum_{k\in K} c_{ik}x_{ik} \tag{4.7a}$$

$$-\sum_{i\in V}\alpha_i\left(1-\sum_{k\in K}\beta_{ik}x_{ik}\right)\left(\boldsymbol{b}^T[I-T(\boldsymbol{x})]^{-1}\right)_i$$

$$\text{subject to} \quad \sum_{k\in K} x_{ik} \le 1, \qquad \forall i \in V, \tag{4.7b}$$

$$x_{ik} = 0, \qquad \forall i \in V, k \in K\backslash K_i, \tag{4.7c}$$

$$\sum_{i\in V}\sum_{k\in K} c_{ik}x_{ik} \le B, \tag{4.7d}$$

$$x_{ik} \in \{0,1\}, \qquad \forall i \in V, k \in K. \tag{4.7e}$$

Table 4.1: Notation of Model (4.7).

Notation	Description
Set	
V	set of locations where plastic travels
K_i	set of types of catching systems that could be chosen at location $i \in V$
K	set of all possible catching systems
Decision variables	
x_{ik}	1 if a catching system of type $k \in K$ is placed at location $i \in V$, 0 otherwise
Input parameters	
$\boldsymbol{b}$	initial distribution vector of plastics
$T(\boldsymbol{x})$	transition matrix between states $i \in V, j \in V \cup \{0\}$
$\boldsymbol{t}_{n+1}(\boldsymbol{x})$	transition vector between states $i \in V, j = n+1$
w	weight that balances costs vs catching more plastics
c_{ik}	cost of placing catching system of type $k \in K$ in location $i \in V$
B	total budget for the area to place catching systems
α_i	impact factor for sensitive areas

5

SOLUTION METHODS

In this chapter, the non-linear model from Chapter 4 is linearized to the final version of the PW-FCLM. The linearization of the objective function allows it to be solved exactly using existing MILP solvers. Then, we show that the PW-FCLM is NP-hard. Therefore, a greedy heuristic is also presented as an alternative solution method.

5.1 LINEARIZATION OF THE MODEL

The following linear model can be derived from our non-linear Model (4.7):

(MIP2)

$$\text{maximize} \quad \sum_{i\in V}\sum_{k\in K}\beta_{ik}v_{ik2} - w\sum_{i\in V}\sum_{k\in K}c_{ik}x_{ik} - \sum_{i\in V}\alpha_i v_{i1} - \sum_{i\in V}\alpha_i\sum_{k\in K}(1-\beta_{ik})v_{ik2} \tag{5.1a}$$

$$\text{subject to} \quad v_{i1} + \sum_{k\in K}v_{ik2} - \sum_{j\in I}q_{ji}v_{j1} - \sum_{j\in V}\sum_{k\in K}(1-\beta_{jk})q_{ji}v_{jk2} = b_i, \qquad \forall i \in V, \tag{5.1b}$$

$$v_{i1} \le M_{i1}\left(1 - \sum_{k\in K}x_{ik}\right), \qquad \forall i \in V, \tag{5.1c}$$

$$v_{ik2} \le M_{ik2}x_{ik}, \qquad \forall i \in V, k \in K, \tag{5.1d}$$

$$\sum_{k\in K}x_{ik} \le 1, \qquad i \in V, \tag{5.1e}$$

$$x_{ik} = 0, \qquad \forall k \in K\backslash K_i, i \in V, \tag{5.1f}$$

$$\sum_{i\in V}\sum_{k\in K}c_{ik}x_{ik} \le B, \tag{5.1g}$$

$$v_{i1}, v_{ik2} \ge 0. \qquad \forall i \in V, k \in K, \tag{5.1h}$$

$$x_{ik} \in \{0,1\}, \qquad \forall i \in V, k \in K. \tag{5.1i}$$

There are several differences with the original model by Berman et al. (1995c) shown in Model (3.10), such as the extra index $k \in K$ in the variables x_{ik}, the extra terms in the objective function for sensitive areas and the extra constraints in (5.1f) for locations $i \in V$ where no catching systems of type $k \in K$ are allowed. Another important difference with the original model are the extra variables v_{ik2}

and parameters M_{ik2} for the different types of catching systems $k \in K$.

Let $\boldsymbol{v}_1 = [v_{11}, ..., v_{n1}]$ and $\boldsymbol{v}_2 \in \mathbb{R}^{N \times |K|}$ be such that $(\boldsymbol{v}_2)_{ik} = v_{ik2}$, then Theorem 2 shows the equivalence between Model (4.7) and Model (5.1). The proof is very similar to the proof of Theorem 1, except for some details regarding the terms containing the impact factor α_i and the extra variables x_{ik} and v_{ik2} and the parameters M_{ik2}.

Theorem 2.

*(a) If $(\boldsymbol{x}^*, \boldsymbol{v}_1^*, \boldsymbol{v}_2^*)$ is an optimal solution to **MIP2** then $\boldsymbol{x}^*$ is an optimal solution to **NLP2**.*

*(b) If $\boldsymbol{x}^*$ is an optimal solution to **NLP2**, then there exist vectors $\boldsymbol{v}_1^*, \boldsymbol{v}_2^*$ such that $(\boldsymbol{x}^*, \boldsymbol{v}_1^*, \boldsymbol{v}_2^*)$ is an optimal solution to MIP1.*

*(c) The optimal values of **MIP2** and **NLP2** are identical.*

Proof. We prove this in two parts.

1. Every feasible solution $\boldsymbol{x}$ of problem **NLP2** has a unique feasible solution $(\boldsymbol{x}, \boldsymbol{v}_1, \boldsymbol{v}_2)$ to **MIP2**.
2. The objective functions of **NLP2** and **MIP2** are equivalent.

Part 1: Suppose we have a location matrix $\boldsymbol{x}$. This uniquely defines the entries of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ because of Constraints (5.1b), (5.1c), and (5.1d). For the following entries, we know that they are zero because of Constraints (5.1c) and (5.1d):

$$v_{i1} = 0, \forall i \in V \text{ when } \sum_{k \in K} x_{ik} = 1, \qquad \text{because of Constraint (5.1c)},$$
$$v_{ik2} = 0, \forall i \in V, k \in K \text{ when } x_{ik} = 0, \qquad \text{because of Constraint (5.1d)}.$$

This means that the entries v_{i1} are zero when there is a facility of some type $k \in K$ in location $i \in V$, and the entries v_{ik2} are zero when there is no catching system of type $k \in K$ in location $i \in V$. We can write a new auxiliary vector $\boldsymbol{z}$ with the remaining n entries of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ that have not been set to zero so far:

$$z_i(\boldsymbol{x}) = \begin{cases} v_{i1} & \text{if } \sum_{k \in K} x_{ik} = 0, \\ v_{ik2} & \text{if } x_{ik} = 1. \end{cases}$$

This means that an entry of z_i is equal to v_{i1} when there is no catching system in location $i \in V$, and equal to v_{ik2} if there is a catching system of type $k \in K$ in location $i \in V$. The vector $\boldsymbol{z}$ can be used to write the n constraints from (5.1b) as the following equation:

$$\boldsymbol{z}^T [I - T(\boldsymbol{x})] = \boldsymbol{b}^T.$$

Since the matrix $T(\boldsymbol{x})$ is substochastic, the inverse $[I - T(\boldsymbol{x})]^{-1}$ is well defined and therefore

$$\boldsymbol{z}^T = \boldsymbol{b}^T [I - T(\boldsymbol{x})]^{-1} \tag{5.2}$$

defines the remaining n nonzero values of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ uniquely.

Part 2: Using the definition of $\boldsymbol{t}_{n+1}(\boldsymbol{x})$ and Equation (5.2), we can see that the first terms of the objective functions are equivalent as shown in the following equation:

$$\boldsymbol{b}^T[I-T(\boldsymbol{x})]^{-1}\boldsymbol{t}_{n+1}(\boldsymbol{x}) = \boldsymbol{b}^T[I-T(\boldsymbol{x})]^{-1}\left[\sum_{k\in K}\beta_{1k}x_{1k},\ldots,\sum_{k\in K}\beta_{nk}x_{nk}\right]^T$$
$$\overset{(5.2)}{=} \boldsymbol{z}^T\left[\sum_{k\in K}\beta_{1k}x_{1k},\ldots,\sum_{k\in K}\beta_{nk}x_{nk}\right]^T = \sum_{i\in V}\sum_{k\in K}\beta_{ik}x_{ik}z_i = \sum_{i\in V}\sum_{k\in K}\beta_{ik}v_{ik2}$$

The last step of the equation follows from the fact that the entries z_i where $x_{ik}\neq 0$ are equal to v_{ik2}.

For the terms of the objective functions containing the impact factor α_i, we use that

$$v_{i1} = \begin{cases} \left(\boldsymbol{b}^T[I-T(\boldsymbol{x})]^{-1}\right)_i & \text{if } \sum_{k\in K}x_{ik}=0, \\ 0 & \text{if } \sum_{k\in K}x_{ik}=1, \end{cases} \tag{5.3}$$

and

$$v_{ik2} = \begin{cases} 0 & \text{if } x_{ik}=0 \\ \left(\boldsymbol{b}^T[I-T(\boldsymbol{x})]^{-1}\right)_i & \text{if } x_{ik}=1, \end{cases} \tag{5.4}$$

which follows from Equation (5.2). Then, we can see that for all $i\in V$:

$$\alpha_i(1-\sum_{k\in K}\beta_{ik}x_{ik})\left(\boldsymbol{b}^T[I-T(\boldsymbol{x})]^{-1}\right)_i = \begin{cases} \alpha_i v_{i1} & \text{if } \sum_{k\in K}x_{ik}=0 \\ \alpha_i(1-\beta_{ik})v_{ik2} & \text{for } k\in K \text{ where } x_{ik}=1, \text{ if } \sum_{k\in K}x_{ik}=1. \end{cases} \tag{5.5}$$

Statements (a), (b), and (c) of the theorem follow directly from these two parts of the proof. □

The interpretation of the decision variables v_{i1} and v_{ik2} and the values of M_{i1} and M_{ik2} is similar to the interpretation given in Section 3.2.2. v_{i1} is the expected number of visits to node $i\in V$ prior to interception by some facility or reaching the destination node 0. v_{ik2} is the expected number of visits to a facility node $i\in V$ with a catching system of type $k\in K$.

Again, M_{i1} is the maximum expected number of visits to node $i\in V$ (under any location vector $\boldsymbol{x}$), and M_{ik2} is the maximum expected number of visits to facility node $i\in V$ with a catching system of type $k\in K$. The values of M_{i1} and M_{ik2} can be computed directly. The maximum expected number of visits to a node $i\in V$ is reached when there are no facilities in the network ($\boldsymbol{x}=\boldsymbol{0}$ and $T(\boldsymbol{x})=T$). Therefore,

$$M_{i1} = \left(\boldsymbol{b}^T[I-T]^{-1}\right)_i. \tag{5.6}$$

The maximum expected number of visits to facility node $i\in V$ with a catching system of type $k\in K$ is reached when there is a catching system of type $k\in K$ in location $i\in V$

($x_{ik} = 1$), but there are no other facilities in the network ($\sum_{k \in K} x_{jk} = 0, \forall j \in V \setminus \{i\}$). In that case, using $B_{ik} = \text{diag}[1, ..., 1, (1 - \beta_{ik}), 1, ..., 1] \in \mathbb{R}^{n \times n}$, we have $T(\boldsymbol{x}) = B_{ik} T$ and

$$M_{ik2} = \left(\boldsymbol{b}^T [I - B_{ik} T]^{-1}\right)_i. \tag{5.7}$$

There are three different parts in the objective function of Model (5.1) that represent the three different objectives:

1. catching as much plastic waste as possible,

2. spending a proportional amount of money per extra amount of plastic flow that is caught with each additional catching system,

3. catching the plastics as early as possible in the total area and catching them before entering a sensitive area or as early as possible if they enter the sensitive area.

We can split up the objective function of Model (5.1) into three parts:

$$f(\boldsymbol{x}) = \underbrace{\sum_{i \in V} \sum_{k \in K} \beta_{ik} v_{ik2}}_{f_1(\boldsymbol{x})} - \underbrace{w \sum_{i \in V} \sum_{k \in K} c_{ik} x_{ik}}_{f_2(\boldsymbol{x})} - \underbrace{\sum_{i \in V} \alpha_i v_{i1} - \sum_{i \in V} \alpha_i \sum_{k \in K} (1 - \beta_{ik}) v_{ik2}}_{f_3(\boldsymbol{x})}. \tag{5.8}$$

We can clearly distinguish the three objectives in the three parts of the objective function, $f_1(\boldsymbol{x})$, $f_2(\boldsymbol{x})$ and $f_3(\boldsymbol{x})$. Note that these terms only depend on $\boldsymbol{x}$ in this notation, since it was shown in Part 1 of the proof of Theorem 2 that there is a unique solution $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ for each location matrix $\boldsymbol{x}$.

The first part, $f_1(\boldsymbol{x})$, is equal to the total proportion of plastic flow that is caught, the second part, $f_2(\boldsymbol{x})$, is a trade-off term for costs and the third part, $f_3(\boldsymbol{x})$, is the trade-off for the leftover flow in the sensitive areas. To properly trade-off the different objectives contained in the objective function, the values of the parameters w and α_i for each location $i \in V$ should be chosen based on practical wishes of the user of the model. This is discussed in Section 6.3.

5.2 COMPLEXITY ANALYSIS

It is mentioned in Berman et al. (1992) that the path-based FCLM is NP-hard, because the formulation of the path-based FCLM is identical to the formulation of the Maximal Coverage Location Model (MCLM). For the MDP-based FCLM proposed by Berman et al. (1995c), no complexity proof was presented in the literature. Since the formulation of the MDP-based FCLM has a different structure than the path-based FCLM and the MCLM, the complexity proof is not trivial. Therefore, we show NP-hardness of our version of the MDP-based FCLM in Model (5.1) by providing a polynomial time reduction from the knapsack problem.

First of all, the knapsack problem is as follows:

$$\text{maximize} \quad \sum_{i=1}^{n} g_i y_i \tag{5.9a}$$

$$\text{subject to} \quad \sum_{i=1}^{n} s_i y_i \leq W, \tag{5.9b}$$

$$y_i \in \{0,1\}, \qquad \forall i \in \{1,...,n\}, \tag{5.9c}$$

where g_i are the profits of the n possible items and s_i are the weights or sizes of the possible items that should fit within the maximum capacity W of the knapsack.

The decision problem form of the knapsack problem is: does a feasible solution $\boldsymbol{y}$ exists such that the objective value of the solution is at least l?

The decision problem form of Model (5.1) is: does a feasible solution $(\boldsymbol{x}, \boldsymbol{v_1}, \boldsymbol{v_2})$ exist such that the objective value of the solution is at least l'?

The decision problem form of Model (5.1) (**MIP2**) is denoted by **D-MIP2**.

Theorem 3. ***D-MIP2*** *is NP-complete.*

Proof. We prove this in three parts.

1. **D-MIP2** is in NP.
2. There exists a polynomial time reduction $Q(\cdot)$ from an instance I of the knapsack problem to an instance $Q(I)$ of **D-MIP2**.
3. For each instance I of the knapsack problem, I is a yes-instance of the knapsack problem if and only if $Q(I)$ is a yes-instance of **D-MIP2**.

Part 1: Given a certificate $\boldsymbol{x}$, there is a unique feasible solution $(\boldsymbol{x}, \boldsymbol{v_1}, \boldsymbol{v_2})$ as shown in Theorem 2. It was shown that the values v_{i1} and v_{ik2}, for all $i \in V$ and $k \in K$ can be calculated using Equation (5.3) and Equation (5.4), which takes polynomial time since the transition matrix $T(\boldsymbol{x})$ can be computed from $\boldsymbol{x}$ in polynomial time and the inverse matrix $\boldsymbol{b}^T[I - T(\boldsymbol{x})]^{-1}$ can also be computed in polynomial time. Then, given all values x_i, v_{i1} and v_{ik2}, for all $i \in V$ and $k \in K$, it takes polynomial time to calculate whether the instance is a yes-instance: $\sum_{i\in V}\sum_{k\in K}\beta_{ik}v_{ik2} - w\sum_{i\in V}\sum_{k\in K}c_{ik}x_{ik} - \sum_{i\in V}\alpha_i v_{i1} - \sum_{i\in V}\alpha_i\sum_{k\in K}(1-\beta_{ik})v_{ik2} \geq l'$.

Part 2: Given an instance $I = (\boldsymbol{g} \in \mathbb{R}_+^n, \boldsymbol{s} \in \mathbb{R}_+^n, W \in \mathbb{R}_+)$ of the knapsack problem, a polynomial time reduction to an instance $Q(I)$ of **D-MIP2** is given as follows. We set the sets V and K to $V = \{1,...,n\}$ and $K = \{1,...,n\}$. In addition, we set

$$w = 0, \quad \alpha_i = 0, \forall i \in V, \qquad \beta_{ik} = 1, \forall i \in V, k \in K,$$
$$q_{ij} = 0, \forall i, j \in V, k \in K, \qquad q_{i0} = 1, \forall i \in V.$$

This means that $T = \mathbf{0}$, $M_{i1} = b_i$ and $M_{ik2} = b_i$. The knapsack problem with profits g_i has the same solution as the knapsack problem with normalized profits $\tilde{g}_i = g_i / \sum_{i=1}^n g_i$. In our polynomial time reduction, we set $b_i = \tilde{g}_i$. Then, we set $K_i = \{i\}$, for all $i \in V$, such that $x_{ik} = 0$, for all $i \neq k$. Finally, we set $c_{ik} = s_k$, for all

$i \in V$ and $k \in K$ and set $B = W$. The knapsack decision problem with parameter l is now reduced to **D-MIP2** with parameter $l' = l / \sum_{i=1}^{n} g_i$.

Part 3: If I is a yes-instance of the knapsack decision problem, then there exists a solution $\boldsymbol{y}$ such that $\sum_{i=1}^{n} g_i y_i \geq l$. If we set $x_{ii} = y_i$, this defines a unique solution $(\boldsymbol{x}, \boldsymbol{v_1}, \boldsymbol{v_2})$. The solution is feasible, because $\boldsymbol{x}$ satisfies the constraints $\sum_{k \in K} x_{ik} = x_{ii} \leq 1$ and $\sum_{i \in V} \sum_{k \in K} c_{ik} x_{ik} = \sum_{i \in V} s_i x_{ii} = \sum_{i \in V} s_i y_i \leq W = B$ and the unique solution $(\boldsymbol{x}, \boldsymbol{v_1}, \boldsymbol{v_2})$ is feasible as shown in Theorem 2. Using Equation (5.4), we find that $v_{ik2} = 0$ if $x_{ik} = 0$, and $v_{ik2} = b_i$ if $x_{ik} = 1$. Note that we do not have to determine v_{i1} since $\alpha_i = 0$ for all $i \in V$ in the objective function. Then, we can see that the objective function of $Q(I)$ becomes $\sum_{i \in V} \sum_{k \in K} v_{ik2} = \sum_{i \in V} v_{ii2} = \sum_{i \in V} b_i x_{ii} = \sum_{i=1}^{n} \tilde{g}_i y_i \geq l / \sum_{i=1}^{n} g_i = l'$, so $Q(I)$ is a yes-instance of **D-MIP2**.

Conversely, if $Q(I)$ is a yes-instance of **D-MIP2**, then there exists a solution $(\boldsymbol{x}, \boldsymbol{v_1}, \boldsymbol{v_2})$ such that $\sum_{i \in V} \sum_{k \in K} v_{ik2} \geq l'$. Then, the solution $y_i = x_{ii}$ is feasible because $\sum_{i=1}^{n} s_i y_i = \sum_{i=1}^{n} c_{ii} x_{ii} = \sum_{i \in V} \sum_{k \in K} c_{ik} x_{ik} \leq B = W$. For the objective value, we know that $\sum_{i=1}^{n} g_i y_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} g_j \right) \tilde{g}_i y_i = \left(\sum_{i=1}^{n} g_i \right) \cdot \left(\sum_{i \in V} b_i x_{ii} \right) = \left(\sum_{i=1}^{n} g_i \right) \cdot \left(\sum_{i \in V} \sum_{k \in K} v_{ik2} \right) \geq \left(\sum_{i=1}^{n} g_i \right) \cdot l' = l$, so I is a yes-instance of the knapsack decision problem.

From the three parts of the proof, it follows that **D-MIP2** is NP-complete. □

5.3 GREEDY HEURISTIC

A simple greedy heuristic was presented in Berman et al. (1995c). In this greedy heuristic, the catching system with the highest increase in objective value is added in each iteration of the heuristic until there is no more budget for new catching systems. This heuristic was slightly adjusted to take the different types of catching systems in the new version of the model into account.

In each step $p \in \{1, 2, \ldots\}$ of the heuristic, we place one extra catching system. We do this by calculating for each location $i \in V$ where no catching system has been placed and for each type of catching system $k \in K_i$ that is allowed in this location, what the objective function would be if catching system $k \in K_i$ is placed at location $i \in V$. Note that we only calculate this for catching systems that can still be added within the budget. When calculating the objective function, we use the objective function in the non-linear form of Model (4.7):

$$f(\boldsymbol{x}) = \underbrace{\boldsymbol{b}^T [I - T(\boldsymbol{x})]^{-1} \boldsymbol{t}_{n+1}(\boldsymbol{x})}_{f_1(\boldsymbol{x})} - \underbrace{w \sum_{i \in V} \sum_{k \in K} c_{ik} x_{ik}}_{f_2(\boldsymbol{x})} - \underbrace{\sum_{i \in V} \alpha_i \left(1 - \sum_{k \in K} \beta_{ik} x_{ik} \right) \left(\boldsymbol{b}^T [I - T(\boldsymbol{x})]^{-1} \right)_i}_{f_3(\boldsymbol{x})}. \tag{5.10}$$

This is equivalent to the objective function of the linear model, but the variables v_{i1} and v_{ik2} are substituted for their corresponding value that we can calculate using $\boldsymbol{x}$. We choose the catching system with the highest increase in parts f_1 and f_3 of the objective function per unit of cost of this catching system and add this to the

solution. Note that we leave out part f_2 of the objective function to ensure that we are not accounting for the costs c_{ik} twice. This ensures that we do not choose the most expensive catching system (the active system) because it has a higher catching probability, if it is possible to catch more plastics by placing several less expensive catching systems (passive systems) within the budget.

The pseudocode of the greedy heuristic is as follows:

Algorithm 1: Greedy heuristic PW-FCLM

Result: $\boldsymbol{x}_p$
$p = 0$
$\text{costs}_0 = 0$
$A_0 = \emptyset$ (set of catching system locations)
$\boldsymbol{x}_0 = \mathbf{0}$ $\left(\text{solution matrix}, \boldsymbol{x}_p \in \{0,1\}^{n \times K}\right)$
while $\exists i \in N \setminus A_p, k \in K_i$ such that $\text{costs}_p + c_{ik} \leq B, f(\boldsymbol{x}_p + e_{i,k}) \geq f(\boldsymbol{x}_p)$ **do**
  $i_p, k_p = \text{argmax}_{i \in N \setminus A_p, k \in K_i} \left(\frac{f_1(\boldsymbol{x}_p + e_{i,k}) - f_3(\boldsymbol{x}_p + e_{i,k}) - f_1(\boldsymbol{x}_p) - f_3(\boldsymbol{x}_p)}{c_{ik}} \right)$
  $\boldsymbol{x}_{p+1} = \boldsymbol{x}_p + e_{i_p,k_p}$
  $\text{costs}_{p+1} = \text{costs}_p + c_{i_p,k_p}$
  $A_{p+1} = A_p \cup \{i_p\}$
  $p = p + 1$
end

where $f(\boldsymbol{x}_p)$ is the objective value of the model given the solution $\boldsymbol{x}_p$, and $e_{i,k}$ is a matrix of size $n \times K$ with a one in entry i, k and zeros in all other entries. Adding a new catching system to the solution corresponds to adding $e_{i,k}$ to the current solution $\boldsymbol{x}_p$. We use the set A_p to keep track of the locations where a catching system has already been placed.

6

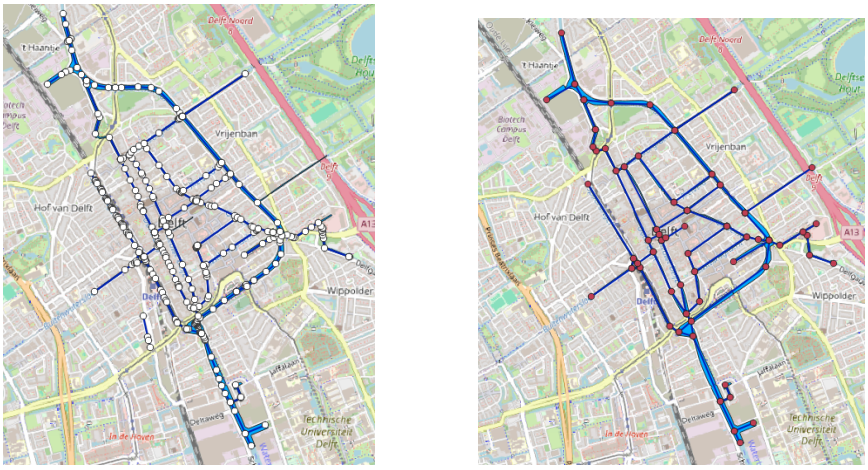
DATA

In this chapter, the preprocessing of data that is used as input to the model is explained. This is separated into three categories of data that are used in the model: network data, flow data and MILP parameters. The network data regards setting up the graph with nodes and edges. The flow data includes all the surrounding factors that influence the flow of plastics. The MILP parameters are the predefined parameters in Model (5.1). The model is applied to the water network of Delft, the Netherlands, but the input data is generated in an automated and general way, such that it can easily be applied to other locations. The data is loaded into QGIS, an open-source Geographic Information System, and is processed using available QGIS plugins and self-written Python scripts. Finally, the variations on the input parameters to create new scenarios for a sensitivity analysis are described.

6.1 NETWORK DATA

The network $G(V,E)$ on which the model is applied, is based on the water network of the area where the problem is solved. First of all, the waterways become the edges E of the network. This is achieved by using available data from Open Street Map to import the water network as a line layer. The line network is naturally divided into edges by the nodes that are located at intersections and locations where a waterway changes its direction. The canals in Open Street Map are very detailed, with small changes in direction of the waterway indicated by a new line separated by a node. Therefore, the second step is to simplify the network using a QGIS processing algorithm called "Simplify" with its tolerance setting set to 20. The algorithm creates geometries with the same features but fewer vertices, such that the distance of the simplified geometries compared to the detailed geometries is less than the tolerance setting of 20 meters. The tolerance setting could be adjusted for different areas to which the model is applied, depending on the size of the area and the level of detail of the waterways. After using this QGIS processing algorithm, some small manual changes are necessary to ensure that a few unnecessary vertices are deleted and the remaining vertices are moved. The result should be a layer with as few vertices left as possible and with edges representing the waterways close enough judged by eye. This should lead to a network with nodes only located at intersections and positions

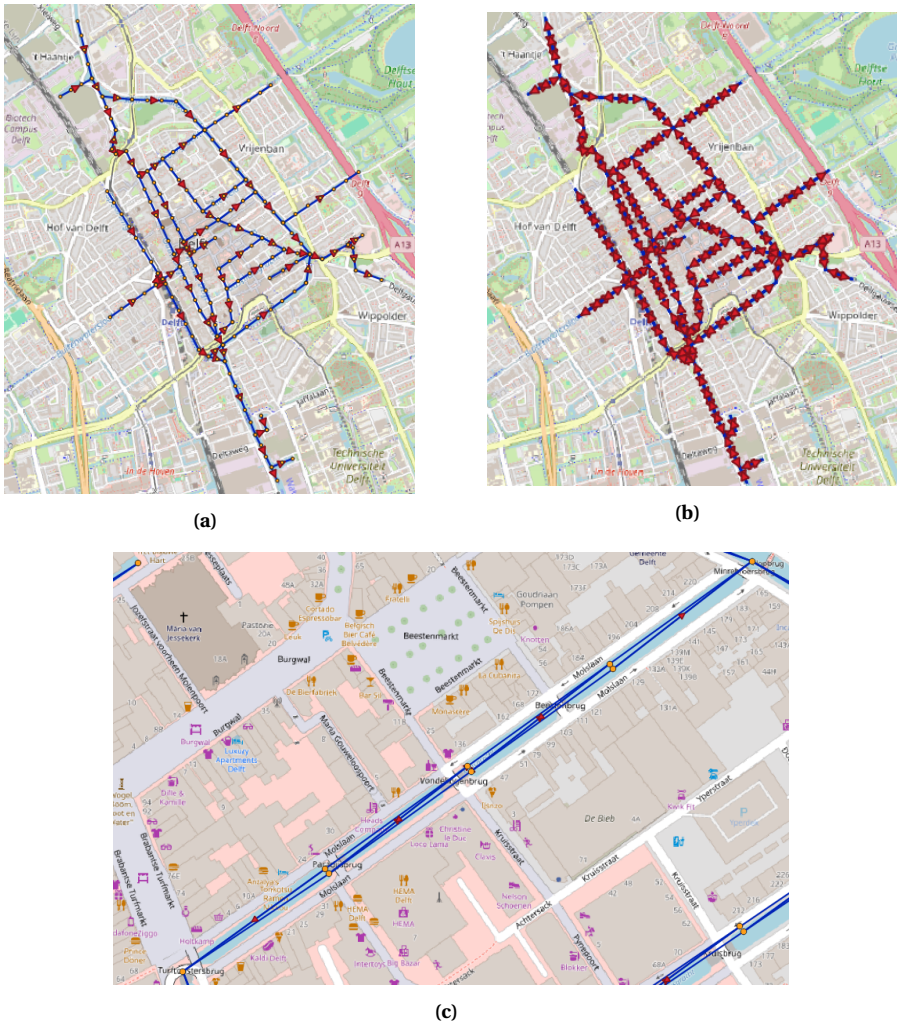
where the water changes direction clearly, as shown in Figure 6.1.



(a) **(b)**

Figure 6.1: The detailed version of the waterways in Delft from Open Street Map is shown on the left with the nodes of the line layer shown in white. On the right, the simplified version of the waterways with fewer nodes (in red) is shown.

The nodes V on the graph represent possible locations where catching systems could be located. Therefore, longer straight segments of the water are separated into more nodes for potential catching systems locations. The maximum distance between two nodes d_{max} is a parameter that can be adjusted to balance the solution space of the model with the preciseness of the locations that are suggested to place a catching system. It is estimated by Noria that a maximum distance of 100m between two possible catching system gives an accurate enough solution. The exact location where the catching system is placed can then be chosen within 100m of the solutions based on practical preferences. However, it is also tested if varying the value of d_{max} influences the solution and objective value of optimal locations suggested by the model. This is further explained in Section 7.1. A Python script interpolates extra nodes on straight segments of the network such that the nodes on a straight segment are equally distributed at a distance of at most d_{max} from each other. The addition of extra nodes on straight segments is shown in Figure 6.2a. The angle of the direction of each edge with respect to north is saved as an attribute of the QGIS line layer such that it can be used to calculate transition probabilities for the plastic flow. Finally, a Python script is used to double the nodes on the straight segments to take the orientation of the systems into account and extra edges are added between the double nodes (as explained in Section 4.2.1). The final network with the double nodes and directed edges is shown in Figure 6.2b.



(a) **(b)** **(c)**

Figure 6.2: In (a), extra nodes (in orange) are added at equal distances less than d_{max} to separate the larger water segments and to consider more locations to place a catching system. Then, the nodes are doubled to take the two possible orientations into account. Extra edges are added between the double nodes to direct the plastic flow to the correctly oriented catching systems. The final network is shown in (b) and a zoomed in version to see the double nodes and directed edges is shown in (c).

6.2 FLOW DATA

When the nodes and edges of the network are established, the next step is to find the flow of plastics on the network. A Markov chain representation of the plastic flow is made. The initial distribution and a transition matrix are necessary to define this Markov chain. To define the full transition matrix of the plastic flow, we need

transition probabilities along the edges between nodes, the probability to get stuck in a node and the probability to be caught by a catching system in a node.

6.2.1 INITIAL PROBABILITY

The initial distribution $\boldsymbol{b}$ is based on sources in the surroundings. Sources are restaurants, public recreational sites such as benches or playgrounds, and markets. The number of sources within a maximum distance $r_{sources}$ of the nodes is used to decide what fraction of the total flow should start at each node. It is estimated that the sources influence plastic waste in an area of radius 100m around their location, therefore $r_{sources} = 100$. The sources are available in GIS and provided by Noria from previous research projects. For each node, the total number of sources within a radius of $r_{sources}$ of a node are counted in the vector $\boldsymbol{b}$ and then $\boldsymbol{b}$ is normalized, such that b_i represents the probability of plastic waste starting in the node $i \in V$.

6.2.2 TRANSITION PROBABILITY

The transition probabilities q_{ij} between nodes $i, j \in V$ are dependent on wind and water geometry. Wind data from the Royal Netherlands Meteorological Institute (KNMI) from the year 2022 is used to calculate the probability distribution of wind in each possible direction. Data from the KNMI weather station (number 344) in Rotterdam was used for the model of Delft, since this is the closest weather station. Previous research from Noria suggests that the wind has the most influence on the direction in which plastics move in the water. The direction in which the plastics move could also be influenced by turbulent and local winds in different directions than the global wind measured in Rotterdam, due to buildings or other height differences in the surroundings. Furthermore, the water flow direction could be different from the global wind direction due to tides, pumping stations, altitude differences, boats or water turbulence. Since GPS measurements done by Noria seem to suggest that the wind has the most influence and there is no data available on the other surrounding factors, it is assumed in this thesis that the wind direction on the water geometry decides in which direction the plastics move. The influence of this assumption on the solution of the model is tested in the sensitivity analysis as explained in Section 7.1.

For each node $i \in V$, the probability that the plastic moves from this node to each of the following nodes $j \in V$ along the edge $(i, j) \in E$ is calculated by counting on how many days of the year the wind direction is closest to the direction of the edge. To do this, we use the angle of the edges with respect to north and the angle of the wind direction with respect to north. If there are two outgoing edges of the node $i \in V$ for which the wind direction of a certain day is within 5 degrees from the direction of the bisector of these two edges, then we count that plastic travels on both edges on that day. An example is shown in Figure 6.3. By normalizing the number of days counted for each edge, the transition probabilities are determined.

6.2.3 PROBABILITY OF GETTING STUCK

There is also a probability q_{i0} that a piece of plastic gets stuck in a certain node. The plastic that gets stuck is directed to the sink node "0" that was added to the

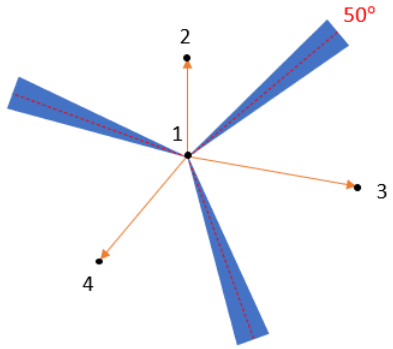


Figure 6.3: Visualisation of how the wind direction is used to calculate the transition probability along an edge. The orange arrows indicate the three outgoing edges (1,2), (1,3) and (1,4) of node 1. The red dashed lines are the bisectors of the three edges. If the wind direction is within ±5° of the bisector (the blue coloured range of wind directions), both edges are counted to have plastic flow due to this wind direction on that day. For example, if the wind direction is within the range $45-55°$ on a certain day, the plastics can travel along both edges (1,2) and (1,3) on that day.

model, as explained in Section 4.2.2. We assume that pieces of plastics get stuck at dead ends, sharp corners, docked boats (such as houseboats) and at places with shore vegetation or water vegetation. To each of these factors, we assign a certain probability of getting stuck.

DEAD ENDS

The probability of getting stuck due to dead ends is denoted by q^d_{i0}. The probability that the wind is blowing in the direction of the dead end is calculated in the transition matrix. We assume that half of the plastics that are moving into the direction of the dead end get stuck, and the other half is allowed to move back when the wind direction is away from the dead end. Therefore, the probability that plastics get stuck at a dead end $i \in V$ is $q^d_{i0} = q_{ji} \cdot 0.5$ for $(j,i) \in E$.

SHARP CORNERS

The probability of getting stuck due to sharp corners is denoted by q^c_{i0}. The probability that passing plastics get stuck in a certain corner is calculated for each corner separately. Then, the probability of getting stuck at a certain node due to the sharp corners that are close to the node is calculated from these probabilities.

The direction and size of each corner is calculated from the geometry of the water polygon layer in QGIS using a Python script. If the size of the corner is less than 110 degrees, the corner is considered sharp enough for plastics to get stuck, as observed in field experiments done by Noria. The probability that the wind direction is blowing into this corner is then calculated by taking the proportion of days of the

year 2022 that the wind was blowing into the corner. This probability is multiplied by 0.5 to calculate the probability of getting stuck in this particular corner, since it is assumed that half of the plastics get stuck in the corner, and the other half is allowed to flow back when the wind direction changes (similar to the dead ends).

To calculate the probability of getting stuck at a node $i \in V$ due to sharp corners, the complements of the probabilities of getting stuck at each of the sharp corners at a distance less than $0.5d_{max}$ of the node are multiplied and subtracted from 1. For example, if there are three sharp corners near node $i \in V$, the probability of getting stuck due to sharp corners is calculated as follows: $q^c_{i0} = 1-(1-q^{c1}_{i0})(1-q^{c2}_{i0})(1-q^{c3}_{i0})$.

BOATS

The probability of getting stuck due to docked boats is denoted by q^b_{i0}. Based on experience at Noria, we estimate that the probability of getting stuck at docked boats is $q^b_{i0} = 0.3$. A layer of houseboats is available in QGIS. If there are houseboats within a distance less than $0.5d_{max}$ of the node $i \in V$, we set $q^b_{i0} = 0.3$.

VEGETATION

We distinguish between a probability of getting stuck due to shore vegetation q^s_{i0} and water vegetation q^w_{i0}.

Shore vegetation grows more quickly in spring and summer than in winter, but it is present throughout the whole year. A line layer of the shore types is available, which indicates whether the shore is made of wall or vegetation. If there is shore vegetation within a distance less than $0.5d_{max}$ of the node $i \in V$, we set $q^s_{i0} = 0.3$, which is estimated based on experience at Noria.

Wall or vegetation are the only two types of shore present in Delft, however, in some parts of Groningen there is a rubble stone shore. This could be added in the future as a third shore type to extend the model.

Water vegetation (such as water lilies or duckweed) is only present between the months april-september, which is half of the year. When there is water vegetation, plastics hardly move from their position. Therefore, a probability of 0.9 of getting stuck due to water vegetation is estimated based on experience at Noria. A polygon layer of the canals in which water vegetation is present is available in QGIS. For this area, we set $q^w_{i0} = 0.45$ as water vegetation is only present half of the year.

These factors are considered independently: if a piece of plastic encounters one of the factors, there is a probability to get stuck. If it does not get stuck, it continues and encounters other factors that could lead to getting stuck. Therefore, the total probability of getting stuck at each node due to any of these factors is calculated by taking the product of the complements of the probabilities for each factor and subtracting this from 1:

$$q_{i0} = 1-(1-q^d_{i0})(1-q^c_{i0})(1-q^b_{i0})(1-q^s_{i0})(1-q^w_{i0}). \tag{6.1}$$

An example for the calculation of the probability to get stuck in a node $i \in V$ due to several different factors is shown in Figure 6.4.

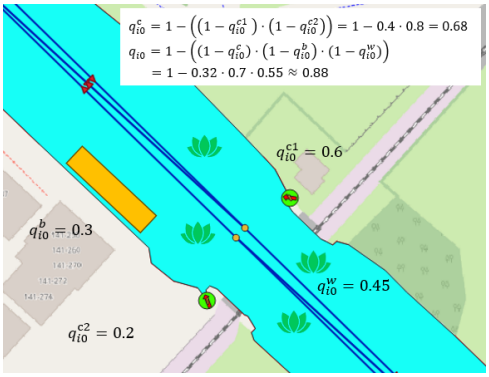


Figure 6.4: Example calculation of the probability of getting stuck due to several different factors together.

6.2.4 CATCHING PROBABILITY

If a catching system is placed in location $i \in V$, we need to calculate the probability that the plastic can be caught by a catching system in this location. To calculate this catching probability, the width of the canal w_i^{canal} and the maximum allowed boat width w_i^{maxboat} at each location $i \in V$ are necessary. Subtracting the maximum boat width from the width of the canal gives the width of the canal that can be blocked by the floating lines of the catching system. This is divided by the width of the canal to obtain a percentage that can be blocked, and then multiplied by the accuracy of the type of catching system a_k to find the catching probability β_{ik} of a catching system of type $k \in K$ in location $i \in V$ as follows:

$$\beta_{ik} = \frac{w_i^{\text{canal}} - w_i^{\text{maxboat}}}{w_i^{\text{canal}}} \cdot a_k. \tag{6.2}$$

The width of the canals at each location $i \in V$ is calculated automatically in QGIS with a Python script. Note that we do not need to know the catching probability at intersections, because in our model we assume that we do not place catching systems at intersections due to the orientation of the opening of the catching system. For each node that is not an intersection, a line perpendicular to the canal is created and clipped by the polygon layer. The length of the remaining line segment is saved as an attribute. The maximum boat width at each location is different for different types of regions in the water network. In the case of Delft, there is a larger waterway where container vessels can sail. These boats are up to 10m wide and should be able to pass from both ways, which is why all nodes on the waterway "Schie" have a maximum boat width of 21m. The canals and bridges in the city center are smaller and boats do not need to be able to pass simultaneously from both ways. The maximum boat width for this area is set to 4m.

Noria offers two types of catching systems as mentioned in Section 3.1.2. Type $k = 1$ is the active catching system, which has accuracy $a_1 = 0.98$ and type $k = 2$ is the passive catching system, which has accuracy $a_2 = 0.85$. The accuracy of the

passive systems is lower because plastics are not lifted from the water by the system and might not always be fully trapped by the system, or plastics could have trouble entering the system if the system is very full.

6.3 MILP PARAMETERS

Finally, there are 4 different types of input parameters for the PW-FCLM in Model (5.1) that still need to be defined: the cost c_{ik} of placing a catching system of type $k \in K$, the total budget B for the area to place catching systems, the weight w that balances costs with catching more plastics and the impact factor α_i of location $i \in V$ for sensitive areas. These are defined based on expert knowledge at Noria from previous research and projects.

The model is tested using fixed costs for the two types of catching systems that Noria offers and a variable budget to test a realistic budget for a variety of possible clients. The costs are not expressed in a specific unit or currency, but as a representative ratio compared to the other type of catching system and the budget. In our tests of the model, we set the costs of each type of catching system equal for all locations $i \in V$. The active catching system has costs $c_{i1} = 1$ and the passive catching system has cost $c_{i2} = 0.2$ for all $i \in V$. The costs of the system in the model are the total costs of placing the catching system. These include purchase costs together with an estimation of the operating costs.

The total budget is varied between $B \in [0.2, 4]$. Since the accuracy of the passive catching systems a_2 is almost as high as the accuracy of the active catching systems a_1 while the costs of passive catching systems are much lower, we expect that the optimal solutions to the tests of the PW-FCLM contain only passive catching systems. That means that for $B = 4$, 20 passive catching systems are placed. In practice, there are other reasons why active catching systems are more suitable than passive catching systems, as explained in Section 3.1.2. For locations $i \in V$ where this is the case, it is possible to set $x_{i2} = 0$ with Constraint (5.1f).

The factors α_i, $i \in V$ and w should be chosen to properly reflect the wishes of the user of the model in the trade-off between the different objectives, based on the following interpretation of these trade-off terms.

6.3.1 TRADE-OFF TERM FOR COSTS

When we fix $\alpha_i = 0$ for all $i \in V$, the trade-off term for costs in the objective function, $f_2(\boldsymbol{x})$, ensures that an extra catching system is only placed when the amount of extra flow that is caught per extra unit of cost is larger than w. We can see this from the following derivations. By setting $\alpha_i = 0$, the objective function is $f(\boldsymbol{x}) = f_1(\boldsymbol{x}) - f_2(\boldsymbol{x}) = f_1(\boldsymbol{x}) - w\tilde{f}_2(\boldsymbol{x})$, where $\tilde{f}_2(\boldsymbol{x}) = \sum_{i \in V} \sum_{k \in K} c_{ik} x_{ik}$ is a term that is equal to the total costs of the solution $\boldsymbol{x}$.

If we compare the objective value of an optimal solution $\boldsymbol{x}^*$ with a different

solution $\boldsymbol{x}$, we see that the following inequalities hold:

$$f(\boldsymbol{x}^*) \geq f(\boldsymbol{x}), \tag{6.3}$$

$$f_1(\boldsymbol{x}^*) - w\tilde{f}_2(\boldsymbol{x}^*) \geq f_1(\boldsymbol{x}) - w\tilde{f}_2(\boldsymbol{x}), \tag{6.4}$$

$$\frac{f_1(\boldsymbol{x}^*) - f_1(\boldsymbol{x})}{\tilde{f}_2(\boldsymbol{x}^*) - \tilde{f}_2(\boldsymbol{x})} = \frac{\Delta\text{caught flow}(\boldsymbol{x}^* - \boldsymbol{x})}{\Delta\text{costs}(\boldsymbol{x}^* - \boldsymbol{x})} \geq w. \tag{6.5}$$

We distinguish between the following four cases of Equation (6.5):

- The numerator and the denominator are both positive. In this case, w is important since in $\boldsymbol{x}^*$ more flow is caught with a higher cost. Then, the fraction is logically bounded from below by w.
- The numerator is positive and the denominator is negative. In this case, both terms $f_1(\boldsymbol{x}^*)$ and $\tilde{f}_2(\boldsymbol{x}^*)$ positively influence the objective value of solution $\boldsymbol{x}^*$ compared to solutions $\boldsymbol{x}$, because it means that more flow is caught with less costs. Therefore, w is not necessary in this case.
- The numerator is negative and the denominator is positive, which means less flow is caught for optimal solution $\boldsymbol{x}^*$ with higher cost. This is not possible when (6.3) holds, therefore the situation does not exist.
- The numerator is negative and the denominator is negative. In this case, the inequality sign is flipped and the fraction is bounded from above by w.

Therefore, the value of w should be equal to the minimum proportion of the total amount of plastics moving in the area that the user of the model wants to catch per unit of cost. In that way, the objective value increases when a new catching system of type $k \in K_i$ is placed at a location $i \in V$ if the extra flow that is caught by the catching system at this location is larger than the minimum amount of plastic flow that we want to catch with the extra costs of the new catching system ($wc_{ik}x_{ik}$). The objective value decreases when an extra catching system does not catch enough extra plastic compared to its extra costs. Therefore, such an extra catching system would not be in an optimal solution of the model in this case.

The value $w = 0.005$ is chosen in the tests of the model, such that an extra active catching system of type $k = 1$ should catch at least 0.5 percent extra of the total amount of plastic in the area and an extra passive catching system of type $k = 2$ should catch at least 0.1 percent extra of the total amount of plastic in the area.

6.3.2 TRADE-OFF TERM FOR SENSITIVE AREAS

To study the interpretation of the impact factor in the third part of the objective function, $f_3(\boldsymbol{x})$, we leave out the second term $f_2(\boldsymbol{x})$ for the trade-off with costs by setting $w = 0$. Furthermore, we assume that the impact factor is equal to $\alpha_i = \alpha$ for all $i \in V_s \subseteq V$, where V_s is the set of nodes in the sensitive area, and $\alpha_i = 0$ in all other nodes $i \in V \setminus V_s$. Then, the objective function can be written as $f(\boldsymbol{x}) = f_1(\boldsymbol{x}) - f_3(\boldsymbol{x}) = f_1(\boldsymbol{x}) - \alpha\tilde{f}_3(\boldsymbol{x})$, where $\tilde{f}_3(\boldsymbol{x}) = \sum_{i \in V_s} v_{i1} - \sum_{i \in V_s}\sum_{k \in K}(1 - \beta_{ik})v_{ik2}$ is a term that is equal to the number of visits of plastic flow in the sensitive area V_s,

given solution $\boldsymbol{x}$. Note that while the flow caught is always less than or equal to 1 (since the model expresses flow in the proportion of the total flow), the number of visits of plastic flow in the sensitive area can be larger than 1 because most plastics can visit several nodes before getting stuck or caught by a catching system.

We compare the objective value of an optimal solution $\boldsymbol{x}^*$ with a different solution $\boldsymbol{x}^+$ which corresponds to the optimal solution of the model with the same inputs except $\alpha_i^+ = 0$ for all $i \in V$. Then, we know that the total amount of flow caught with solution $\boldsymbol{x}^*$ is lower than or equal to the total amount of flow caught with solution $\boldsymbol{x}^+$, $f_1(\boldsymbol{x}^*) \leq f_1(\boldsymbol{x}^+)$. We also know that the total number of visits of flow in the sensitive area with solution $\boldsymbol{x}^*$ is lower than or equal to the total number of visits of flow in the sensitive area with solution $\boldsymbol{x}^+$, $\tilde{f}_3(\boldsymbol{x}^*) \leq \tilde{f}_3(\boldsymbol{x}^+)$.

If we compare the objective value of an optimal solution $\boldsymbol{x}^*$ with a different solution $\boldsymbol{x}^-$ which corresponds to the optimal solution of the model with the same inputs except $\alpha_i^- \geq \alpha_i$ for all $i \in V$, then, we know that the total amount of flow caught with solution $\boldsymbol{x}^*$ is higher than or equal to the total amount of flow caught with solution $\boldsymbol{x}^-$, $f_1(\boldsymbol{x}^*) \geq f_1(\boldsymbol{x}^-)$. We also know that the total number of visits of flow in the sensitive area with solution $\boldsymbol{x}^*$ is higher than or equal to the total number of visits of flow in the sensitive area with solution $\boldsymbol{x}^-$, $\tilde{f}_3(\boldsymbol{x}^*) \geq \tilde{f}_3(\boldsymbol{x}^-)$.

6

By comparing an optimal solution $\boldsymbol{x}^*$ to the solutions $\boldsymbol{x}^+$ and $\boldsymbol{x}^-$, we find that:

$$f_1(\boldsymbol{x}^*) - \alpha \tilde{f}_3(\boldsymbol{x}^*) \geq f_1(\boldsymbol{x}^+) - \alpha \tilde{f}_3(\boldsymbol{x}^+), \tag{6.6}$$

$$\frac{f_1(\boldsymbol{x}^*) - f_1(\boldsymbol{x}^+)}{\tilde{f}_3(\boldsymbol{x}^*) - \tilde{f}_3(\boldsymbol{x}^+)} = \frac{\Delta\text{total flow caught}(\boldsymbol{x}^* - \boldsymbol{x}^+)}{\Delta\text{flow visits in sensitive area}(\boldsymbol{x}^* - \boldsymbol{x}^+)} \leq \alpha. \tag{6.7}$$

Note that the sign of the inequality is flipped, because $\tilde{f}_3(\boldsymbol{x}^*) - \tilde{f}_3(\boldsymbol{x}^+)$ is negative. The bound is positive because the numerator and denominator are both negative. The impact factor α should be chosen equal to the maximum amount of total flow caught that the user of the model is willing to "trade" per unit of fewer flow visits in the sensitive area.

Using similar reasoning, we find the following inequality for $f(\boldsymbol{x}^*) \geq f(\boldsymbol{x}^-)$:

$$\frac{f_1(\boldsymbol{x}^*) - f_1(\boldsymbol{x}^-)}{\tilde{f}_3(\boldsymbol{x}^*) - \tilde{f}_3(\boldsymbol{x}^-)} = \frac{\Delta\text{total flow caught}(\boldsymbol{x}^* - \boldsymbol{x}^-)}{\Delta\text{flow visits in sensitive area}(\boldsymbol{x}^* - \boldsymbol{x}^-)} \geq \alpha. \tag{6.8}$$

This means that solution $\boldsymbol{x}^*$ is better than solution $\boldsymbol{x}^-$ if the amount of extra flow that is caught in total is at least α per unit of extra flow visits in the sensitive area.

In the following, we study three ways of using the impact factor in the model: catching the plastics as early as possible in the whole area, a sensitive area in the city center and a sensitive node at the edge of the study area.

CATCHING PLASTICS EARLY IN THE WHOLE AREA

To ensure that plastics are caught as early on their paths as possible, we set α_i to a small positive value for all $i \in V$. It was shown in Section 5.1 that $\beta_{ik} M_{ik2}$ is

the maximum amount of flow that can be caught in location $i \in V$ with a catching system of type $k \in K$. Therefore, if we choose $\alpha = \min_{i \in V, k \in K} \beta_{ik} M_{ik2}$, we ensure that the objective function is not dominated by the term $f_3(\boldsymbol{x})$. In that case, the model only considers catching the plastics early in the case of tiebreakers, but does not choose to catch plastics as early as possible when it means that less plastic is caught in total. If we choose $\alpha = \max_{i \in V, k \in K} \beta_{ik} M_{ik2}$, we expect that part f_3 of the objective function starts to dominate, which means that less flow is caught in total. It is also interesting to explore intermediate values of α, for example the mean of $\beta_{ik} M_{ik2}$ for all $i \in I$ and $k \in K$.

The effect of setting the impact factor $\alpha_i = \alpha$ to a small positive value for all $i \in V$ is tested for

$$\alpha \in \{0, \min_{i \in V, k \in K} \beta_{ik} M_{ik2}, \frac{1}{|V||K|} \sum_{i \in V, k \in K} \beta_{ik} M_{ik2}, 0.5 \cdot \max_{i \in V, k \in K} \beta_{ik} M_{ik2},$$
$$\max_{i \in V, k \in K} \beta_{ik} M_{ik2}\} = \{0, 0.00003, 0.005, 0.022, 0.044\}$$

and plotted in Figure 6.5. The total flow caught is plotted for each value of α for several budgets, the total number of visits of flow in the sensitive area is plotted, and it is verified whether Equation (6.7) and Equation (6.8) hold. From the two plots at the top of Figure 6.5, we see that the total flow caught does not decrease a lot for $\alpha = 0.5 \cdot \max_{i \in V, k \in K} \beta_{ik} M_{ik2}$, while the total flow in the sensitive area decreases visibly for most budgets. Therefore, it is chosen to set $\alpha_i = 0.5 \cdot \max_{i \in V, k \in K} \beta_{ik} M_{ik2}$ for all $i \in V$ in the model to catch plastics as early as possible. We can also verify from the bottom two plots in Figure 6.5 that Equation (6.7) and Equation (6.8) hold except from dividing by almost zero for $B = 2$, $\alpha = 0.5 \cdot \max_{i \in V, k \in K} \beta_{ik} M_{ik2}$.

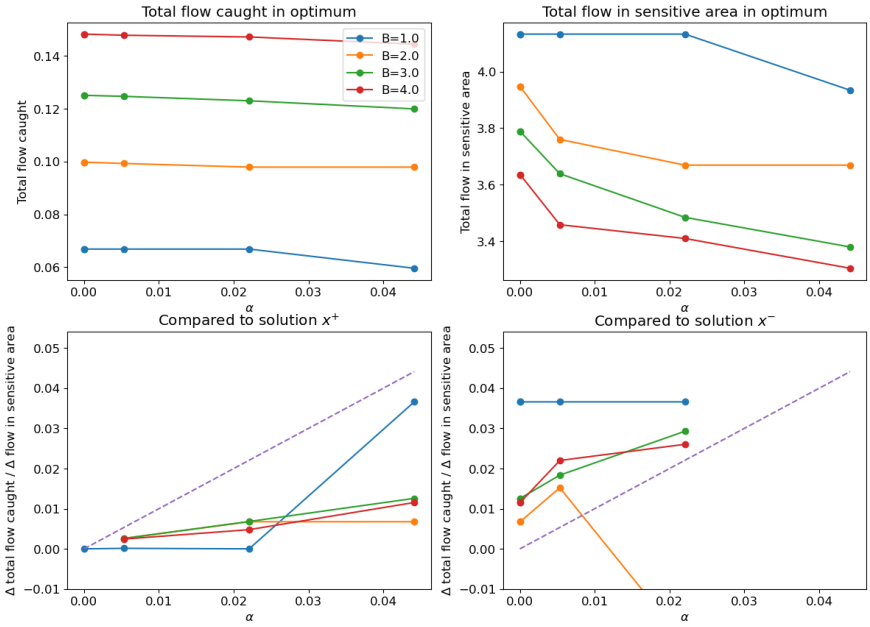


Figure 6.5: Flow caught and flow left in sensitive area plotted on the left and right top, for different small values of α_i at all nodes for catching early. Δ total flow caught / Δ flow visits in sensitive area is plotted compared to solution $\boldsymbol{x}^+$ (corresponding to the optimal solution for the case $\alpha_i^+ = 0$) on the left bottom and to solution $\boldsymbol{x}^-$ (corresponding to the optimal solution for the case where $\alpha_i^- = 0.044$) on the right bottom. Using the purple dotted line $y = x$, we verify that Equation (6.7) holds on the left bottom plot and that Equation (6.8) holds on the right bottom plot.

SENSITIVE AREA

To test for a sensitive area, we set α_i to a higher value for nodes in the city center of Delft. In Figure 6.6, the total flow caught is plotted for several budgets, the total number of visits of flow in the sensitive area is plotted, and it is verified that Equation (6.7) and Equation (6.8) hold for $\alpha_i \in \{0.5, 0.7, 0.9\}$ for $i \in V_s$ where V_s are the nodes in the city center of Delft. We see one exception in the right bottom plot for $B = 1$ and $\alpha_i = 0.7$ due to a division by almost zero.

SENSITIVE NODE AT THE EDGE OF THE AREA

It was also tested how the model responds when $\alpha_i \in \{1, 2, 3\}$ for the most northern node in the area of Delft, to reflect the case where it is important that the plastic flow does not leave the area towards the North Sea. In Figure 6.7, the total flow caught is plotted for several budgets, the total number of visits of flow in the sensitive area is plotted, and it is verified that Equation (6.7) and Equation (6.8) hold for $\alpha_i \in \{1, 2, 3\}$ for the most northern node in the area of Delft.

In all computational results shown in Chapter 7, we only use the impact factor for catching early.

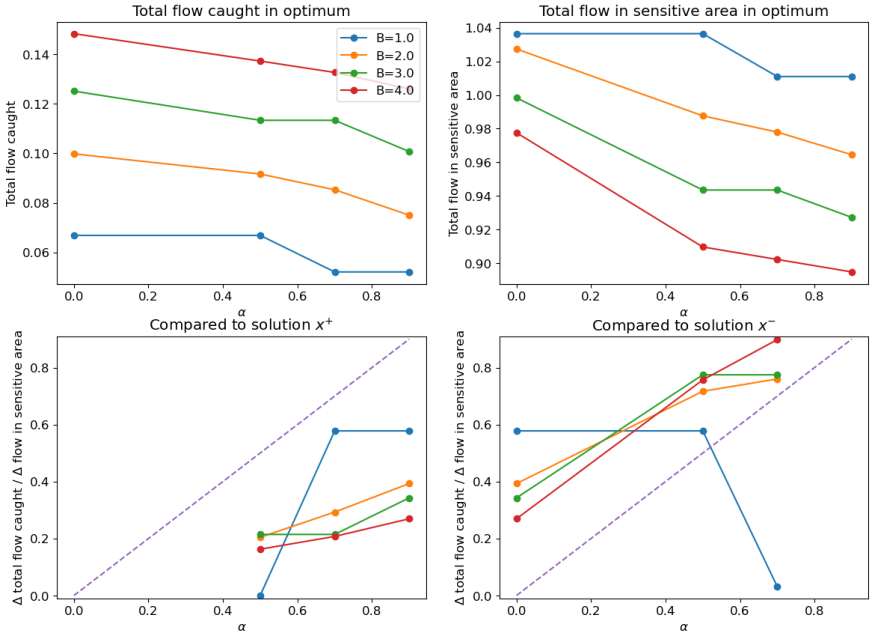


Figure 6.6: Flow caught and flow left in sensitive area plotted on the left and right top, for different values of α_i for a sensitive area V_s in the city center. Δ total flow caught / Δ flow visits in sensitive area is plotted compared to solution $\boldsymbol{x}^+$ (corresponding to the optimal solution for the case $\alpha_i^+ = 0$) on the left bottom and to solution $\boldsymbol{x}^-$ (corresponding to the optimal solution for the case where $\alpha_i^- = 0.9$ for all $i \in V_s$) on the right bottom. Using the purple dotted line $y = x$, we verify that Equation (6.7) holds on the left bottom plot and that Equation (6.8) holds on the right bottom plot.

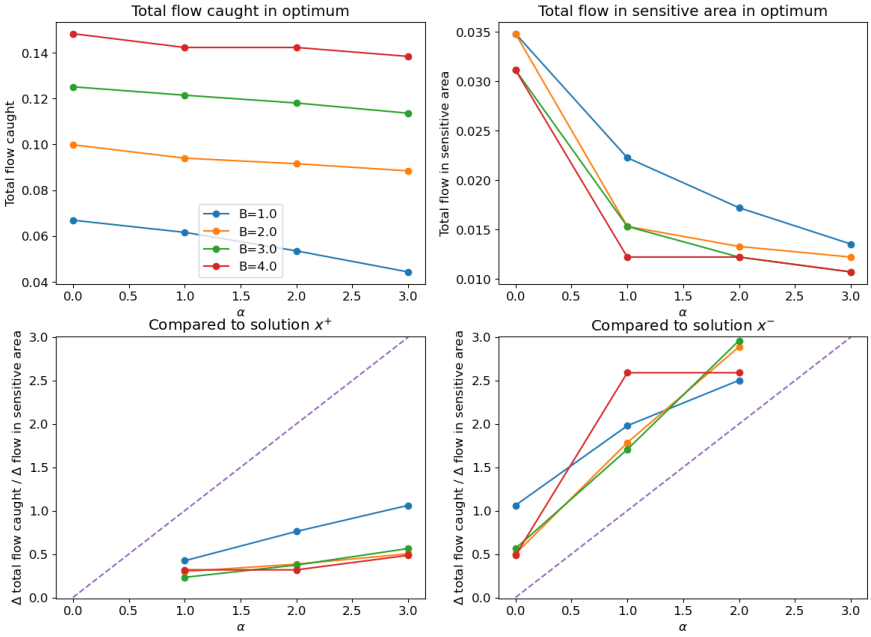


Figure 6.7: Flow caught and flow left in sensitive area plotted on the left and right top, for values of α_i at a sensitive node at the edge of the area. Δ total flow caught / Δ flow visits in sensitive area is plotted compared to solution $\boldsymbol{x}^+$ (corresponding to the optimal solution for the case $\alpha_i^+ = 0$) on the left bottom and to solution $\boldsymbol{x}^-$ (corresponding to the optimal solution for the case where $\alpha_i^- = 3$ for the sensitive node) on the right bottom. Using the purple dotted line $y = x$, we verify that Equation (6.7) holds on the left bottom plot and that Equation (6.8) holds on the right bottom plot.

6.4 SCENARIOS FOR SENSITIVITY ANALYSIS

After designing the PW-FCLM in Model (5.1) and pre-processing all the data as input, it is valuable to perform a sensitivity analysis to check which input parameters largely influence the solution and the corresponding amount of plastic flow that is caught. For the sensitivity analysis, six different types of input parameters are adjusted:

1. The distance between nodes d_{max}, which directly influences the number of nodes n.
2. The impact factor α_i for $i \in V$ for catching early and sensitive areas. The variations of α_i and the corresponding results are shown in Section 6.3.2.
3. The initial distribution $\boldsymbol{b}$.
4. The transition probabilities q_{ij} for $i, j \in V$ due to the wind.
5. The probability of getting stuck q_{i0} for all $i \in V$, due to the five different factors that influence this probability: dead ends q_{i0}^d, sharp corners q_{i0}^c, docked boats q_{i0}^b, shore vegetation q_{i0}^s and water vegetation q_{i0}^w.
6. The catching probability and accuracy of different types of catching systems β_{ik} for $i \in V, k \in K_i$.

For each type of input parameters, we describe how the input is changed to create new scenarios for the sensitivity analysis.

6.4.1 NUMBER OF NODES

It was estimated that a distance of $d_{max} = 100$ between the possible locations as a solution is practical. However, it was also tested whether putting the nodes closer or further apart together resulted in a very different solution to the model. Therefore, we compared the base scenario with $d_{max} = 100$ with new scenarios with $d_{max} \in \{50, 75, 125, 150\}$. Note that the steps in Section 6.1 and Section 6.2 need to be repeated fully to create these new scenarios, because a new network with a different distance between the nodes needs to be set up, and the transition probabilities need to be calculated for different locations. For the remaining types of input parameters, it is not necessary to repeat all steps in Section 6.1 and Section 6.2. It suffices to change only the vector of the corresponding input parameter.

6.4.2 INITIAL DISTRIBUTION

The initial distribution of the base scenario as calculated according to the method explained in Section 6.2.1 is varied to obtain 4 new scenarios. We create 3 new scenarios where $\boldsymbol{b}$ from the base scenario is perturbed in each location with a random variation of $\pm 10\%, \pm 20\%$ or $\pm 30\%$ compared to the base scenario. For example, if we perturb the base scenario with 10%, we generate a random vector of length n using the numpy package in Python, where each entry of the random vector has a value of 1.1 or 0.9, each with probability 0.5. We multiply this random

vector entry-wise with the initial distribution vector $\boldsymbol{b}$ from the base scenario and normalize this to obtain the initial distribution vector for the new scenario. The average result of 10 runs for each new scenario is plotted in the sensitivity analysis. We also test the scenario where the initial distribution is uniform: $b_i = 1/n$ for all $i \in V$.

6.4.3 TRANSITION PROBABILITIES

The transition probabilities of the base scenario are based on the wind directions measured during the year 2022, at a KNMI weather station closest to the test area, as explained in Section 6.2.2. To test the sensitivity of the model to wind data from different years, we test 3 new scenarios where the transition probabilities are calculated based on wind directions from the years 2020, 2021 and 2023. We also test one scenario where the transition probabilities to move in each possible direction from each node are uniform.

6.4.4 PROBABILITY OF GETTING STUCK

For the probability of getting stuck, we vary the probabilities of the five individual factors that influence the probability of getting stuck separately. The probability of getting stuck at each node $i \in V$ for the new scenario is then calculated using Equation (6.1), where one of the factors is changed and the other four factors are equal to the base scenario. We test each factor separately to test how sensitive the model is to the probabilities that we assigned to each factor as explained in Section 6.2.3, since it is possible for the user of the model to change these values.

GETTING STUCK DUE TO DEAD ENDS

In the base scenario, the probability of getting stuck due to dead ends q^d_{i0} is calculated based on the wind directions compared to the direction of the dead end, as explained in Section 6.2.3. We vary q^d_{i0} with the following percentages: $-20\%, -10\%, +10\%, +20\%$. That means, we create four new scenarios where the probability of getting stuck due to dead ends is equal to $q^d_{i0} \cdot 0.8$, $q^d_{i0} \cdot 0.9$, $q^d_{i0} \cdot 1.1$ and $q^d_{i0} \cdot 1.2$.

GETTING STUCK DUE TO SHARP CORNERS

In the base scenario, the probability of getting stuck due to sharp corners q^c_{i0} is calculated based on the wind directions compared to the geometries of the sharp corners as explained in Section 6.2.3. We vary the q^c_{i0} with the following percentages: $-20\%, -10\%, +10\%, +20\%$. That means, we create four new scenarios where the probability of getting stuck due to dead ends is equal to $q^c_{i0} \cdot 0.8$, $q^c_{i0} \cdot 0.9$, $q^c_{i0} \cdot 1.1$ and $q^c_{i0} \cdot 1.2$.

GETTING STUCK DUE TO DOCKED BOATS

In the base scenario, the probability of getting stuck due to docked boats q^b_{i0} equals 0.3 when there are docked boats near node $i \in V$. We test the new scenarios where we set $q^b_{i0} \in \{0.1, 0.2, 0.4, 0.5\}$ when there are docked boats near node $i \in V$. If there

are no docked boats near node $i \in V$, q^b_{i0} equals 0 in both the base scenario and the new scenario.

GETTING STUCK DUE SHORE VEGETATION

In the base scenario, the probability of getting stuck due to shore vegetation q^s_{i0} equals 0.3 when there is shore vegetation near node $i \in V$. We test the new scenarios where we set $q^s_{i0} \in \{0.1, 0.2, 0.4, 0.5\}$ when there is shore vegetation near node $i \in V$. If there is no shore vegetation near node $i \in V$, q^s_{i0} equals 0 in both the base scenario and the new scenario.

GETTING STUCK DUE WATER VEGETATION

In the base scenario, the probability of getting stuck due to water vegetation q^w_{i0} equals 0.45 if there is water vegetation near node $i \in V$. We test the new scenarios $q^w_{i0} \in \{0, 0.05, 0.3, 0.6, 0.9\}$ when there is water vegetation near node $i \in V$. If there is no water vegetation near node $i \in V$, $q^w_{i0} = 0$ in both the base scenario and the new scenario. The scenario q^w_{i0} equals 0 corresponds to the scenario in winter when there is no water vegetation, and the scenario $q^w_{i0} = 0.9$ corresponds to summer.

6.4.5 CATCHING PROBABILITIES AND ACCURACY

In the base scenario, the accuracy of the type 2 (passive) catching systems a_2 is 0.85 compared to $a_1 = 0.98$. Since the price difference is $c_{i1} = 1$ and $c_{i2} = 0.2$ for all $i \in V$, we expect to reach a break-even point of the accuracy compared to costs (a_k/c_{ik}) for the active and passive catching system if we set $a_2 = 0.98 * 0.2 = 0.196$. If the accuracy of type 2 catching systems a_2 is below this break-even point, we expect that the optimal solution to the model contains active catching systems if the budget allows it. If $a_2 > 0.196$, we expect that the optimal solution to the model contains active systems. Therefore, we vary $a_2 \in \{0.096, 0.196, 0.296, 0.396\}$ and test whether the solution contains only passive catching systems or also active catching systems. When varying a_2, we recalculate β_{ik} for all $i \in V, k \in K$ accordingly and use this as an input to the model.

The results of the sensitivity analysis with these new scenarios are shown and discussed in Section 7.1.

7

COMPUTATIONAL RESULTS

In this chapter, the computational results of the thesis are given. First of all, we present a sensitivity analysis of six different types of input parameters. Then, an analysis of the computational runtime is shown. Runtimes of different solution methods and different problem sizes are tested for different budgets, to provide an advice for the user of the model which problem sizes and budgets can be solved efficiently. Finally, a case study of a single run of the model is presented.

Experiments were run on a computer with an Intel Core i9-9900K processor running at 3.60GHz with 16.0 GB of RAM. The code of the PW-FCLM and all results is available on Github.

7.1 SENSITIVITY ANALYSIS

In this section, the results from the different new scenarios for the sensitivity analysis are shown and discussed. We vary 6 types of input parameters to create the new scenarios as explained in Section 6.4. The goal of the sensitivity analysis is to find out which inputs largely influence the solution of the model. It is important that the data for these input parameters accurately represent reality to ensure that the optimal solution we find is in fact optimal for the most realistic scenario. The results are plotted for the study area of Delft, the Netherlands. Based on these results, the sensitivity analysis for several types of input parameters is also performed on a second study area in Groningen, the Netherlands, to see if the sensitivity varies when the environment of the study area is different.

To test the sensitivity of each type of input parameters, we compare a new scenario to a "base scenario". This base scenario has the input parameters as explained in Chapter 6. In each new scenario, we vary one type of input parameters and leave the rest fixed at the base scenario values. To compare the optimal solution of the new scenario to the optimal solution of the base scenario, we look at two different measures: the distance between the solutions and the relative difference between the flow caught with the solution to the base scenario when used in the new scenario and the optimal flow caught in the new scenario with perfect knowledge. To illustrate this, we show how a base scenario (in green) and a new scenario (in

pink) are compared using an example in Figure 7.1.

(a) (b)

Figure 7.1: The optimal solution of the base scenario (green crosses) and the optimal solution of the new scenario (pink crosses) are shown. In Figure 7.1a, the distances between the two solutions are shown. In Figure 7.1b, the input flow in the base scenario (in green) and new scenario (in pink) is also shown. Note that Figure 7.1b is zoomed in on a part of Figure 7.1a. The input flow in the new scenario is different because $\boldsymbol{b}$, q_{ij}, q_{i0} or β_{ik} are different in the new scenario. We can see that we can catch less pink flow when we use the green catching systems than when we use the pink catching systems.

In Figure 7.1a, we see the optimal solution of the base scenario in green and the optimal solution of the new scenario in pink. The solution shows which locations are the optimal locations to place the catching systems. The distance between the chosen locations in the two solutions is calculated by first making a minimum weight perfect matching between the two solutions, where the weights are equal to the geographic distance between two locations. Then, we determine the distances between the matched locations of the two solutions as shown in Figure 7.1a. We plot the maximum distance between the solutions, which is 400 in this example, and the mean distance between the solutions, which is 500/3 in this example. This allows us to see if the catching systems of the optimal solution for the new scenario are placed far away from the catching systems of the optimal solution for the base scenario, without having to plot them on the map for all scenarios. We chose to consider absolute distance and not the distance of the shortest path via the network, since it is more efficient to calculate and provides enough information.

In Figure 7.1b, an example is shown of different input flow in the base scenario (in green) and in the new scenario (in pink). The flow in a new scenario is different

because we change the variables $\boldsymbol{b}$, q_{ij}, q_{i0} or β_{ik}, which changes the transition probabilities. The pink crosses represent the optimal solution of the new scenario. The flow caught in the new scenario with perfect information is equal to $f_1^{\text{new}}(\boldsymbol{x}^{\text{new}})$. Note that the total flow caught, determined by function f_1 in the objective function, has the superscript "new" because this also changes when the plastic flow changes in the new scenario. $\boldsymbol{x}^{\text{new}}$ corresponds to the location vector of the pink crosses, which is the optimal solution for the considered new scenario. $f_1^{\text{new}}(\boldsymbol{x}^{\text{new}})$ is equal to how much of the flow in the new scenario (the pink flow in Figure 7.1b) is caught using the solution $\boldsymbol{x}^{\text{new}}$. In the visualization, this is equal to how much pink flow can be caught using the pink catching system locations. The flow caught with the base scenario optimal solution $\boldsymbol{x}^{\text{base}}$ (the green crosses) in the new scenario is equal to $f_1^{\text{new}}(\boldsymbol{x}^{\text{base}})$. In the visualization, this is equal to how much pink flow can be caught using the green catching system locations. We can see that this is not optimal, since one of the green catching systems is in a location where there is no pink flow.

We plot the relative difference between the flow caught with the base scenario optimal solution when used in the new scenario, compared to the optimal flow caught in the new scenario with perfect knowledge: $\frac{f_1^{\text{new}}(\boldsymbol{x}^{\text{base}})-f_1^{\text{new}}(\boldsymbol{x}^{\text{new}})}{f_1^{\text{new}}(\boldsymbol{x}^{\text{new}})}$. We use only part f_1 of the objective function, corresponding to the total flow caught, because it allows us to calculate relative differences more easily. The objective value can be negative when α_i is large enough, which makes it difficult to calculate relative differences for the whole objective function. It is more insightful to plot the relative differences because the total amount of flow caught differs for different budgets. Relative differences are also more interesting because the absolute differences are proportions of the total flow, which seem like small numbers while they can be relatively large compared to the total proportion that is caught.

Showing these two measures allows us to see from the plot how different the locations are in the optimal solution to the base scenario and in new perturbed scenarios, but also whether using the optimal solution to the base scenario in a new scenario would result in a much lower caught flow than the optimum of the new scenario.

All instances for the sensitivity analysis are solved with the Gurobi solver using an academic license. Due to limited time for computational tests, the solver was terminated after 3600 seconds, which is why some scenarios were tested with lower budgets than others.

7.1.1 NUMBER OF NODES

We first compared the base scenario with $d_{max} = 100$ with new scenarios with $d_{max} \in \{50, 75, 125, 150\}$. We chose $d_{max} = 100$ based on expert knowledge. The results are plotted in Figure 7.2.

The solution to the base scenario is far from optimal in the new scenarios ($> 15\%$ for all budgets in the new scenario with $d_{max} = 50$ and $> 30\%$ from optimal for the lowest budgets $B \in \{0.2, 0.4\}$ which are most realistic for clients). Therefore, it was decided to change the number of nodes in the base scenario. From some tests, it

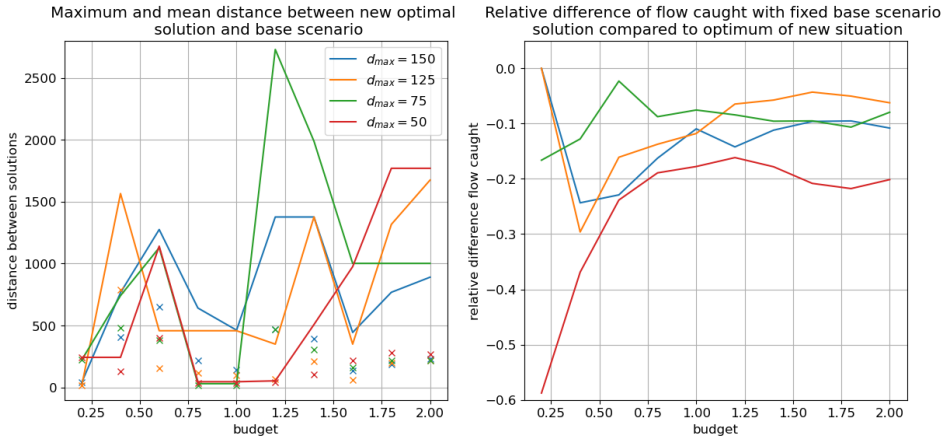


Figure 7.2: Sensitivity analysis of the distance between nodes d_{max} compared to a base scenario with $d_{max} = 100$. On the left, the maximum distance between the locations of the solutions are plotted as a line and the mean distance is plotted as a cross. The relative differences in flow caught compared to perfect information are plotted on the right. The legend is the same for both plots.

seemed that $d_{max} = 60$ was more suitable as a base scenario.

The results from a comparison of the base scenario with $d_{max} = 60$ with new scenarios with $d_{max} \in \{30, 40, 50, 70, 80, 90, 100\}$ are plotted in Figure 7.3. The solutions to the base scenario are now $< 20\%$ from optimal for all budgets compared to new scenarios with a smaller distance between the nodes ($d_{max} \in \{30, 40\}$). Therefore, the base scenario was adapted to $d_{max} = 60$ in the rest of the sensitivity analysis.

It seems that the main reason why the distance between the number of nodes is important, is because nodes located at wider locations of the waterways have a higher catching probability. This means that a location where the waterway is slightly wider is optimal in scenarios with a small distance between the nodes, while there is no node in this precise location in the scenarios with a larger distance between the nodes. Therefore, the level of detail of the network with $d_{max} = 100$ was too low.

In Section 3.2.2 we explained that the optimal locations of flow capturing facilities could be on edges instead of nodes for some instances of the MDP-based FCLM. Adding extra nodes to the network is similar to having the possibility to place facilities on edges, which could lead to an optimal location that was not available with fewer nodes in the network.

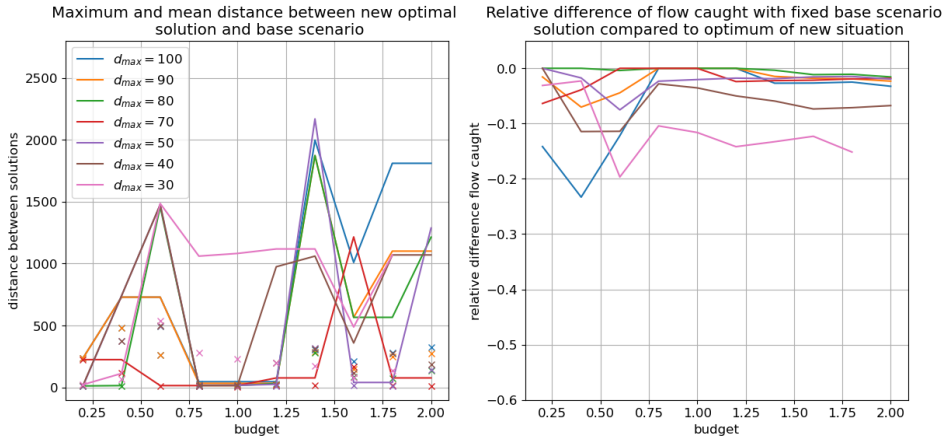


Figure 7.3: Sensitivity analysis of the distance between nodes d_{max} compared to a base scenario with $d_{max} = 60$. On the left, the maximum distance between the locations of the solutions are plotted as a line and the mean distance is plotted as a cross. The relative differences in flow caught compared to perfect information are plotted on the right. The legend is the same for both plots.

7.1.2 CHOOSING THE IMPACT FACTOR

In Section 6.3.2, the influence of choosing different values for the impact factor is shown for $d_{max} = 100$. Since we changed the distance between the nodes to $d_{max} = 60$, we show the total flow caught and the number of visits of flow in the sensitive area for different values of α_i for the new network. Again, we show three different cases: catching the plastics as early as possible, a sensitive area or a sensitive node at the edge of the area.

In Figure 7.4, we show the results of tests for several small positive values of α_i for all $i \in V$ for catching early. This is initially tested for α calculated using β_{ik} and M_{ik2} in the same way as in Section 6.3.2:

$$\alpha \in \{0, \min_{i\in V,k\in K}\beta_{ik}M_{ik2}, \frac{1}{|V||K|}\sum_{i\in V,k\in K}\beta_{ik}M_{ik2}, 0.5\cdot\max_{i\in V,k\in K}\beta_{ik}M_{ik2},$$
$$\max_{i\in V,k\in K}\beta_{ik}M_{ik2}\} = \{0, 0.00002, 0.004, 0.023, 0.046\}$$

Since we see that the total flow caught already decreases by 1% between $\alpha = \frac{1}{|V||K|}\sum_{i\in V,k\in K}\beta_{ik}M_{ik2} \approx 0.004$ and $\alpha = 0.5\cdot\max_{i\in V,k\in K}\beta_{ik}M_{ik2} \approx 0.023$ for budget $B = 1.2$, we also test $\alpha = 0.25\cdot\max_{i\in V,k\in K}\beta_{ik}M_{ik2} \approx 0.011$. It seems that $\alpha = 0.25\cdot\max_{i\in V,k\in K}\beta_{ik}M_{ik2} \approx 0.011$ is the most appropriate choice for catching early, since we see a clear decrease in the flow left in the sensitive area and only a small decrease in the total flow caught for this value of α_i.

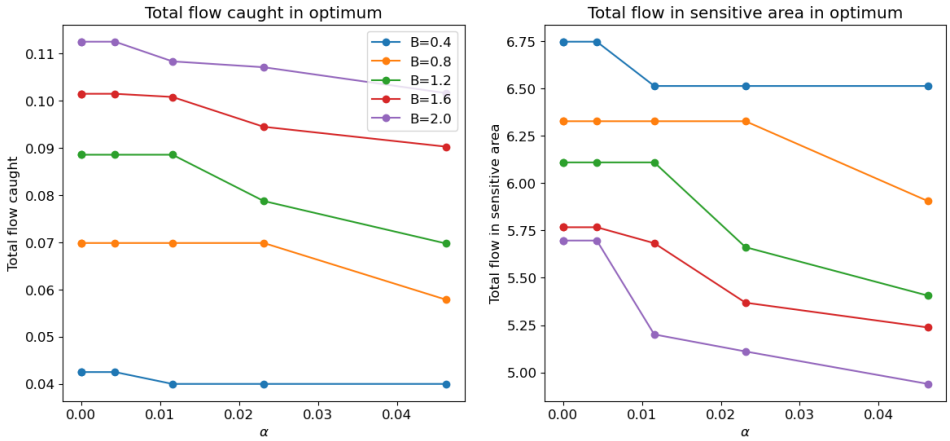


Figure 7.4: Total flow caught and flow left in entire study area, plotted against different small values of α_i at all nodes for catching early. The legend is the same for both plots.

In Figure 7.5, we show the results for tests with $\alpha_i \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ for all nodes i in the city center of Delft. We added extra tests with $\alpha_i \in \{0.1, 0.3\}$ that we did not do for the case with $d_{max} = 100$ in Section 6.3.2. This is because the total flow in the sensitive area already decreased strongly with $\alpha_i = 0.5$ for some budgets.

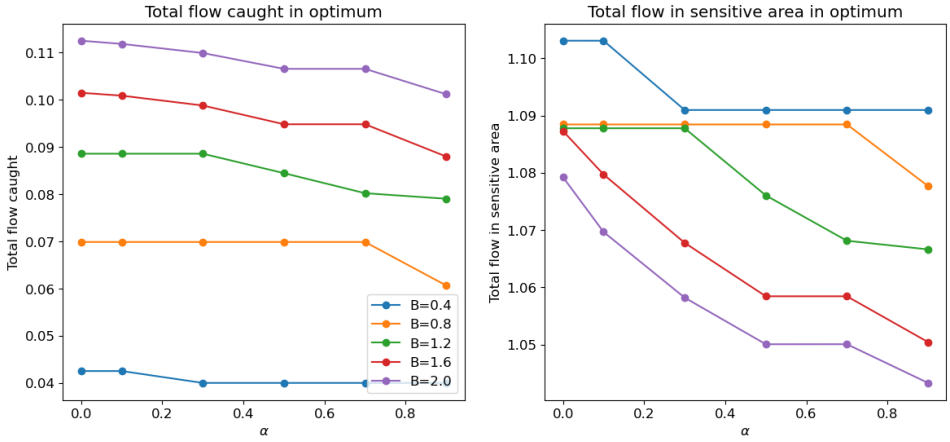


Figure 7.5: Total flow caught and flow left in the sensitive area, against different values of α_i at the nodes in the city center of Delft. The legend is the same for both plots.

In Figure 7.6, we show the results for tests with $\alpha_i \in \{1, 2, 3, 4, 5, 6\}$ for the most northern node in the area of Delft. We added extra tests with $\alpha_i \in \{4, 5, 6\}$ that we did not do for the case with $d_{max} = 100$ in Section 6.3.2. This is because the total flow in the sensitive area did not decrease for budgets $B = 0.4, 0.8, 1.2$ for lower α_i.

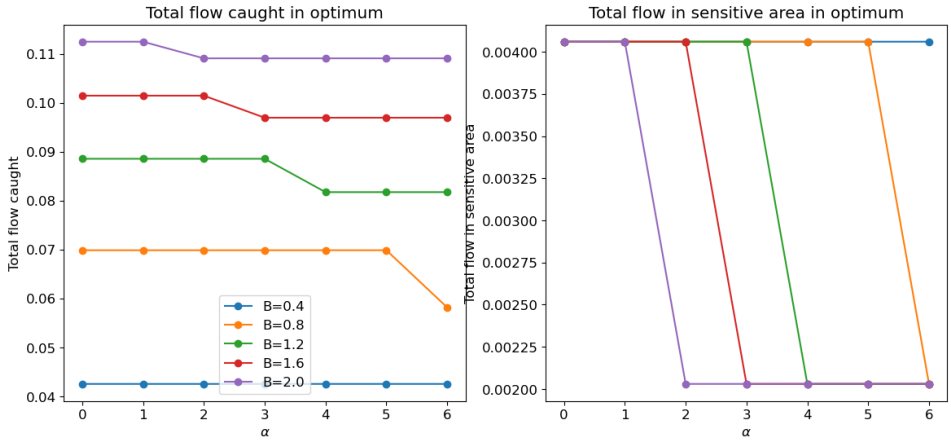


Figure 7.6: Total flow caught and flow left in the sensitive area, against different values of α_i at a node at the edge of the area. The legend is the same for both plots.

From the tests for these three cases of using the impact factor, we see that it depends on the instance how much a certain value of α_i influences the total flow caught and the total flow left in the sensitive area. When using the model, it is advised to plot the total flow caught and the total flow left in the sensitive area for different values of α_i for the chosen budget B before choosing the final value of α_i. The rest of the sensitivity analysis is run with an impact factor for catching early equal to $\alpha = 0.25 \cdot \max_{i \in V, k \in K} \beta_{ik} M_{ik2} \approx 0.011$ for all $i \in V$.

7.1.3 INITIAL DISTRIBUTION

To test the sensitivity of the model to the initial distribution, the initial probability from the base scenario is manipulated entry-wise by ±10%, ±20% and ±30%. A uniform initial probability is also tested. The results of these four new scenarios are plotted in Figure 7.7.

Even though the distance between the solutions of the new scenarios compared to the base scenario is quite large for some budgets, the model does not seem very sensitive to small perturbations of ±10%, ±20% and ±30% of the initial distribution $\boldsymbol{b}$. When using a uniform distribution as the initial distribution, the model is slightly more sensitive, however, the flow caught with the base scenario is never less than 90% of the flow caught with perfect information.

7.1.4 TRANSITION PROBABILITIES

To test the sensitivity of the model to different wind data input, we compare the base scenario of wind directions from the year 2022 to wind directions data from the years 2020, 2021 and 2023. We also compare the base scenario to turbulent wind with a uniform probability to transition from each node to its neighbours. The results of these four new scenarios for the transition probabilities are plotted in Figure 7.8.

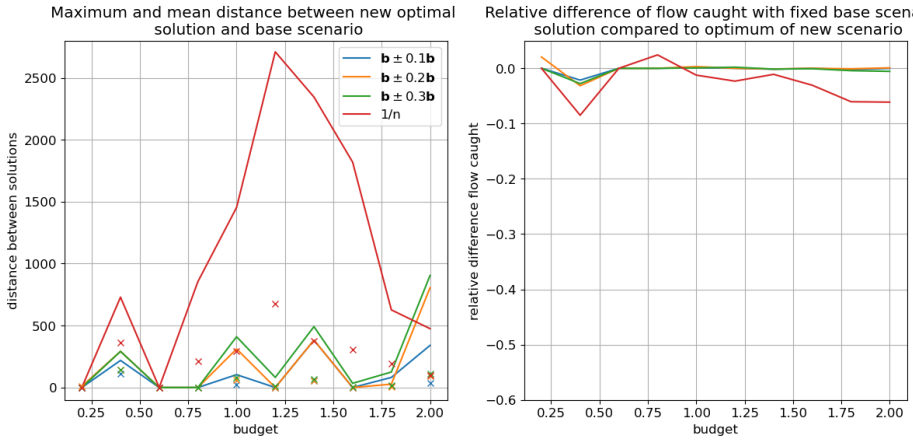


Figure 7.7: Sensitivity analysis of the initial distribution $\boldsymbol{b}$. On the left, the maximum distance between the locations of the solutions are plotted as a line and the mean distance is plotted as a cross. The relative differences in flow caught compared to perfect information are plotted on the right. The legend is the same for both plots.

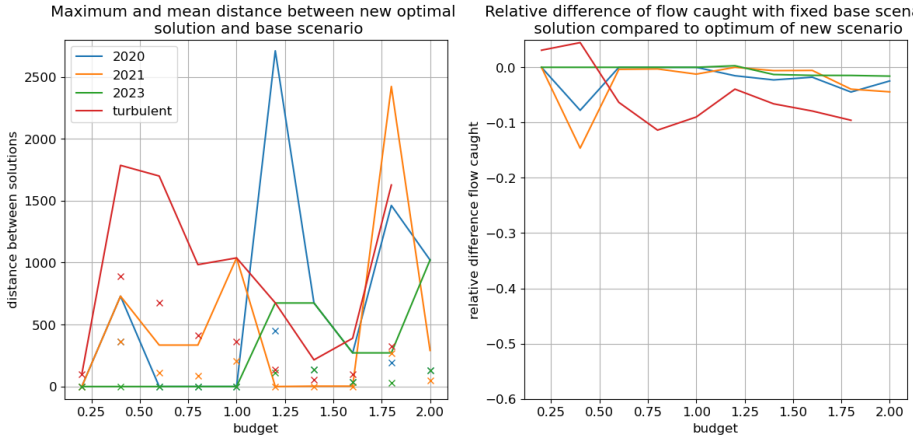


Figure 7.8: Sensitivity analysis of the transition probabilities q_{ij} with data of wind directions from different years. On the left, the maximum distance between the locations of the solutions are plotted as a line and the mean distance is plotted as a cross. The relative differences in flow caught compared to perfect information are plotted on the right. The legend is the same for both plots.

Again, the distance between the solutions of the new scenarios compared to the base scenario is quite large for some budgets, but the model does not seem very sensitive to wind input data from different years. For budget $B = 0.4$ and years 2020 or 2021 as wind data, the flow caught with the base scenario is approximately 10% less than the amount of flow caught with perfect information. When using

turbulent wind, the model is slightly more sensitive for most budgets, however, the flow caught with the base scenario is never less than 85% of the flow caught with perfect information.

7.1.5 PROBABILITY OF GETTING STUCK

GETTING STUCK DUE TO DEAD ENDS

The probability of getting stuck due to dead ends q^d_{i0} is varied with the following percentages: $-20\%, -10\%, +10\%, +20\%$. The results of these four new scenarios are shown in Figure 7.9. Even though there are a few scenarios and budgets where the distance between the optimal solution of the new scenario and the solution of the base scenario is large, the relative difference in flow caught compared to perfect information is very small. The model does not seem sensitive to changes in the probability of getting stuck due to dead ends.

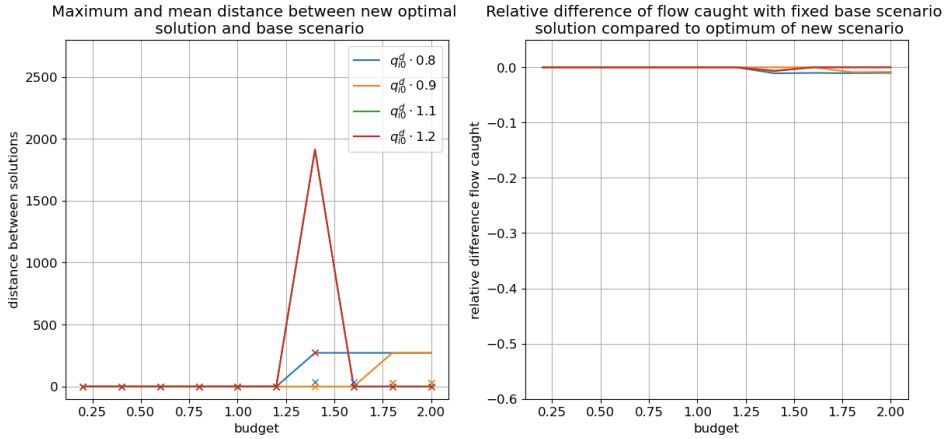


Figure 7.9: Sensitivity analysis of getting stuck due to dead ends q^d_{i0}. On the left, the maximum distance between the locations of the solutions are plotted as a line and the mean distance is plotted as a cross. The relative differences in flow caught compared to perfect information are plotted on the right. The legend is the same for both plots.

GETTING STUCK DUE TO SHARP CORNERS

The probability of getting stuck due to sharp corners q^c_{i0} is varied with the following percentages: $-20\%, -10\%, +10\%, +20\%$. The results of these four new scenarios are shown in Figure 7.10. Even though there are a few scenarios and budgets where the distance between the optimal solution of the new scenario and the solution of the base scenario is large, the relative difference in flow caught compared to perfect information is very small. The model does not seem sensitive to changes in the probability of getting stuck due to sharp corners.

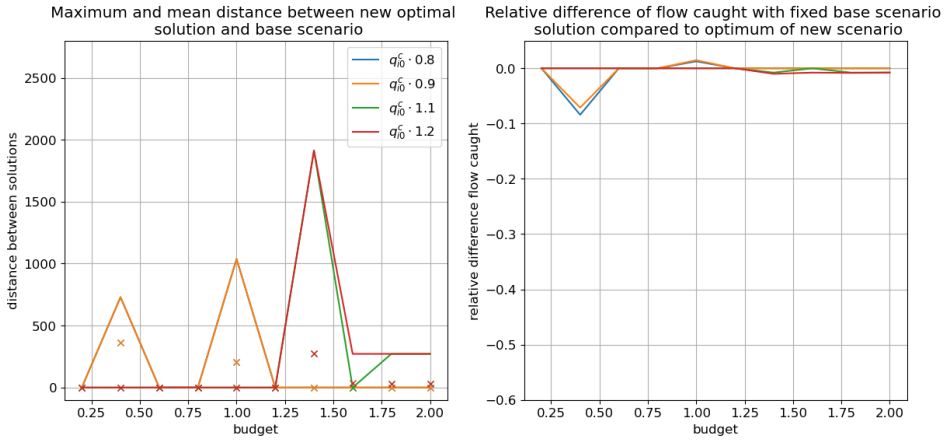


Figure 7.10: Sensitivity analysis of getting stuck due to sharp corners q^c_{i0}. On the left, the maximum distance between the locations of the solutions are plotted as a line and the mean distance is plotted as a cross. The relative differences in flow caught compared to perfect information are plotted on the right. The legend is the same for both plots.

GETTING STUCK DUE TO DOCKED BOATS

We vary the probability of getting stuck due to docked boats to $q^b_{i0} \in \{0.1, 0.2, 0.4, 0.5\}$ compared to the base scenario $q^b_{i0} = 0.3$. The results of these four new scenarios are shown in Figure 7.11. Even though there are a few scenarios and budgets where the distance between the optimal solution of the new scenario and the solution of the base scenario is large, the relative difference in flow caught compared to perfect information is very small. The model does not seem sensitive to changes in the probability of getting stuck due to docked boats.

GETTING STUCK DUE SHORE VEGETATION

We vary the probability of getting stuck due to shore vegetation to $q^s_{i0} \in \{0.1, 0.2, 0.4, 0.5\}$ compared to the base scenario $q^s_{i0} = 0.3$. The results of these four new scenarios are shown in Figure 7.12. The catching system locations in the optimal solution of the new scenario and the optimal solution of the base scenario are exactly the same. Therefore, the relative difference in flow caught compared to perfect information is zero. The model is not sensitive to changes in the probability of getting stuck due to shore vegetation. This could be because there are few locations in Delft with shore vegetation.

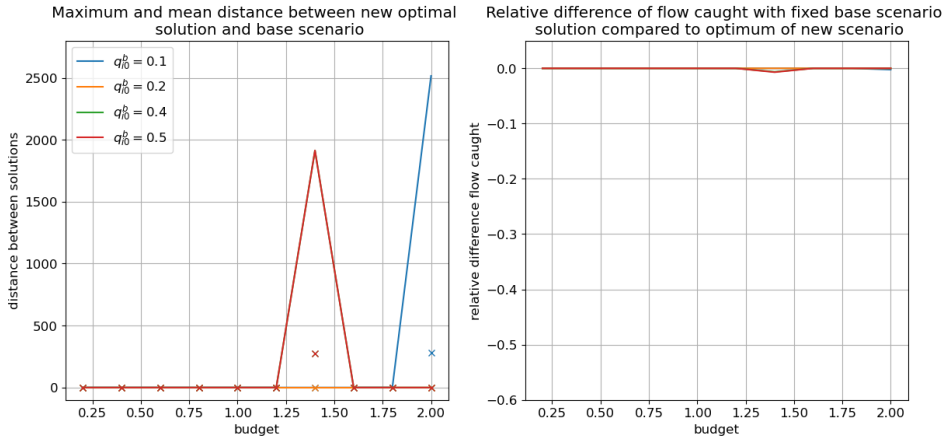


Figure 7.11: Sensitivity analysis of getting stuck due to docked boats q^b_{i0}. On the left, the maximum distance between the locations of the solutions are plotted as a line and the mean distance is plotted as a cross. The relative differences in flow caught compared to perfect information are plotted on the right. The legend is the same for both plots.

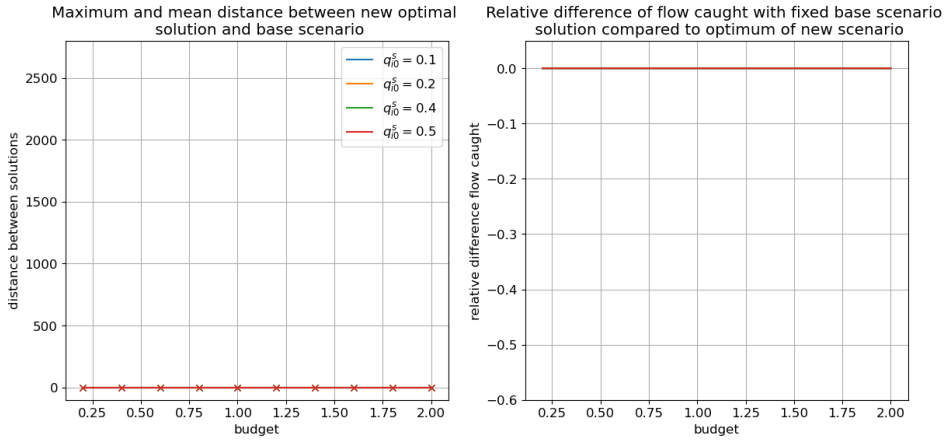


Figure 7.12: Sensitivity analysis of getting stuck due to shore vegetation q^s_{i0}. On the left, the maximum distance between the locations of the solutions are plotted as a line and the mean distance is plotted as a cross. The relative differences in flow caught compared to perfect information are plotted on the right. The legend is the same for both plots.

GETTING STUCK DUE WATER VEGETATION

We vary the probability of getting stuck due to water vegetation to $q^w_{i0} \in \{0, 0.05, 0.3, 0.6, 0.9\}$ compared to the base scenario $q^s_{i0} = 0.45$. The results of these four new scenarios are shown in Figure 7.13. The model does not seem sensitive to increases in the probability of getting stuck due to water vegetation, however, decreasing the probability of getting stuck due to water vegetation to a probability

close to zero is very sensitive. This is likely because the water vegetation is the main reason why the plastics get stuck in Delft, since $q_{i0}^{w} = 0.45$ is the highest contributing factor of plastics getting stuck of all five factors and there is a lot of water vegetation in Delft. When $q_{i0}^{w} = 0$, plastics are able to move and spread out in the area a lot more, which could explain why the solution and the caught flow is a lot different. The runtime of the Gurobi solver was a lot longer for the new scenario with $q_{i0}^{w} = 0$, which is why it was only tested up until a budget of $B = 1.4$.

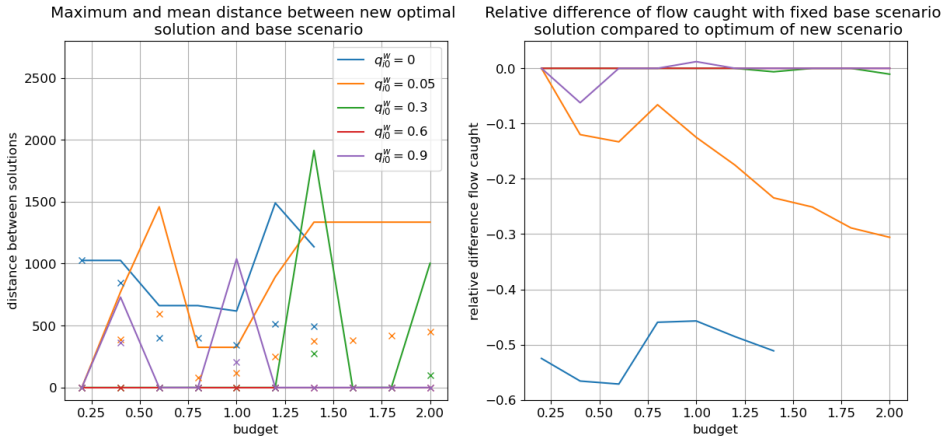


Figure 7.13: Sensitivity analysis of getting stuck due to water vegetation q_{i0}^{w}. On the left, the maximum distance between the locations of the solutions are plotted as a line and the mean distance is plotted as a cross. The relative differences in flow caught compared to perfect information are plotted on the right. The legend is the same for both plots.

7

7.1.6 CATCHING PROBABILITIES AND ACCURACY

We vary the accuracy of the passive catching systems to $a_2 \in \{0.096, 0.196, 0.296, 0.396\}$ and plot the number of type $k = 1$ and type $k = 2$ catching systems in a stacked bar chart for each value of a_2 and each budget, as shown in Figure 7.14. We expect to see active catching systems (type 1) when the budget allows it for a_2 below the break-even point $a_2 < 0.196$. This is the case for $a_2 = 0.096$ as expected. For $a_2 = 0.196$ (exactly on the break-even point) we see that no active catching systems are in the optimal solution for $B = 1$ and only one instead of two active catching systems for $B = 2$. This could be caused by the trade-off term with the costs f_2.

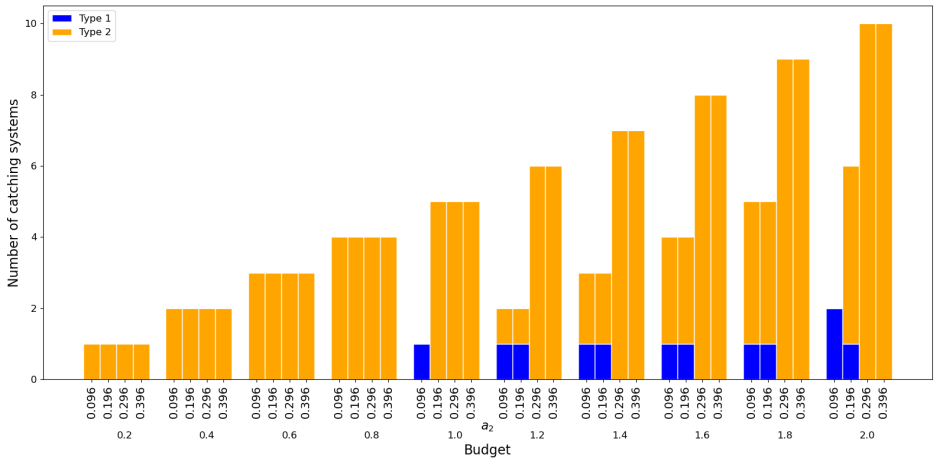


Figure 7.14: Stacked bar chart of the number of catching systems of type 1 and 2 for different values of a_2 and different budgets B.

7.1.7 GRONINGEN

For the most sensitive types of input parameters we plot the relative difference of flow caught compared to perfect knowledge for Delft and Groningen next to each other to compare the sensitivity under different environmental conditions. In particular, we show the senstivity analysis for Delft and Groningen together for the number of nodes, the initial probability and the transition probability due to wind, because these were sensitive input parameters for Delft. The maximum distance between the nodes and the value of the impact factor for catching early for Groningen were chosen the same way as Delft: $d_{max} = 60$ and $\alpha_i = 0.25 \cdot \max_{i \in V, k \in K} \beta_{ik} M_{ik2}$.

In Figure 7.15, we see that the model is more sensitive to variations in the distance between the nodes. When looking at the map, it seems that this is also because the nodes are in different locations where the waterways are much wider for some values of d_{max}. We see in Figure 7.16 that the model is equally sensitive to small perturbations of the initial distribution $\boldsymbol{b}$, but is more sensitive when the initial distribution is uniform in Groningen. This could be explained by the fact that the initial distribution of Groningen differs more from a uniform distribution than the initial distribution of Delft, due to the different distributions of sources in the areas. In Figure 7.17, we see that the model is equally sensitive to different years or turbulent wind as an input.

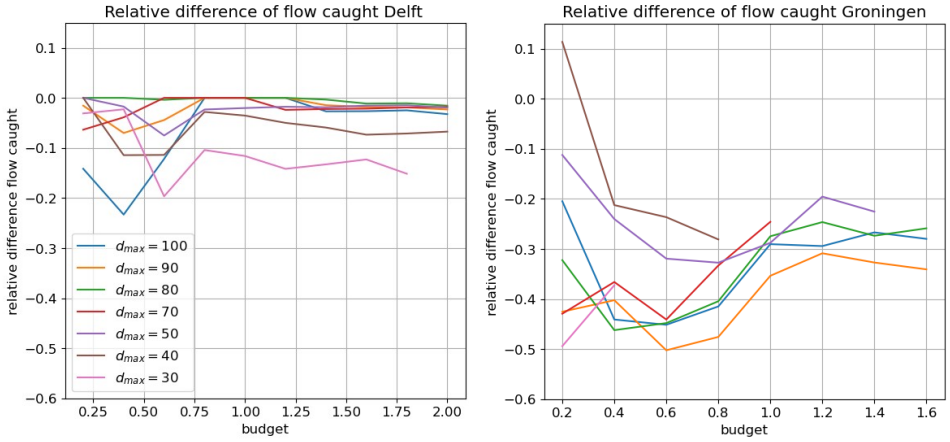


Figure 7.15: Comparison of the sensitivity analysis of the distance between the nodes d_{max}, for Delft on the left and Groningen on the right. For both cities, the relative differences in flow caught compared to perfect information are plotted. The legend is the same for both plots.

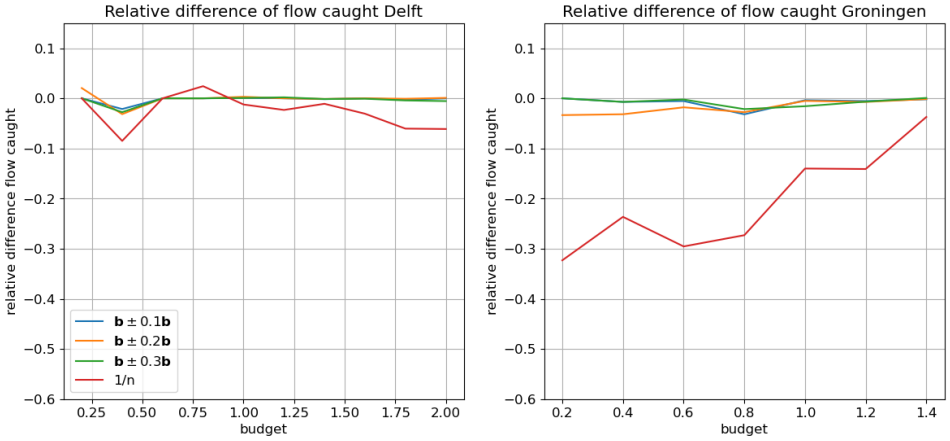


Figure 7.16: Comparison of the sensitivity analysis of the initial distribution $\boldsymbol{b}$, for Delft on the left and Groningen on the right. For both cities, the relative differences in flow caught compared to perfect information are plotted. The legend is the same for both plots.

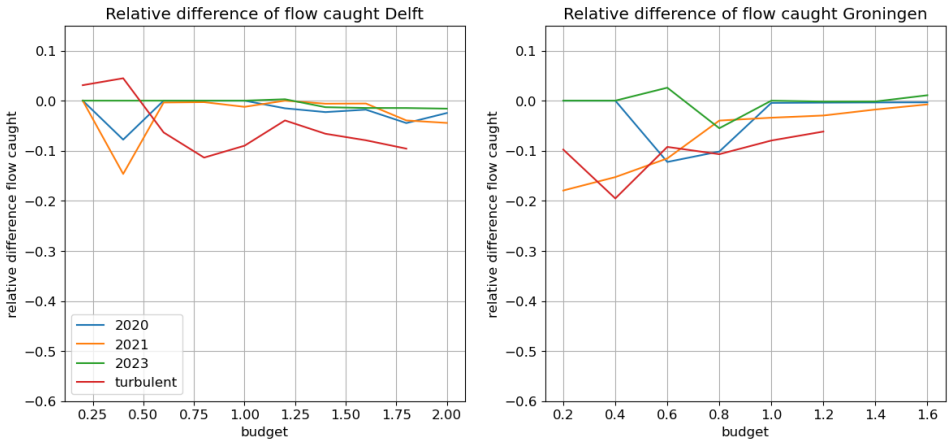


Figure 7.17: Comparison of the sensitivity analysis of the transition probabilities q_{ij} due to wind, for Delft on the left and Groningen on the right. For both cities, the relative differences in flow caught compared to perfect information are plotted. The legend is the same for both plots.

We also plot the stuck probability due to docked boats and the stuck probability due to shore vegetation, because there are more houseboats in Groningen and there is more shore vegetation. Note that the limits of the y-axis are different for these plots compared to the plots of the more sensitive types of input parameters. This is done such that we can see the difference between Delft and Groningen clearly for these less sensitive types of input parameters. From Figure 7.18 and Figure 7.19, we see that the model is more sensitive to changes in the probability of getting stuck due to docked boats or shore vegetation when there are more docked boats or there is more shore vegetation in the area. However, the total amount of flow caught in these new scenarios is still more than 92% of the flow caught with perfect information.

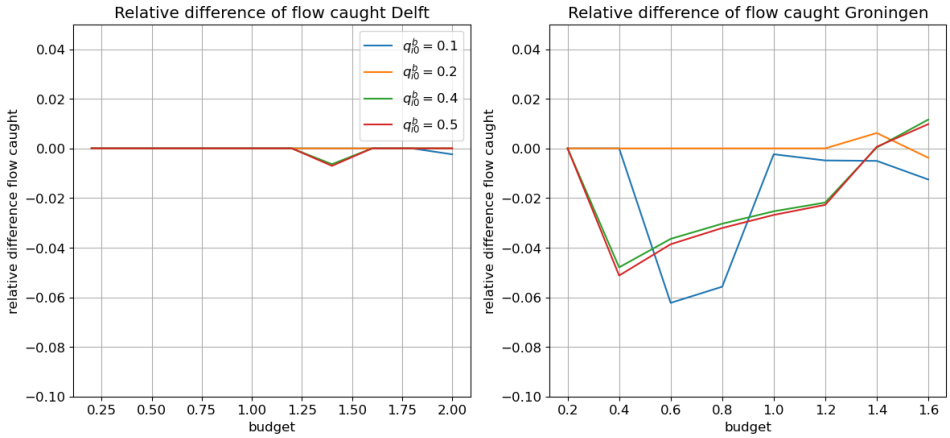


Figure 7.18: Comparison of the sensitivity analysis of the probability of getting stuck due to boats q^b_{i0}, for Delft on the left and Groningen on the right. For both cities, the relative differences in flow caught compared to perfect information are plotted. The legend is the same for both plots.

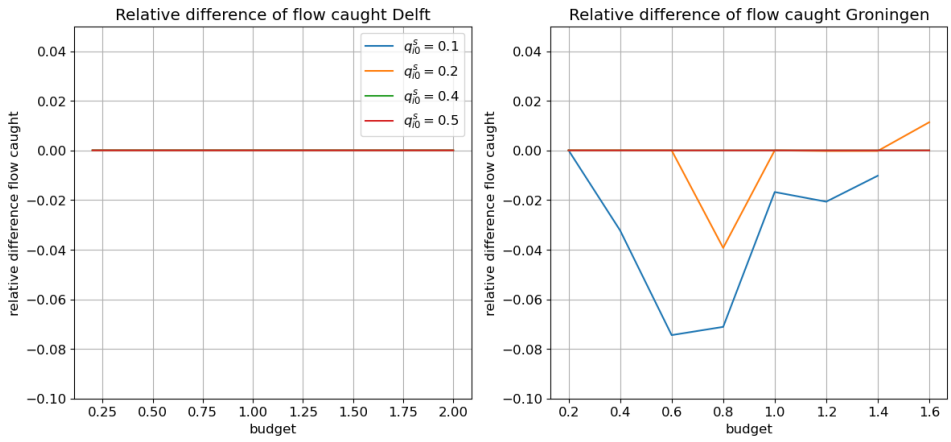


Figure 7.19: Comparison of the sensitivity analysis of the probability of getting stuck due to shore vegetation q^s_{i0}, for Delft on the left and Groningen on the right. For both cities, the relative differences in flow caught compared to perfect information are plotted. The legend is the same for both plots.

7.2 COMPUTATIONAL TIME

From the results of the sensitivity analysis in Section 7.1, it is clear that the model is sensitive to the distance between the nodes d_{max}. If the distance between nodes is decreased, there are more nodes that are considered as potential locations in the model which leads to more variables x_{ik} in the PW-FCLM. Since it was proven in Section 5.2 that the PW-FCLP is NP-hard, we expect an exponential increase in runtime when the number of variables in the model increases. Additionally, we expect that the runtime exponentially increases with the budget B. Therefore, we analyse the computational time to solve the PW-FCLM using different solution methods as presented in Chapter 5.

We try to find the exact solution to the PW-FCLM in Model (5.1) using the Python library PuLP with the Gurobi solver using an academic license and PuLP's default open source CBC solver. While performing the sensitivity analysis, the Gurobi solver with an academic license was able to find the exact solution to the problem with $n = 308$ number of nodes and a budget of $B = 4$ efficiently (within 15 minutes). For a larger number of variables, $n = 522$, we performed the sensitivity analysis for most new scenarios up until a budget of $B = 2$ within 60 minutes. However, it is expensive for organizations that focus on catching plastic waste to purchase a license to the commercial solver. Therefore, it is important to test the computational runtime of the open source solver. We also tested whether an initial solution as a warm start to the open source and commercial solver reduced the runtime. An open source solver is significantly less efficient, therefore a slightly adjusted version of the greedy heuristic presented by Berman et al. (1995c) is also applied to solve the problem.

7.2.1 COMPARISON OF RUNTIMES OF DIFFERENT SOLUTION METHODS

Given the initial estimation that a maximum distance of $d_{max} = 100$ gives a sufficient level of detail for practical use of the PW-FCLM, the number of nodes in the network of the test area in Delft is $n = 308$. This is within the bounds of the maximum problem size allowed by the Gurobi solver without an academic license. The maximum problem size is 2000 variables and 2000 linear constraints when using the Gurobi solver without an academic license. A realistic maximum budget of a client is $B = 2$, which means that we expect a solution with 10 passive catching systems. However, since the Gurobi solver solved the instances with a smaller number of nodes within a few minutes, we test up to a maximum budget $B = 4$ in these cases. The solver is terminated after 3600 seconds because of limited time for these computational tests. We compare the runtimes of the Gurobi solver and the CBC solver with a cold start and a warm start from the heuristic solution for the instance with $d_{max} = 100, n = 308$ in Figure 7.20. We see that the Gurobi solver finds the optimal solution to all instances with budgets up to $B = 4$ within 10 minutes, while the open source solver can only solve budgets up to $B = 2.6$ within an hour. The difference in runtime between the warm start and cold start is small.

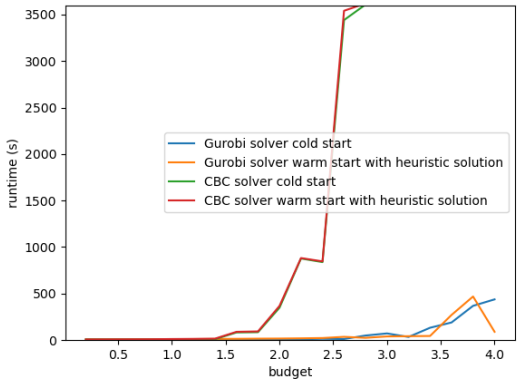


Figure 7.20: Comparison of runtimes plotted against budgets for 4 different solution methods.

7.2.2 COMPARISON OF RUNTIMES OF DIFFERENT PROBLEM SIZES

Since it was estimated that a maximum distance of $d_{max} = 50$ is the largest possible level of detail that the PW-FCLM could need, this problem size is also tested. The runtimes for $d_{max} \in \{50, 75, 100, 125, 150\}$ using the Gurobi solver with a cold start are plotted in Figure 7.21. The solver is terminated after 12 hours for finding the solution to each instance with a new budget. We can see that the runtime sharply increases with the number of nodes in each instance. From these graphs, we see that instances with $n \leq 412$ and $B \leq 2$ can be solved efficiently within a few hours for both the Gurobi solver and the CBC solver, but for larger budgets or more nodes the runtime quickly increases to over 12 hours.

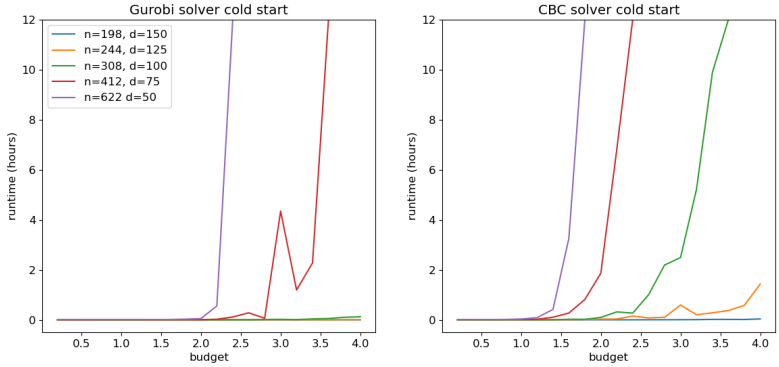


Figure 7.21: Comparison of runtimes plotted against budgets for instances with different problem sizes due to the number of nodes for the Gurobi solver on the left and the CBC solver on the right. The legend is the same for both plots.

7.2.3 COMPARISON OF HEURISTIC AND EXACT SOLUTION

The greedy heuristic has been tested for several different input instances and different budgets. These input instances are based on the most sensitive scenarios from the sensitivity analysis in Section 7.1.

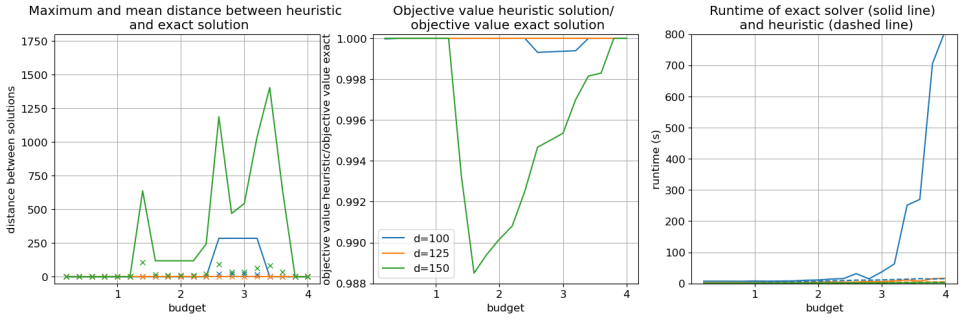


Figure 7.22: Comparison of greedy heuristic and exact solution for instances with different numbers of nodes d. The legend is the same for all three plots. The distance between the solutions, the optimality gap and the differences in runtimes are shown.

For the instances shown in Figure 7.22, Figure 7.23, Figure 7.24 and Figure 7.25, we can see in the middle plots that the heuristic performs within 1.5% of the optimum objective value. We can also see that the heuristic is faster for some instances for the higher budgets. The Gurobi solver is very efficient for these instances with $n = 308$ for small budgets, which is why the heuristic finds a solution with similar runtimes compared to the Gurobi solver for budgets $B \leq 2$. While the heuristic finds very close to optimal solutions in these test instances, it is not guaranteed that this is the case on different instances.

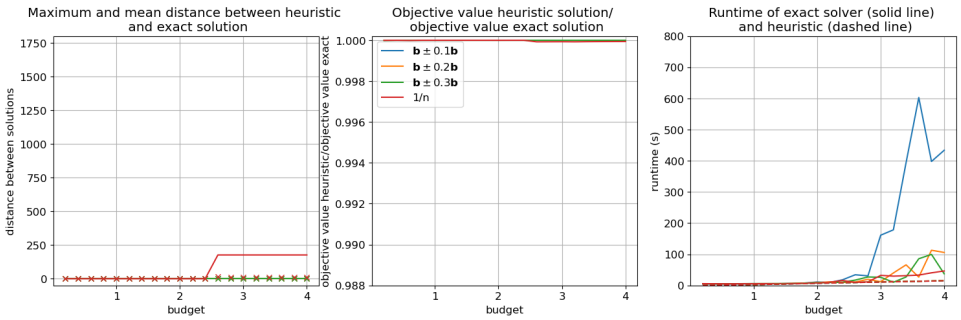


Figure 7.23: Comparison of greedy heuristic and exact solution for instances with different initial distributions with respect to the base scenario initial distribution $\boldsymbol{b}$. The legend is the same for all three plots. The distance between the solutions, the optimality gap and the differences in runtimes are shown.

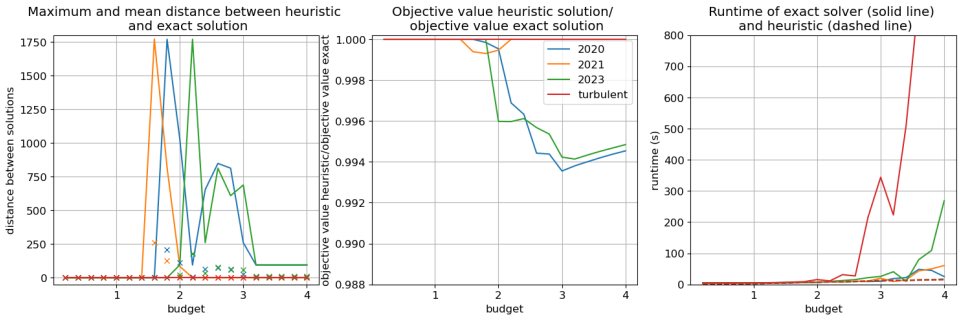


Figure 7.24: Comparison of greedy heuristic and exact solution for instances with different wind direction data. The legend is the same for all three plots. The distance between the solutions, the optimality gap and the differences in runtimes are shown.

The heuristic could be improved by adding a local search to the found solution. However, we can see from Figure 7.22, Figure 7.24 and Figure 7.25 that the distance between the catching systems in the optimal solution and the solution from the heuristic can be larger than 1000 meters, which means that the local search heuristic would have to search many options to find the optimal solution. Additionally, the runtime of the heuristic could be improved by investigating a faster calculation of the inverse that is necessary calculate the caught flow in the objective function when testing the new locations and types of catching systems in each step of the heuristic. The matrix $[I - T(\boldsymbol{x})]$ for which we need to compute an inverse changes by one row when we add a catching system in each step of the heuristic. The Sherman-Morrison formula could perhaps speed up the calculations of the inverse, since we only need the inverse of $[I - T]$ which we permute by vector calculations in each step instead of computing the inverse again.

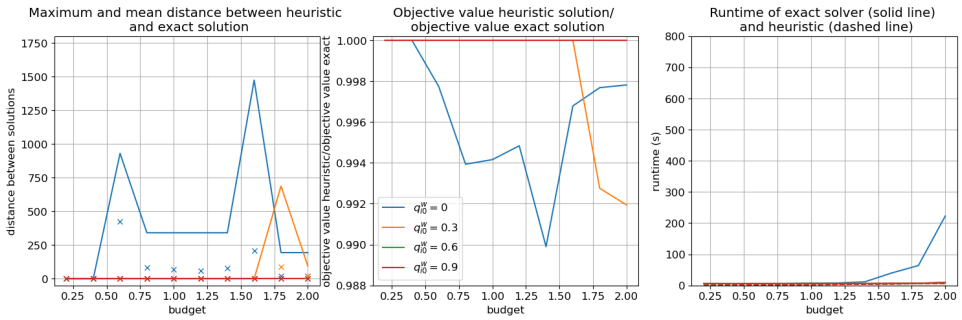


Figure 7.25: Comparison of greedy heuristic and exact solution for instances with different probability of getting stuck due to water vegetation. The legend is the same for all three plots. The distance between the solutions, the optimality gap and the differences in runtimes are shown.

7.2.4 QUALITY OF THE SOLUTIONS

We shortly investigate the quality of the found solutions of the runtime tests in Figure 7.21 where the solver was cut off. For the following cases, the CBC solver was cut off after 12 hours of runtime:

- $n = 308 (d = 100), B = 3.6$,
- $n = 412 (d = 75), B = 2.4$,
- $n = 622 (d = 50), B = 1.6$.

The Gurobi solver was cut off after 12 hours of runtime for the instances:

- $n = 412 (d = 75), B = 3.6$,
- $n = 622 (d = 50), B = 2.4$.

We show the objective value after cutting off the solution and compare this to the objective value of the heuristic solution in Table 7.1. We see that the gap between objective values of the solutions after cutting off the solver at 12 hours is very small for the CBC solver. The cut off solutions from the Gurobi solver have a larger gap.

Table 7.1: The objective value of the solutions from the CBC solver after cut off at 12 hours are compared to the optimal solutions of the Gurobi solver and the heuristic solutions.

n	d_{max}	B	CBC solver objective value after 12 hours	Gurobi solver objective value optimal solution	heuristic solution objective value
308	100	3.6	0.138732143	0.138732144	0.138732144
412	75	2.4	0.112281697	0.112282056	0.112281697
622	50	1.6	0.116068729	0.116068729	0.116068729

Table 7.2: The objective value of the solutions from the Gurobi solver after cut off at 12 hours are compared to the heuristic solutions.

n	d	B	Gurobi solver objective value after 12 hours	heuristic solution objective value
412	75	3.6	0.137066623	0.131829372
622	50	2.4	0.136881322	0.131577573

From the comparison of the runtimes of these different solution methods, it seems most suitable to use the Gurobi solver and try to keep the number of nodes at a maximum of $n \approx 375$. For this number of nodes, the Gurobi solver does not need an academic license and instances with a budget of maximum $B \leq 2$ can be solved efficiently within a few hours. For instances with larger budgets, it seems most suitable to use the heuristic.

7.3 CASE STUDY: DELFT

The sensitivity analysis and tests of the computational runtime of the model are used to decide how the final model should be used. In this section, we show how a single run of the PW-FCLM works for a case study in Delft.

First of all, the data from Chapter 6 should be pre-processed such that it can be used as an input to the model. In Appendix A, a block diagram of all the different types of input and their processing is shown. A manual for the pre-processing of the data is also available.

The impact factor α_i is only used for catching early in this case study, and is equal to $0.25 \cdot \max_{i \in V, k \in K} \beta_{ik} M_{ik2}$ for all $i \in V$. Then, we run the complete model. A solution file is written that can be plotted on the map in QGIS, such that the locations, types and orientations of the catching systems are shown on the map. The proportion of the total amount of plastic flow caught by each type of catching system is also shown by a label on the map. An example for a case study where $n = 366$ ($d_{max} = 85$) with $B = 0.6$ is shown in Figure 7.26. The proportion of the total amount of plastic flow through each node is shown on the map by a graduated color. This run is from a winter scenario, where the probability of getting stuck due to water vegetation is equal to zero for all nodes $i \in V$. The other input parameters are as described in Chapter 6.

In Figure 7.26, we see that the chosen locations are at nodes where the plastic flow is high. The plastic flow seems highest in several nodes at intersections of the waterways, however, it is not possible to place a catching system in these locations. The total flow caught is equal to 0.12, which means we are catching 12% of the plastic flow in the area.

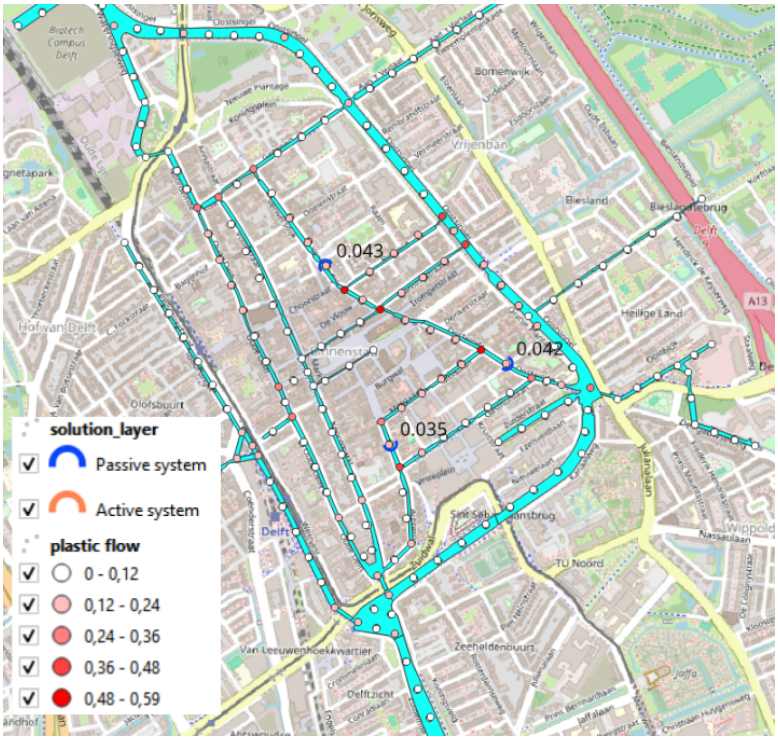


Figure 7.26: Example of a run of the model with $n = 366$ ($d_{max} = 85$) and $B = 0.6$ in Delft. The catching systems in the solution are shown as half circles with their opening oriented in the same way as the catching system should be oriented. The blue color represents a passive catching system. The amount of plastic flow caught by the catching systems is printed on the map next to each system. For each node, the amount of input flow through the node (when there are no catching systems in the area) is shown by graduated colors.

8

Discussion

From the computational results in Chapter 7, we see that the PW-FCLM designed in this thesis is most sensitive to the distance between the nodes d_{max} and the probability of getting stuck due to water vegetation q_{i0}^{w} when testing the model for the study area of Delft. The initial distribution $\boldsymbol{b}$ and the transition probabilities q_{ij} due to the wind are slightly sensitive when comparing to a new scenario with uniform probabilities. However, this situation seems far from reality. To verify the input data for the initial distribution, it could be useful to do a field study of counting the number of plastics found per location soon after a certain area has been manually cleaned very thoroughly, such that all plastics found must have recently entered the water. Furthermore, the correlation between the wind direction and the movement of plastics could be further investigated by performing and studying more GPS routes of moving plastics. While it is not the focus of this thesis, Noria could benefit from more research to verify the probabilities we assigned to each type of input parameters, such that the input represents reality most accurately. Then, PW-FCLM finds the optimal locations for the most realistic scenario.

When comparing the sensitivity analysis for a different area, i.e., Groningen, the same types of input parameters seemed most sensitive. Here, the model seemed slightly more sensitive to the probability of getting stuck due to more houseboats and shore vegetation. However, only a small decrease could be seen in the relative difference of flow caught with the base scenario optimal solution compared to perfect information.

The sensitivity of the model to the distance between the nodes d_{max} is most important to further develop the PW-FCLM. The reason why the model is so sensitive to this parameter seems to be because the nodes in the network end up in different locations such that the waterways are slightly wider or more narrow depending on the location. This can largely influence the catching probability of a node. While it is true that locations where the canal is wider have a higher catching probability, it should not depend on the chosen distance between the nodes whether the model considers these locations. This could be mitigated by calculating the maximum width of the waterway on the whole segment of the waterway from one node to the next. The current way of calculating the width of the waterway using the polygon

layer in QGIS does not allow us to do this, but can be developed to improve the model.

The sensitivity of the model to the probability of getting stuck due to water vegetation is also interesting to examine more carefully. Since this factor is the main reason why plastics get stuck in Delft, setting this probability to a lower value means that plastics can travel a lot further in the network, which changes the flow and the corresponding optimal solutions for catching systems a lot. We used $q_{i0}^w = 0.45$ to average for the half of the year where there is a lot of water vegetation where we expect that $q_{i0}^w = 0.9$ and the other half of the year where we expect that $q_{i0}^w = 0$. We could consider expanding the model to allow for moving the catching systems to a different location seasonally.

The Markov chain representation of the flow should be further investigated. In this version of the model, we do not take the influence of time into account in the plastic flow, since it does not matter exactly how long it takes for plastics to travel a certain way for them to be caught by a catching system along the way. However, depending on the distance between the number of nodes, plastics travel less far in the network if the probability of getting stuck remains the same but there are more nodes per distance of the waterways. Perhaps, the probabilities of getting stuck should be adjusted, such that the average distance that the plastics travel does not depend on the chosen distance between the nodes. Furthermore, it is assumed that plastics that get stuck can not continue moving through the network later on. While this is true for some plastics, there are also plastics that eventually continue moving, for example when the wind direction changes. For future research, it is recommended to find a way to incorporate this effect in the model and to perform tests with GPS trackers to find out more about the probability that this happens.

Additionally, sharp corners, boats and shore vegetation are counted if they are within a radius of the node that we are considering. In this thesis, it was not investigated whether the optimal solution changes when we precisely take into account that plastics only get stuck due to sharp corners, boats and shore vegetation if they traveled along an edge that passed any of these factors. However, this means that the probability of getting stuck depends on which edge the plastics traveled in the previous step, which violates the Markov property of "memorylessness", which means we could no longer use the MDP theory that we used to design the model.

It is important to think about which decisions are difficult to make using human reasoning and benefit from the use of the model, and which decisions we over-complicate by including them in the model. For example, we could argue whether the model should make the trade-off between the two types of catching systems. Right now, it seems that the passive system has the best accuracy compared to its costs. The reason why an active system could be more suitable for certain locations is when there is a large water flux, for example due to tides, or when there is a large amount of plastics flowing by the locations. The active system has a larger compartment to store caught plastics and can be emptied more easily. The difference between the two types of catching systems could be included in the

model more precisely if the capacity of the catching systems and the corresponding expected costs for emptying the catching systems are also included.

When the model is used in a very large area, it is necessary to put the nodes at a larger distance d_{max} because of the runtime of the model. It is recommended to further explore what distance is most appropriate for larger areas such as provinces or even countries. It is also interesting to see whether the distance between the nodes should be the same for each type of waterway. For example, we could place fewer nodes on a straight waterway and more nodes on more detailed waterways or on waterways with many factors that influence the probability of getting stuck. Furthermore, it is interesting to look into the representation of the amount of plastic flow in the model. Currently, the flow is represented as a proportion of the total flow in the area. However, this does not say much about the total number of pieces of plastic or their volume or weight, while this might be interesting and more intuitive for the user. This would especially help in choosing parameters w and α_i.

For all considerations, it is important to keep in mind whether adding extra features to the model really makes the model more realistic and accurately helps the decision making, or if the model is being over-complicated and the runtime is unnecessarily increased.

Finally, the results of our heuristic showed that the performance of the heuristic is at least 98% of the optimum objective value. While it seems that the heuristic works well for our test instances, we can not guarantee its performance on new instances yet. A worst case performance proof of $f_{\text{greedy}}/f_{\text{opt}} \geq 1 - e^{-1} \approx 0.632$ was shown for a similar greedy heuristic by Berman et al. (1995c). It would be interesting to investigate whether a worst case performance bound can be proven for our version of the MDP-based FCLM with a greedy heuristic. Furthermore, an extension of the heuristic might provide even better results. We noticed that a small distance between the nodes was necessary for the tests in this thesis and the runtime was shown to increase quickly with extra nodes. Therefore, it seems promising to focus on finding a close to optimal solution if it decreases the runtime, especially to make the model more suitable for larger areas.

9

Conclusion

In this thesis, we developed a Plastic Waste Flow Capturing Location Model to find the optimal locations to catch as much plastic waste from waterways as possible. The plastic waste flow is represented by a Markov chain with an initial distribution based on plastic waste sources in the surroundings, transition probabilities based on wind directions and water geometry and a probability of getting stuck due to dead ends, sharp corners, docked boats, shore vegetation and water vegetation. A way of processing this data into input parameters for the Markov chain is described in this thesis.

The existing Markov Decision Process based Flow Capturing Location Model is extended with the possibility to have different types of catching facilities, taking sensitive areas into account and having a specific orientation for the catching facilities to catch flow. The equivalence of the linearized version of our extension of the model is shown and a proof of NP-hardness of the problem is given. A greedy heuristic is also presented.

The sensitivity analysis of the input parameters for a case study in Delft showed that the model is most sensitive to the distance between the nodes that are considered as possible locations and to the probability of getting stuck due to water vegetation. The same parameters seemed most sensitive in a different study area of Groningen.

The model seems sensitive to the distance between nodes because the nodes in the network end up in different locations such that the waterways are slightly wider or more narrow depending on the location, which can largely influence the catching probability of a node. It is valuable to examine a way to mitigate this, for instance, by calculating the width of the canal that can be blocked by the catching system in a different way. It is also recommended to further research the Markov chain representation of the flow, since in the current version of the model, the plastics travel less far in the network when the distance between the nodes is decreased and the probability of getting stuck remains the same.

When using the model for new case studies, it seems from the analysis of the computational runtimes that it is most suitable to set the distance between the catching systems such that $n \leq 375$ for budgets $B \leq 2$. This allows the user to solve

the model exactly using the Gurobi solver without an academic license. If it is desired to have a more detailed network with more nodes or the budget is higher, the heuristic seems most suitable. For further research, it is recommended to study how the model should be used on much larger regions as case studies, such as provinces or entire countries.

BIBLIOGRAPHY

Averbakh, I. and Berman, O. (1996). Locating flow-capturing units on a network with multi-counting and diminishing returns to scale. *European Journal of Operational Research*, 91:495–506.

Berman, O., Bertsimas, D., and Larson, R. C. (1995a). Locating discretionary service facilities, ii: Maximizing market size, minimizing inconvenience. *Research*, 43:623–632.

Berman, O., Krass, D., and Wei Xu, C. (1995b). Locating discretionary service facilities based on probabilistic customer flows. *Transportation Science*, 29(3):276–290.

Berman, O., Krass, D., and Wei Xu, C. (1995c). Locating flow-intercepting facilities: New approaches and results. *Annals of Operations Research*, 60:121–143.

Berman, O., Krass, D., and Wei Xu, C. (1997). Generalized flow-interception facility location models with probabilistic customer flows. *Communications in Statistics. Stochastic Models*, 13(1):1–25.

Berman, O., Larson, R., and Fouska, N. (1992). Optimal location of discretionary service facilities. *Transportation Science*, 26(3):201–211.

Boccia, M., Sforza, A., and Sterle, C. (2009). Flow intercepting facility location: Problems, models and heuristics. *J. Math. Model. Algorithms*, 8:35–79.

Borrelle, S. B., Ringma, J., Law, K. L., Monnahan, C. C., Lebreton, L., McGivern, A., Murphy, E., Jambeck, J., Leonard, G. H., Hilleary, M. A., Eriksen, M., Possingham, H. P., Frond, H. D., Gerber, L. R., Polidoro, B., Tahir, A., Bernard, M., Mallos, N., Barnes, M., and Rochman, C. M. (2020). Predicted growth in plastic waste exceeds efforts to mitigate plastic pollution. *Science*, 369(6510):1515–1518.

Brotcorne, L., Laporte, G., and Semet, F. (2003). Ambulance location and relocation models. *European Journal of Operational Research*, 147(3):451–463.

Church, R. and ReVelle, C. (1974). The maximal covering location problem. *Papers of the regional science association*, 32(1):101–118.

Grubesic, T. H., Wei, R., and Nelson, J. (2017). Optimizing oil spill cleanup efforts: A tactical approach and evaluation framework. *Marine Pollution Bulletin*, 125:318–329.

Hodgson, M., Rosing, K., Leontien, A., and Storrier, G. (1996). Applying the flow-capturing location-allocation model to an authentic network: Edmonton, canada. *European Journal of Operational Research*, 90(3):427–443.

Hodgson, M. J. (1990). A flow-capturing location-allocation model. *Geographical Analysis*, 22(3):270–279.

Kemeny, J. G. and Snell, J. L. (1960). Finite markov chains. *Springer*.

Kuby, M., Lines, L., Schultz, R., Xie, Z., Kim, J. G., and Lim, S. (2009). Optimization of hydrogen stations in florida using the flow-refueling location model. *International Journal of Hydrogen Energy*, 34:6045–6064.

Lebreton, L., Slat, B., Ferrari, F., Sainte-Rose, B., Aitken, J., Marthouse, R., Hajbane, S., Cunsolo, S., Schwarz, A., Levivier, A., Noble, K., Debeljak, P., Maral, H., Schöneich-Argent, R., Brambini, R., and Reisser, J. (2018). Evidence that the great pacific garbage patch is rapidly accumulating plastic. *Scientific Reports*, 2018.

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

Sherman, P. and Sebille, E. V. (2016). Modeling marine surface microplastic transport to assess optimal removal locations. *Environmental Research Letters*, 11.

Tan, J. and Lin, W.-H. (2014). A stochastic flow capturing location and allocation model for siting electric vehicle charging stations. *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, pages 2811–2816.

Wagner, M., Scherer, C., Alvarez-Muñoz, D., Brennholt, N., Bourrain, X., Buchinger, S., Fries, E., Grosbois, C., Klasmeier, J., Marti, T., Rodríguez-Mozaz, S., Urbatzka, R., Vethaak, A., Winther-Nielsen, M., and Reifferscheid, G. (2014). Microplastics in freshwater ecosystems: What we know and what we need to know. *Environmental Sciences Europe*, 26:12.

Wu, F. and Sioshansi, R. (2017). A stochastic flow-capturing model to optimize the location of fast-charging stations with uncertain electric vehicle flows. *Transportation Research Part D: Transport and Environment*, 53:354–376.

Yang, J., Xiong, J., Liu, S., and Yang, C. (2008). Flow capturing location-allocation problem with stochastic demand under hurwicz rule. *2008 Fourth International Conference on Natural Computation*, 7:169–173.

You, F. and Leyffer, S. (2011). Mixed-integer dynamic optimization for oil-spill response planning with integration of a dynamic oil weathering model. *AIChE Journal*, 57:3555 – 3564.

Ypma, S. L., Bohte, Q., Forryan, A., Naveira Garabato, A. C., Donnelly, A., and van Sebille, E. (2022). Detecting the most effective cleanup locations using network theory to reduce marine plastic debris: a case study in the galapagos marine reserve. *Ocean Science*, 18(5):1477–1490.

A

APPENDIX: FLOWCHART OF PW-FCLM

An online link of the flowchart is available via Drawio. A pdf image of the flowchart is also attached below.