Synthesis Project 2021

Finding the plastic hotspots with (GIS) data

Team members

Lisa Geers
Pinelopi-Eirini Kountouri
Katrin Meschin
Konstantinos Pantelios
Androniki Pavlidou
Simon Pena Pereira

TU Delft

NORIA SUSTAINABLE INNOVATORS

FINDING THE PLASTIC HOTSPOTS WITH (GIS) DATA

Synthesis Project 2021

MS.c Geomatics for the Built Environment

18 June 2021

Team members:

Lisa Geers
Pinelopi-Eirini Kountouri
Katrin Meschin
Konstantinos Pantelios
Androniki Pavlidou
Simon Pena Pereira

Supervisors TU-Delft:

Giorgio Agugiaro
Ken Arroyo Ohori

Client:

Noria:
Rinze de Vries,
Sophie Broere

Abstract

The plastic pollution of aquatic environment is undoubtedly an emerging environmental risk, as it negatively affects ecosystems globally to a great extent. To prevent the plastic soup from growing even further, a Delft-based start-up Noria has developed plastic collectors, to remove plastic from rivers and canals before it reaches the ocean. In order for these devices to give maximum positive effect, they need to be installed in areas where plastic is more likely to accumulate - the plastic hotspots. Taking into consideration various natural attributes that affect the movement of the plastic waste in the water, such as wind direction, water flow, canal geometry, vegetation and man made structures in waterways; potential hotspots can be predicted in a model which would allow more efficient coordination of the cleaning process. Thus, this project aims to locate plastic accumulation zones in the city of Delft in a (semi-) automated manner using open spatial data analysed in GIS and a network simulation model.

The methodology developed in this project results in the visualisation of potential plastic hotspots where Noria's collectors could be placed in order to remove and recycle the plastic. The potential hotspots suggested by the model were compared with ground truth data collected. The final result yielded only 20% accuracy and therefore did not meet the initial expectation. An evaluation of the shortcomings was made with suggestions for future research.

Acknowledgements

This report is documenting the work that has been accomplished in order to find the plastic hotspots with spatial data within the context of the Synthesis project. The current project was conducted by a team of six first year students of the Master of Geomatics for the Built Environment at Delft University of Technology. In this project, the Master students put their effort and knowledge together for 10 weeks to implement a topic given by Noria - a Delft-based start-up focusing on innovative sustainability solutions. Throughout the project the students cooperated with Noria to gain further knowledge and skills in processing, analysing and visualising spatial data.

The assistance and direction from our supervisors and client were central for this project's progress. Their contribution is greatly appreciated by all the team members. We would like to express our gratitude to Giorgio Agugiaro and Ken Arroyo Ohori for the support and feedback they provided us during the meetings and thank also our clients, Rinze de Vries and Sophie Broere, for sharing with us the innovative methodology they already use and providing us guidance for our work during our weekly meetings.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Context

The accumulation of plastic waste in water resources is threatening the ecosystem, biodiversity and human health globally. It has been estimated that over 5 trillion pieces of plastics that weight more than a quarter million tons floats in the ocean around the world (Eriksen et al. 2014). The majority of plastics originate from inland and are transported to the sea by rivers, which act as conveyor belts collecting more and more plastics as they move downstream (Parker 2019). When plastic is not removed from the environment, it decomposes into small particles and stays in the environment as micro-plastic, making it almost impossible to collect and dispose. The abundant micro-plastics in the oceans has proven to interact with the ecosystem, thus disrupting biogenic flora and fauna (Kandasubramanian and Issac 2021). New innovative solution for plastic removal from the water network has been developed by Delft-based start-up Noria. Their plastic collectors can remove up to 95% of the plastic floating past the device. In order to get the maximum benefit of such plastic collectors, they need to be installed in plastic accumulation zones, called the plastic hotspots. While thus far such hotspots have been detected by field work which is time consuming, more efficient automated way is needed for hotspot prediction. Such automated model, however, requires a good understanding of plastic transport in Dutch urban environment.

Currently research on plastic hotspots has dominantly been conducted for oceans as the highest plastic concentration in surface waters is considered to be in subtropical gyres (Eriksen et al. 2014). However, as it is suggested that 4.8 to 12.7 mega tons of plastic enters the oceans each year largely through rivers (Jambeck et al. 2015), the new branch of research on riverine plastic debris transport is considered as relatively young science (Emmerik and Schwarz 2020). The main research topics in this field have concentrated on quantifying riverine plastic flow or general modelling of total plastic transported by rivers. Common research method for quantifying plastic transport in smaller study area are labor intense, involving field work for counting plastics in freshwater environment either by mapping (Tasseron et al. 2020), by collecting surface plastic (Gasperi et al. 2014) or collecting subsurface plastic with nets (Morritt et al. 2014). Such methods are time-consuming, expensive and spatially limited. A quicker, cost-effective and semi-automatic analysis that would improve our understanding on plastic transport and accumulation in the environment is needed. However, several studies have highlighted the knowledge gap in understanding of the riverine plastics.

This report aims to contribute to the understanding of the plastic transport in freshwater environment as well as provide an experimental method for automated tool that predicts plastic hotspots and builds on factors affecting plastic transport described by research done so far. A workflow is presented, incorporating a spatial data analysis using Geographic Information System (GIS) and agent-based network model for simulating plastic transport and hotspot generation in Dutch urban environment. This model can be used by Noria for getting a better overview of plastic transport in the urban environment and find areas where plastic collection would have the most positive effect, without having to conduct time-consuming and labor-intense field work.

1.2 Problem definition and scope of this project

For automatic plastic hotspot detection, one of the approach used in ocean environment has been pixel-based image analysis from airborne or spaceborne images. Examples of such are hotspot maps of marine plastic debris in Hawaii (Moy et al. 2018) and usage of SWIR spectral signatures of plastics in the ocean (Shungudzemwoyo P Garaba et al. 2018). Considering freshwater environment, Jakovljevi et al. (2019) developed an experimental algorithm for detecting floating plastic in river environment using remote sensing data. However, it was concluded, that due to the nature of freshwater bodies, with mud, turbidity, suspended solids and phytoplankton - the algorithm did not work as well as it would in the open ocean (Jakovljević, Govedarica, and Taboada 2019). This leaves a gap for quick detection and modelling of plastics in fluvial environment as the methods used in ocean environment cannot be used. Concerning this issue, this report proposes a new experimental method aiming to create a simulation of plastic transport and accumulation in freshwater network, building on findings from previous research on parameters affecting plastic transport in inland waterbodies.

There is strong evidence from around the world for the correlation between high quantities of plastic waste in freshwater systems and high population density (Best 2019). Tasseron et al. (2020) studied the hotspots in Dutch cities of Leiden and Wageningen by counting and categorising the number of plastics in water. The difference in the distribution of the plastic hotspots in the two Dutch cities was suggested to be due to the concentration of potential sources and proximity of the canals to the city center. Plastic accumulation was noticed around locations where water flow was obstructed, such as dead ends, houseboats, quay walls, bridges and in vegetation. Lastly, it was noted, that in both of the cities most of the plastic was of Multilayer and PO-soft type, associated with food wrappings and plastic bags, making the fast food restaurants, market places and shops as one of the key plasticc sources (Tasseron et al. 2020). Indeed, similar trend was found in River Seine, France, where significant proportion of plastics collected by a network of floating debris-retention booms consisted of food wrappers and containers and plastic cutlery (Gasperi et al. 2014). Using this knowledge, the plastic sources and potential accumulation zones could be mapped and analysed in the urban environment, forming the basis of the assumptions for our GIS spatial analysis.

In the present project we seek to implement a semi-automatic method that would illustrate the plastic movement in Dutch urban water network and identify locations where plastics would potentially accumulate. The study focuses on the city of Delft in the province of South Holland in the Netherlands. The method is uses spatial data analysis for parameters accounting for plastic movement and accumulation, and network analysis for simulating plastic transport in water network in Delft. In more detail, taking into account the wider problem of plastic accumulation and its extent, in order to develop a strategy for their detection in water resources (rivers/ channels) of the study area, we considered three key aspects of plastic transport:

- The sources from where plastic gets into the water;
- Plastic movement according to the dominant wind and flow direction;
- Plastic accumulation in hotspots;

Based on these three aspects, we aim to answer the main research question: *How can potential plastic hotpots be identified in a (semi-) automatic way in freshwater bodies in the city of Delft?* To create algorithms for an automatic or a semi-automatic model that will present the potential locations, it is necessary to use relevant spatial data (see Chapter 2.1). Specifically, the main objective is to combine the spatial information and characteristics from different data sets (layers), with the ultimate goal to develop a procedure that will return automatically the locations where plastics are accumulated.

Due to the lack of previous research with similar aim, our process is divided into two phases, evolving from testing the effectiveness of different parameters and datasets separately to a more comprehensive network model combining different parameters in agent-based network model (see Chapter 2.3 and 2.4).

The main steps taken for that purposes are shown in flowchart below:

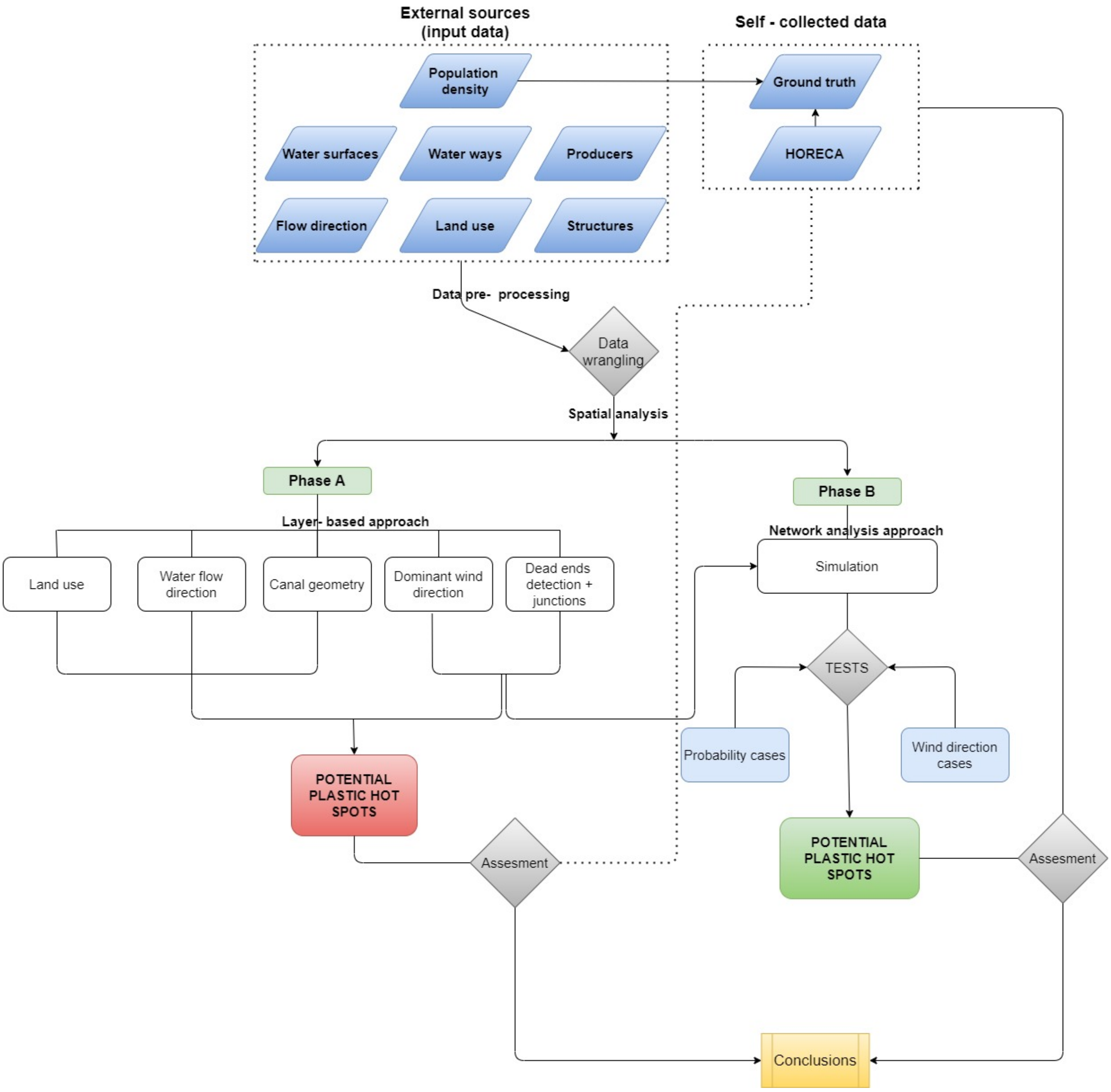


Figure 1.1: Flowchart of the research approach

1.3 Stakeholders

It is clear that in order to solve a global problem, collective efforts are needed to address it and Noria actively contributes to this effort. The company, located in the city of Delft in the Netherlands, seeks a method to identify and remove plastics from the water using their smart solutions, addressing the problem with innovation. In more details, by using GIS data and field work for estimating where more plastic may be found, Noria is placing their specially designed plastic collectors in order to remove the floating plastics from the water. The way the company operates is based on three steps:

- The analysis of the problem in the perspective study area. In this stage the extent of the problem is analysed in the working region by trying to answer questions like: *"How much plastic flows in your water?"*, *"What are the most ideal locations to collect plastic?"* and *"Which parties must be jointly involved in the solution?"*
- The deposition of plastics. In this part, the company places, monitors and maintains their system at strategic locations that autonomously filters the floating plastic from the water.
- The re-use of the collected plastics. By re-using the collected plastic, it is possible to make new high-quality products, therefore saving resources while raising awareness of the issue. Noria has already produced several products, like waste grabbers, from the plastic collected from the water.

Our connection with Noria is made through the Technical University of Delft and as our client, the company commissioned us to devise the project with the theme ***"Finding the plastic hotspots with (GIS) data"***. The implementation of the project is done under the help and guidance of our two supervisors, Dr. Giorgio Agugiaro and Dr. Ken Arroyo Ohori, members of the 3D Geoinformation Group. They contributed actively to both the understanding of the problem and the techniques developed for its completion. Ultimately, our team is responsible for implementing the theory that characterizes the problem of detecting plastics at points of potential accumulation.

1.4 Results

The method developed in this project creates a model which performs network analysis connecting the plastic sources, parameters affecting plastic transport and the potential plastic accumulation zones. For plastic sources, the locations of restaurants and markets derived from the Google Maps as well as population data were included. The plastic transport was modelled along the water line features which made up the network in our network model. In addition, simplified parameters of wind with constant velocity and direction as well as water flow with constant velocity are affecting the movement of plastic objects in the model. Lastly, the model considers canal geometry and dead ends of canals where flow is obstructed in relation to wind direction to determine if the plastic may start to accumulate in such areas or not.

Utilising the information provided in combination with the available tools and softwares, we developed a simulation model which identifies the potential places where plastics tend to accumulate in the water bodies of the study area. After implementing our approach, we came to the conclusion that the model is not sufficiently representing the complexity of the urban environment, given the fairly large set of assumptions taken into account. The accuracy of our simulation reaches 20 %, which shows the inability of the model to approach reality. However, lessons learned from this project with recommendations for future studies may lead the way for improving the outcome of such automatic modelling of plastic hotspots.

1.5 Reading guide

The report is structured as follows: Chapter Two provides the experimental methods and methodology followed, as well the information regarding the data used in this project. In Chapter Three the results are presented that are derived from the developed methods and algorithms, accompanied by their limitations. The evaluation of the developed approaches are illustrated in Chapter Four, while in Chapter Five the conclusions are provided with recommendations for future work.

2 | Methods

2.1 Datasets

For the purpose of developing any form of models and algorithms using GIS data, it is of vital importance to have a good knowledge of the data as well as of the corresponding content (metadata, attributes). Especially in cases of complex problems (i.e. the present project), where a combination of different datasets must be used, their quality and completeness (metadata and attributes information) are determinant factors that directly affect the final result.

In order to end up with the most appropriate, dataset to use for each case, the user has to consider many factors that directly affect the models and algorithms that are under construction, as well as the final outcome. The importance of the data quality and its completeness, in our case, was clear during Phase A (2.3), both during the algorithms' construction as well as in the final results, the derived models showed to be affected directly from the input datasets. When talking about datasets, we do not solely focus on the information presented on the map, but on their attributes as well. Having the knowledge of the attributes and their corresponding meaning, the user will gain knowledge of the importance of investing in information technology. This is something that can be explained with the DIKAR (Data, Information, Knowledge, Action, Results) model (2.2), in which the direct connection and interaction between information, knowledge and the final users action, in order to solve a problem is presented (Ward J. 2016). The theory behind the model and each intermediate part is that the main problem is divided into smaller ones (named as gaps) and the solution of them leads to the solution of the initial problem. Having a good (and/or complete) knowledge about the used data and their attributes will positively influence both the decision-making ability and the quality of the final results (since the user is able to judge what information is useful or not, the user will take the action critically) (Murray 2002). In our approach, we followed the steps shown in the DIKAR diagram, starting with the data provided to us, continuing to decode the data into the corresponding information that can be derived from them, as well as their most in-depth knowledge, so that we can act accordingly, through the development of our models, to reach the final result regarding the detection of plastic accumulation.

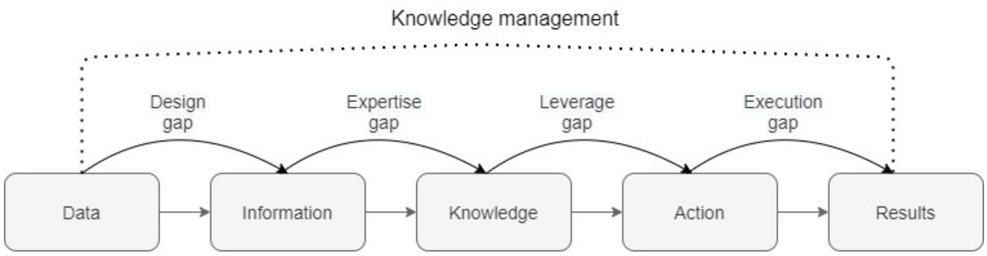


Figure 2.1: DIKAR model

For the chosen approach, our project needed datasets for a water network, the flow direction of the channels in this network, plastic sources such as restaurants and population data, locations of physical structures in the water that may obstruct the water flow and for vegetation. Therefore, before starting the implementation of our approach, it was necessary to first collect the datasets relevant for our problem. The main sources used for datasets collection are shown in Figure 2.2. It should be noted that all datasets used are of vector type.

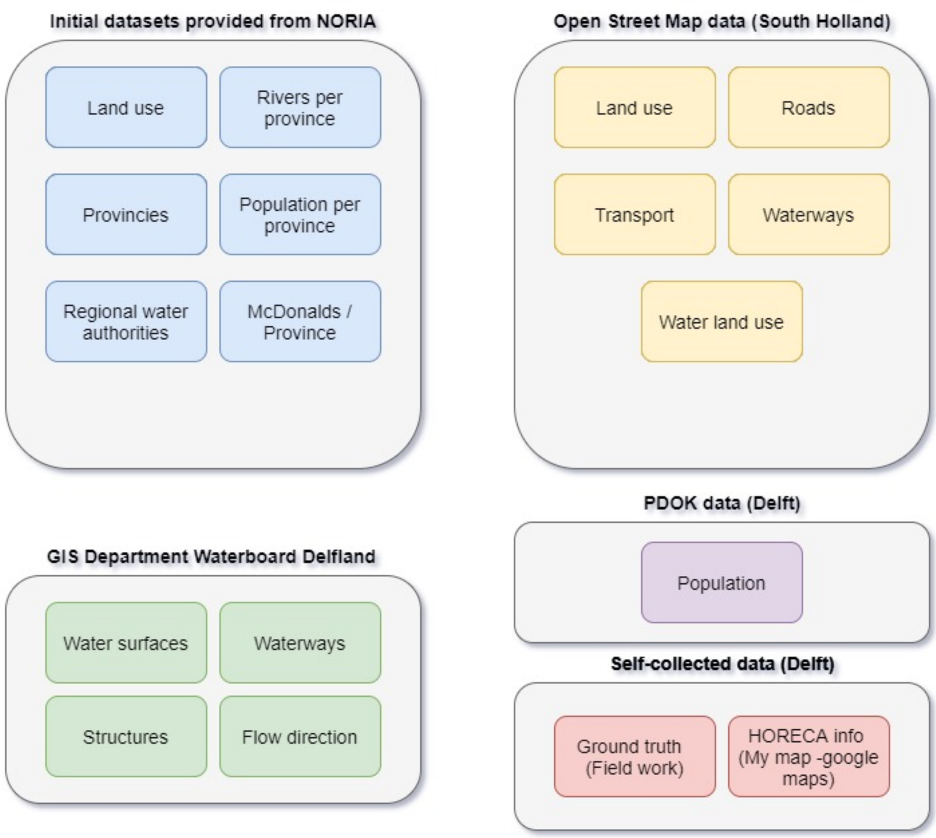


Figure 2.2: Data sources

When conducting a GIS analysis, the quality and integrity of data has major impact on the final results. It is important to note, that not all the datasets shown in Figure 2.2 were available for us from the start of the project and extra time was put into finding and requiring the needed datasets from the data owners, such as the Waterboard of Delftland. From the start of the project, only the initial datasets shown in Figure 2.2 were available for us. Due to the need for more accurate data, particularly for the water network, Open Street Map (OSM) data was taken into use. Finally, by the time we had already reached Phase B in our development, we received more detailed data from Waterboard Delftland, also incorporating water flow direction as well as physical structures.

In order to later assess our model and validate our results, ground truth data to test our model was needed. For this, field work was conducted as we separated the study area into smaller regions of interest based on the distribution of water network and our first results extracted from our models in Phase A (2.3). The duration of the field work was four days during which the weather was mostly windy with the wind blowing from South and South-West, which is the dominant wind direction for the Netherlands. In this procedure, we took into account the number of plastics found in the visited places as well as the factors that may lead to plastic accumulation in the specific location. The covered ground truth area is shown in Figure 2.3.

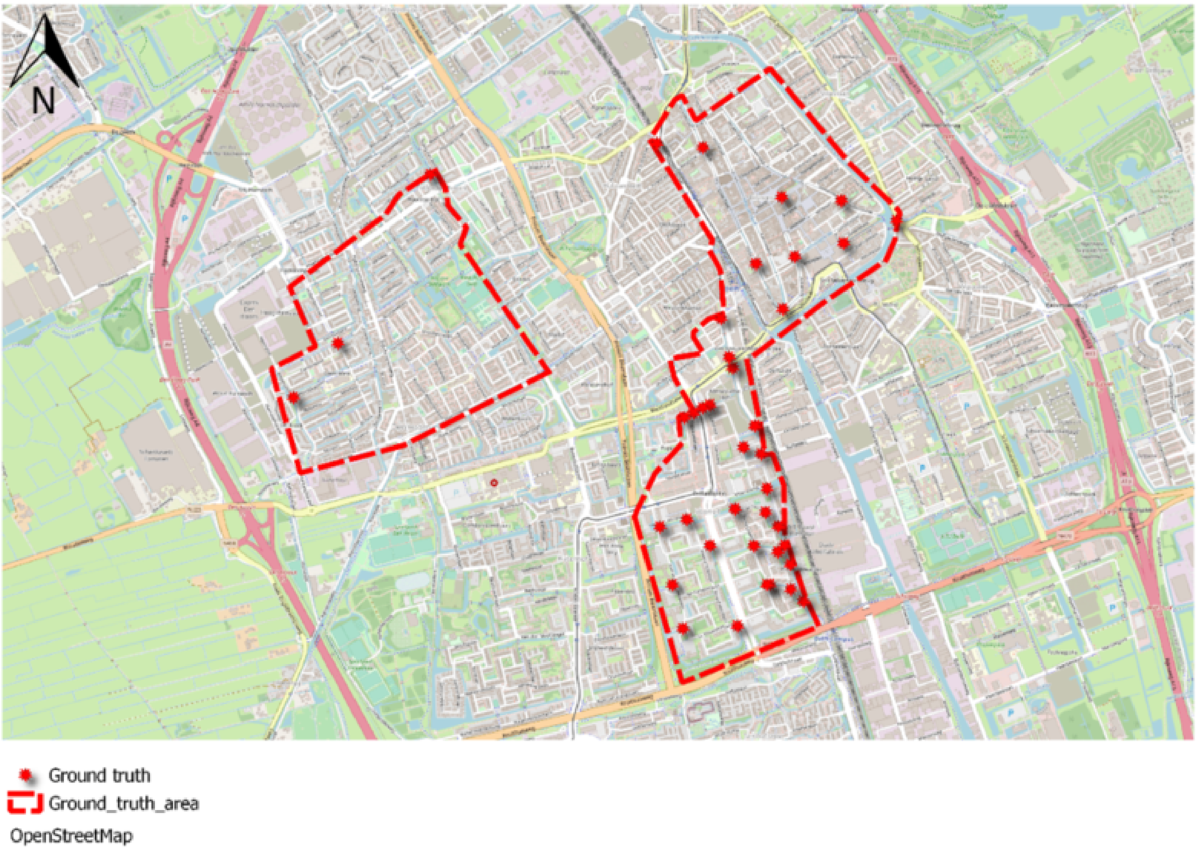


Figure 2.3: Ground truth area

Based on the studies that relate higher plastic pollution to higher population density and number of fast food restaurants (Best 2019; Tasseron et al. 2020; Gasperi et al. 2014), a dataset on population density in Delft was included. To account for fast food restaurants, additionally to the layer showing all the McDonald's fast food restaurants in the study area provided by our client Noria, another dataset was created to include more comprehensive potential plastic sources. It was assumed that other important plastic sources in Delft city center exist, such as the markets, restaurants and other fast food chains. A layer was created about the HORECA (leaving out the hotels) using the Google Maps tool My-Map. All the available restaurants, bars, cafes and markets in the study area were included in the dataset, as well as the main squares in the city center of the city of Delft (Figure 2.4).

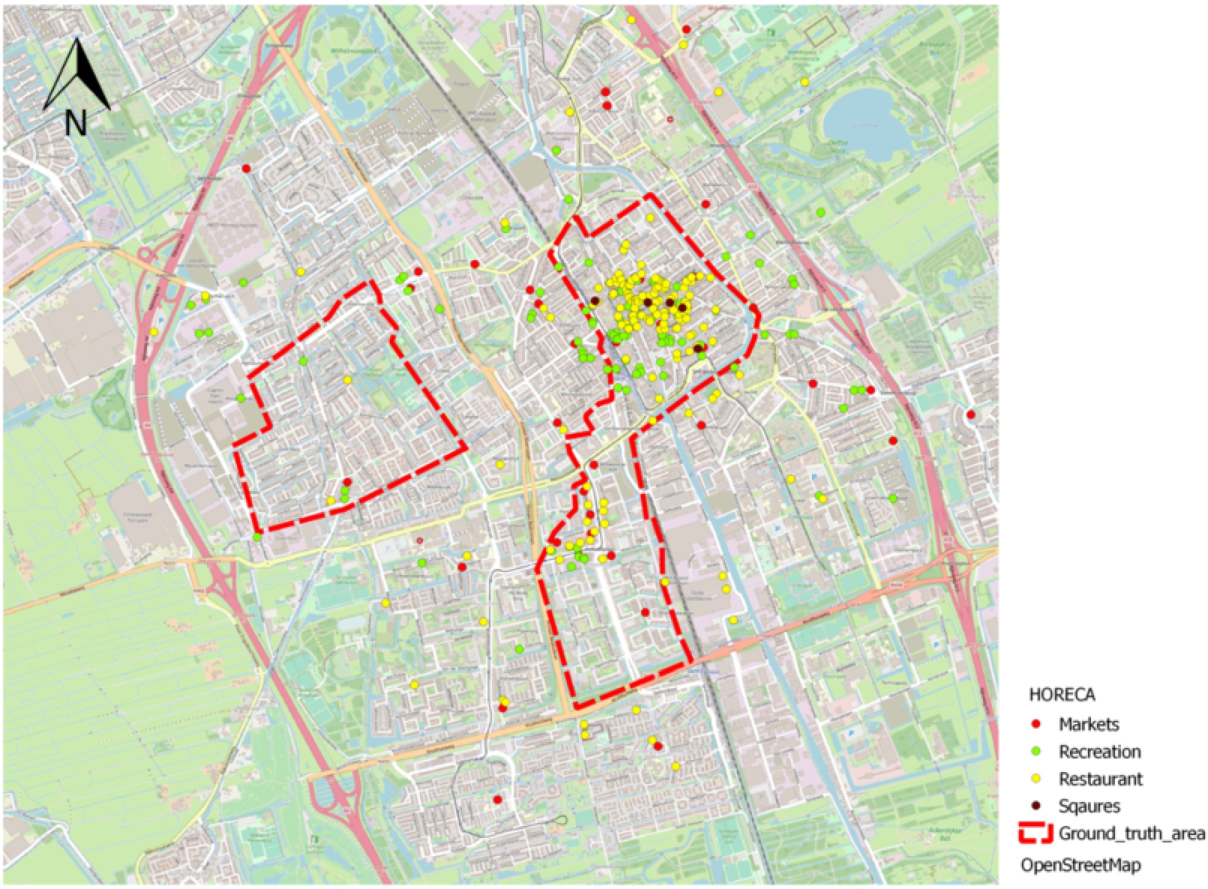
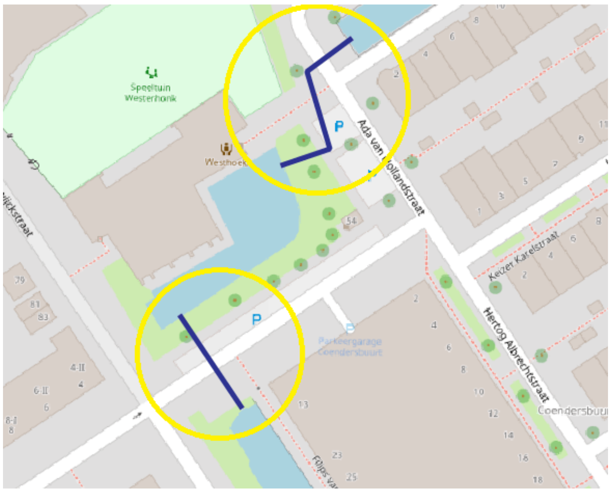


Figure 2.4: Horeca information

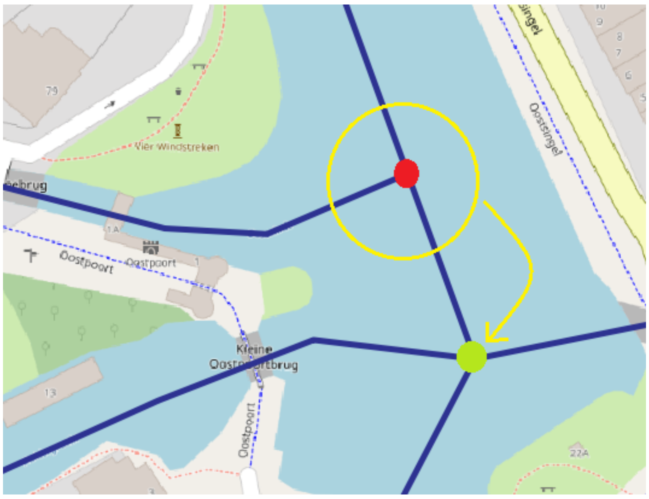
Throughout the development of the model for plastic hotspot detection, it became increasingly more clear what was needed from the input data for the automatic analysis to work. In the following section, we bring out some of the key requirements identified for the input data. All the datasets presented in Figure 2.2 were tested and analysed during this project. However, only the ones shown in Table 2.1 were used in our final approach. The following section gives insights of how the choice was made and which data inconsistencies were encountered.

2.1.1 Requirements for data

As mentioned above, the basis in our proposed method is the data of the water network. This dataset connects all the parameters used the model and missing data can have a negative impact on the final result, because our model assumes that the water network is the same as in reality. Some examples of the inconsistencies that were apparent in our initial datasets are shown in Figures 2.5a, 2.5b and 2.6.



(a) Wrongly representation of water surfaces



(b) Redundant information

Figure 2.6: Topological inconsistencies

In the first case, two connections of water surfaces (through water line segments) emerge, which should not exist. The two water surfaces are likely to be connected underground, but for our approaches these connections create problems, as the branches (Figure 2.5a blue lines) cannot be traversed by plastics. On the contrary, the information that is needed is missing. This concerns the water body itself (i.e. should be presented with a line segment in the middle of the surface), as it is likely for the plastics to be accumulated on the surface of the water .

In the second example (Figure 2.5b), the yellow arrow illustrates the location of the node (red) and where the node should be (green node). The green node represents the converge point where all the water line segments should end up. That undesirable detail creates redundancies, which are not beneficial for our approach. The potential plastic hotspots that we are looking for are based on the location of the vertices related to the water line segments. Having redundant information can lead then, to wrongly identifications. Moreover, in the same figure, it can be noticed the existence of a more complex representation of the water line segments (sharp angles), could be avoided (simplify line segments, reducing the wrongly provided information).

Finally, in the last case (Figure 2.6), a topological inconsistency can be observed. Although the water surface ends at a specific place (wall), the line representing it continues until it reaches a corner on the opposite side. It is noted that there may be underground flow, but for the analysis of the network and its further utilization for our work, underground flow is not considered, because underground water features are not reachable for plastics in most cases.

To conclude, it is crucial that the dataset of the water networks is an accurate representation of the reality and is free of redundancies. After trial and error with conducting preliminary analysis on the datasets available for us, for the implementation of our approach, we decided to include only the datasets shown in Table 2.1. It should be noted that data on water network by OSM was used in Phase A, whereas improved data from Delft waterboard was used in Phase B.

Table 2.1: Used data

<table>
<tr><th></th><th></th><th>DATA</th><th>ATTRIBUTES</th></tr>
<tr><td rowspan="7">PHASE B</td><td rowspan="5">PHASE A</td><td>Ground truth</td><td>The layer includes information about the (checked) plastic hotspots detected in the study area (city of Delft). This information is accompanied with details concerning the wider population in the area found, the wind direction the day we detected them and their distance from the nearest plastic sources (next layer).</td></tr>
<tr><td>HORECA</td><td>The layer contains information about markets, bars, cafés, restaurants, bakeries and other relevant sources that produce plastics (apart from hotels).</td></tr>
<tr><td>Population</td><td>The layer includes the population density (per m2) in the study area for the year 2020.</td></tr>
<tr><td>Water surfaces (OSM)</td><td>These surfaces are polygons that represent the available water information for the study area (South Holland). However, it is not a complete dataset, while insufficient.</td></tr>
<tr><td>Waterways (OSM)</td><td>This layer represents the main water network of the study area, but also in this case, there was missing, misleading or wrongly depicted information.</td></tr>
<tr><td rowspan="2"></td><td>Water surfaces (Waterboard)</td><td>More complete dataset, providing information about the water surfaces in the form of polygons.</td></tr>
<tr><td>Waterways (Waterboard)</td><td>This is also a more detailed vector layer that provides information about the water network existing in the study area.</td></tr>
</table>

2.1.2 Data preparation

In order to achieve more accurate results, further modifications were applied on the datasets used. These modifications concern mainly spatial operations that helped us to fix the topology and/or geometry of the datasets. However, in the next step and given the available time, in Phase B 2.4 we used the new datasets supplied by the Waterboard, as they are more complete. A different processing procedure was followed, based on the network analysis approach. Specifically, we enforced some modifications to the waterways and water surfaces layers, in order to *clean* them from useless and/or redundant information.

The main objective of the data cleaning for the network approach was to extract the corresponding water-line segments from the water-representing polygons. Although the datasets from the waterboard were more complete, there were cases in which the provided information was unnecessary and/or irrelevant. Our goal was to maintain those line segments that are traversable by plastic (i.e. disregard line segments containing information about steel grids). To do so, we implemented a series of spatial operations during which several assumptions had to be made, in order to deal with inconsistencies of the data. In more details, some examples of the main problems that were faced and needed to be solved were:

[1]: Illustration of pipes-dangling branches: In the waterways dataset, apart from the depiction of the water element, there were several cases where pipes were presented that connect the different river or canal branches with each other (Figure 2.7). Given that for network analysis we wanted to maintain only the information about the line segments derived from the intersection of water surfaces and waterways, we had to extract that pipes, as well the remaining branches (if any) that had been artificially added to the line feature dataset for connecting the pipe with the centreline of the water. The inclusion of pipes in the dataset would induce two issues. First, the model could predict a plastic hotspot inside a pipe, where the cleanup would not be possible. Second, in most cases the plastic is not able to flow through the pipes as the pipe openings are covered with grid stopping larger objects entering the pipe.

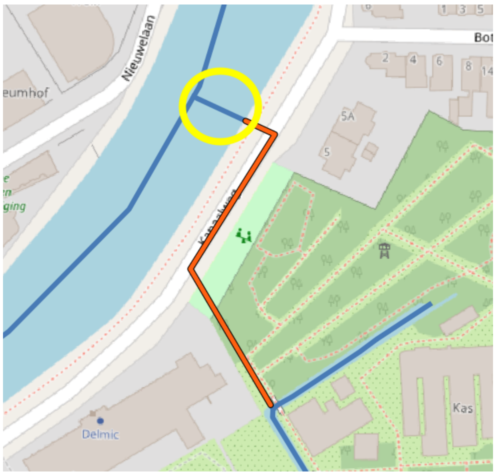


Figure 2.7: Example of pipe connecting two waterways. In orange is presented the pipe line, while in the yellow circle is shown the remaining branch

[2]: Gaps: In the water surfaces dataset, there were some gaps between continuous waterbodies, which lead to incorrect information about the waterway continuity (Figure 2.8). This discontinuity hampered the data correction process and spatial operations (i.e. snapping) required.

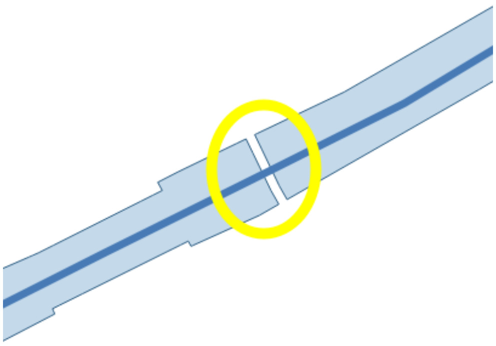
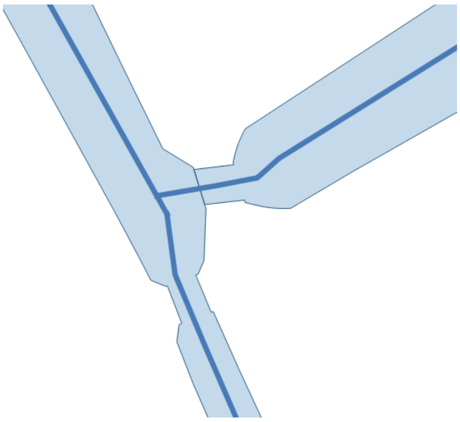
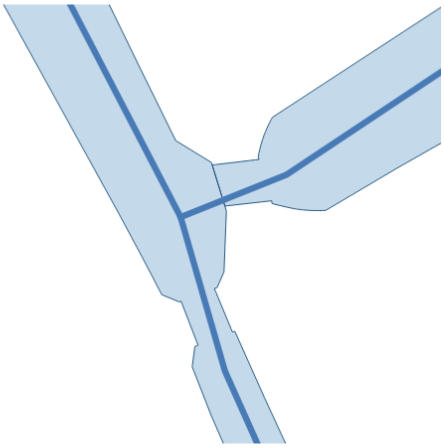


Figure 2.8: Example of disconnected polygons

[3]: Simplification: The above mentioned procedures were followed by simplification techniques in order to improve/fix the topology of the dataset and to avoid redundant information (Figure 2.9a and 2.9b). After trial and error in the simplification methods, we achieved to reduce the number of the waterline vertices from 236348 to 83601, preserving the topology of the waterline network. In the end, the attribute table of the final dataset, consists only the relevant information of the waterlines (i.e. length, depth, width).



(a) Example of non-simplified waterlines



(b) Example of simplified waterlines

It should be mentioned that during the procedure we tried to clean and fix the topology of the provided data as much as possible. However, not all inconsistencies were repaired since there were special cases, where a general approach cannot cover. Moreover, given that we worked on a common database, during cleaning procedure, spatial indexes were used, in order to reduce queries' execution time and make the procedure more efficient.

2.1.3 Softwares and tools

In our analysis we used two open softwares, QGIS and Visual Studio Code as well as the free open source relational Database Managent System PostgreSQL/PostGIS. Specifically, for the first one, in order to both process the data as well as to implement our algorithms, the use of QGIS plugins was necessary. As plugin is defined an extension to QGIS which provides additional functionality that is not included in the core package, it can run within the QGIS environment and allows the interaction with QGIS interface. In our case, from the plugins directory we used Database Manager to integrate and manage our spatial database (PostgreSQL/PostGIS). The advantage of using database connection within QGIS is that we can store all layers with their own style and share the final project online without sending files (shapefiles) for each new layer created during the process. We also used a core plugin for spatial data processing framework (Processing). From Processing Toolbox were selected the appropriate algorithms (*Providers*) fitted in our needs such as the vector analysis, vector general, vector overlay and vector selection. Then, from the above mentioned algorithms specific functions were selected (i.e intersection, buffer) and executed within QGIS. Finally, the graphical modeler was especially helpful as it allowed us to create complex models using a simple interface.

The programming was done in Visual Studio Code using Python programming language. The agent-based network analysis was done by using NetworkX - a free, open source Phyton library for network science, distributed under Modified BSD Licence.

Other Phython libraries used include math, random, matplotlib and numpy.

2.2 Synthetic Procedures - Methodology

Due to lack of existing similar research on detecting plastic hotspots in fluvial environment in semiautomatic way, our first approach was to look separately at several possible factors that may predict where plastic in Dutch rivers and canals accumulates. These factors were based on the field experience of Noria as well as analysis of our ground truth data points from Delft. The first assumptions were the following:

- Plastic accumulates in the dead ends of canals where there is no water outflow (Tasseron et al. 2020);
- Plastic carried in a water stream gets stuck in vegetation in water, therefore accumulates in vegetated banks;
- Plastic is pushed by wind, therefore moves downwind and accumulates in bends and closed areas;
- Plastic is moved by water flow;
- More plastic gets into the water around restaurants, markets and densely populated areas (Best 2019; Gasperi et al. 2014).

Initially, each of the above cases were looked at separately in order to detect automatically places where conditions are met for plastic accumulation. We call this phase in our project Phase A where datasets were analysed separately. The outcomes of this were used in Phase B when more comprehensive network analysis was developed.

2.2.1 Algorithm assumptions

Starting with the more detailed analysis of the problem and its subdivision into smaller tasks, we faced several obstacles, which directly concern the behavior of plastics in the water, as well as the factors lead them to into the water bodies. In particular, these barriers concern:

1. **Plastic sources**: About plastic sources, we took into account possible sources producing plastics, focused mainly on restaurants, cafés, bars and markets. We focused on that sources considering that they are the most effective plastic producers. Here, it should be mentioned that for all sources we assumed an one-to-one approach, which means that each source produce exactly one plastic. Moreover, in our sources are absent other possible individual sources like humans. A human being could also be considered a source of plastic, although not a direct producer, it is a means of transporting plastics. More specifically, a person can hold a plastic and deposit it away from a source that we have already took into account (i.e.

restaurant, cafeteria). This movement of deposition, can be considered as a habit mainly in cases of specific routes that are regularly followed by people (paths, frequent routes to and from work, pedestrian traffic). Additionally, waste receptacle (bins) were not included in our approach, neither as a source of plastic nor as a means of its transportation. Although we looked for (possible) available datasets, this information where also absent.

2. **Natural phenomenon**: Regarding the main physical parameters that affect plastic behavior, we had to converge at a point, for the variables concerning the natural flow (if any) of the canals, as well as the direction of the wind that may lead them to a specific location. Given the complexity of natural phenomenon, it is obvious that it is difficult to take into account and model all the different parameters that determine both themselves as well as the influence they exert on plastics and their transport. Specifically, regarding the direction of the wind, it is noted that we considered it constant, with direction South-West (SW), which is the predominant direction of the wind in the study area ((Windfinder n.d.)). We also considered that the wind is evenly distributed in the space and is not affected by obstacles such as buildings, the height difference of the wall (surrounding the channels) and the water level. On the other hand, regarding flow direction, it is noted that it is considered constant in one direction, without taking into account tides, possible turbulent flow, underground pumping stations and other physical (or not) factors that may affect the natural flow of the canal.

3. **Plastic hotspots**: As far as the potential places where plastics tend to accumulate are concerned, it is mentioned that we assumed them to be the dead ends of the canals. In particular, we took under consideration both man-made ends (i.e. brick walls Figure 2.10a) as well as natural ends (corners with vegetation Figure 2.10b).

It is mentioned that after our field work, we observed that couple of our hypotheses were negated. For example, while in the first in-situ research our assumptions seemed to be verified (algorithms' outcome), in a subsequent check the points that initially turned out to be hotspots, no longer corresponded to our allegations.These assumptions affect directly the outcomes of our methods and algorithms (see Chapter 3 and 4).

(a) Example of man-made canal dead end

(b) Example of natural canal dead end

2.3 Phase A – Layer-based approach

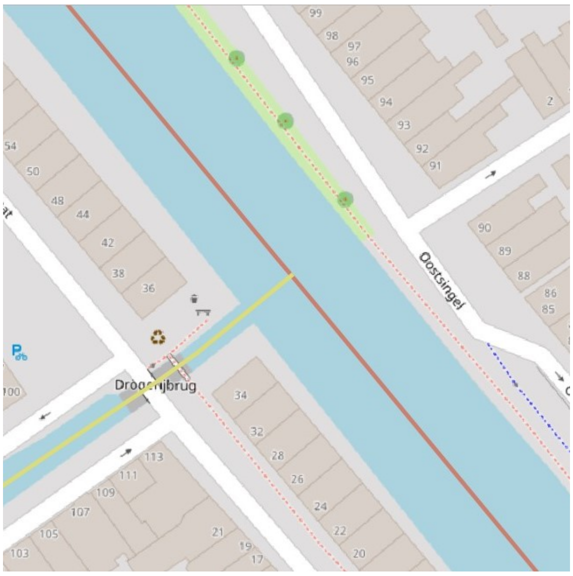
While there is research about the accumulation and detection of plastic hotspots in the oceans and marine environment, examples of models for plastic hotspots in complex urban canal network as is in the Netherlands, are rare. Due to the original and complex nature of the study, the development plan was divided into two phases. In the first part of the development work, each of the datasets were looked at separately to see their potential for semi-automatic plastic hotspot detection. In this report, we call this the Phase A where the key objectives were to understand the nature of plastic accumulation, test the quality of the datasets and assess the assumptions made about the factors that cause plastics to accumulate. Once a method exists for each of the parameters to be analysed (semi-) automatically, the project moved on to Phase B where the selected parameters were added into one comprehensive model.

2.3.1 Dead ends and junctions detection

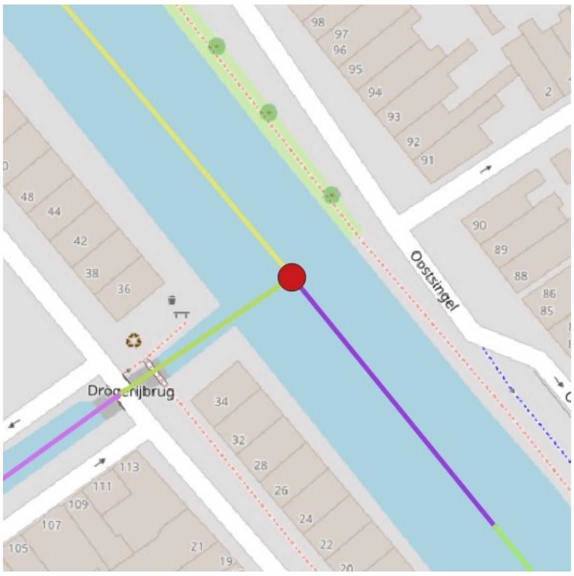
We consider waterway junctions as an irregularity in the water flow that might cause plastic accumulation along the junction's riverbank. On the other hand, we take into account dead ends as places with a high probability to be plastic hotspots due to their closing geometry which obstructs the water flow (Tasseron et al. 2020). Our basic assumption was that plastic gets constantly pushed to the very end of dead ends where it eventually accumulates. Although plastic hotspots exist at the very end of dead ends, we observed that dead ends are rather stagnant than moving waters. Therefore, we had to regard another driving force for the plastic movement, that reasons the existence or non-existence of hotspots in dead ends. For example, a favourable wind direction that pushes the water as well as the plastics on the water surface into the dead end. That procedure of pushing and pulling wind forces is explained in more detail in the subchapter of dominant wind direction (see Chapter 2.3.3). For the purpose of incorporating all relevant waterway obstacles in Phase B, the objective of this algorithm was the detection and classifications of all dead ends and junctions in the waterway network.

The detection of dead ends is solely based on the geometry of the waterways. Although the quality of the algorithm's output strongly depends on the data input's quality, the workflow of detecting dead ends in the given network does not need to be modified for different datasets, which makes it adaptable for other waterways with different attributes. The implementation roughly divided to two main steps and is applicable on single-line strings as well as on multi-line strings, where each river/canal segment is represented by its own (multi-) linestring. In the first step, the start and end nodes of each (multi-) linestring are extracted from the waterline network. Then, only those nodes are identified as dead ends that do not share a position with a start or end node of another (multi-) linestring. By doing this the algorithm makes sure that the selected start and end nodes of (multi-) lines are not connected to further line segments and are truly dead ends.

In order to detect junctions, the same approach as for dead ends is carried out. First, start and end nodes are detected. Then, the junctions are detected by the number of (multi-) linestrings intersecting with their start and end nodes at the same position. In this implementation, junctions are classified as nodes in which start or ends nodes of at least three or more line segments overlap. This constraint makes the algorithm only working for strongly segmented waterlines where three line segments cross although they represent only two waterways, as shown in Figure 2.11a. This is the case for our data, due to the cleaning process. For a bigger junction where at least three different waterlines cross, a junction is detected with both datasets, OSM and our cleaned data (Figure 2.12a & 2.12b).

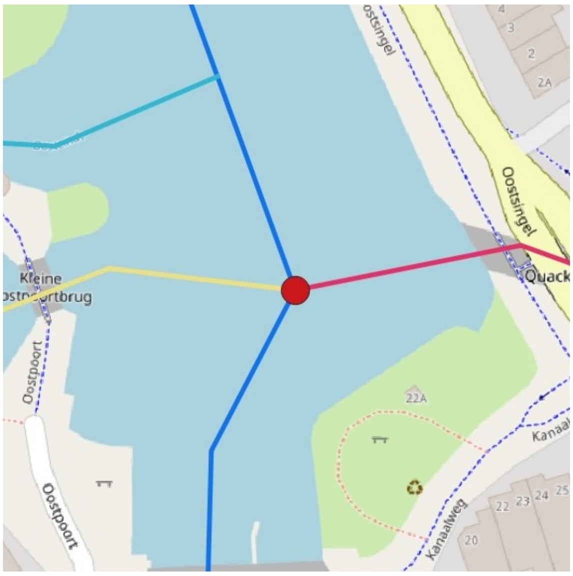


(a) No junction with two OSM waterline segments

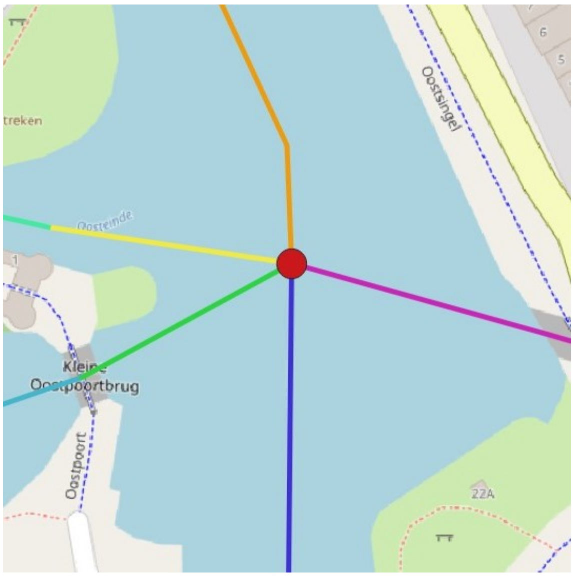


(b) Junction in red based on three waterline segments

Figure 2.11: Algorithm not applicable for junctions with less than three crossing line segments



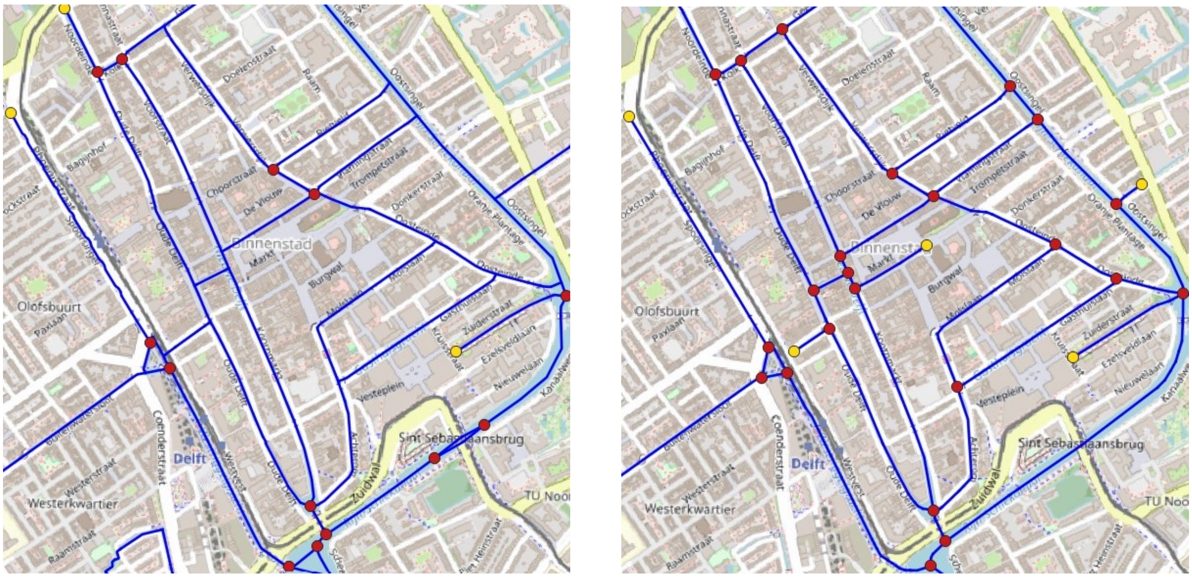
(a) Big OSM junction in red



(b) Big junction in red

Figure 2.12: Algorithm is applicable for junction with three or more crossing line segments

The final output of dead ends and junction in the city centre of Delft is shown in the following Figure 2.13b in comparison with the output based on OSM data in Figure 2.13a.



(a) Output based on OSM data

(b) Output based on cleaned data

Figure 2.13: Comparison of dead ends (yellow) & junctions (red) based on different datasets

2.3.2 Canal geometry

By looking for further irregularities in waterlines the focus can also be set on the bending of canals. This is done by considering the intermediate nodes of a waterline between start and end node. While dead ends and junctions detection works with (multi-) linestrings, this approach only works with multilinestrings, since it calculates the relative angle of a node based on its own coordinates as well as with the previous and following nodes' coordinates. Therefore, at least two line segments in one multilinestring are needed to calculate a relative angle. To extract only those angles that actually represent a bending, only nodes with an angle smaller than 165° or greater than 195° are extracted from the waterline (Figure 2.14). These steps were executed by a python script that was integrated in the QGIS Graphical Modeler. Further steps are needed to provide a more comprehensive output of waterway angles. For example, the entirety of a curve in the multilinestring could be detected by analysing the number of nodes on the string with the corresponding angles. Furthermore, sharp angles representing turnings in the waterways can be detected by only considering smaller and greater angles ($< 90°$, $> 270°$). Nevertheless, for including the entirety of curve in Phase B, it would be rather beneficial to know the orientation of the bending. Thus, a differentiation between cut bank and point bar could be made, which is crucial for simulating the impact of water flow and wind direction on natural river curves. Since this differentiation and the curve's orientation could not be incorporated in the algorithm due to lack of existing studies on accumulation phenomena in canal's curves and due to the time scope of this project, the output of this algorithm is not passed to Phase B.

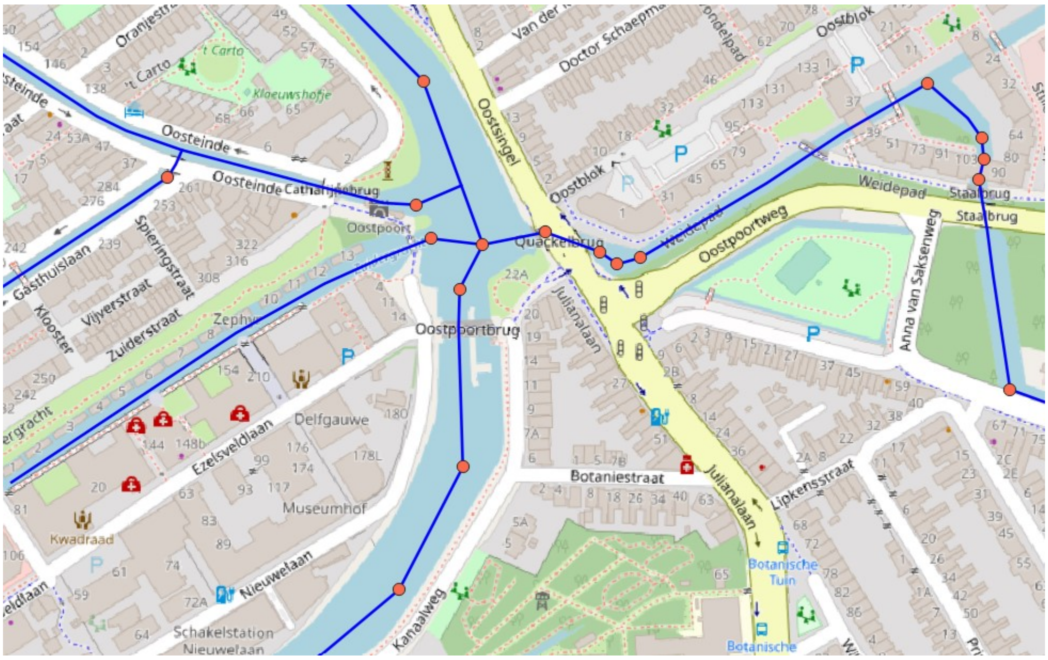


Figure 2.14: OSM waterlines; Nodes (in orange) with angles < 165° or > 195°

2.3.3 Dominant wind direction

Together with the flow of the water, wind plays a key role in the transport of plastic in water. The movement of plastics in water, as well as on land, is affected by both the speed and the direction of the wind. Regarding the study area and the Netherlands more broadly, it is noted that the combination of strong winds and the predominantly flat landscape allows to assume a strong influence of the wind parameter on the transport of plastic.

Taking into account the influence of wind to the plastics' behavior, we aimed to develop an algorithm that would detect the potential location where plastics tend to accumulate (hotspots), based on the wind agent. Specifically, it is mentioned that we considered that the wind direction dominates in the study area in order to reduce the complexity of the specific factor and the corresponding affects. In more detail, after research (yearly weather data measurements: (Services n.d.)) as well as feedback from our clients, we considered that the wind direction in our case study is stable and its direction is from South-West (SW).

For the wind model we used multiline datasets acquired from Open Street Map and Delfland Waterboard GIS department. The datasets, while similar in features and geometry, were of different quality resulting in the need of cleaning operations. Additionally, differences in the attributes of the two datasets required different implementation of the model that needed to be changed manually. Such differences can be located in the attribute table of the layers concerning mainly the name and nature of ID. In OSM there was a different definition of unique objects (waterlines) than the Waterboards data. While we can assume that waterboards' datasets can be considered as the most accurate one, it was found out that the model worked better with the OSM dataset. This remark can be accounted to the fact that in the model's case, high level of detail causes conflicts with the notion of reality that it tries to represent. An example of this is going to be thoroughly explained after the model's description (Figure 2.18).

In a few words, the physical phenomenon that the model tries to represent is the pushing of floating objects (plastics on water) due to the wind's influence. The most simple observed case is when the wind pushes the object towards its direction. However, this simple observation starts to become more complicated considering that in most cases, the wind direction intersects with canals' geometry. The result of this intersection is that the plastics are being pushed to the banks or sides of the water bodies based on the relation between the wind direction and the waterbody's geometry. So, the current wind model tries to represent this behaviour using the following operations. First of all we are interested in waterlines that have specific boundaries starting from a node and ending to another regardless of any other intermediate nodes. Next, we consider the wind direction to be 45 degrees (SW direction) based on the dominant wind direction in the Netherlands. Moreover, we took an assumption, that any object that is being pushed against the sides of a water body can slide forward or backwards depending on a theoretical inclination that the wind forms with the edge of that side. The approach checks consecutive water line segments to determine if plastics are being pulled towards the starting node of the edge or pushed towards the finishing node of it.

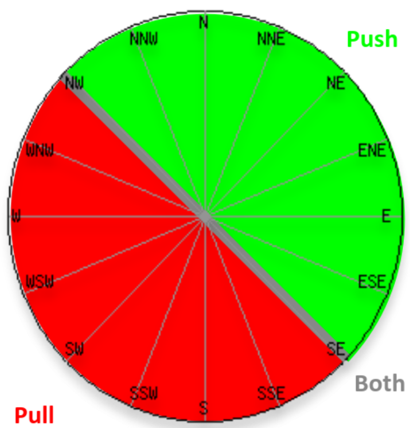


Figure 2.15: Plastic behavior based on relative angle of edge and wind direction.

In the above figure the waterline edge is placed on the gray lines with the starting node placed in the middle and ending node placed on the circle to represent the direction. Additionally, based on the SW wind direction, all edges that lie in the green semi-circle push the plastics that they contain to their finishing node. On the other hand, edges that lie in the red semi-circle, pull all of their plastic to the edge's starting node. Having this in mind, we can determine where all plastics can accumulate in a unique water line system. Note, however, that there is also a special case of a perpendicular intersection of the wind and the waterline edge. In reality, this situation can have different outcomes that all depend on the various factor like the wind velocity, the surface material, plastic's shape and size, surface off-water geometry and others. The different outcomes could be that plastic are pushed exactly against the wall so they either stay put or some move towards a specific direction and others to the opposite direction. Due to the randomness and the rareness of this phenomenon we chose to assume that in these situations the plastics move forward.

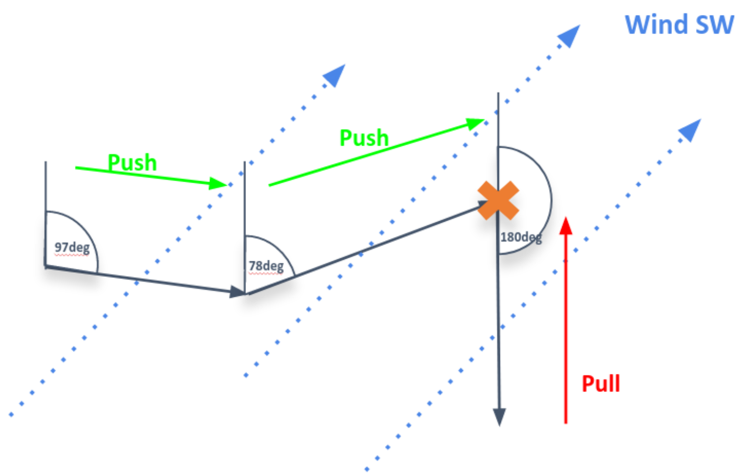


Figure 2.16: Practical example of the wind model functionality.

The above example illustrate the way that the wind model works in practice. Starting from the first waterline segment of a unique system, the model calculates the clock-wise angle from the North, its starting node and its finishing node. This angle is being compared against the notion derived from Figure 2.15 in order to determine the direction of the moving plastics. On the first segment, the angle of 97 degrees falls in the green semi-circle so the plastics move to the head of the arrow. In the next edge, the angle is 78 degrees falling again to the green semi-circle which means that all plastics move again to the head of the arrow (this includes plastics from the previous segment). Lastly, the angle of the last edge is 180 degrees falling in the red semi-circle which means that all plastics move to the tail of the arrow. So in the end all plastics accumulate the point marked with the orange 'X'.

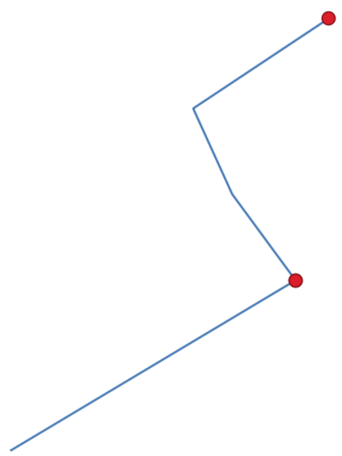


Figure 2.17: Example from dataset

The above figure illustrates the result of the wind model for a waterline segment. The red points are identified as positions where plastic could accumulate (assuming that the wind is the only factor affecting them).

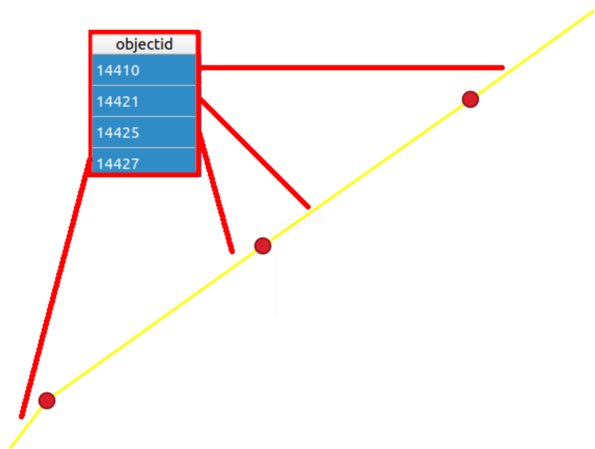


Figure 2.18: Example from dataset were the model fails.

The above figure illustrates many cases were the wind model fails due to data configuration. As already mentioned above, the water bodies dataset provided by the Delfland Waterboard a high level of detail that posed a hindrance for our approach. To elaborate, a unique water line system could contain many unique identifications for different segments of it. The issue arose due to the fact that the model recognizes unique systems based on their unique IDs. As a result, a unique system with many different IDs is being recognized as n number of different unique systems where n is the number of unique ID inside the system. In the figure, a continues line is being presented, however each line segment has its own unique ID which makes the model to consider them as their own different systems.

2.3.4 Water flow direction

Water in rivers, streams and canals can have flow, which can for example be caused by difference in elevation or influences of pumping stations. The direction of the water flow can greatly influence the direction in which floating and submerged pieces of plastic go. For this reason it is important to incorporate the flow direction parameter into the model. The first information that was gathered about the flow direction of inland waters in Delft originated from Rijkswaterstaat. They stated that only the big rivers have a clear flow direction, which is also influenced by the tides near the sea. Because Rijkswaterstaat did not have a dataset available that contained the flow direction, the Open Street Map Waterways dataset was used. In order to avoid doing too much manual work, two big generalisations were made: there is only flow in features classified as "river" and all rivers float from East to West. The line features classified as river in the OSM dataset can be seen in 2.19 in blue for the whole of South-Holland.

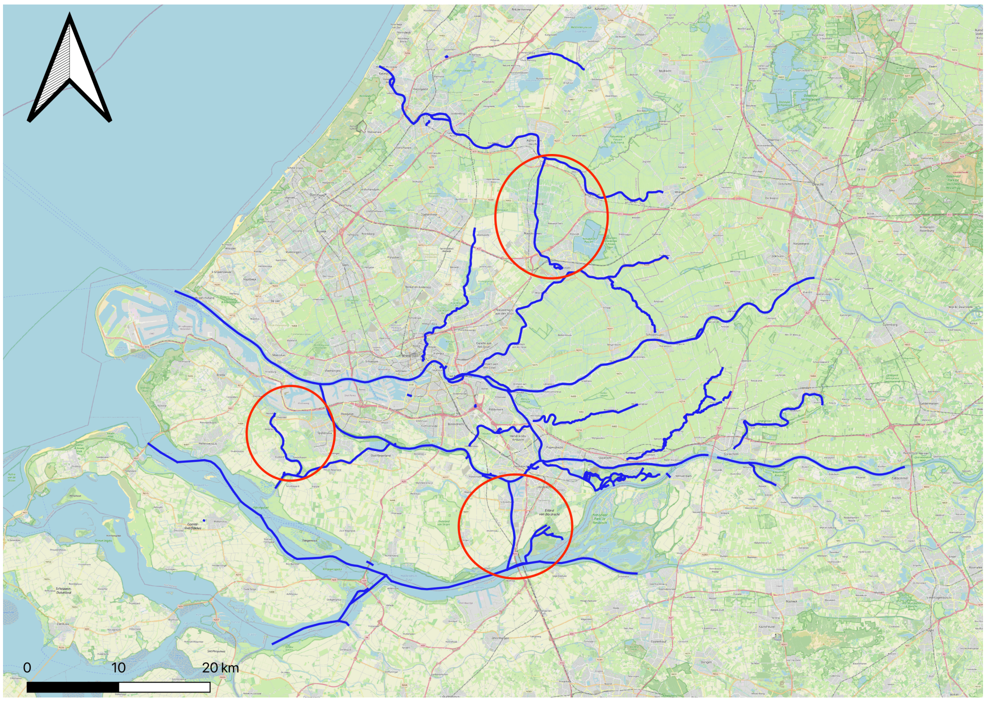


Figure 2.19: Lines classified as river in the OSM waterways layer (in blue). Examples of areas where the assumption that water flows westwards does not hold true are circled in red.

Using the canal geometry, the West side of each line feature in a river could be identified. This was done using a QGIS model, which could also be exported as Python-file. The goal was to combine this model with the other hot-spot identification models to identify which hot-spots were also in line with the flow direction. However, as could be seen in figure 2.19, the assumption that all rivers flow westwards does not hold true in all cases. Because of the great influence of flow direction on the movement of plastics, it was decided that more accurate assumptions about flow direction were needed. Firstly, the line geometry of the OSM Waterways dataset was investigated. The direction of the lines are displayed in 2.20. However, the direction of the lines were not found to be corresponding with reality. Moreover, as can also be seen in 2.20 not all channels are included in the dataset and a lot of them are disconnected. Because the quality of the data can greatly influence the results of all models, we approached the GIS-department of the Waterboard Delfland to ask if they could provide more accurate data.

Figure 2.20: The Open Street Map waterways features displayed in blue. The direction of the geometry is displayed in black.

The waterboard provided a very detailed line dataset covering all the water in their jurisdiction. It can be seen in 2.21. It is clear that this dataset is more accurate than the OSM dataset the was previously used, more features are included and these datasets are all connected.

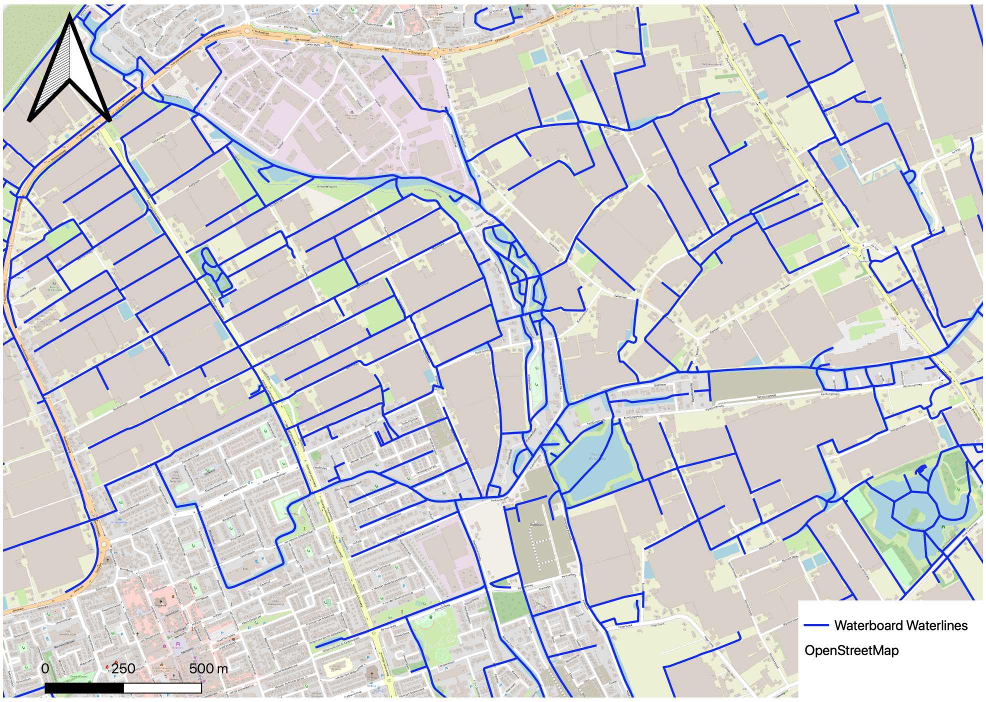


Figure 2.21: Waterline dataset provided by Waterboard Delfland displayed in blue.

The waterboard also provided us with a line dataset where the direction of flow is stored in the geometry of the line (water flows from the first to the last node). It can be seen in 2.22. Because the waterline and flow direction datasets were provided while the project was in Phase B, they were mostly used as input for the simulation and network analysis.

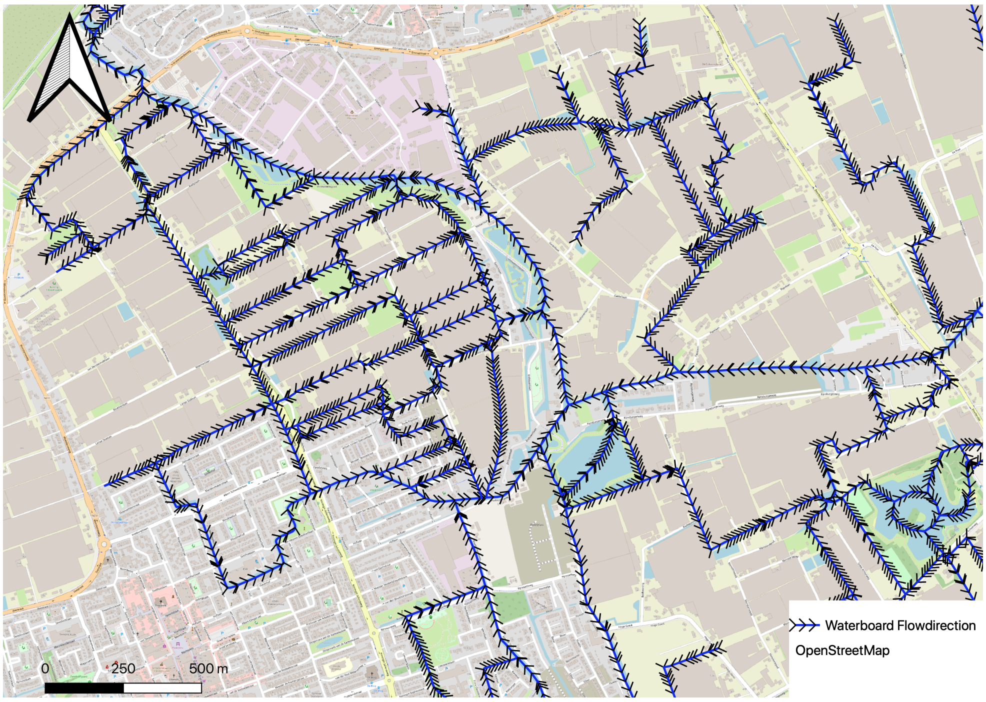
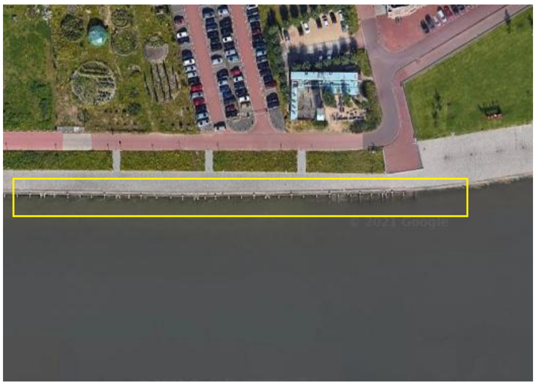


Figure 2.22: Flow direction dataset provided by the Waterboard Delfland. Lines are displayed in blue, flow direction is displayed in black.

2.3.5 Land use

As stated above, one of the assumptions of our methodology which was taken from field experience by our client Noria was that plastic tends to gets stuck in places where there is vegetation in water. If the canal has hard shores of concrete wall then the plastics do not easily get stuck and instead follow the water flow or are carried on by the wind (Figure 2.23a). On the other hand, if vegetation like bushes or reeds is present in the canal/river and the shore is not man-made, there is higher probability for plastics to accumulate (Figure 2.23b).

Provided by the client, we had polygon feature dataset on the land use, classifying the land into 13 categories, including built (bebouwd), wet natural terrain (nat natuurlijk terrein), recreation (recreatie), semi-built (semi bebouwd) and water. This layer was used in order to find the vegetated areas in the cities which are surrounded by the built environment, marking the place where man-made channel shore turns into natural bank. A python script was made for faster processing and testing the method in QGIS environment using the Python console. The script lets the user to select which categories from the land use polygon layer would be considered as vegetated areas, as water and as built environment. The script would first do a validation test for the polygon features, which proved to be necessary for the land use dataset. It would then select the vegetated areas which are by the water and also in a built environment through series of intersection test. The model was tested, first only selecting the wet natural terrain, built and water categories for the model. Thereafter adding semi-built land use category for the built environment and recreational land use category for the vegetated areas to compare the outcome.



(a) Example of hard shore (in yellow) is highlighted the exampled area-Delfshaven

(b) Example of vegetation in water (in yellow-Lageveld) is highlighted the exampled area

The outcome of the above mentioned procedure did not meet our expectations. The vegetated areas detected were mainly only outside the urban areas (Figure 2.24 and 2.25) and were intersecting with the built environment feature mainly due to the large size of the polygons. Furthermore, it was noted, that the land use dataset does not show small patches of canal sides that are vegetated which were identified during field work and using Google Street View. Additionally, much of the green areas in the cities are classified as recreational rather than wet natural terrain. However, under the recreational land use are channel shores which are both natural or man-made, causing a lot of noise and false identifications in the model. It was therefore concluded that the land use dataset is not suitable for finding vegetated shores in urban environments.

Figure 2.24: Examined area (presented with the red solid line)- study area (presented with the red dashed line). In purple are presented the vegetation polygons

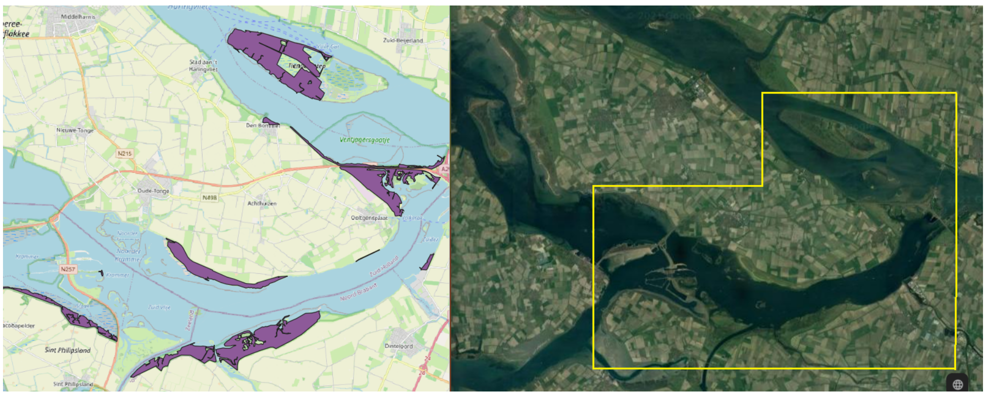


Figure 2.25: Example of shoreline plants located away from the urban fabric (in yellow) is highlighted the exampled area)

2.3.6 Plastic sources

For understanding how plastic moves and accumulates in water network, it is important to consider where the plastic gets into the water. Study by Tasseron et al. (2020) showed that more plastic hotspots were found in Dutch canals near markets, shops and parking lots, with the majority of plastic deriving from food wrappings and plastic bags. Data of such places were retrieved from Google Maps, where the location of restaurants, markets and bars in the city of Delft were extracted as point layer using Make Map function. As the layer was imported to QGIS, the aim was to select the points that are close enough to water for the plastic to get into the channels. As plastic on land travels mostly due to wind, the aim was for the model to consider that plastic can travel further down wind.

The model was built in QGIS graphical modeller, where two buffers are created around each point of plastic source. The first buffer is surrounding the point 360 degrees, the second buffer is only covering a 90 degree angle towards azimuth direction set by the user. The aim is to give the user an option to set the buffer further down wind. The tool then checks which buffers intersect with water layer and adds a point to the intersection between the polygon buffer and the water line features. By this method the assumption that more plastics will get into the water near restaurants and market places is entered into the model.

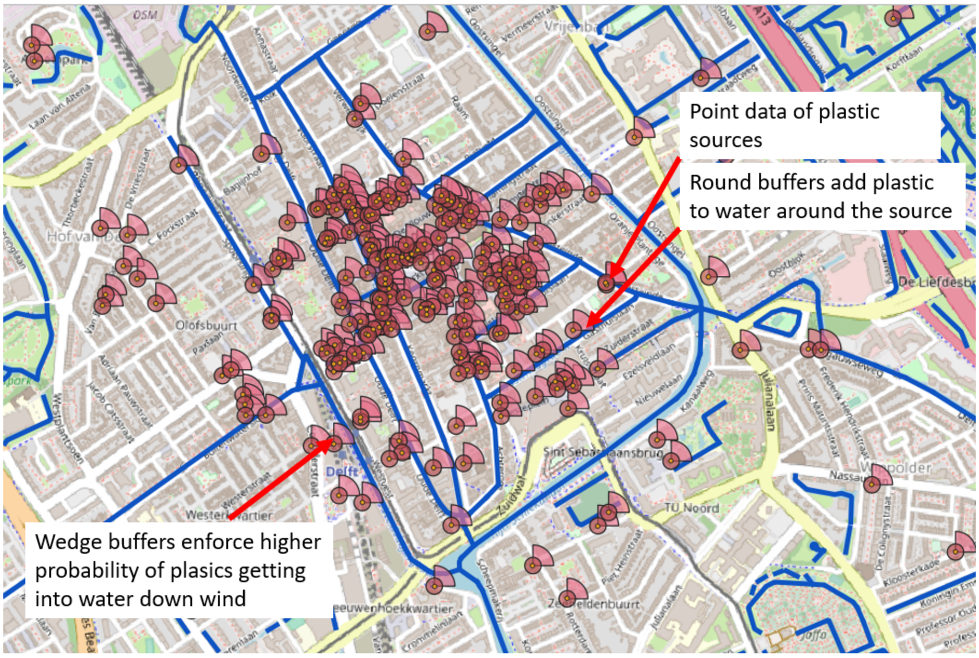


Figure 2.26: Example illustrating the double buffer approach, giving the user an option to enforce plastic transport from plastic sources to the water network down the wind whereas plastic will travel less in other directions. In this example, the wind is set to northeast.

In addition to these plastic producers, we also took into consideration the population density of the area. To implement this we used the dataset of $500m^2$ population density of 2019 derived from CBS (Figure 2.27) and an assumption that 1 plastic is being randomly littered for every 2000 people. In the end, the plastics produced from the abovementioned locations were enhanced by adding the plastic produced from the population density.

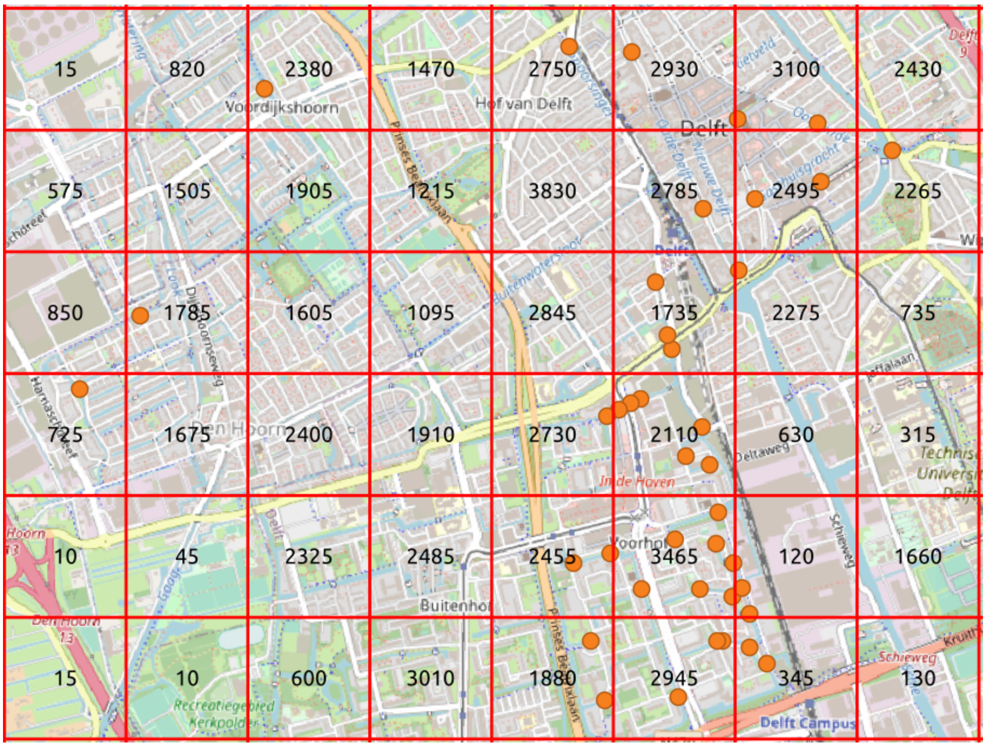


Figure 2.27: Ground truth overlayed by $500m^2$ population density

2.3.7 Conclusions for layer-based approach

The layer-based approach allowed us to explore the parameters which help to detect potential plastic hotspots with GIS analysis as well as to understand the requirements for the input data. The considerable effect of the quality and completeness of input data to the results was evident in all the parameters investigated. Methods were developed to analyse the canal geometry and detect in a semi-automatic way the dead ends of water networks and places where plastic can accumulate due to wind given that the input water line features have consistent feature IDs. It was concluded that the land use dataset is not suitable for determining where plastics may accumulate in vegetation due to the simplification of the land use categories and low spatial resolution of the classified areas. Lastly, a model was developed to determine where on the input water line features plastic may get into the water, considering the surrounding population, restaurants and markets.

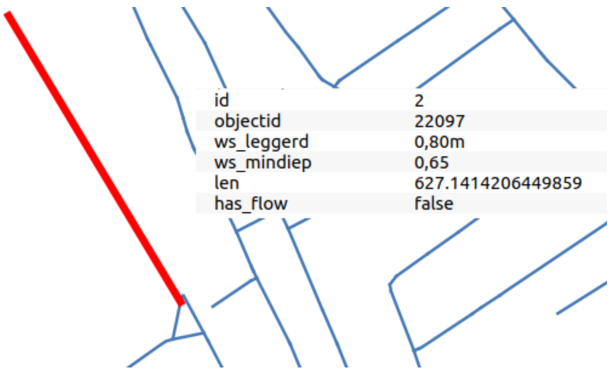
Each of the separate models outputs a number of potential hotspots - places where plastic may accumulate based solely on the single phenomena analysed. However, a more sophisticated approach is needed to determine which of all those potential hotspots are real. This is addressed in more comprehensive model in Phase B, where the dead end detection, canal geometry analysis, flow direction data, location of plastic sources and the assumption of wind direction is used in network analysis. However, no working model for vegetation detection of satisfactory outcome was found and excluding vegetation is a limitation for our approach. It is important to highlight that the model also is oversimplifying the wind direction parameter. Specifically, we assume that the wind has a constant direction (dominant wind direction) from South-West (SW), for the whole study area. However, in dense urban areas, the complexity of the built environment creates complex wind directories which could be recovered in 3D model analysis, which is however out of scope for this project.

2.4 Phase B – Network analysis and simulation

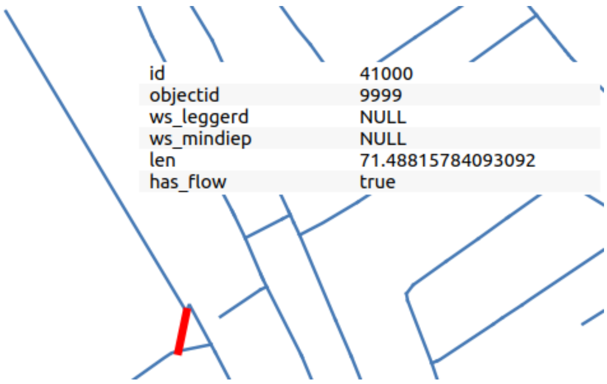
Network science dates back to the graph theory originated from 18th century Swiss mathematician Leonhard Euler, who solved the Königsber bridge problem. Nowadays, network analysis is used in different fields from social research to engineering to analyse connections between different phenomena. As an effective tool for analysing spatial connections, it was chosen as the approach for modelling the plastic movement in urban environment from the sources to the accumulation zones in water. Simply put, in our network analysis the network is the water network in which plastic objects move as agents under the influence of external parameters.

2.4.1 Data Input

The data used in the simulation can be considered as the inputs and separated into two distinct categories. The first category is the edges of the network and the second is the nodes. These are the most important concepts that we are considering for the network analysis operation and the simulation. For the edge dataset, we use the waterlines dataset derived from Delfland Waterboard (GIS department) after the cleaning process (Chapter 2.1.2). Note that this dataset was extensively modified in order to include the attribute of flow direction. As such, we replaced all intersecting features from the waterline dataset that had the flow direction embedded to their geometry. However, in order to be able to distinguish between geometries that were digitised based on flow direction and those based on the random choice of the GIS technician drawing them, we included a *boolean* field (Figures 2.28a and 2.28b). Additionally, manual editing was conducted in cases of disconnected network (snapping geometries) that the automatic cleaning process introduced or could not capture.



(a) Waterline feature that its geometry does not relate to real water flow



(b) Water feature that its geometry does relate to real water flow

It is important to note that while we are only interested in the edges of the waterline dataset there actually exists a node network defined by the boundaries of the lines. These nodes are being used by the simulation but are considered as irrelevant nodes, as they only exist for feature geometry purposes and hold no other meaningful information.

The node dataset that the simulation uses, consists of relevant nodes that we pick from all those irrelevant nodes based on some attributes that they possess. In more detail, we consider as relevant nodes, those vertices that were deemed as dead ends (based on dead-end model's output) and junctions. Dead ends, as mentioned in Chapter 2.3.1, describe vertices connected to only one edge and represent the starting or ending point of a water line. Junctions are vertices that are connected to three or more edges. These nodes are important to the simulation as they signal that plastic could move to different directions based on different attributes happening at the node. Additionally, these relevant nodes pose as gateways where plastic is being inserted to the simulation as explained below in the simulation procedure. This field was based on proximity to water explained at Chapter 2.3.6.

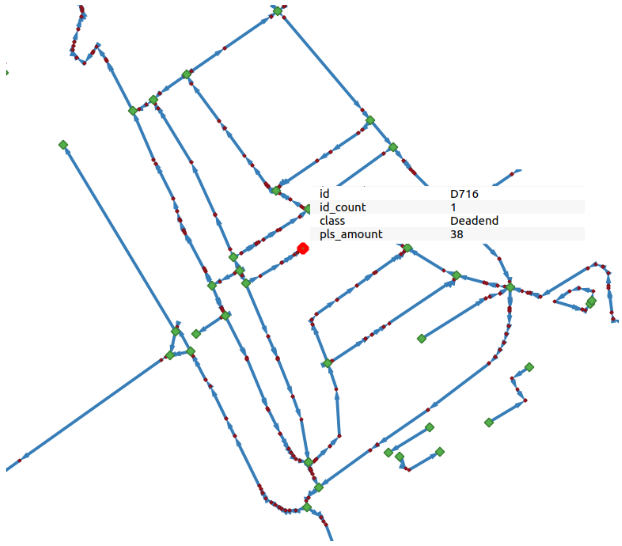


Figure 2.29: Visual representation of the network in the city center of Delft. Relevant nodes, Irrelevant nodes, Edges

Moreover the simulation gives the option for user-defined wind direction input in degrees. Note also, that in the wind direction parameter an additional 15 degrees are added corresponding to the leeway drift that is observed on floating objects (Hackett, Breivik, and Wettre 2006). In the simulation description, when referring to wind direction, the leeway drift is accounted for. We choose to allow only for that parameter as is the most important one, which influences the results of the simulation. Additionally, some other other important technical assumptions made is that the wind velocity is constant at 10.5m/min and the flow speed at 4m/min. The wind velocity was calculated using the basic rule of thumb from (ibid.) that roughly 3% of the wind's speed is translated to movement of small floating objects and that the average annual wind velocity in the Netherlands varies around 21km/h. For the flow speed, the assumption was made based on average measurements of published sensors (available online at: `https://www.rijkswaterstaat.nl/`) scattered throughout the South-Holland province. More information about the results are going to be provided in Chapter 3.

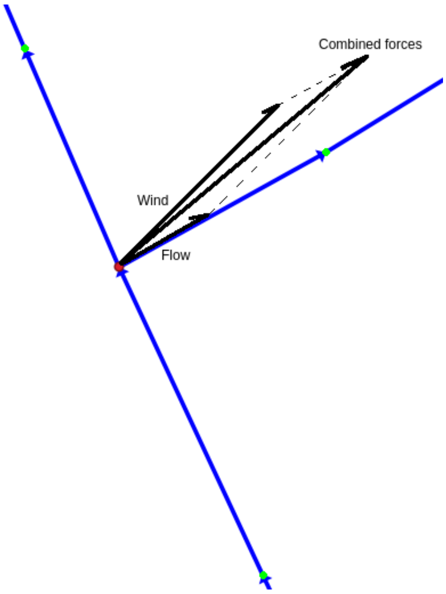
2.4.2 Algorithm procedure and basic steps

The algorithm aims to model the movement of plastic objects using an agent-based approach with Object-Orientated programming in Python. However, due to the fact that we are modelling inanimate objects without clear interactivity between them, we assume that the agents are isolated from each other and are being affected individually only by their environment. To further elaborate, the agents are plastic units with attributes for their unique ID, for flow velocity parameter, wind velocity parameter and the time passed in simulation. The wind and flow velocity attributes relate to the plastics' translated velocity based on those forces. There are additional attributes which are more relevant to the inner computational functions of the simulation rather than to the practical comprehension of the simulation.

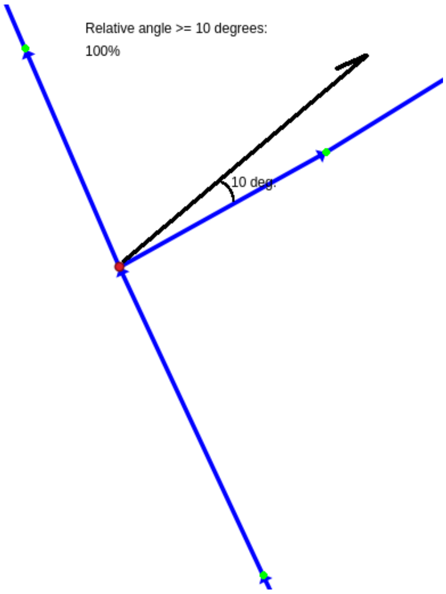
The initialisation of the simulation happens at the stage where all plastic objects are being inserted to the system. It was decided, that plastics are inserted to the relevant nodes, and the amount of objects depends on the corresponding field value derived from the *Plastic sources* model (Chapter 2.3.6).

Figure 2.30: Visual representation of the network's initial state

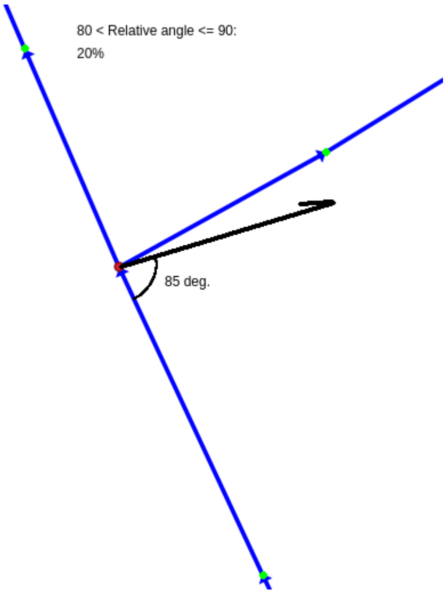
After the initial stage, all plastics that are inside nodes make a movement decision based on the attributes of the node that they lie inside and their neighboring nodes (if any). For every passing time unit of the plastic, neighbor checking happens to assess the probability of the plastic to move towards them. The decision is made based on probabilities that we manually define and are hard-coded into the simulation. The most important attribute is the plastic's directional velocity, involving the flow and wind parameters, against the direction of the water line. Having knowledge about that relative angle between the combined forces and the water line, we can determine the probability of the plastic to move towards the neighboring node. A simple demonstration is that a relative angle of 180 degrees is not going to move the plastic to that neighboring node as its velocity is opposite of that particular direction. However, all of the nodes are going to be checked for possible routes before determining that the plastic stays put. The other attribute that the algorithm considers is the category of the node. If the plastic has reached a *dead-end* then it already reached a terminal point and is removed from the simulation and stored. Note, here that in the case where a dead-end is the plastic's initial node, the simulation forces the plastic to pass through the previous procedure of neighbor checking.



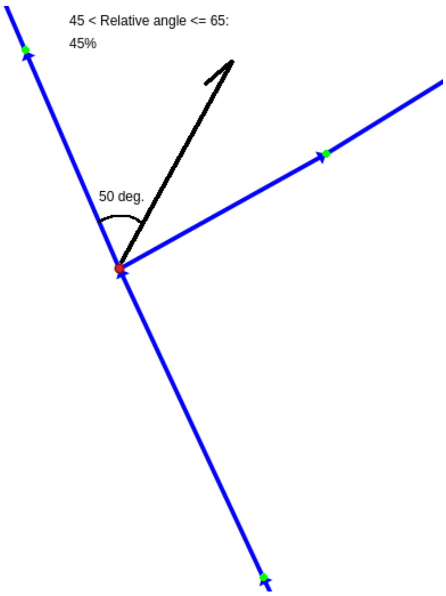
(a) Definition of the combined vectorial forces that affect the plastic in middle node.



(b) Determining the probability of the plastic moving towards the first neighbor upper right.



(c) Determining the probability of the plastic moving towards the second neighbor lower right.



(d) Determining the probability of the plastic moving towards the third neighbor upper left.

The above figure, illustrate in a practical example how the decision of plastic movement is defined based on the natural forces wind and flow direction. Note that in the cases the data for flow direction is absent, the *combined forces* vector only includes the wind element.

The order of the 'neighboring node checking' is based on a *first come first served* decision, meaning that regardless the different probability values of all the nodes, the plastic moves towards the first one to pass the test without giving the option for the rest to try. Additionally, it is important to note that the process of the plastics checking all of the neighboring nodes and the decision making are all happening in the same time instance.

After a plastic leaves the node based on the made decision, it moves with velocity derived from the magnitude of the combined forces. All measurements are based on meters per minutes and the time instances pass for each minute. This helps the simulation determine the traveling distance of the plastic in each instance and figure out the time of arrival to the designated neighboring node. After the arrival to that node, the procedure repeats and a new decision has to be made to determine the next location where the object is going to move (if any).

Lastly, when all plastic objects are removed from the system, the simulation terminates. There is no particular output of the simulation itself, however the object-orientated code implementation has objects of nodes and plastics with interconnected relations to each other. This in practice means that we are able to find into which node each plastic ended up, measure the absolute amount of plastics in a node and even find the exact time instance of when the plastic was removed from the system ("got stuck in a potential hotspot"). The code implementation illustrates the meaningful outcome that is relevant to the customer in a figure containing the initial stage and finishing stage of the simulation.

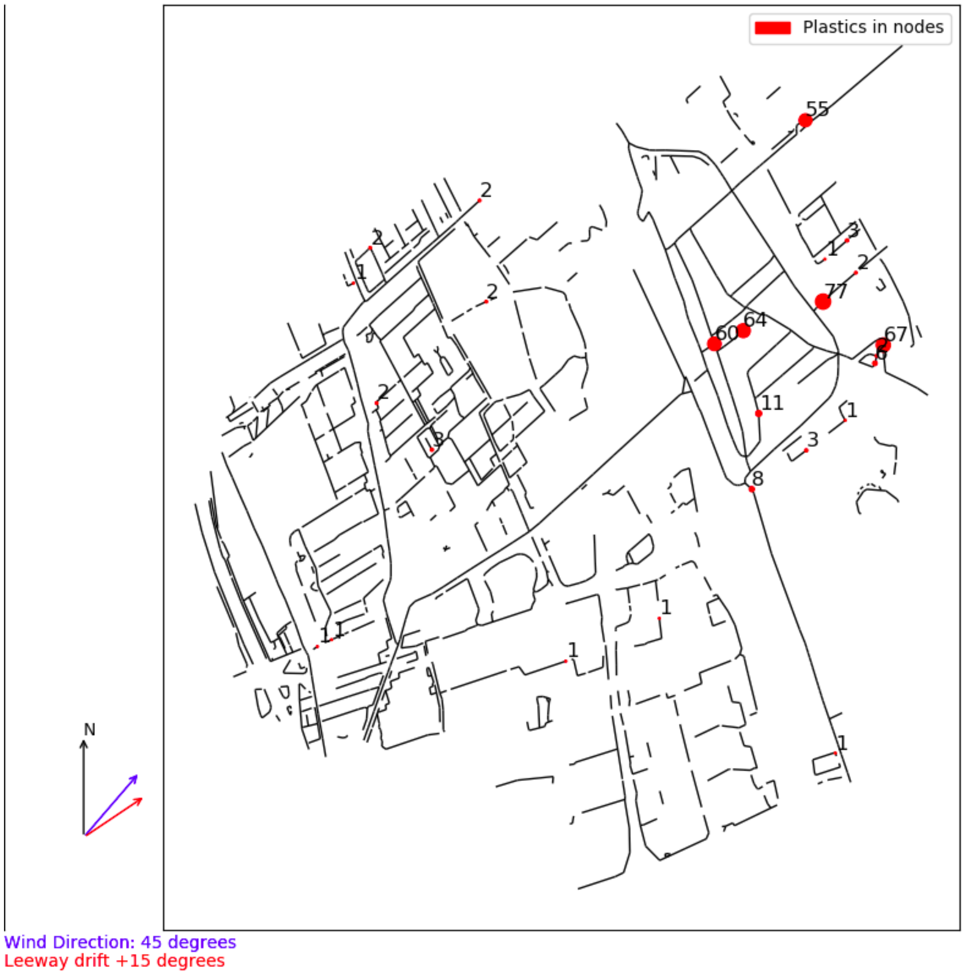


Figure 2.32: Visual representation of the network's finishing stage

The code implementation exports the nodes containing plastics in .shp format. The fields included are the IDs of the nodes, the amount of plastic that each contains and the class of the node (Figure 2.33).

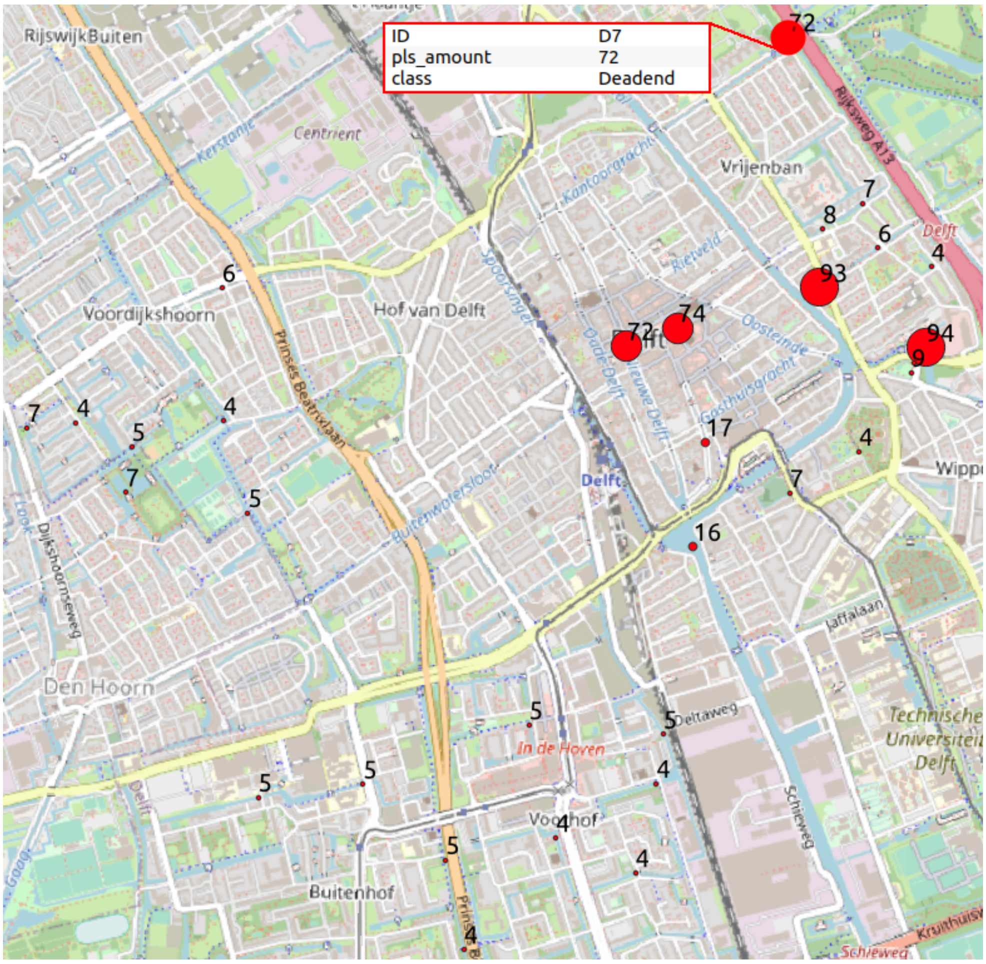


Figure 2.33: Visual representation of potential hot-spots derived from the simulation

2.4.3 Simulation testing approach

As the simulation follows a non-deterministic approach based on probabilities, it is expected to present different output each time. Consequently, in order to come to a stable assessment conclusion about the potential hotspot and the robustness of the simulation, we conducted a series of tests using slightly different parameters each time. Two different configurations were tested, each five times with each of the seven different wind directions chosen, producing 70 test runs in total. The number of tests for each parameter change was chosen on one hand due to time limitation and on the other hand due to the small deviation presented in the outputs.

Tests that aim to include wind direction deviations from the absolute South-West using the **first probability configuration** (Table 2.2).

- 5 tests for wind direction of 30 degrees

- 5 tests for wind direction of 35 degrees
- 5 tests for wind direction of 40 degrees
- 5 tests for wind direction of 45 degrees
- 5 tests for wind direction of 50 degrees
- 5 tests for wind direction of 55 degrees
- 5 tests for wind direction of 60 degrees

Tests that aim to include wind direction deviations from the absolute South-West using the **second probability configuration** (Table 2.3).

- 5 tests for wind direction of 30 degrees
- 5 tests for wind direction of 35 degrees
- 5 tests for wind direction of 40 degrees
- 5 tests for wind direction of 45 degrees
- 5 tests for wind direction of 50 degrees
- 5 tests for wind direction of 55 degrees
- 5 tests for wind direction of 60 degrees

Relative angle (forces ∠ edge)	**Probability to move**
0 <= angle <= 10	100%
10 <angle <= 20	95%
20 <angle <= 45	60%
45 <angle <= 80	45%
80 <angle <= 90	20%
90 < angle	0%

Table 2.2: Relative angle (in degrees) probabilities - 1st configuration

Relative angle (forces ∠ edge)	**Probability to move**
0 <= angle <= 10	95%
10 <angle <= 20	80%
20 <angle <= 45	50%
45 <angle <= 80	20%
80 <angle <= 90	5%
90 < angle	0%

Table 2.3: Relative angle (in degrees) probabilities - 2nd configuration

From the above tests, we end up with 70 datasets containing potential hotspots. The frequency of each node's appearance in these datasets is calculated. Nodes with low frequency of presence are discarded from the final output while the others are considered as potential hotspots to be validated against ground truth data (Chapter 3). Sufficient ground truth data could be used to optimize the probability parameters as in this stage they are defined only by personal assumptions.

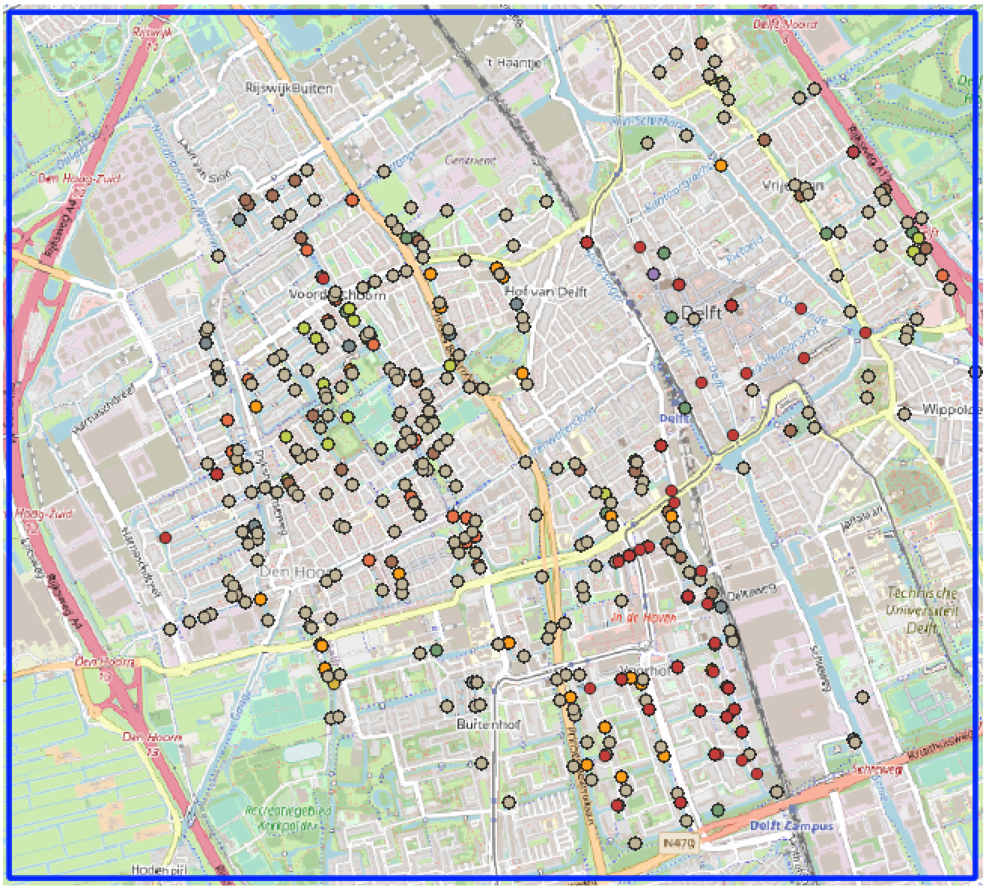


Figure 2.34: Visualisation of all hotspot tests output. Simulation extents

3 | Validation and results

3.1 Simulation Assessment

To assess the simulation approach we made use of the tests that were mentioned in Chapter 2.4.3. The assessment was made against the ground truth dataset that was derived from our field work (Figure 3.1). Note, however, that due to the difference between the simulation extent and the ground truth area, we only took into account those predicted hotspots that could be found inside the ground truth area. In few words, the assessment followed two approaches, a more general ground truth approach for overview and the test approach for specifically assessing the accuracy of our method. In the first case, the objective is to find out the proportional amount of times that a ground truth hotspot was present in the tests while simultaneously exposing other semantic information. In the second one, the objective is to determine the percentage of correctly identified hotspots in the most frequently present nodes from the test.

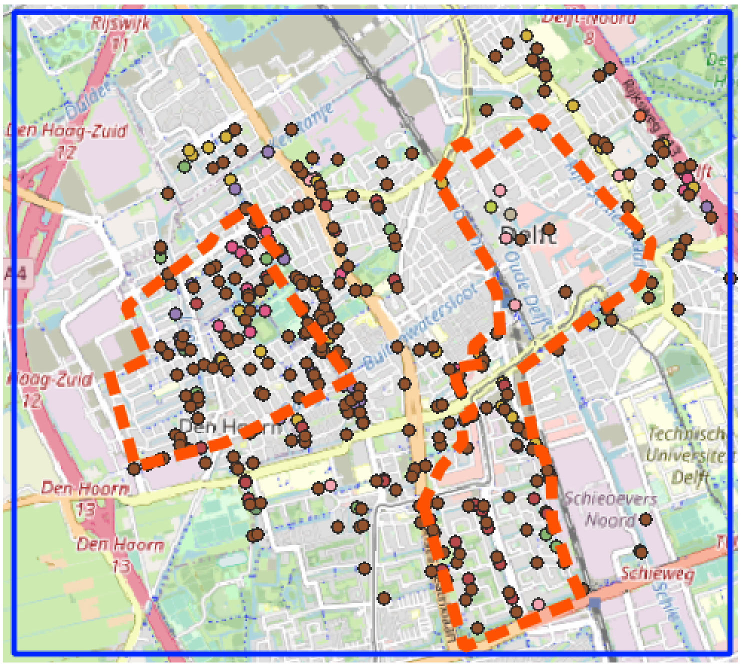


Figure 3.1: Predicted Hotspots from the 70 tests in Simulation extent and Ground truth area

3.1.1 Ground truth approach

For this approach, a simple model was developed that calculates the amount of test nodes that fall within each of the ground truth hotspots. The percentile of this value of the total number of tests, is the final output and is presented in the figure below. (Figure 3.2)

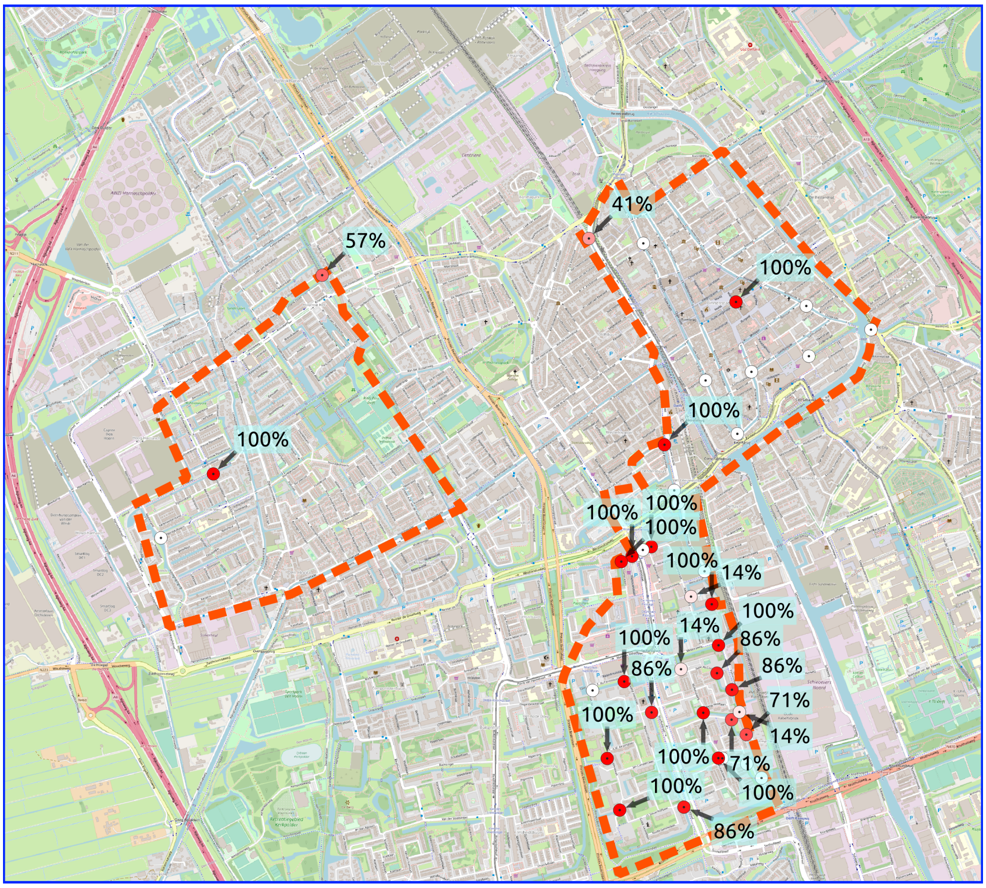


Figure 3.2: Ground truth approach results. Simulation Extent, Ground truth area, Ground truth hotspots

In this figure, white hotspots presented a 0% accuracy which means in practice that no layer (out of 70 tests) was able to identify them.

	Ground truth	Percentage	Plastic Amount	Node classes
Identified (100%)	13	31%	99	Dead-end. Irrelevant
Identified (99% - 80%)	4	10%	7	Dead-end, Irrelevant
Identified (79% - 60%)	2	5%	4	Dead-end, Irrelevant
Identified (59% - 40%)	1	2%	2	Irrelevant
Identified (39% - 20%)	1	2%	6	Dead-end
Identified (19% - >0%)	3	7%	7	Dead-end, Irrelevant
SUM	24	57%	125	Dead-end, Irrelevant
DIFF (0%)	18	43%	-	Junction
OVERALL	42	100%	-	Dead-end, Irrelevant, Junction

Table 3.1: Results from Ground truth assessment approach.

3.1.2 Tests approach

This approach can be considered the opposite of the previous one (Chapter 3.1.1) but is different to its core. We chose to include it as it shows very different and interesting results representing more the final assessment of the simulation. The hotspots gathered from the test were found to be much more in quantity than the actual hotspots located from the field work. This means that there is also another accuracy measure that we need to take into account. This measure is the number of correctly predicted hotspots validated from the ground truth dataset. It is also important to note that the combined tests passed through a filter that discarded outliers corresponding to nodes with low frequency (nodes with <10% presence).

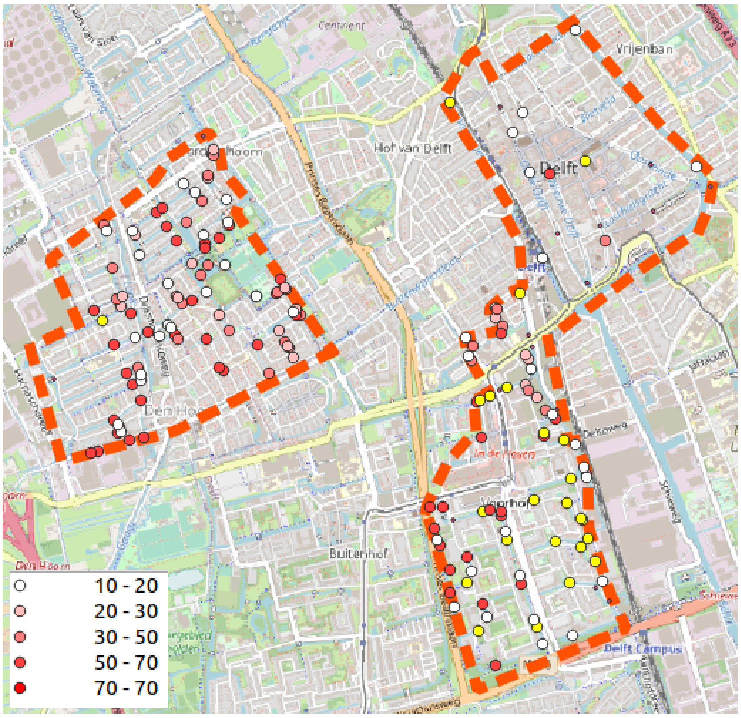


Figure 3.3: Predicted hotspots from tests (frequency adjusted using red scale) and correctly identified hotspots from ground truth (in yellow)

The final result of this approach is that 24 out of 168 predicted were correctly identified, giving an accuracy of 14.2%. However, in order to make the assessment better representative to the reality, we made some adjustments to the predicted hotspots. To elaborate more, a filter was put to discard nodes with less than 4 plastics in them, as their size cannot justify the "hotspot" definition.

The result of this cleaning process can be seen in the below figure. (Figure 3.4)

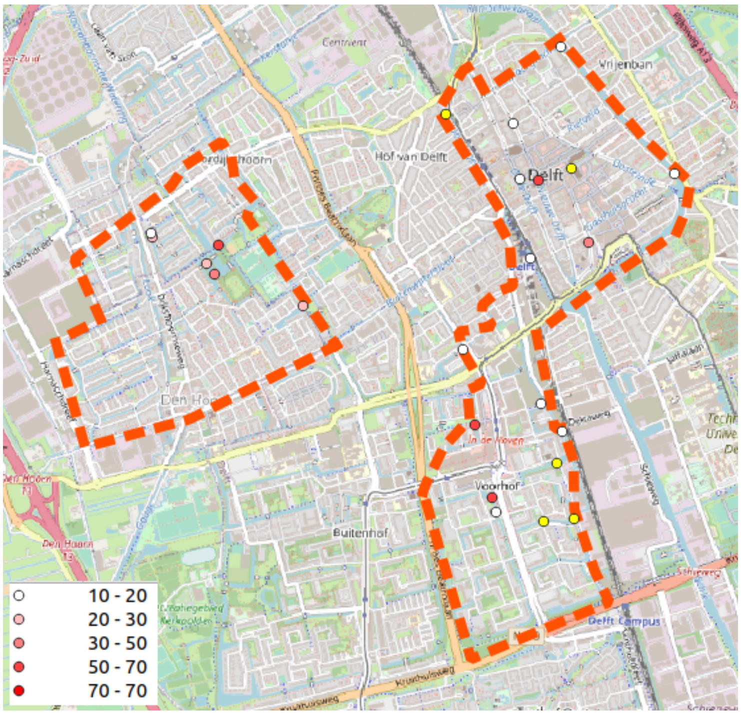


Figure 3.4: Predicted (and specifically defined) hotspots from tests (frequency adjusted using red scale) and correctly identified hotspots from ground truth (in yellow)

From the above figure we finally determine that 5 out of 24 predicted hotspots were correctly identified, giving an overall accuracy of 20.8%.

4 | Evaluation and discussions

The results obtained from both the individual models as well as the network analysis are reasonable and acceptable, taking into account the assumptions made during their development and entailing the simplification necessary for modelling the natural environment. As a consequence of the assumptions made at the beginning of the analysis, various constraints have been created, which oblige the model to operate under idealized conditions and data. More specifically, these limitations are due to the fact that we were not fully aware about them a-priori and therefore it would not possible to include them in our approach. These limitations concern eight aspects discussed below:

The way in which the urban fabric is structured. This means that the buildings and/or other man-made structures pose obstacles both in the movement of plastics before they reach and/or fall into the water as well as an obstacle that does not allow wind to move them when in water (Figures 4.1a and 4.1b). In both cases, the location of the plastics would be different from the one our simulation detects, since the proximity to water parameters is not constant and reliable. This could be solved if a 3D model of the city was available in combination with a more sophisticated algorithm and thus the proximity could be more representative to reality.

(a) Multi-storey building *blocks* the wind　　(b) Brick wall where plastics are obstructed

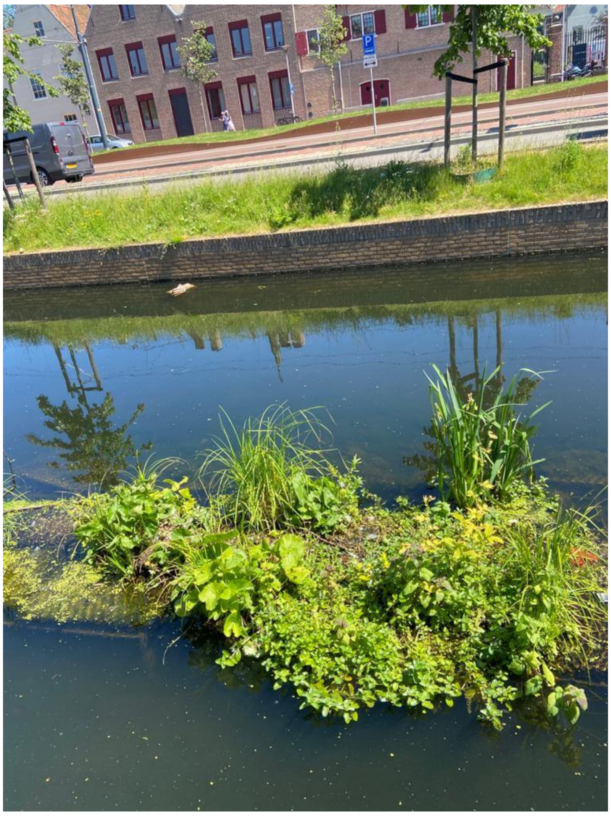
Figure 4.1: Cases where man-made objects pose obstacles for plastics before reaching the water

Partly enclosed canals We took into account the possible concentration of the plastics, considering that they could be transported into the water, as well as that the water level is at the same height as the ground level. Therefore, cases of protection railings, brick walls raised around the perimeters of the canals or other types of mechanical structures that can be placed above the inlet to catch any debris that may become en-trained (Sheng 1990), which prevents access and contact with humans and therefore plastic were omitted (Figures 4.2a and 4.2b). Based on the above, it can be understood that a plastic may never reach the water resource that our algorithm identifies as a possible hotspot.

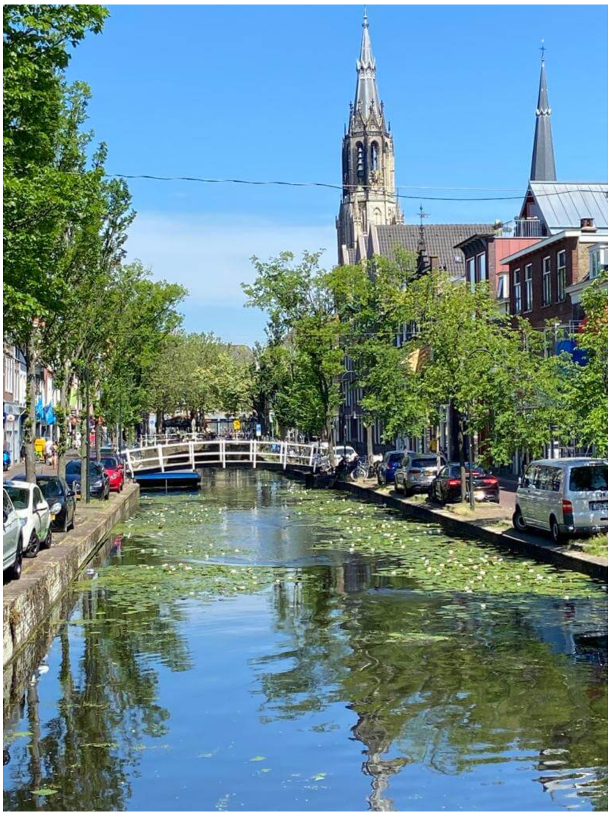
(a) (b)

Figure 4.2: Examples of human intervention for canals protection from debris

The weather conditions. It is noted that depending on the season, the places where plastics tend to accumulate differ (Emmerik, Strady, et al. 2019). A typical example for the study area is the vegetation (lily pond) within the canals that grows during the summer months (Figures 4.3a and 4.3b). This vegetation is an intermediate obstacle to which the plastics clung and therefore never reach the point identified by our simulation. In addition, in these months when there is greater mobility of people, the production of plastics increases following exponential rates and their transfer to different places is done randomly without relying solely on the sources we took into account when creating our models. In contrast, in the winter months, when precipitation is prevalent, plastics do not necessarily adhere to the dead ends, as due to the increase in water level, they are likely to be swept away by the flow and settle in other places. In more extreme weather conditions where ice is also apparent, the movement and accumulation of the plastics is not possible to be detected since other physical phenomenon like sliding occurs.



(a) Vegetation into the canal. **Date: 14-06-2021**



(b) Lily pods overgrowth **Date: 14-06-2021**

Figure 4.3: Examples of vegetation into the water during summer month

The limited data available One of the most important parameters that affect our results is the available information and its degree of completeness. Specifically, in addition to the obvious data (water resources network, population data, provincial borders), relevant data such as the location of waste bins, metal bars on the dead ends, the existence of underground pipes, land use related to vegetation in canals/ rivers and how it changes based on the time they were not available so it was not possible to include them in our approach. In addition, there were cases where the data did not correspond to reality, given the change in the natural environment due to human intervention, and the characteristics of the data were not updated.

The knowledge of the study area. It is noted that in our case, given the size of the area, we focused on further analysis, the on site control of the current situation was possible. Nevertheless, there were changes in our observations regarding the removal of plastics from the positions identified as hotspots, as it can be seen in Figures 4.4, 4.5 and 4.6). From the figures it is clear that plastic accumulation is not a stable condition, since even during 24 hours their location is changed or the spot is cleaned. This makes the modeling more complicated since even in reality there is not a certain place where we can assure plastics attendance. It is noted that due to the knowledge of the study area, several modifications could be made to the data in order to correspond to reality and to avoid limiting the performance of the simulation to more general cases and networks. However, this requires extensive knowledge of the study area and field work that contradicts the automatic nature of our approach.

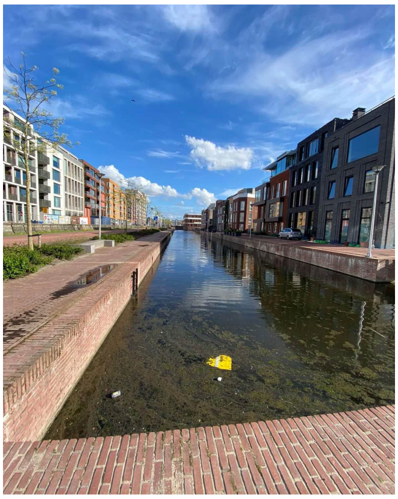


Figure 4.4: Identified hotspot; ground truth. **Date: 21-05-2021**

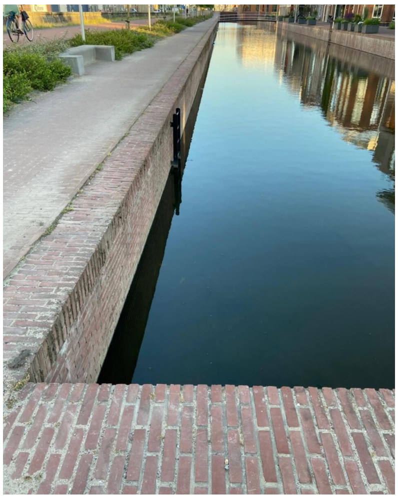


Figure 4.5: Cleaned water corner (hotspot). **Date: 13-06-2021**

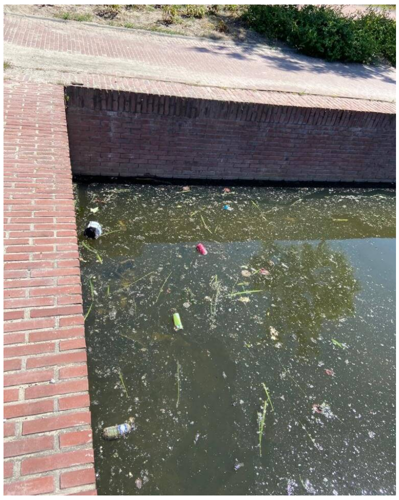


Figure 4.6: Plastic appearance. **Date: 14-06-2021**

Figure 4.7: Example of changed conditions

The reliability of the ground truth data. As mentioned in the above section, the plastic accumulation that was found as part of the ground truth dataset is not a stable condition. Because the plastic objects do not stay at the locations that they were found, they must either be removed from the water or moved to another location in the water. Because tracking the individual pieces of plastics through the system was not possible during the timespan of this project, we investigated if plastics get taken out of the water by cleaning activities. To find out the areas and frequency of the cleaning activities in the waters of our fieldwork area, we had a meeting with the area manager of municipality of Delft. He stated that the canals in the city centre are cleaned up with nets from the sides by an external contractor on Mondays, Wednesdays and Fridays. The area covered can be seen in Figure 4.8. The canals in the city centre are also being cleaned on Tuesdays and Thursdays by the Stunt foundation with electric boats. The remaining waters outside of the city centre are cleaned around six times a year. He also stated that the flow speed in the city centre is not very high, so plastics do not have much time to move through the canals. This brings us to the conclusion that the hotspots found in the field might not be accumulation areas, but areas where a lot of plastic enters the water. Because the flow speed is not very high in the city centre of Delft and the plastics are removed quickly, they do not have much time to move around in the system to reach such an accumulation area. In order to create a more representative fieldwork dataset in the future, ideally all cleaning activities in the area would need to stop. In this way the plastic that is in the water system has the time to move around and reach a hotspot where it accumulates.

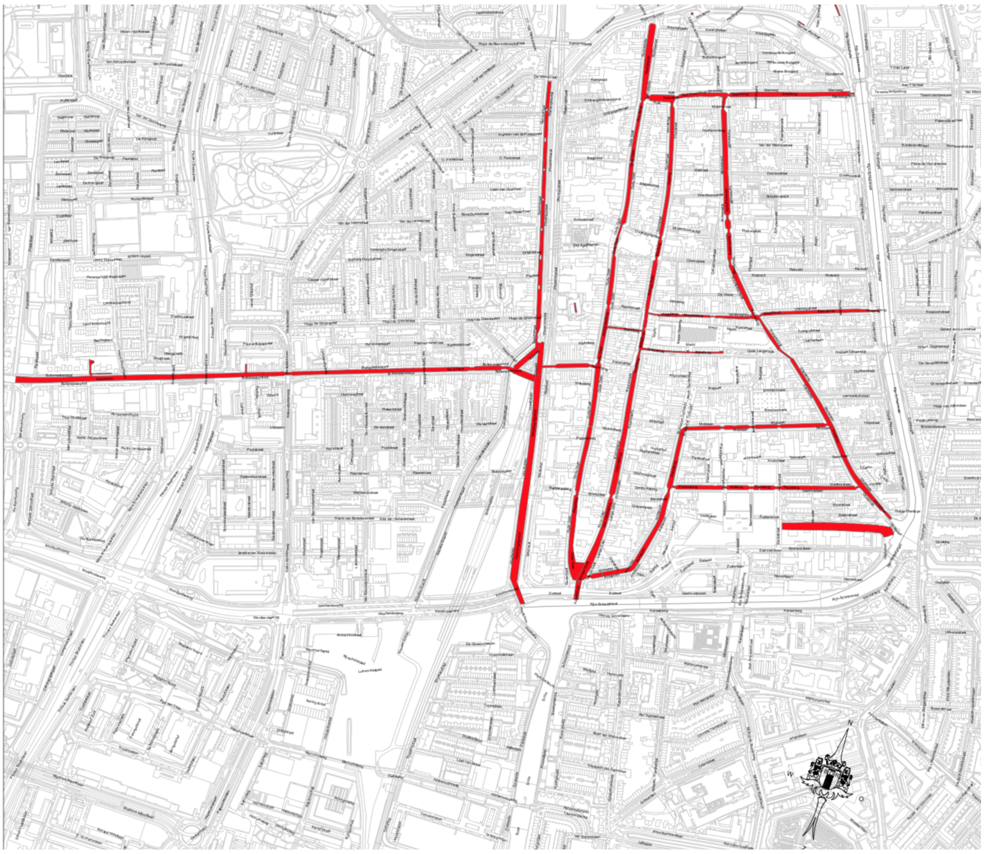


Figure 4.8: Map of the Municipality of Delft displaying where the external contractor cleans the water (displayed in red).

The effectiveness of the method outside the study area. The scope of this project only focused on modelling plastic transport in Delft city center. Due to the time limitation, no further tests were done in evaluating the model in other places as new set of ground truth data would be needed. It is assumed that the conditions in other Dutch cities are similar to the water network in Delft. However, the model is possible not work in places where the transport by water flow dominates over transport by wind. The Netherlands has a unique water network made of canals and rivers whereas in most countries canals are not that common. It is therefore assumed that the developed method would not give satisfactory results if tested in study areas outside of the Netherlands.

The oversimplification of the problem, based on several assumptions (Chapter 2.2.1). Given the complex nature of the problem and the fact that plastics' behavior is affected by various parameters (natural or not) we had to make several assumptions in order to simplify the problem and adapt it to more reliable results. Specifically, since it was not possible to capture the complexity of the topic we took under consideration the major factors that affect plastic behavior in the water, omitting the ones mentioned above. Although these considerations contributed to the improvement of the technical part of our approach (simulation), they cost its accuracy and the possibility that it could be used in more general cases.

From all the above it is concluded that in order for the already existing model to become more efficient and reliable (reflect reality), all the above parameters should be taken into account and included in the corresponding simulation. In this way, the problem is more defined while arbitrariness in the results are avoided.

5 | Conclusions and Future Work

5.1 Conclusions

This project aimed to create a semi-automatic model simulating the transport and accumulation of plastic debris in Dutch urban water networks. The purpose of this model was to aid the prediction of plastic accumulation zones in order to find where it would be the most efficient to install plastic collectors by our client Noria. Although there is a knowledge gap in research on plastic transport in the freshwater environment, parameters affecting plastic transport cited in literature and noted from the experience of our client were included in a method incorporating a GIS analysis and network simulation. First, a GIS analysis is done where locations of potential plastic sources, dead ends and canal geometry where water flow is obstructed are detected. Using the acquired knowledge and data from this analysis, the network simulation identifies places with the highest probability of plastic accumulation based on our assumptions. The results, however, showed large over-simplification of the problem. Out of the 24 hotspots predicted by the model, 5 'true' hotspots are found in reality.

The complexity of the environment which was being modelled was evaluated. Many simplifications of the real world were introduced for the sake of decreasing the complexity of the model and due to the time available for the project. Such simplifications include constant velocity of water flow, constant wind direction and no attribution to changing of the environment, vegetation cover and weather conditions were given. It was also noted that the model is highly dependent on the quality and completeness of the input data. Limitations were introduced by data not being available, such as detailed data on vegetation in the water network, location of trash bins in the city and civil structures in the canals. Finally, the water in reality was assessed against the ground truth collected during field work. However, it was discovered only at the end of the project that the canals in the study area are cleaned several times a week, creating a possible negative impact on the assessment, because potentially some accumulation areas were not mapped as they had been cleaned.

Overall it is believed that the model can assist our client in choosing the most efficient location for their trash collector. As the model does not produce a high accuracy, field research is still needed to determine the 'real' hotspots out of all that are predicted by the model. However, added value comes from pre-selection of areas where field research is needed, removing the need of mapping the plastic in the entire area of interest.

5.2 Future work

The project was implemented during a period of two months and the short time available caused several limitations to our final results. For further analysis to improve the already existing results, it would be advisable to use machine learning in order to classify the plastics detected into different categories according to their type, based on the identification given from the Society of plastics industry (SPI) (Products 2015), or size. This could be achieved by building classification algorithms that uses training data in order to predict the probability that one object belongs to one of predetermined categories. In our case the plastics detected fall into one of the seven categories mentioned above.

Another possible approach for the detection of potential locations where plastics tend to accumulate is to use real time data, collected through monitoring procedures. In more details, for our study case, bridge-mounted cameras could be used in order to spot not only the places where plastics are gathered but the quantity of them as well, skipping visual counting. More analytically, with the use of video camera technologies, images from different time periods can be produced, consisting the input data in a deep learning analysis. Then, with the use of deep learning object-recognition algorithms, plastics' type can be characterized, making their recognition from the images easier. As in any method, also in this case it is necessary to make certain assumptions and determine parameters that affect the operation of the algorithm that is going to be followed. For example, the location of the camera in relation to the study area (distance from plastic, possible obstacles in front of the lens, quality of the lens etc.) can play an important role in the detection (Lieshout et al. 2020).

In literature, there have already been proposed methods for monitoring plastic movements using UAVs or spaceborn remote sensing for multispectral images (Emmerik and Schwarz 2020). Using satellites for this purpose involves addressing problems related to the distribution, types, quantities and sources of plastics (Knaeps 2020). In this case, the aim is to extract information about plastics, using spectral measurements. However, as mentioned in the introduction, it is not effective to use this method for smaller size water bodies (canals/rivers), so we could benefit from this approach by acquiring information about the vegetation existed into the water bodies, which is an information missing from our approach. To do so, first it should be selected a trained area as well as some target materials, which in our case will be the vegetation. Then, image classification process can be implemented in order to categorize and label groups of image's pixels, into one of several land cover classes, that would follow user defined rules (relevant to the target materials) (Topouzelis, Papakonstantinou, and Shungudzemwoyo P. Garaba 2019). With image classification it is aimed the identification and depiction of the features occurring in the image used, in terms of the specified target materials existed on the ground.

Using an image processing software, it is possible to develop a statistic characterization of the reflection for each information class (vegetation) on each spectral band, using indicators such as normalized difference vegetation index (NDVI). Once this is achieved, the user is able to examine the reflectance for each pixel and decide about which of the signatures presented it resembles the most. Taking into account the reflectance of the basic ground elements that is known, the vegetation can be detected as the objects that their signature would differ significantly from the one corresponds to water (Science Chulalongkorn University n.d.). In this case, it is noted that in order to obtain satisfactory results, high geospatial resolution images are required, in which it will be possible to distinguish the various vegetation patches that exist in the study area. Even more attention should be paid to the date that the aerial photograph was taken, as physical factors (i.e. atmospheric condition, clouds, river condition-water flow) directly affect the results of the process mentioned.

Bibliography

Best, Jim (2019). "Anthropogenic stresses on the world's big rivers". In: *Nature Geoscience* 12.1, pp. 7–21.

Emmerik, Tim van and Anna Schwarz (2020). "Plastic debris in rivers". In: *Wiley Interdisciplinary Reviews: Water* 7.1, e1398.

Emmerik, Tim van, Emilie Strady, et al. (2019). "Seasonality of riverine macroplastic transport". In: *Scientific reports* 9.1, pp. 1–9.

Eriksen, Marcus et al. (2014). "Plastic pollution in the world's oceans: more than 5 trillion plastic pieces weighing over 250,000 tons afloat at sea". In: *PloS one* 9.12, e111913.

Garaba, Shungudzemwoyo P et al. (2018). "Sensing ocean plastics with an airborne hyperspectral shortwave infrared imager". In: *Environmental science & technology* 52.20, pp. 11699–11707.

Gasperi, Johnny et al. (2014). "Assessment of floating plastic debris in surface water along the Seine River". In: *Environmental pollution* 195, pp. 163–166.

Hackett, Bruce, Oyvind Breivik, and Cecilie Wettre (2006). "Forecasting the Drift of Objects and Substances in the Ocean". In: *Ocean Weather Forecasting: An Integrated View of Oceanography*. Ed. by Eric P. Chassignet and Jacques Verron. Dordrecht: Springer Netherlands, pp. 507–523. ISBN: 978-1-4020-4028-3. DOI: `10.1007/1-4020-4028-8_23`. URL: `https://doi.org/10.1007/1-4020-4028-8_23`.

Jakovljević, Gordana, Miro Govedarica, and Flor Álvaret Taboada (2019). "Remote sensing data in mapping plastics at surface water bodies". In: *Conference: FIG Working Week*.

Jambeck, Jenna R et al. (2015). "Plastic waste inputs from land into the ocean". In: *Science* 347.6223, pp. 768–771.

Kandasubramanian, Balasubramanian and Merlin N. Issac (2021). "Effect of microplastics in water and aquatic systems". In: *Environmental Science and Pollution Research* 28.16. ISSN: 19544–19562. DOI: `10.1007/s11356-021-13184-2`. URL: `https://link.springer.com/article/10.1007/s11356-021-13184-2#citeas`.

Knaeps, Els (2020). *DETECTING MARINE PLASTICS! AN EXPERIMENT*. URL: `https://blog.vito.be/remotesensing/detecting-marine-plastics` (visited on 06/04/2021).

Lieshout, Colin van et al. (2020). "Automated River Plastic Monitoring Using Deep Learningand Cameras". In: *Earth and Space Science* 7.8. DOI: `10.1029/2019EA000960`. URL: `%5Chttps://doi.org/10.1029/2019EA000960`.

Morritt, David et al. (2014). "Plastic in the Thames: a river runs through it". In: *Marine Pollution Bulletin* 78.1-2, pp. 196–200.

Moy, Kirsten et al. (2018). "Mapping coastal marine debris using aerial imagery and spatial analysis". In: *Marine pollution bulletin* 132, pp. 52–59.

Murray, Peter (2002). *KNOWLEDGE MANAGEMENT AS A SUSTAINED COMPETITIVE ADVANTAGE*. URL: https://iveybusinessjournal.com/publication/knowledge-management-as-a-sustained-competitive-advantage/ (visited on 06/05/2021).

Parker, Laura (2019). *The world's plastic pollution crisis explained*. URL: https://www.nationalgeographic.com/environment/article/plastic-pollution (visited on 05/25/2021).

Products, Scranton (2015). *Different Types of Plastics and SPI Codes Used to Classify Them*. https://www.scrantonproducts.com/different-types-of-plastics-and-the-spi-codes/. Accessed: 2021-06-4.

Science Chulalongkorn University, Faculty of (n.d.). *Environmental Remote Sensing (Courseware 2309507) -Image Classification*. http://www.sc.chula.ac.th/courseware/2309507/Lecture/remote18.htm. Accessed: 2021-06-05.

Services, Meteorological (n.d.). *Weather Online*. https://www.weatheronline.co.uk/Netherlands/Amsterdam/Wind.htm. Accessed: 2021-06-12.

Sheng, T.C. (1990). *Watershed Management field manual: Watershed survey and planning*. 1st Edition. FAO CONSERVATION GUIDE 13/5.

Tasseron, Paolo et al. (2020). "Plastic Hotspot Mapping in Urban Water Systems". In: *Geosciences* 10.9. ISSN: 2076-3263. DOI: 10.3390/geosciences10090342. URL: https://www.mdpi.com/2076-3263/10/9/342.

Topouzelis, Konstantinos, Apostolos Papakonstantinou, and Shungudzemwoyo P. Garaba (2019). "Detection of floating plastics from satellite and unmanned aerial systems (Plastic Litter Project 2018)". In: *International Journal of Applied Earth Observation and Geoinformation* 79, pp. 175–183. ISSN: 0303-2434. DOI: https://doi.org/10.1016/j.jag.2019.03.011. URL: https://www.sciencedirect.com/science/article/pii/S0303243419301163.

Ward J., J. Peppard (2016). *The Strategic Management of Information Systems Building a Digital Strategy*. 4rth Edition. Wiley.

Windfinder (n.d.). *About Windfinder*. https://www.windfinder.com/about/. Accessed: 2021-05-10.

A1 | Appendix

A1.1 User Manual

In this section the user manual of the final product is given. The process is visualised in A1.2. It consists of three main steps: data preparation, data generation with QGIS models and running the simulation. The user is expected to have possession over a computer or laptop were at least QGIS 3.10 and Python 3.7 can be installed. The device also needs sufficient storage capacity to store the necessary datasets. The product should be able to run on any operating system.

A1.1.1 Data preparation

Because the quality of the input data can greatly influence the accuracy of the results, it is important to check if the input data is up to standard.

To be able to run the models and simulation the following datasets are needed:

- The **water features** as a **line** dataset. The dataset should be available as a shape-file for it to be usable in both QGIS and the simulation. The features should be as connected as possible for the water network in the simulation to work. The features should also be very close to reality to ensure that the output is as accurate as possible.
- The **water features** as a **polygon** dataset. The dataset should be available as a shape-file and should overlap the line water features dataset.
- The **water flow direction** as a line dataset. The dataset should be available as shape-file and the direction of the flow should be stored in the geometry of the lines. This means that the water flows from the first node to the last node of the specific line feature. Naturally the flow direction of this dataset should be the same as in reality.
- The **sources of plastic** as a point dataset. This should also be available as shape-file.
- The **population density** as a polygon dataset.

All datasets should only cover the study area to prevent unnecessary computing.

A1.1.2 QGIS models

The next step is to run the QGIS models. In order to do this the user needs to have QGIS 3.10 or higher installed. We assume that the user has some basic knowledge about GIS-systems and knows how to load in the shape-file layers.

The following QGIS models need to be downloaded:

- The data cleaning model
- The relevant nodes model
- The plastic insertion model

Once this is done, the models can be run in QGIS.

1. The first step is open a new QGIS project and load in all the necessary layers.
2. The first model that must be used is the data cleaning model.
   (a) The user can open models in QGIS by clicking on the toolbox in the attributes toolbar, followed by clicking on the models icon. The user needs to select "Open Existing Model" or "Add Model to Toolbox". Another way to open a model in QGIS is to drag the model file into the QGIS window.
   (b) In the data cleaning model window the user should select the water polygon feature layer as the *Water bodies parameter*
   (c) The user should select the water line feature layer as the *Water line parameter*
   (d) The output of the model is a cleaned water line features layer.
3. Then we use the relevant nodes model.
   (a) Open this model the same way as before.
   (b) Select the cleaned water line feature layer (output from data cleaning model) as the *Waterlines Network parameter*.
   (c) The output of the model is a point layer containing the relevant nodes for the water network.
4. Finally, we use the plastic insertion model
   (a) Open this model the same way as explained in 2(a).
   (b) Select the plastic sources point layer as the *Point layer of plastic sources parameter*.
   (c) Give the *radius of the 360 degree buffer* in metres. See figure A1.1

(d) Give the *radius of the wind buffer* in metres.

(e) Select the *ID column from the relevant nodes layer* (output from relevant nodes model).

(f) Select the cleaned water line feature layer (output of data cleaning model) as the *Water lines parameter*.

(g) Give the *Width of the wind buffer* in degrees.

(h) Give the *Wind direction* in degrees.

(i) Select the population density layer as the *Population density parameter*

(j) Select the relevant nodes layer (output from relevant nodes model) as the *nodes for the networkx graph parameter*.

(k) Three layers are outputted. The two buffer layers are not needed and can be discarded. The final nodes layer should be kept and stored.

5. To be able to use the layers created by the models, they need to be stored as a shape-file on the device.

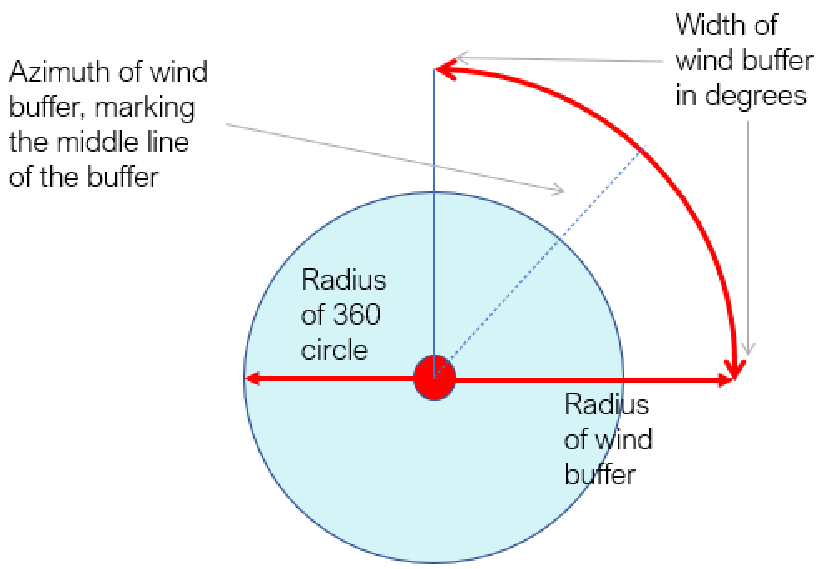


Figure A1.1: Explanation of values inserted into the plastic insertion model

A1.1.3 Simulation

The simulation is developed using the programming language 'Python'. The version on which it was developed is 3.8.5. Additionally, the code uses some 3rd party packages that need to be installed that may also require extra dependencies:

- NetworkX (`https://pypi.org/project/networkx/`)
- Numpy (`https://pypi.org/project/numpy/`)
- Matplotlib (`https://pypi.org/project/matplotlib/`)

The code is segmented into 3 python files:

- main.py (Core and Plots)
- classes.py (Classes,methods and functions)
- simulation.py (Algorithm)

The code requires to include a 'Data' directory into its own directory containing a shape-file of edges and one of nodes that are the inputs of the simulation. The exact name of the shapefiles is currently hardcoded into a relative path. Users that want to use other datasets than the one provided by us, must change the relative path from within the code itself (in main.py). The shape-file of the edges is the output shape-file of the cleaned data model, processed according to the steps in 2.4.1 to improve the data and to include the flow direction. The nodes are in the output shape-file of the plastic insertion model. It is important for the node dataset to contain fields of ID, class, and plastic amount (pls_amount) as this information is crucial to the code's functions. The steps to run the simulation are described below:

1. Install all the dependencies mentioned above and download the simulation folder.
2. Add the line and node shape-files to the Data directory (currently contains the datasets used in this project).
3. To execute the code, the file main.py needs to be run either from the terminal or from an IDE.
4. After that, the user is asked to insert the input in the terminal that defines the wind direction (CW from North in degrees).
5. Before the code finishes its execution, a figure pops up presenting the initial and final stage of the simulation as shown in Figures 2.30 and 2.32. Closing the figure's window will result to the termination of the execution.
6. Finally, the code writes the resulting nodes as a shape-file into the "./Data/Plastics" directory.

A1.1.4 User manual flowchart

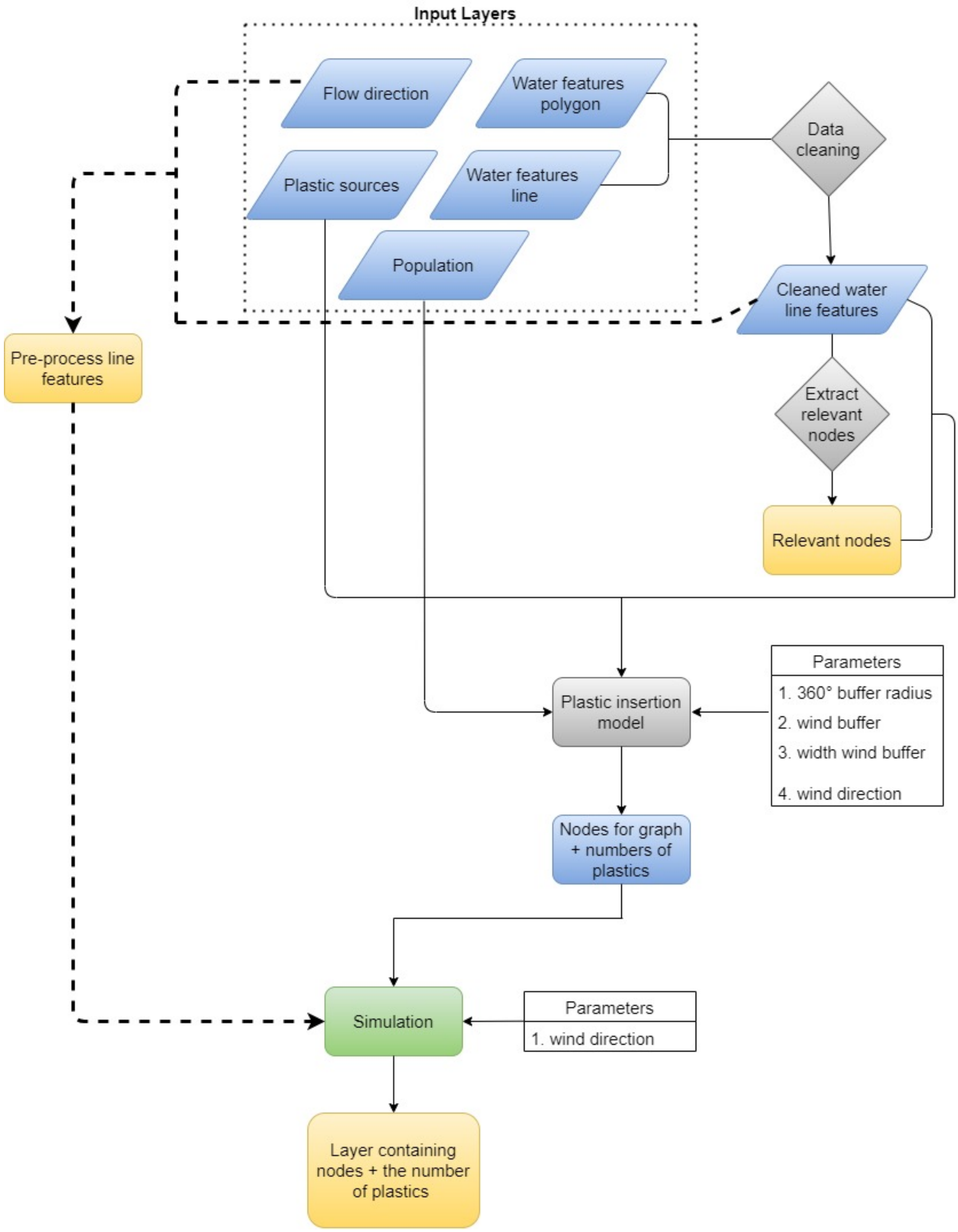


Figure A1.2: Flowchart of user manual

A1.2 Simulation pseudo-code

Algorithm 1 Plastic Movement Simulation

```
Input Waterlines, Relevant_nodes, WindDirection (default: 45)
Output Potential_Hotspots (Point Shapefile)
IrrelevantNodes ← Waterlines.nodes
N ← Relevant_nodes ∪ Irrelevant_nodes
W ← Wind Direction + Leeway_Drift(default: 15 degrees)
A ← Relevant_nodes['plastic_amount'] {'A' is a list containing all plastic objects that
are active}
for all A do
  A ← Updated active plastic list
  if plastic in N then
    Neigh ← neighboring nodes of node
    for all Neigh do
      Relativeangle ← Neigh→∠Forces→ {W + Flow}
      P ← Probability of plastic to move to Neigh node {Relativeangle}
      if P then
        Plastic moves
        Remove plastic from node
        plastic.velocity ← ||Forces||
      else
        Plastic stays put
      end if
      Deactivate plastic {Plastic exits the system from node}
    end for
  else
    if plastic.distancetonode <= 0 then
      Insert plastic in node {plastics has reached the node}
    else
      plastic.distancetonode − = plastic.velocity
    end if
  end if
end for
return Write (in shapefile) the nodes from N that have plastics in them
```