Reverse household waste flow mapping

Davey Oldenburg

July 2019

TUDelft

REVERSE HOUSEHOLD WASTE FLOW MAPPING

A thesis submitted to the Delft University of Technology in partial fulfillment
of the requirements for the degree of

Master of Science in Geomatics for the Built Environment

by

Davey Oldenburg

July 2019

The work in this thesis was made in the:

REPAiR Project, AMS institute
Department of Urbanism
Faculty of Architecture & the Built Environment
Delft University of Technology

Supervisors: ir. Rusnė Šilerytė
               Prof. dr. Jantien Stoter

Co-reader: Dr. ir. Stefan van der Spek

ABSTRACT

Our linear way of producing has to come to an end. More and more companies and government institutions see the value of circularity of raw materials. The municipality of Amsterdam published several reports to clarify the directions towards circularity. A report on waste-chains highlights the importance of organic residual flows for circular economy transition. This thesis focuses on organic residual streams, the lack of understanding of these streams has been mentioned as an obstacle to planning by many of regional stakeholders. The topic of this thesis is formulated within the context of the REPAiR project by the AMS Institute. The starting point of this thesis is the waste collection location in public space and follows the waste back to the retailers of food. The households are identified, after which the used trashcan and the retailers of food are identified.

To create the geographical connection between the three nodes in this thesis, a related works research is conducted. The main source of food waste is food retail and the main disposal is collection of residual waste. Since the municipality of Amsterdam is not collection organic waste separately, residual waste is considered. Multiple researches added spatial dimensions to their material flows, trying to create a better insight in sinks and flows of the chains. Household level is not reached. This leads to the research question: *How can household waste flows be mapped in reverse based on their environment?*

To create the mentioned connections, two main methods are developed. First, the method household-to-trashcan and second, the method household-to-retail. The first method is simulating the pedestrian path from a household towards their used trashcan. A least cost raster is created from the BGT, the direct environment is translated into walkability values representing the willingness to walk on a certain object. By comparing the paths to different trashcans, the trashcan reachable with the smallest least cost value is picked. Another factor is inserted, considering that walking to the trashcan is mandatory trip, executed in as less time as possible, city hotspots are inserted. The hypothesis suggests that people are willing to walk further, if the residents can combine their trip to the trashcan with their trip to, for example, the supermarket. By changing the values in the least cost raster, this direction is made preferable. After the validation of this method is concluded, that a stronger influence of the city hotspot (among the tested values), results in a higher positive correlation of the number of households per trashcan and the mass of waste in the trashcan in kg.

The second method household-to-retail, is a demonstration, due to the uncertainly and the many involved personal factors. Food retailers in different isochrone areas are given importance values based on their distance towards the considered household and the floor size of the retailer.

After establishing the connections from trashcan to household and from household to food retail, reverse distribution of the waste measure in the trashcan can be executed. In this way is known which household contributed how many kg per year to the trash composition of the used trashcan. When adding the other step, the contribution by the food retailers can be given.

This thesis generates the first step in mapping residual waste flows touching household level. Promising positive correlations between the number of households using a trashcan and the kg measured inside this trashcan are shown. The method household-to-retail cannot be validated and is therefore only a demonstration. The outcome of this thesis is a step in the right direction to be able to follow waste through the waste chain, by analyzing results in trash composition and knowing which households and retailer contribute to this composition.

ACKNOWLEDGEMENTS

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

ACRONYMS

1 | INTRODUCTION

Our linear way of producing has to come to an end. The world population is estimated to grow to 9 billion people in 2050 (Searchinger et al., 2013). All these individuals have different needs and to answer those needs, raw materials are required. Raw materials are the base of many products used daily, but they are finite. The dependence on these materials result in fluctuations in prices and scarcity. More and more companies and government institutions see the value of circular production lines, recognizing the opportunities of innovative production techniques and a reduction of dependency on raw materials, which might result in cost minimization (Accorsi et al., 2015; Murray et al., 2017). For example, the impact of circular economy on electronic waste is estimated on saving 500 million euros in Europe alone (MacArthur E., 2013). The Dutch research organization TNO estimates the added value of circularity to the Dutch metal and electronics sector to be 7.3 billion euros (Bastein et al., 2013). To encourage this, the municipality published several reports on circularity aiming to clarify directions for improving sustainability in the Amsterdam metropolitan region (Municipality Amsterdam, 2015b).
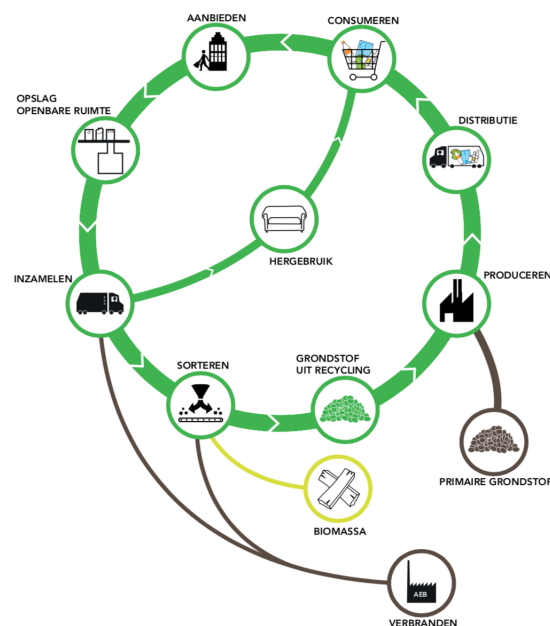
Figure 1.1: Ideal waste chain by the municipality of Amsterdam [source: (Municipality Amsterdam, 2015a)]

One of these reports focuses on the waste chains within the municipality (Municipality Amsterdam, 2015a). This report provides calculations on waste separation, composition and trends, and includes insight in specific waste chains for different types of waste (e.g.plastic or glass) and provides a more circular approach to waste management. Figure 1.1 visualizes the ideal waste chain in Amsterdam, but at this moment, the percentage of of waste that is reused is just 26 present (Municipality Amsterdam, 2015a).

The report highlights the importance of two chains for the Circular economy transition: organic residual flows and the construction and demolition value flow (Geldermans et al., 2016; Municipality Amsterdam, 2015a). Instead of addressing the waste problem when the waste has already been produced, waste flow mapping creates the possibility to follow waste reverse in the chain and therefore the materials can be addressed closer to the source. Kurdve et al. (2015) acknowledges the value of spatial dimension to flows, analyzing and understanding the chains, by stating that these mapping results can make the analysis more accurate and therefore give more insight in waste generation and material waste handling activities. Mapping of waste flows is used in multiple applications to identify and map nodes where different waste management operations are executed.

In the vision and action agenda *Circular Amsterdam*, Municipality Amsterdam (2016) states that by making more (open) data available, the municipality can stimulate innovation in the city. This open data platform can offer a transparent overview of the supply, demand and use of organic residual streams. It can address the uncertainty in the market by improving the matching of supply and demand. The lack of understanding has been mentioned as an obstacle to planning by many of regional stakeholders. The mapping of the waste flows can contribute in a better understanding of the obstacle of matching supply and demand. Possibly resulting in suited waste prevention instead of waste reaction. The aim of this thesis is to map household waste flows and include household level.

The topic of this thesis is formulated within the context of the REPAiR project by the AMS Institute. The REPAiR project studies the implications of possibilities of Circular Economy (CE) in the management of waste flows, aiming to fill a gap in the research on consumer behaviour and sustainable urban development (REPAiR, 2018). This thesis is building upon research and data provided by the H2020 project "REPAiR - Resource Management in Peri-Urban Areas. Going Beyond Urban Metabolism" led by prof. Arjan van Timmeren. The research by REPAiR focuses, among other, on visualizing steps of the waste chain and mapping them for the Amsterdam Metropolitan Area from the neighbourhood to the waste treatment plant.

To continue the research on urban metabolism, this thesis is adding new nodes from the waste chain to the initiated research by REPAiR. The starting point of this thesis is the waste collection location in public space (this was also the start point in the REPAiR project) and follows the waste back to the retailers selling the food which can become the waste. The households are identified, after which the used trash disposal and back to its producers. The identified steps in the waste chain result in the possibility to follow flows (reverse) and address waste closer to the source.

1.1 PROBLEM STATEMENT

Waste flow mapping creates a possibility to follow waste reverse in the waste chain and therefore materials can be addressed closer to the source. Making more (open) data available would result in a transparent overview of the supply, demand and use of organic residual streams. By mapping waste flows on household level, specific households and their food retailers can be targeted for development towards a CE.

However, data to make this specific targeting possible is not available. Mapping waste flows with a household level of detail has never been done before and this data is therefore missing. Waste policies are implemented, but household behaviour stays unknown, the bottlenecks of the implementations stay hidden. Municipalities are collecting waste data for monitoring purposes, but this data is not used to develop methods to provide this missing insight.

In the response to this problem, this thesis does the first attempt in mapping household waste flows. Since household flows are highly depending on human

decisions, this research predicts human behaviour based on their environment. The result contains steps in the household waste chain, with their location. The connections are established between the household and their food retailer and the household and their used trashcan.

1.2 DEFINITIONS

Since multiple technical terms are used throughout this research, this section provides an overview of what these terms represent in the context of this thesis. Nodes and links are used to describe the steps and connections within the waste chain, only represent topological information. Vertices and edges are used to describe parts of a graph and contain geometric data (in addition to topological data). For the graphical representation two different terms with different intentions are used. Firstly, visualization is a graphical representation without a spatial dimension. Second, mapping is a graphical representation with a spatial dimension. This thesis is working towards a mapping representation of the found links between the nodes in the waste chain.

Term	Explanation
Node	Step in waste chain (e.g.trashcan, household, retailer)
Link	Connection between nodes in waste chain
Vertex	Node containing geometric data
Edge	Link containing geometric data
Trashcan	Step in waste chain (also referred to as: waste storage in public space)
Retailer	Step in waste chain (also referred to as: supermarket, supplier)
Visualization	Graphical representation **without** spatial dimension
Mapping	Graphical representation **with** spatial dimension

Table 1.1: Definitions of often used terms

1.3 RESEARCH QUESTIONS

The main research question for this thesis is:

How can household waste flows be mapped in reverse based on their environment?

The goal of this research is to map the waste flows, with Amsterdam as test area. The main question is researched through the following sub-questions:

1. **Theory:**
   Why is it necessary to map waste flows? How has the mapping of (waste) flows been done in the past? How are waste flows visualized/mapped?

2. **Design:**
   How can methods be developed to add a spatial dimension to household waste flows?
   a) Which datasets are needed to develop such methods?
   b) Which trashcan is used by a household based on factors of their environment?
   c) Which food retailers are used by a specific household based on factors of their environment and characteristics of the retailers?

3. **Validation:**
   How can the municipality dataset on trashcan weight be used to validate the spatial dimensions of the waste flows?

4. **Mass of flow:**
   How can the connections in the waste chain nodes be used to follow waste mass through the household waste flow, based on the municipality dataset on trashcan weight?

1.4 SCOPE

Geographic scope

Since the data needed for this research is available for the area of Amsterdam metropolitan region, the implementation of the methods are using Amsterdam as a test area. By selecting two different building blocks the methods can be tested. The methods are tested on a selection of two different building blocks, that represent typical Amsterdam housing as well as typical Dutch housing. The method tested to a specific geographic scope can be applied to other areas, in which case additional data may be required.

Material scope

Since this thesis is building upon the REPAiR H2020 project and taking into account the municipality reports, the material in the flow is focused on food waste. Amsterdam is not collecting organic waste separately and therefore the residual trashcan is used as the node in the waste chain. Therefore this method (main method household-to-trashcan) is not necessarily only applicable to food waste, but also to general waste. Therefore, household waste is the material scope. For the method household-to-retailer, the focus lies on food waste, since most of the household waste comes from food retailers (CREM Waste Management, 2016).

Flow scope

Whereas the REPAiR project mapped the link between two nodes in the waste chain, this thesis is focusing on two more nodes. Figure 1.2 highlights the three nodes that are added within this thesis. The nodes are the connections between the collection of *non hazardous waste*, *the preparation and consumption in private households* and *the retail and markets*. Figure 1.2 is focusing on food waste, but as stated in the material scope, the methods can be applicable to household waste.

Out of scope

The development of methods is executed on two testing areas within the relatively dense urban context of the city of Amsterdam. Applying these methods to less dense or rural areas within or outside the Amsterdam metropolitan region is out of the scope of this thesis. The three nodes highlighted in figure 1.2 are the main focus of this thesis and other nodes are not considered.

1.5 RESEARCH METHODOLOGY

To answer the questions mentioned in chapter 1.3, the research methodology applied in this research is shown in figure 1.3. The following paragraphs elaborate on the different components of this research.

1. Literature review

Conversations with researchers related to the REPAiR project are held to find the gab in their research. Afterwards, the problem definition of this thesis is subtracted
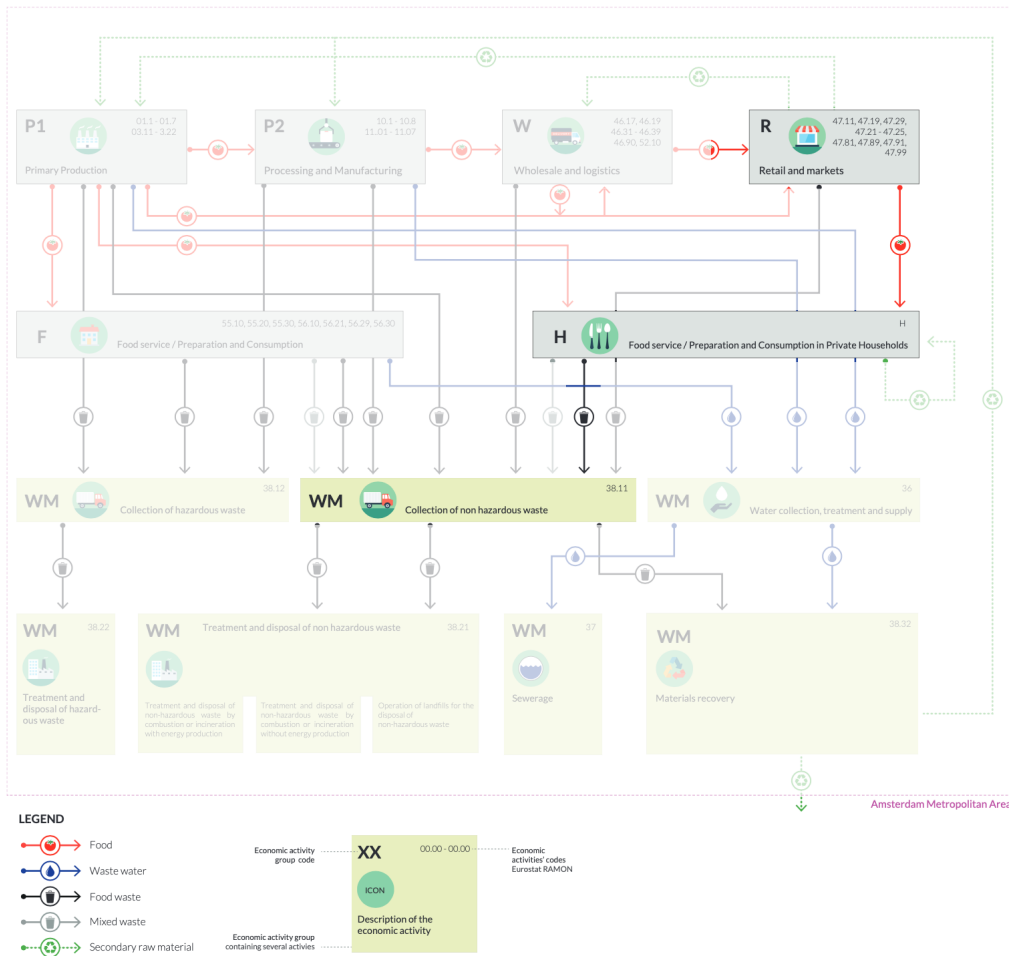
Figure 1.2: Flow scope of this research. [based on source: (Geldermans et al., 2016)]
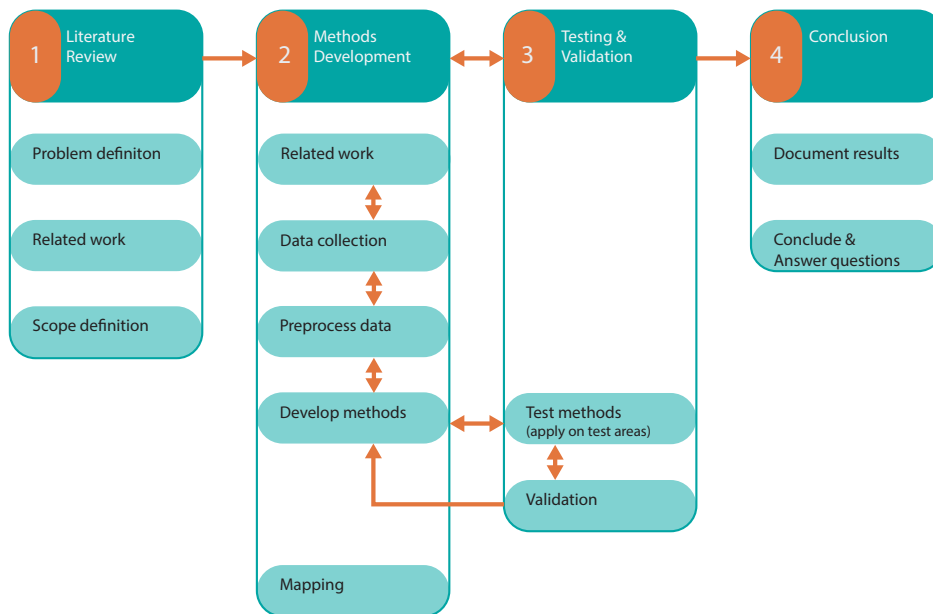
Figure 1.3: Research methodology of this thesis.

from literature by researching attempts to add a geographic dimension to waste flows. In addition to these researches, literature on food waste is added. The scope definition is formulated after finding the research gab based on the conversations as well as within the literature review.

2. Methods development

After the literature review, there is an indication of the needed data to create the methods. To know which technical approaches are needed to create the methods, an extra related work section, regarding these methods is created. The potential of the data are discussed, as well as its limitations. Since the data is collected from different sources, the datasets are preprocessed for further use.

After the collection of the data, the methods can be developed and tested in an iterative manner. Since preprocessing is dependent on the method, preprocessing is integrated in the method development. By mapping the results on the test areas, the results of the implementations are visualized.

To develop the methods, the following approach is implemented. First a hypothesis based on related work and experience of the writer of this thesis is formulated. This hypothesis is implemented in the method and afterwards tested with a validation dataset. The outcome of the testing is used to prove, disprove and improve the methods.

As shown in figure 1.3, there is a strong connection between the methods development phase and the testing & validation phase. The results of the validation are used as improvement for the development. This is an iterative process.

3. Testing & validation

Within the methods development, the results are assessed against the needs of this thesis. This is a continuous process that is needed to optimize the results. The main part of the phase is to test the method on a specified testing area. The methods are validated and optimized with the validation data.

4. Conclusion

In this phase the results are documented, the research questions answered, discussion of results elaborated and the conclusion formulated. This thesis report is a documentation of the conducted research.

1.6 READING GUIDE

The structure of this thesis is as follows:

- **CHAPTER 2** contains the related work, and therefore creates an overview of the related research of this thesis topic.

- **CHAPTER 3** contains the methodology behind the developed methods. In this chapter, four different methods are discussed.

- **CHAPTER 4** contains the execution of the methods described in Chapter 3 and the results of these methods.

- **CHAPTER 5** contains the conclusion of this thesis by answering the research questions, reflecting and discussing the outcomes and recommendations & future works.

2 | RELATED WORK

Despite the fact that food waste is not the main scope of this research, the topic is introduced and elaborates why household waste is considered in the approach of this thesis (section 2.1). Because the level of detail of this thesis is not present in research, a more general overview of flows & diagrams is given, starting with the visualization of waste flows. First, in section 2.2 Material Flow Analysis (MFA) is addressed. In section 2.3 a visualization method for flows is presented. Both of these examples are mainly focused on the size of the flows and the connections. Thirdly, section 2.4 elaborates on the addition of a spatial dimension to the previously mentioned flows visualizations. Fourthly, the research of REPAiR is addressed. Finally, a conclusion is presented in section 2.6.

2.1 FOOD WASTE

This thesis focuses on household waste household level. Food waste can be seen as a significant part of household waste CREM Waste Management (2016). Food waste is currently a widely discussed topic due to the awareness of the environmental impact and the positive consequences of reducing this type of waste at household level (Quested et al., 2013). Quested et al. (2013) also acknowledged the difficulty of comprising food waste behaviours, but provides multiple approaches to give insight, such as a survey to understand people's behaviours as well as a framework to understand prevention of food waste at home. According to the survey, wasting food that could still be eaten increases a feeling of 'guilt' among the respondents. However, the respondents give more value to eating as healthy as possible than reducing their environmental impact.

Parfitt et al. (2010) focuses on the growth of the earth population and the food waste problems that rise with this. Parfitt et al. (2010) states that households are one of the biggest contributors of food waste. In figure 2.1 different sources and disposal routes of household food and drinks in the UK are shown. This is something similar to what is shown in the REPAiR research in figure 1.2, this figure is further discussed in section 2.4.
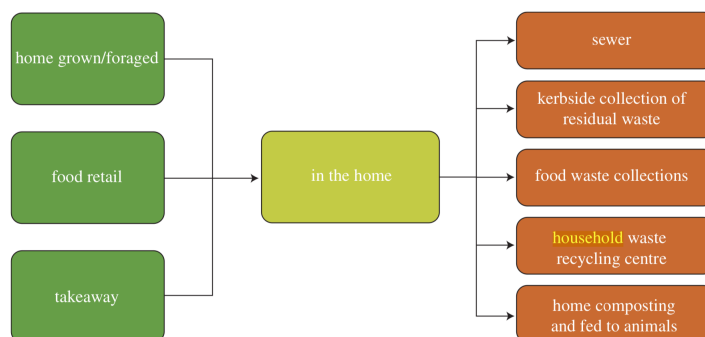
Figure 2.1: Sources and disposal routes of household food and drink in UK homes. [source: (Parfitt et al., 2010)]

Parfitt et al. (2010) is focusing on food waste in the United Kingdom, however The Netherlands has done multiple researches regarding food waste. One the or-

ganizations focusing on waste management is CREM (CREM Waste Management, 2016). This research is written in Dutch and therefore freely translated by the author of this thesis. The aim of the CREM research is to give insight in the size and composition of food waste via residual waste and organic waste in 2016. One of the major statements is the lack of organic waste collection in Amsterdam. All the organic waste is collected together with the residual waste. Therefore organic waste and residual waste are treated the same until it reaches the waste treatment plant. The rest of the research is related to the composition of the food waste, which is beyond the scope of this thesis.

2.2 MATERIAL FLOW ANALYSIS

A way of analyzing waste flows is via MFA. MFAs allows the tracking of specific material flows through a socio-economic system and provide a method to identify sources of generation and account for hidden flows and sinks that could remain unexplained, when only looking at the end of the flow (Brunner and Rechberger, 2004). Therefore the definition of MFA is: "systematic accounting of the flows and stocks of materials within a system defined in space and time. It connects the sources, the pathways and the intermediate and final sinks of a material"(Brunner and Rechberger, 2004). Thus, this analysis can give insight in the flow of waste through a city, while also identifying sinks and underlying flows that might have stayed hidden when only checking on a (neighbourhood) level. With MFA the key steps for prevention or recycling policies can be identified (Font Vivanco et al., 2012). Bulkeley and Gregson (2009) acknowledged the value by mentioning that households waste behaviour is leading when suggesting appropriate recycling schemes. To achieve this, it is important to integrate the flows and stocks' societal dimensions into the MFA (Wallsten, 2015). Significantly, the social context of material flows studies is often forgotten in the research. This has been recognized by Wallsten (2015).

2.3 SANKEY DIAGRAM

An approach to visualize a MFA is a Sankey diagram. These visualizations tend to focus on the size of waste flows. These Sankey diagrams make it possible to (1) represent visual flows related to a given functional unit or period of time and (2) the link widths represent the size of the flow (R.C. Lupton, 2017). An example is shown in figure 2.2. Sankey Diagrams are developed over 100 years ago and are created to analyse the thermal efficiency of steam engines and has since been applied to depict the energy and material balances of complex systems (Schmidt, 2008). Therefore the Sankey diagrams is 'an important aid in identifying inefficiencies and potential for savings when dealing with resources'(Schmidt, 2008). Flows are representing a movement of material or energy, but the endpoint of a flow does not always have the same meaning. Some approaches show a clear difference between stocks and processes, or places and transitions. This can mean either transformation, transportation and/or storage (Brunner and Rechberger, 2004). Between the beginning and the end of a Sankey diagram, different nodes represent the intermediate steps.

Since the waste chain is a complex system where a Sankey diagram can add information, Parizeau et al. (2015) used Sankey diagrams to illustrate the economic and material flows from food provisioning to food consumption. After illustrating the flows, multiple conclusions on food loss in the waste chain are exposed. The use of the Sankey diagram seems therefore effective for this purpose.
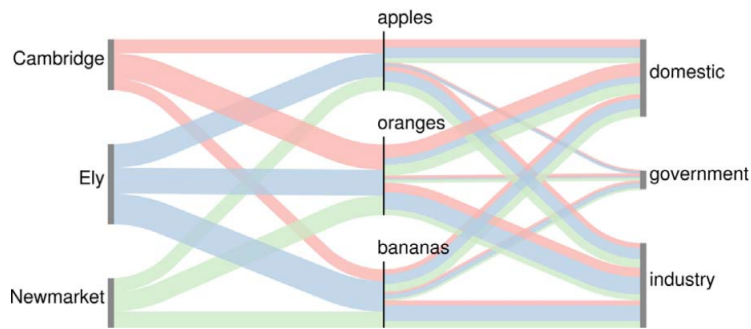
Figure 2.2: An example of a Sankey diagram. [source: (R.C. Lupton, 2017)]

2.4 ADDING A SPATIAL DIMENSION

This thesis is mapping waste flows, therefore multiple researches that attempt to add special dimensions to flow analysis and visualizations are elaborated in this section.

2.4.1 SAMFA

Adding a spatial dimension to a MFA has been done before. For instance, Roy et al. (2015) developed a model to measure flows of construction material at a spatial level, the Spatial Allocation of Material Flow Analysis (SAMFA). The SAMFA model draws on the concepts of urban metabolism and links these with an ability to visually represent the implications of spatially and temporally differentiated scenarios related to the accumulation of construction material flows and associated energy use in the housing sector. The spatial aspect of the SAMFA model is mostly on an electoral divisions scale. The model, that consists of three pieces, is able to estimate and spatially allocate construction material flows at the electoral divisions level. MFA mostly shows a top down approach, which makes them most reliable on national level. The SAMFA shows a clear bottom up approach, which is more reliable on a regional level, this can also be considered more useful than down-scaling. However, Roy et al. (2015) only shows values for county level.

2.4.2 SMFA

Subsequently, Font Vivanco et al. (2012) developed a model for Spatial Material Flow Analysis (SMFA), for building waste with transport intensity indexes. The spatial dimension of this research is provided by tracking all actors based on their longitude latitude coordinate. Afterwards a network analysis is performed using the road network to calculate the pair-wise distance for the optimal road route between all the agents that were connected in the flow.

2.4.3 Waste flow mapping

Another possibility is Waste Flow Mapping (WFM). Kurdve et al. (2015) used a WFM method which is synthesized to be used by both waste management researchers and practitioners. The method relies on proven tools and methods to analyze the current state and find improvement potentials with regard to material losses and inefficiencies in the handling of materials and waste. Something that sounds familiar with the earlier mentioned terms/methods. Kurdve et al. (2015) used an implementation of seven steps for the WFM method (figure 2.3).
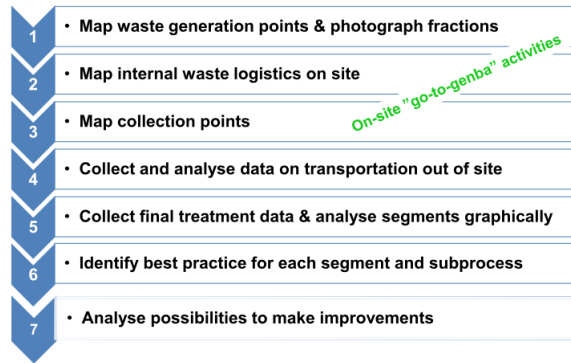
Figure 2.3: The wasteflow mapping (WFM) method implemented in seven steps. [source: (Kurdve et al., 2015)]

2.5 REPAIR PROJECT

The report 'D3.3 Process model for the two pilot cases' from the REPAiR project gives a start of visualizing waste flows in the Amsterdam Metropolitan region (Geldermans et al., 2016). This visualization is shown in figure 2.4, and looks similar to what is mentioned in the MFA explanation earlier, it shows different nodes and links. In this figure the flow of food is also visualized before it is classified as waste. Figure 2.4 is shown before in section 1.4 to visualize the scope of this thesis.
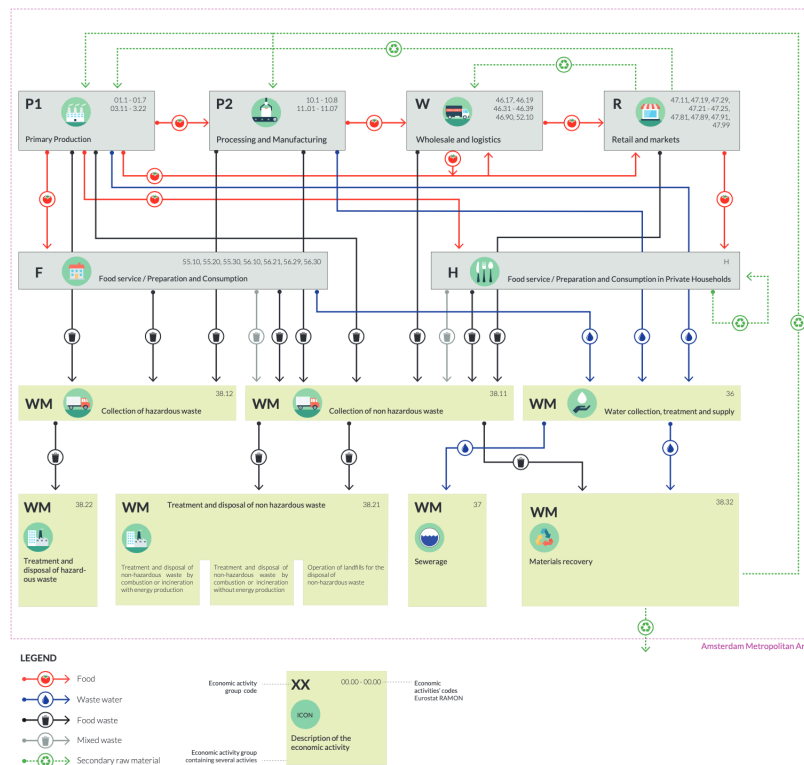
Figure 2.4: System diagram of activities and flows. [source: (Geldermans et al., 2016)]

The REPAiR program used a Spatial Sankey Diagram to map waste flows. By knowing the size of the food waste for every neighbourhood, the underlying network and the destination of the waste treatment plant for every neighbourhood, the diagram can be made. In figure 2.5, the food waste going to specific treatment plans is mapped. Because it is a Sankey diagram, a thicker line width means more waste coming from this area.
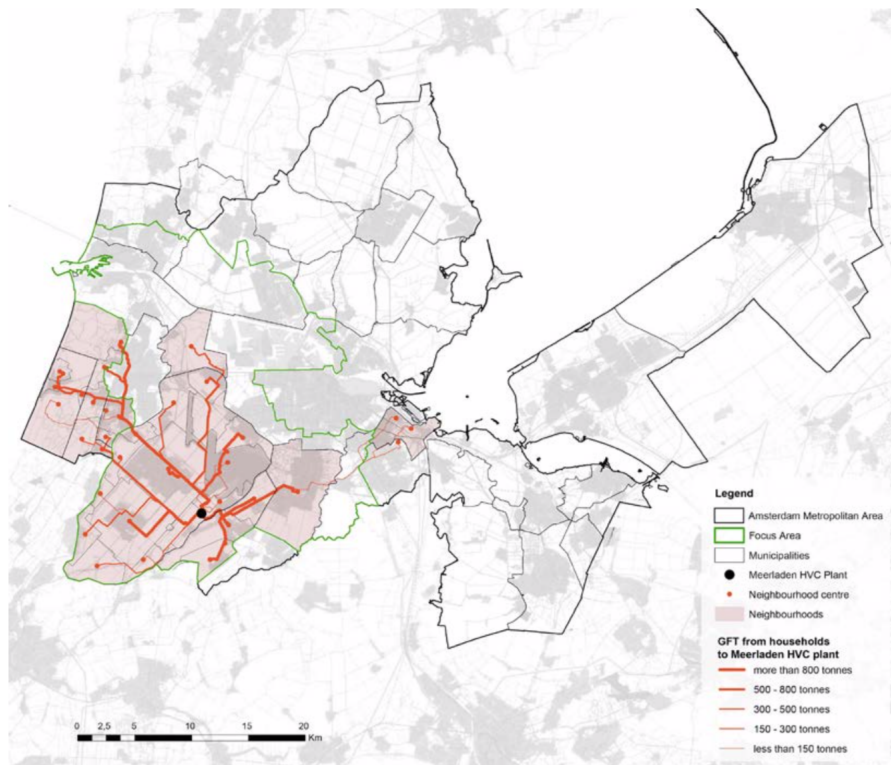
Figure 2.5: The material flows collected as organic waste in the composting facility in kg per year, per neighbourhood and its locations in the Metropolitan Area. [source: (Geldermans et al., 2016)]

2.6 CONCLUSION

Giving insight in waste behaviour is something that is missing in literature, the importance of food waste research is mentioned in section 2.1. This section elaborates the different sources and disposals of households, therefore is figure 2.1 similar to nodes treated in this thesis. In this research food retail is the source of materials which become waste at households, since this is the biggest input of materials in the households. The main disposal output is the collection of residual waste, because the lacking of organic waste collection in Amsterdam. The other disposal outputs, such as via the sewer or home composting can be seen as negligible due to small amounts these outputs represent.

Multiple spatial visualization methods are shown which are capable of visualizing waste flows. SMFA visualize sinks and flows of the material, while a Spatial Sankey Diagram focusses on the size of the flows.

Figure 2.3 creates a structure to implement the WFM method. The steps *map waste generation points*, *map internal waste logistics on site* and *map collection points* are within the scope of this thesis. The term WFM is chosen to be the title of this thesis, because it gives the clearest explanation of the applied method and ideal result.

Since the goal of this thesis is to map nodes of the waste chain, it gives an extra dimension to the unexpected hidden flows and sinks that Bulkeley and Gregson (2009) mention. Household level is touched and therefore results in a better understanding of human waste behaviour. When more insight in the behaviour in households is obtained, specific materials can be followed through different steps of the waste chain. The result of usage of different materials in different food retailers can be followed until the composition of waste in the trashcans.

3 | METHODOLOGY

In figure 3.1 the overview of the coming chapters is presented. The focus are the three different methods that are developed, with the main focus on household to trashcan and household to retailer. The latter is creating the input for the afterwards described mass of the flow.

In this chapter the methodology of the different methods are presented. First in section 3.1 the description of the two main methods is presented. Second, in section 3.2 the test area is described. The previously mentioned section are identical for both of the main methods. Where after the sections are specified for one of the two main methods. This is elaborated in section 3.3 and section 3.4. The development of the additional method is presented in section 3.5. The methods are applied in chapter 4.
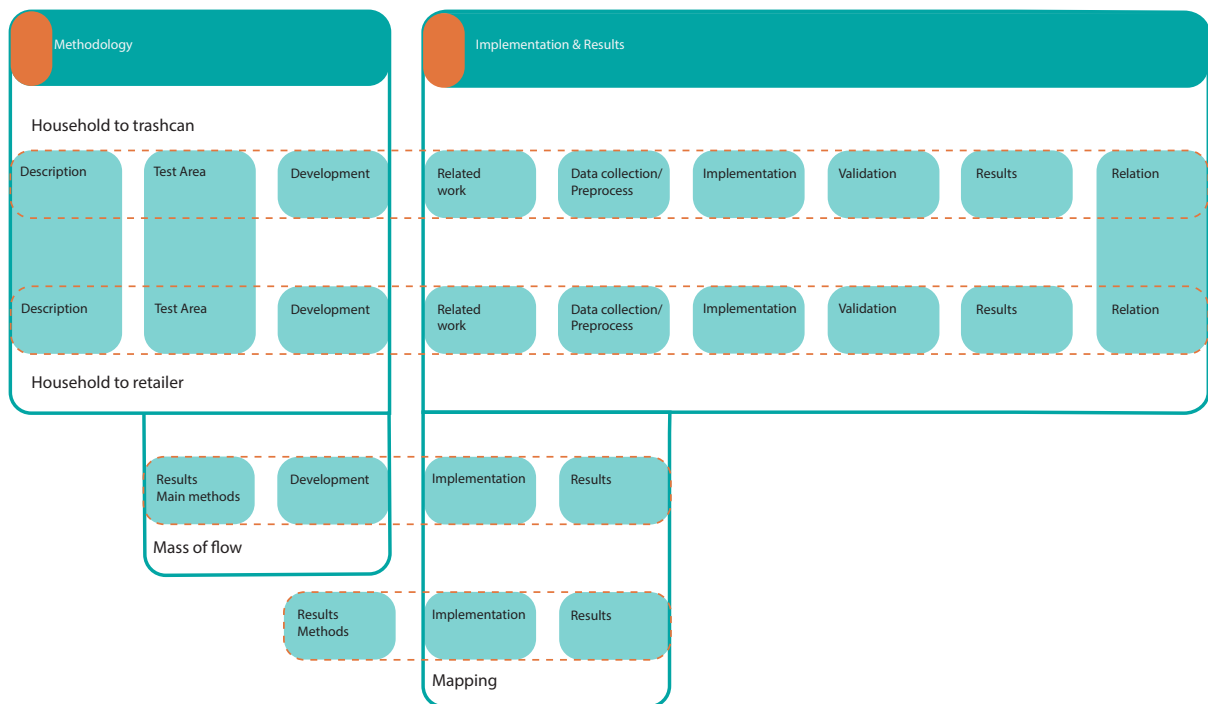
Figure 3.1: Scheme of this chapter and chapter 4, addressing the different sections.

3.1 DESCRIPTION OF DEVELOPED METHODS

For the development of the methods, the system diagram of REPAiR is used. By simplifying the specific flow that is within the scope of this thesis, figure 3.2 is created. As concluded in section 2.6, the main input of household waste to a household are the food retailers and the main output of household waste from a household is the kerbside collection of residual waste and food waste collections (as visualized in figure 2.1). Since the municipality of Amsterdam is not collecting food waste sep-

arately (Municipality Amsterdam, 2015a), the major output is kerbside collection of residual waste.

The connections of the nodes in the waste chain are as displayed in figure 3.2. This figure visualizes the connections between the nodes that are within the scope of this thesis (section 1.4), the figure is a simplification of figure 1.2. Two different connections and one additional method is developed. From the households, the trashcans which are used by this household are found (1). From the households the used retailers are selected (2). The outcome from the two different connections are used to give an indication of the mass of the flow over the previously discussed relations (3). The mapping is visualizing the results (4).
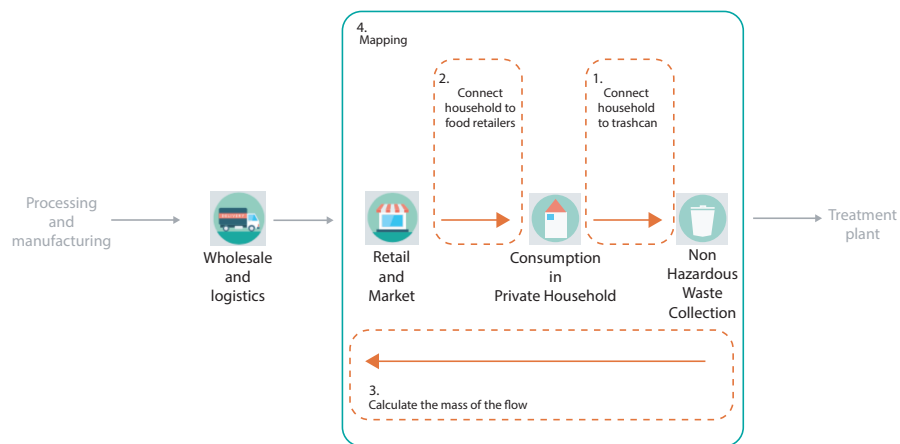
Figure 3.2: Simplification of figure 2.4, showing which connection is considered which method.

The two main developed connections are numbers 1 & 2. Important for this thesis is the difference between the two. To start with, in the connection household to trashcan, the material in the flow is classified as waste. While for the connection household to retailer, the material is up until then not classified as waste. This turning point is the household itself. This distinction is important for the development of the fundamental methods. Since the connections are both highly dependent on human behaviour, both relations have different objectives from the residents. Where method household-to-trashcan is a prerequisite to transport waste, the execution is carried out as less time consuming as possible and always starting at the household itself. According to the model of Hoogendoorn and Bovy (2004) mandatory trips are mostly explained by minimizing the trip (creating the shortest path). Method household-to-retailer, is finding the retailer as contributor of food waste for a household. Notwithstanding, this method is still a mandatory trip, but people are willing to travel further for specific retailers such as cheap supermarkets or supermarkets with better quality food (Dunkley et al., 2004). This creates a bigger uncertainty, since more factors are involved. The distinction of the methods is important when the development takes place.

Considering the nodes are connected in the first two methods, (3) is giving a mass to this flow. Data about the mass of the waste in a particular trashcan is known, thanks to a municipality dataset (section 4.2.1 is elaborating more on this dataset). The two main methods are used to create the connection between the nodes, the method: mass of the flow is defining the mass (3) of every link, the mass is mapped in the umbrella method (4).

(a) Test area Nassaukade

(b) Test area Prinshendriklaan

Figure 3.3: Maps of the two test areas

3.2 TEST AREA

To be able to apply the methods from figure 3.2 a test area is specified. Since both main methods are having households as starting point, the development is executed on different types of households. When an idea is implemented in the methods, the results are projected on both of the areas. In this way it can be excluded that the result is influenced by the specific area. The areas are specified to represent the most common house typologies in the Netherlands, both are situated in Amsterdam (due to the municipality dataset). The most households in the Netherlands are one family homes, typically with a back and a front yard, three levels and in a row (rowhouses) (CBS, 2019). The most common households in Amsterdam are multiple family homes, without a front yard, mostly also without a back yard, sometimes with a balcony (CBS, 2019).

Nassaukade

This test area is chosen because the author of this thesis lived here, short enough not to be biased in the research, but long enough to be able to better validate the outcomes. The area is situated in the Old-west area in Amsterdam. The street visualized in figure 3.4a, is rather busy and difficult to cross. Since this area is located in the centre of the city, multiple hotspots are located around this building block. This area is representing a typical Amsterdam living environment (figure 3.3a).

(a) Test area Nassaukade

(b) Test area Prinshendriklaan
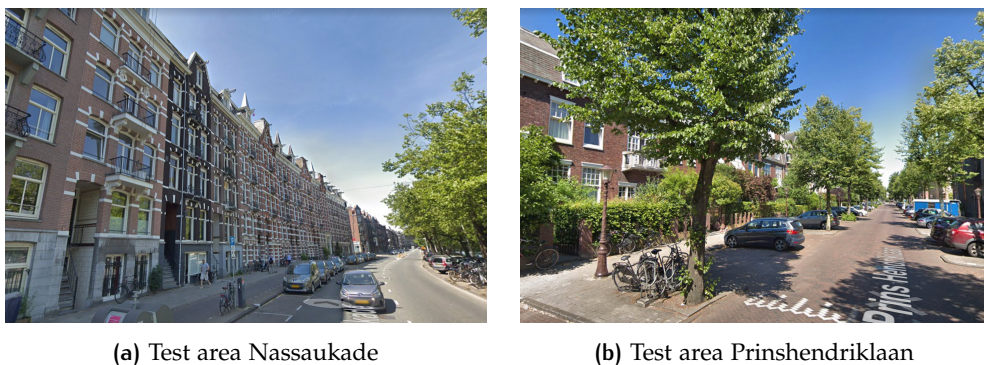
Figure 3.4: Streetview of the two test areas [source: (Google Maps, 2019)]

Prinshendriklaan

This test area is chosen because it is one of the few places in Amsterdam where single family homes are present. The area is situated in the Old-south area in Amsterdam. The street visualized in figure 3.4b, is spacious and looks like a Dutch

village. Since this area is located relatively far from the centre, only one hotspot is located close to this building block. This area is representing a typical Dutch living environment (figure 3.3b).

	Test Area Nassaukdade	**Test Area Prinshen-driklaan**
Dimensions	80m*180m	180m*65m
Area	2550 m^2	2060 m^2
Typology type	Multiple family homes	Family homes
Number of households	374	55
Average floorsize	58 m^2	235 m^2
Building block shape	Elongated	Elongated
Location close trash-cans	Some without crossing the street	Only after crossing the street
Hotspots around	Multiple	One
Travel intensity adjacent roads	11000-15000 daily objects [a]	not covered therefore lower than 4000
Represents	Typical Amsterdam house	Typical Dutch house
Neighbourhood name	Da Costa buurt	Willemsparkbuurt

Table 3.1: Test area comparison

[a] source: Municipality Amsterdam (2019)

Comparison

In table 3.1 the comparison is made between the two areas, the general aspects of both building blocks are written down. The number of households in the the Nassaukade area are significantly higher than the households in the Prinshendriklaan, but this was expected due to the different typology the location represents. At the Prinshendriklaan non of the trashcans are reachable without crossing the street, at the Nassaukade some on the trashcans are reachable without crossing the street. Since the travel intensity of the adjacent roads of the Nassaukade is higher than the Prinshendriklaan, pedestrians probably have to wait longer to be able to cross the road. Both of the building blocks have an elongated shape, but for both of the blocks the trashcans are located all around the block.

3.3 DEVELOPMENT: HOUSEHOLD TO TRASHCAN

The purpose of the first method household-to-trashcan is described in this section. First a summary of what the method should contain is presented. Where after the needed input to retrieve the needed output is discussed.

The start of the method is to consider what the purpose of the method is. The development of the method is executed in such a way, that the method is as accurate as possible. Improvements which speed up processing time are only implemented if they do not influence the accuracy of the method. Related works are used to implement specific technical approaches (described in the next chapter), but since a method like method:household to trashcan is not developed before, retrieving the most accurate result, with available information, is the main priority.

As mentioned in section 3.1, this method focuses on the movement of waste from the household to the collective waste collection. Borgers and Timmermans (1986) states that 52 percent of pedestrian routes can be explained by only minimizing the length of the routes from starting point to destination. The goal is to find the used trashcan to be able to connect the nodes, this connected trashcan can differ from the

trashcan that is the closest, or the one that the municipality placed for this specific household.

Based on the author's own behaviour, the behaviour analysis of people close to the author and the related work (such as Borgers and Timmermans (1986)) the following hypothesis is stated for this method:

Residents of a household choose the closest residual waste trashcan to throw away their household waste, but are willing to walk further if the walk is in the direction of a city hotspot.

Input

To be able to test the hypothesis stated above, the certain input datasets are needed. The inputs are representing either the nodes of the waste chain, or influences mentioned in the hypothesis. Specific datasets which are needed to perform the technical methods are discussed in section 4.1.1. First a dataset containing all the different households in the test areas are needed. The households are representing one of the nodes in the waste chain. Secondly, a dataset with the trashcans is needed, the municipality has open-data containing this information. Finally, a procedure to locate hotspots within the urban context needs to be developed.

The hypothesis suggests an exception to picking the closest trashcan. A part in the processing requires to force a path in the direction of a hotspot. Additional paths are created to other locations of waste containers. The additional results are different paths to clusters of containers where the other types are present. Since this thesis is focused, but not limited to household waste, the paths to other waste types are also created. The residents want to minimize the paths, they are likely to combine the residual waste with the location of the e.g.plastic waste (table 3.2). Therefore they are walking a longer distance, but minimize the number of trips. Due to the missing data about the number of times and other type than residual is used, the paths are added as additional paths.

Types of trashcans	Residual	Plastic	Paper	Glass

Table 3.2: Amsterdam waste container types [Source: (Municipality Amsterdam, 2015a)]

Goal

When the inputs are processed, the outcome is a path from the household to the used trashcan. All households are containing a maximum of four paths (when all the waste types are at separate locations) and a minimum of one path (when all the waste types are present at one location).

Validation

Finally, the hypothesis is validated. The municipality of Amsterdam delivered a dataset containing the mass of the trashcan in kg per year. More households using a trashcan results in more kg of waste per year, a strong positive correlation is expected between the kg per year and the sum of the households using this trashcan. By changing the impact of the hotspot, the influence of this input is shown in the correlation. It is expected that the correlation is higher when the hotspot influence is applied.

3.4 DEVELOPMENT: HOUSEHOLD TO RETAILERS

The development of the second method household-to-retailer is described in this section. This section elaborates on the inputs and output of this method.

The start of the method is to consider what the purpose of the method is. The development of the method is executed in such a way, that the method is as accurate as possible. Improvements which speed up processing time are only implemented if they do not influence the accuracy of the method. Related works are used to implement specific technical approaches (described in the next chapter), but since a method like method household-to-retailer is not developed before, retrieving the most accurate result, with available information, is the main priority.

Based on the author's own behaviour, the behaviour analysis of people close to the author and the related work (such as: Benoit and Clarke (1997)) the following hypothesis is stated for this method:

Food stores sell products that afterwards turn into waste and therefore contribute to house-hold waste. Bigger and closer food stores have a bigger contribution.

Input

To be able to test the hypothesis stated above, the following input datasets are needed. First all the households in the test area (same as implemented in the previous method). Secondly, a dataset containing all the stores selling food. Thirdly, the floor size of the food stores. Finally, the hotspot of the city.

Goal

Since a household is not only using one specific food store, values of the expected contribution of different stores are generated. This importance value is calculated for all stores which are considered for this specific household. Two different datasets are created, first, a list of retailers to get an indication of which retailer is influencing the specific household. Second, a list of retailers with their importance value for a specific household.

Validation

Due to missing data, no validation method is developed for this method. If data was available on for instance the revenue for a store, or the mass of the produced waste, a similar correlation could be executed as described in the previous method. A higher total of importance values (form all the households around the store) should have resulted in a positive correlation with the generated revenue or waste mass. For now, this hypothesis stays invalidated.

3.5 DEVELOPMENT: MASS OF THE FLOW

Since the ultimate goal of mapping waste flows is to be able to follow waste through the waste chain, a start is made by giving the connections researched in this thesis a mass. Due to the availability of the mass of the waste collected in the trashcans, the ability rises to distribute this over the connections. This section elaborates on this matter.

Input

To be able to assign the mass to the connections, the following input is considered. First, the outputs of the previously mentioned methods is used as input. By knowing which waste is coming from where, the connections can be added. Subsequently, the most important input of this development is the kilograms of waste in a specific waste container.

Goal

The output of this method is the kilograms of waste going through the connected nodes by showing household waste mass and retailer contribution to this mass.

4

IMPLEMENTATION & RESULTS

This chapter describes the implementation of the methods from chapter 3. The structure of figure 3.1 is continued. The two main methods are elaborated in a parallel way, due to overlap in some of the sections.

First section 4.1.1 elaborates the related work of the technical aspects of the method household-to-trashcan, afterwards 4.1.2 discusses the technical aspects of the second method. Section 4.2 discusses for the two main methods how the input data is gathered, from which source it is retrieved and how the preprocessing is completed. Afterwards, section 4.3 elaborates on the implementation of the development from sections 3.3 & 3.4. Fourthly, section 4.4 presents the validation of the outcome of the implementation. When the outcome is optimized and validated the results are visualized in section 4.5. To conclude the main methods, section 4.6 provides description of the relation of the outcome with the stated problem definition.

The methods building on the outcomes of the previously mentioned methods, are described afterwards. First section 4.7 elaborates on the implementation of the mass of the flow. Followed by section 4.8 presenting the results. Finally, section 4.9 shows the implementation of the mapping, afterwards section 4.10 presenting the some mapping results

4.1 RELATED WORK

In this section, the related work needed to implement the methodology is discussed. First, the related work regarding the method household-to-trashcan is elaborated, thereafter the work needed for the method household-to-retailer is addressed.

4.1.1 Household to trashcan

In this section an overview of the literature for the applied method household-to-trashcan is given. Waste flow mapping on household level is not featured in the current literature, therefore, a wider literature research for not directly related topics is exectued. Since the connection between the nodes described in 1.4 is mainly based on choices of people, a review of literature in pedestrian behaviour is done. Thereafter, different methods to create routes between start- and end-points are assessed. Finally, the needed inputs for the implementation concluding from the related work is given.

Pedestrian behaviour

Research in behaviour of pedestrians can have various applications, for example congestion in pedestrian traffic (Hoogendoorn and Daamen, 2005) or approaches to validate designs (Daamen and Hoogendoorn, 2007). Pedestrian behaviour is challenging in theoretical and practical problems, with pedestrian route choice and activity scheduling as the most interesting and difficult ones (Hoogendoorn and Bovy, 2004). For this thesis the behaviour and movement between a known start point and end point is the most important, since these are influencing the shortest path to the nodes in the waste chain.

Route choice

When a pedestrian route choice model with multiple intermediate steps is created, only minimizing the distance between the steps is predicting 52 percent of the observations correctly (Borgers and Timmermans, 1986). Pedestrians have endless routes to choose from, not even taken into account that there are multiple possibilities within the same route. 'Key elements in predicting pedestrian behaviour are activity scheduling (choice to execute a particular activity, the order in which activities are performed, the location where to perform an activity), route choice between activities in the two-dimensional continuous space, and multi-directional walking behaviour' (Hoogendoorn and Bovy, 2017). Hoogendoorn and Bovy (2017) creates a framework for pedestrian behaviour. When the walker is about to perform a prior activity, the theory states that, pedestrians make a simultaneous route-choice and activity schedule decision, therefore minimizing subjective uncertainty. This is dependent on external factors (e.g.presence of obstacles, stimulation of the environment) and internal factors (e.g.time-pressure and attitude of the walker). Including pedestrian traffic conditions, this results in the pedestrian walking activity. In turn, pedestrian walking behaviour is the result of predicted cost minimization at the operational level, where pedestrians take into account costs due to digress from their intended route and frequent or obvious accelerations and delays (Hoogendoorn and Bovy, 2017). Figure 4.1 shows the pedestrian behaviour framework by Hoogendoorn and Bovy (2017), visualizing inputs of pedestrian behaviour as discussed above.
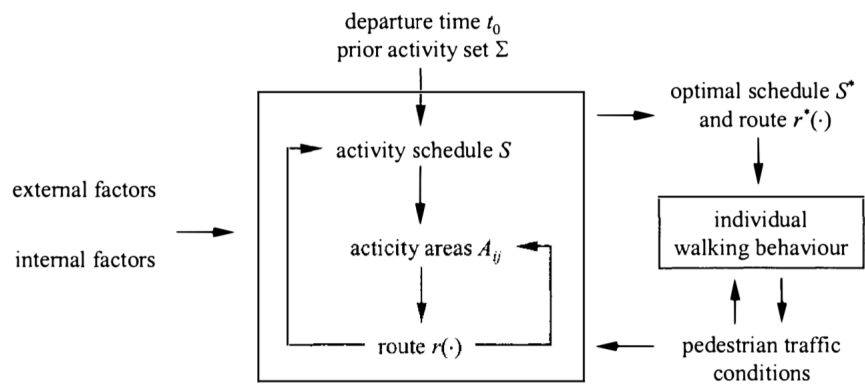
Figure 4.1: Pedestrian behaviour framework. [source: (Hoogendoorn and Bovy, 2017)]

Distance measuring methods

As described in the previous section, one of the most important factors to predict route choice for pedestrian movement, is route minimization. There are various approaches to measure a route length, Ammerlaan (2016) wrote a Master thesis containing different methods for distance measuring related to waste routes. All of these methods are executed on the same testing area and the results are compared. In figure 4.2 the difference between the methods is visualized, where Euclidean distance Measure (EM) describing the direct distance in Two Dimensional (2D) space. This approach is always the shortest, because it is not influenced by any environmental factor. Traditional Network Method (TNM) is using an underlying network to calculate the distance values by using an OSM network. Thereafter the LCP is using an underlying raster to create a more realistic walking pattern, based on the BGT. Finally, Adapted Network Method (ANM) is a combination of the TNM and the LCP. Vreugdenhil et al. (2016) concluded that all methods except for the EM are a realistic representation of reality.
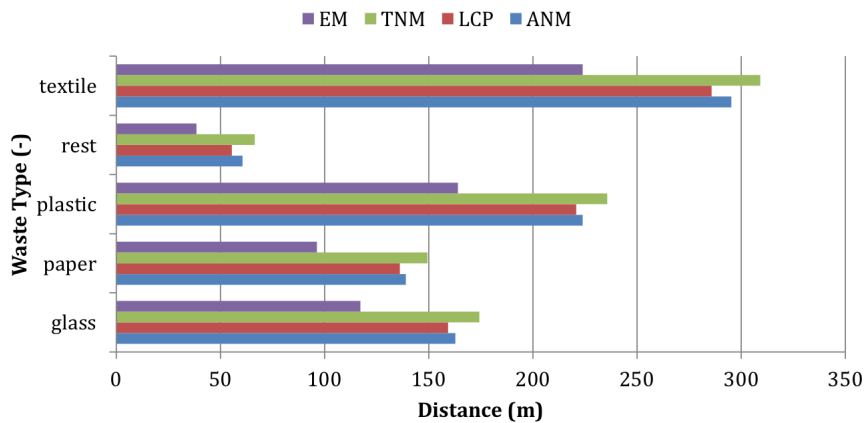
Figure 4.2: The average walk distances calculated per method for each waste type for the neighbourhood IJburg in Amsterdam. [source: (Vreugdenhil et al., 2016)]

To find which one of the methods is the most suitable for this thesis, the following aspects are taken into account. The EM is unrealistic and is therefore not considered, pedestrians are following walkable spaces and this method is neglecting these. The TNM is using an underlying network and therefore the result is not realistic as a pedestrian path, because pedestrians do not necessarily follow a road network. Another major drawback is that network analysis is only available on the roads, therefore the start of the analysis is always a node in the network. To find the node corresponding to a household, the closest vertex is found. This vertex can be behind the household and accordingly unrealistically connected. Therefore is concluded that the LCP is the best option. It uses built environment data from the BGT and it allows shortcuts if they make the route significantly faster, thus results in a more realistic pedestrian path to measure.

Least cost path analysis

The following paragraphs are giving a better understanding of the LCPA and how it is preformed. First a general application of the analysis is presented, afterwards a different application of the LCPA is discussed.

General LCPA application

LCPA is implemented in literature to find the best path of the linear feature given the topography and start & end-point (Collischonn and Pilar, 2000). The goal of the LCPA is to find the path from a start to end-point, with the least cost. When the LCPA is applied to create a route through the mountains, the least cost represents the shortest path with the least slope. One of the inputs for the LCPA is a Digital Terrain Model (DTM) raster, a DTM is a model of the earth ground surface, having a height value for every cell of the raster. This digital representation can be used as an input to calculate the slope for every pixel. The cost path raster is a combination of the costs for every pixel and the slope for every pixel. This raster can be used to draw the least-cost-path from start to end-point.

To draw this path, a combination of different algorithms is used (e.g. Dijkstra). Traditionally, Dijkstra is used to solve traditional network problems. In figure 4.3(a) a traditional method is visualized. Different nodes are linked with different links containing different weights. Figure 4.3(b) can be seen as a network representation of a raster data structure where the centres of the cells are the nodes and the links are the possible connections between them. The nodes are containing a specific cost value, coming from the previously mentioned combination of the slope and the DTM. The algorithm is able to draw a path through the cost raster, going from

start to end, where the cost is the lowest. In an example of a path in the mountains, the path with the lowest slope and a minimized distance is found (figure 4.4).
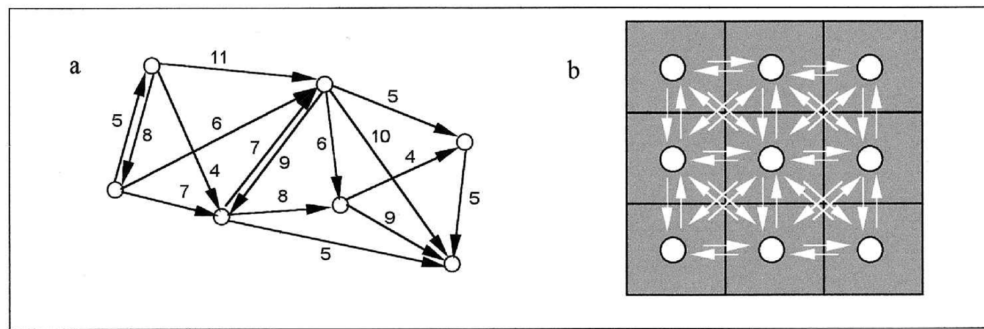
Figure 4.3: (a) Classical network with weighted links connecting nodes; (b) Network representation of raster data structure with eight links entering and eight links leaving each cell. [source: (Collischonn and Pilar, 2000)]
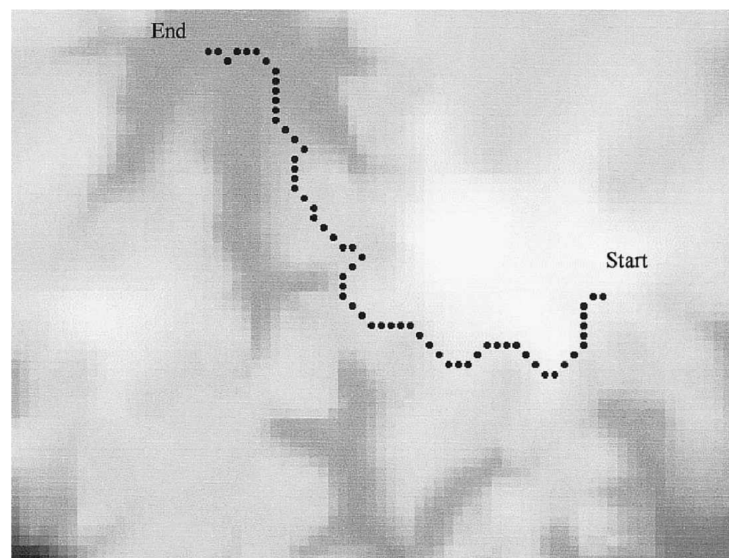
Figure 4.4: Example of a path through a cost raster. [source: (Collischonn and Pilar, 2000)]

Other LCPA applications

LCPA is not something limited to planning routes for roads and canals as shown in the application by Collischonn and Pilar (2000). Multiple sources use the analysis to predict animal behaviour, based on other factors together with the regular factors like the slope. Vignieri (2005) found a strong correlation between the LCP on a raster minimizing elevation and maximizing riparian forest cover, with the gene flow in the Pacific jumping mouse. Where the LCP (in landscape genetics), is now representing the mobility of the species (Wang et al., 2009). The cost raster created in the paper of Wang et al. (2009) is not containing slope values, but different habitat types, such as: grassland, chaparral, or oak woodland. The result is the prediction of movement of the California tiger salamander.

Not only animal behaviour is predicted, but also human behaviour in ancient times. Herzog (2013) states that 'In prehistory, paths are often the result of centuries of experience with movement within a certain landscape. According to the implicit model of least-cost path analysis, the routes are improved gradually so that finally

optimal paths result.' However, he mentioned that the input raster is mostly fully based on the slope cost.

Input for technical method

According to Ammerlaan (2016) the LCPA has never been applied to pedestrian behaviour within microscale built environment. After research of the author of this thesis, this can be validated. This seems like a missed opportunity based on the successful application on animal behaviour mentioned above. Herzog (2013) is already applying the method on human behaviour, but on a macroscale environment and only by using the slope factor. For this thesis the approach from Ammerlaan (2016) is used to predict pedestrian walking paths, based on the promising results from this paper and because it takes into account environmental factors deriving from the BGT.

In figure 4.5 the wanted output of this method is visualized and number 1 until 4 represent the different fundamental inputs. The output is the goal of this method, the walking route from household the used residual trashcan. (1) and (2) represent the earlier mentioned nodes in the waste chain. (3) represents the hotspots. As concluding from this section the LCPA is used to predict pedestrian walking paths. Therefore (4) is representing the raster needed to implement this analysis.

The data representing these inputs are discussed in the section 4.2.1. Since the inputs are already determined, data is found to represent the elaborated inputs as suitable as possible. Therefore source of the data is influenced by the process of this method. The goal is known and the needed inputs to achieve this goal are discussed in this section. Accordingly, the data selection is influenced by the method of this research.
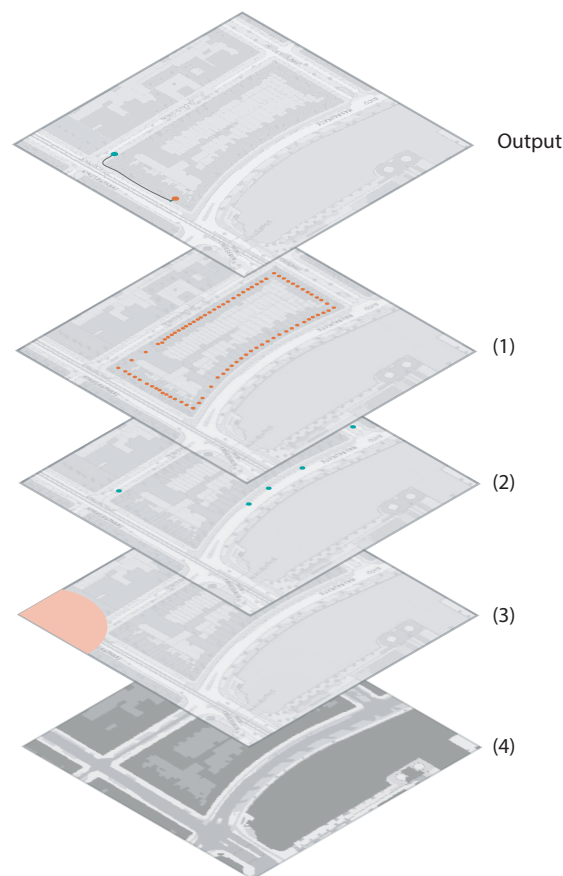
Figure 4.5: Overview of the goal and the needed inputs for method household-to-trashcan.

4.1.2 Household to retailer

In this section an overview of the literature for the applied method household-to-retailer is given. Because this method is based on personal preference, influential factors that are known for affecting retail preference are discussed. Afterwards an approach of determining the distance of the stores is elaborated. Finally, the needed inputs for the implementation concluding from the related work are given.

Location selection for retail

Chen and Tsai (2016) creates a data mining framework based on rough theory to improve location selection decisions. For this research the location selection is inverted, by looking at the attractiveness of a store for a specific household instead of looking at the households that could be attracted by the store. Chen and Tsai (2016); Benoit and Clarke (1997) state that a higher population size and a higher population growth rate have positive influence on the revenue of a store. Since the population size and growth do not relate to specific households these theories are neglected. However, the store size is declared as a predictor of the store revenue, a bigger floor size results in a higher revenue. This can be implemented in the method household-to-retailer. Benoit and Clarke (1997) states a possible solution to identify used stores based on the distance to the household: 'One solution may be to identify primary, secondary or tertiary buffers and allocate more revenue from the former than from the latter'. Meaning that the considered retailers for a household are filtered in three different distances and the closest distance is considered having the most influence, the latter the least. The availability of a parking space is also positively correlated with the store revenue. Since this thesis has Amsterdam as testing location, a parking space is considered not influential, since the bike is the most important mode in Amsterdam (section 4.3.2).

Isochrone area

To determine how far the retailers are from the household an isochrone area is created. An isochrone is used instead of a buffer because it represents a more accurate travel time based approach. Isochrone is defined as: 'a line joining a set of points at equal travel time from a specified location' (O'Sullivan et al., 2000). And an isochrone area refers to 'an area of all points contained within an isochrone which are reachable in the specified time or less'.

Store size together with the distance from the household to the store in question creates an influence factor of how much the store is expected to be used by this residence. By using this approach, further stores are possibly neglected. This is in contrast with the announcement of Dunkley et al. (2004); Parizeau et al. (2015), who state that people are willing to bypass the closest store if the food in the further store is better quality. Therefore a method is developed, where closer stores have more influence, whilst not excluding further stores from their possible impact.

Input for technical method

In figure 4.6 the wanted output of this method is visualized and number 1 until 5 represent the different inputs. The output is the goal of this method: find the retailers which are most likely to be used and enrich them with an importance value. (1) & (2) represent the different nodes in the waste chain. (3) is extra data for the retailers to add a floor size. (4) is the same hotspot as used in the previous method. Finally, the last input (5) is the street network to be able to create the three different isochrone areas.

The data representing these inputs are discussed in the section 4.2.2. Since the inputs are already determined, data is found to represent the elaborated inputs as suitable as possible. Therefore the data is influenced by the process of this method.

The goal is known and the needed inputs to achieve this goal are discussed in this section. Accordingly, the data is influenced by the process of this research.
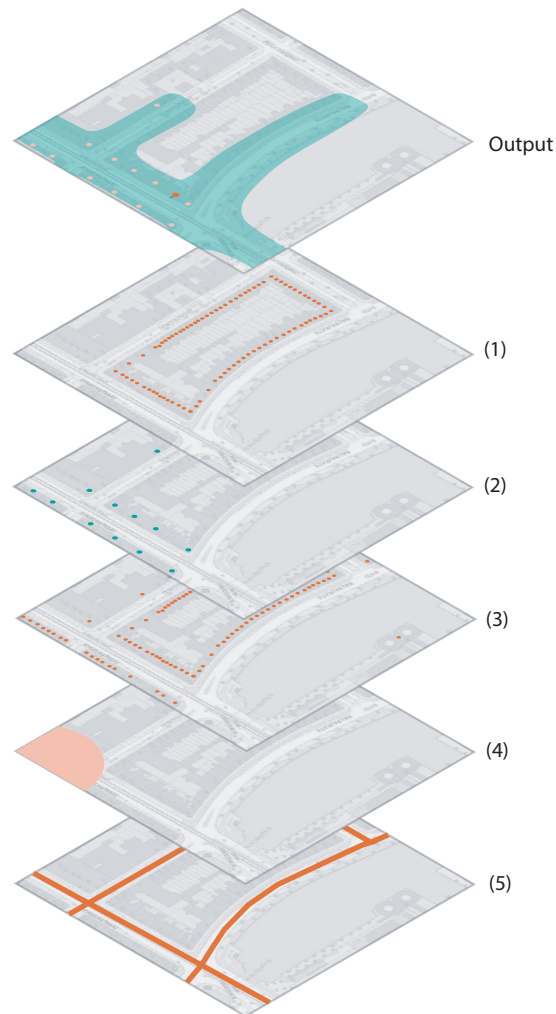
Figure 4.6: Overview of the goal and the needed inputs for method household-to-retailer.

4.2 DATA COLLECTION/PREPROCESSING

To structure this section, every input visualized in figure 4.5 & 4.6 is analyzed one by one. For every input the sources are described afterwards, the preprocessing steps are elaborated based on a flowchart created for every processing. The flowchart assesses first the source, secondly the transformations and finally the outcome of the preprocessing (which is the input for the method). The preprocessing is executed with Feature Manipulation Engine (FME), unless another tool is mentioned. The Coordinate Reference System (CRS) used in this thesis is (since the application is entirely in The Netherlands) Rijksdriehoek (RD) coordinates new.

4.2.1 Household to trashcan

To make the implementation of the method household-to-trashcan possible, the inputs from figure 4.5 are required. The inputs one until four are treated respectively.

Some of the transformations are rather simple to implement other ones require more literature research or improvement within the transformation.

Input 1: preprocess households

The purpose of this preprocessing is to achieve accommodation objects within the testing area. These objects represent the households required for the implementation. Since households in the same building footprint present the same used trashcans, the *id* of the footprint is added to every household. Figure 4.7 shows the flowchart for this preprocessing, the numbering in the figure is the same as used below.

Figure 4.7: Overview of the needed steps to achieve input 1.

Source

The sources to achieve the required result, are accommodation objects and footprints, both retrieved from the Basis Administratie Gebouwen (BAG). The BAG is maintained by the Dutch Cadastre and represents all the buildings in the Netherlands. The BAG consists of three different parts, first are the accommodation objects (points), second the addresses of these objects (points) and last the footprint of the buildings containing these points (polygons). The open data portal of the Netherlands (Publieke Dienstverlening Op de Kaart [Public Service On the Map] (PDOK)) provides a Web Feature Service (WFS) for the BAG data, due to the limitation of 5000 features, this service is not applicable for this thesis. Therefore an *.mxd* is received containing the three BAG parts for the entire Netherlands. In the view of the fact that this is a massive file together with the unfavorable that *.mxd* files are only to open with ArcGIS, only the accommodation objects are retrieved in this matter. The building footprints are obtained via the main supervisor and are fully preprocessed for the Amsterdam region in *.shp* extension.

Transformation 1: extract

As mentioned, the first transformation aims to extract the different parts from the massive BAG file. ArcGIS is only running on a Windows operating system and the author of this thesis is working on a Mac operating system, this transformation is therefore extra time consuming. After running ArcGIS on Windows OS in a virtual machine the accommodation objects are extracted. The building footprint part of the BAG is significantly bigger because it consists of polygons, this set is retrieved from a different source (as described in the previous paragraph).

Transformation 2: filter

When the BAG is retrieved from its source it contains elements which are no longer existing. They have an *enddate*, which is the date the elements are deleted. By filtering on the elements that have this *enddate*, they are removed from the dataset.

Transformation 3: spatial relation

To find out in which building footprint the household is located, a spatial relation transformation is preformed. By adding the *buildingID* to the accommodation

objects within the building footprint, the footprint for all the objects is initiated. Households inside the same building footprint use the same door to exit the house, therefore it is assumed that these households use the same trashcan.

Outcome

The goal of this preprocessing is to achieve the accommodation object dataset, containing information about the building footprint the objects lies in.

Input 2: preprocess trashcans

The purpose of this preprocessing is to retrieve information about the trashcan locations in the city of Amsterdam. The municipality of Amsterdam provides an open data portal a dataset containing all containers inside the city. After contact with the municipality they provided this research with a dataset containing both the locations of the trashcans and the mass of the waste during the emptying of the trashcans. The goal is to retrieve all the trashcan locations, their types and the mass of the waste for a specific time frame. Figure 4.8 shows the flowchart for this preprocessing, the numbering in the figure is the same as used below.

Figure 4.8: Overview of the needed steps to achieve input 2.

Source

The dataset is a Microsoft Excel file, containing a line for every measuring moment, restulting in a massive database file with over 400.000 measurements. Every line gives information of the mass of the waste in the trashcan, the location of the trashcan, the type of waste to be found at this location and the date & time it is emptied.

Transformation 1: convert

The *.xlsx* format cannot be imported in FME. The spreadsheet document contains multiple tabs and FME is not able to process these tabs. Therefore the tab containing the measurements is converted into an open source *.csv* format. This transformation is executed within the Excel environment.

Transformation 2: filter

The measurements contain two values for the mass of the waste, one before the trashcan is emptied and one afterwards. The difference of these two measurements is the mass of the trash inside the container. Some of the measurements have negative difference values, this is impossible and therefore these measurements are considered errors.

Transformation 3: format

Some trash locations are emptied only twice in a year. The total mass of this location is not representative in comparison with trashcans that are emptied multiple times a week. To be able to retrieve information based on the day the trashcan is emptied, the date format from the source file is made consistent. The source format

is $dd/mm/yyyy$ and this transformed into $yyyymmdd$. In this structure sorting the date files becomes possible.

Transformation 4: add geometry

The measurements are only text and do not contain a geometry. The longitude and latitude coordinates are written in two separate columns. These are used to put the location of the trashcan on the map, afterwards the coordinate reference system is transformed from *WGS*84 to RD coordinates.

Transformation 5: split

Every measurement can represent a different kind of trash type. To be able to calculate the mass of the trash, first is specified which type of trash is measured in this line (see table 3.2 for the different types of waste in Amsterdam).

Transformation 6: calculate

As every waste type is separated, the total amount of trash per type per year can be calculated. The calculation is therefore grouped by as well the *locationid* as by the waste type the results is the amount of waste per type per year per location.

Transformation 7: aggregate

To calculate the mass per waste type all the measurements are needed. Now that these are determined, the aggregations based on location can be executed. The result is the total mass of waste per type per year per location.

Transformation 8: create

To be able to identify which locations have a residual container. A list is generated containing all the waste types at the location. If a location does not contain a certain type, the mass specification is empty.

Transformation 9: specify

After trial and error is specified that some trash locations contain significantly less mass than others. While checking the source data is observed that these locations are emptied significantly less than others. This means that some of the trashcans are not representative for a full year. Therefore the following is adapted, when a trashcan is emptied (at least once) in all four quarters of a year, the location is considered valid for as well the validation as the rest of the methods.

Outcome

The outcome of this preprocessing is a new shape file, containing the locations of the trashcans in Amsterdam. With additional information about the waste types at this location and the kg of waste per type per year.

Input 3: preprocess hotspot

The purpose of this preprocessing is to retrieve information about the hotspot locations in the city of Amsterdam. These hotspot are representing a direction preference based on interesting locations within the city. By finding clusters of shops, public transport stops or other objects that represent an attractive place, these preferences are found. Google uses a similar approach to find areas of interest to color them on their map service. On their blog they state the following: "We determine "areas of interest" with an algorithmic process that allows us to highlight the areas with the highest concentration of restaurants, bars and shops"(Li and Bailiang, 2019). To create the hotspots in this thesis the same approach is used, by taking the food actor dataset and creating clusters out of them, a similar map is be created. The food actor dataset is an extract from the ORBIS database that shows all businesses engaged in economic activities under the NACE group "I - Wholesale

and retail trade, transportation and storage, accommodation and food service activities" registered in the municipality of Amsterdam. This dataset has been provided by the REPAiR project under the consent to be used only for the purpose of this research and not made available to the third parties (Bureau Van Dijk, 2019) (see figure A.1 in the appendixes). Figure 4.9 shows the flowchart for this preprocessing, the numbering in the figure is the same as used below.
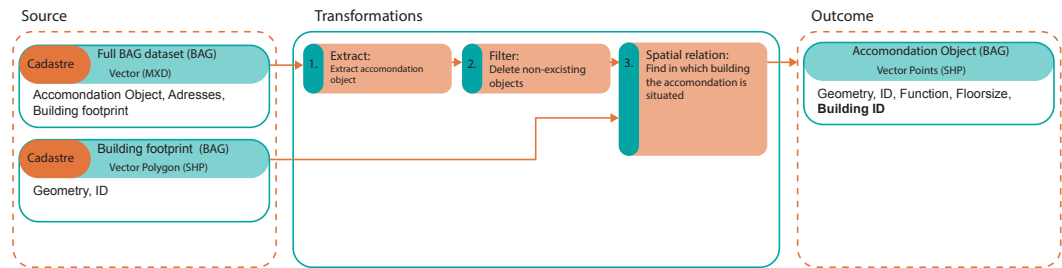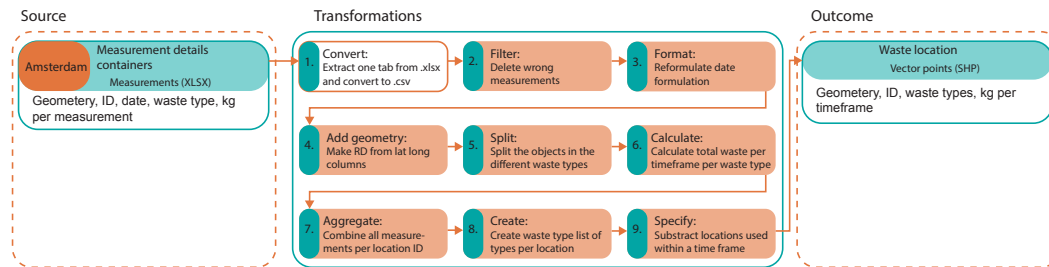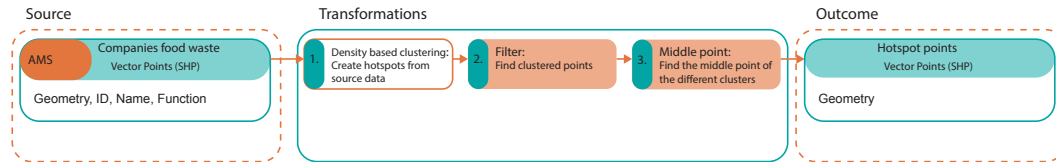

Figure 4.9: Overview of the needed steps to achieve input 3.

Source

The source dataset used to created the hotspot is the companies food waste dataset provided by the REPAiR project. The dataset contains all actors taking part in the food waste chain. Besides food retailers like supermarkets, also hotels are present. Since hotels are mostly located in interesting areas, this dataset can be a worthy representation of the interesting areas in Amsterdam.

Transformation 1: density based clustering

This transformation is performed within ArcGIS, because it provides a tool to manually adjust density based clustering. The goal is to create a clustering which is similar to the 'area of interest' approach of Google. The clustering transformer requires input point features, where after the clustering method can be chosen. To be able to adjust as much as needed, the chosen method is: defined distance. If the *minimum features per cluster* is set on 30 and the *search distance* is 100, the result from figure 4.10a is obtained. The inputs are manually changed until the result is comparable with the Google Maps result. Important hotspots in the city such as: Leidseplein, De Pijp, De Hallen and Haarlemmerdijk are found by the clustering method, therefore this result is concluded as sufficient to represent the hotspot areas in Amsterdam.

Transformation 2: filter

After finding the clusters the dataset is inserted in FME, where the clustered points are retrieved. The saved file with *.shp* format from ArcGIS is containing all the points and their cluster number. When the cluster number is 0 the point is not clustered. The clustered points are aggregated into multipoints per cluster.

Transformation 3: middle point

Since the aggregated clusters are multiple points, the direction preference can not be performed, therefore the clusters are transformed into a single point. Because the direction preference is only a direction, the middle point of the cluster is sufficient. This is the only necessary output, a list with coordinates of the cluster middle points.

Outcome

The outcome of this preprocessing is a shape file with all the middle points of the generated hotspot areas in Amsterdam.

(a) Result of trial and error of density based clustering [source: ArcGIS]

(b) Result intention (red represents the areas defined as interesting by Google) [source: (Google Maps, 2019)]
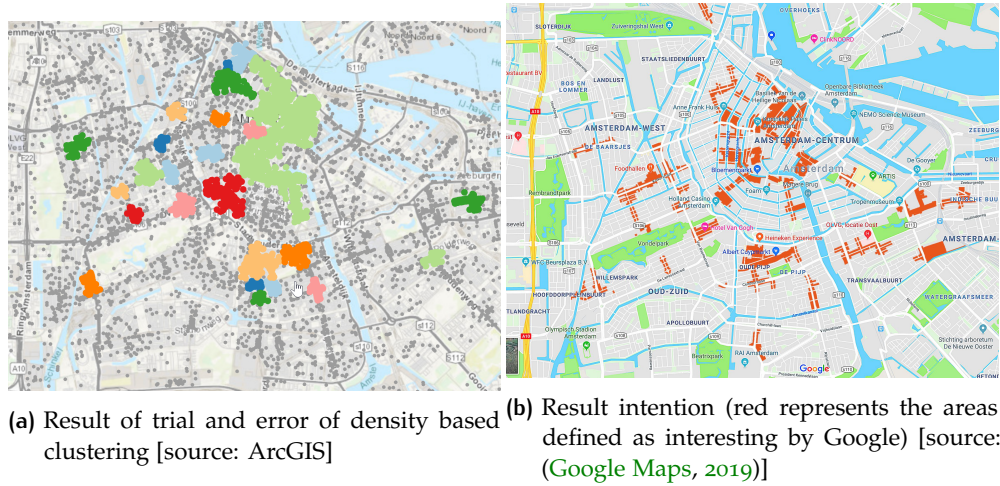
Figure 4.10: Comparison of the density based clustering in ArcGIS and the example of Google Maps

Input 3: preprocess least cost raster

The purpose of this preprocessing is to create a least cost raster to preform the LCPA. The values within the raster are representing a walkability value, since the LCPA is searching or the least cost, the objects of the BGT with the highest preference of pedestrians is represented by the lowest value in the raster. Which surfaces are represented by which walkability value is based on previous research by Ammerlaan (2016) and Alfonzo (2005). Ammerlaan (2016) uses the paper by Alfonzo (2005) to obtain the values, in this section the approach is criticized. The goal is to achieve a raster with dimensions big enough to cover the test areas and the raster is able to generate realistic walking paths for pedestrians. An additional requirement of this raster is that it cannot contain any holes. Figure 4.11 shows the flowchart for this preprocessing, the numbering in the figure is the same as used below.
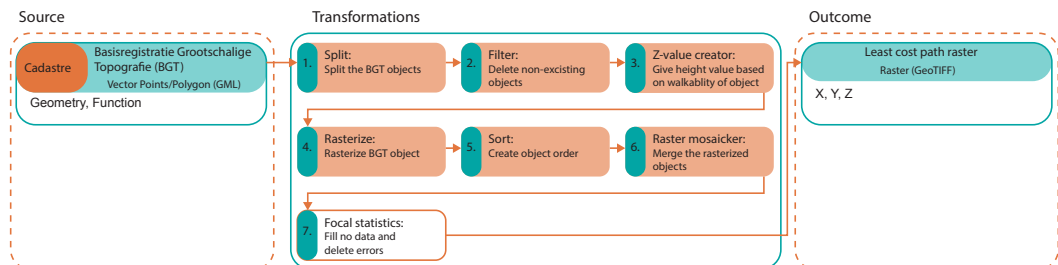
Figure 4.11: Overview of the needed steps to achieve input 4.

Source

As described in section 4.1.1 the cost values are based on the BGT dataset by the Dutch Cadastre. This dataset can be downloaded from PDOK by selecting specific areas you want to download. Figure 4.12 shows an example of the BGT in the corner of one of the test areas. Green represents the element: traffic area, this element contains, for instance, main roads, pedestrian paths, parking spaces and tram roads. This element contains the best walking area (pedestrian path) and relatively easy traversal elements (like roads). The light blue represents water and is not able to be crossed. Only the elements which are polygons are subtracted from the source dataset, since the rasterization of line elements is not possible (due to resolution shortcomings).

Transformation 1: split

Figure 4.12: Example of the BGT dataset, visualizing traffic area, water element, private gardens, etc

The BGT consists of multiple elements, the objects which have different elements containing different walkability values are extracted here. Some elements have the same cost value for every object, such as *waterbody*, it does not matter which waterbody, it is not able to be crossed by a pedestrian. More difficult is *trafficarea*, pedestrian path is having the lowest walkability value, while a highway is also a part of this element and not able to be crossed. Table 4.1 is visualizing the different useful objects within the elements, this table is a result of trial and error by trying to retrieve a connected raster without holes. Other elements like: *waterbody*, *buildings* and *separations* do not need any splitting they are processed directly.

BGT element	objects
TrafficArea	pedestrian path, pedestrian area, residential area, bicycle path, parking space, driveway, public transport path, local car way, regional car way, highway, train tracks
PlantCover	grass, unknown
BareTerrain	property, unknown, vegetation, closed pavement

Table 4.1: BGT elements and their useful objects

Transformation 2: filter

Just as with the previous preprocessing of basic registrations in the Netherlands, the BGT contains elements that do no longer exist . Therefore this transformation deletes the non-existing elements.

Transformation 3: z-value creator

Since a raster is representing a continuous area in a 2.5D space, the representation is a 2D map with a z-value representing the height. In the case of this cost raster, the z-value is not an actual height value, but the walkability value. By adding a z-value to the elements filtered in the previous transformation, the cost raster can be created. To assign the z-values, the research by Alfonzo (2005) is used and the cost raster by Ammerlaan (2016) is a working example. Alfonzo (2005) presents five levels of walking needs, the first one is feasibility, in the case of bringing away your trash, it is always feasible to walk. This need is therefore always satisfied. The second need is accessibility, he states: 'accessibility factors may include the presence

(a) Resulted paths based on different cell sizes, red = 0.5, blue = 1 and purple = 2

(b) Impossible path created with cell size 2
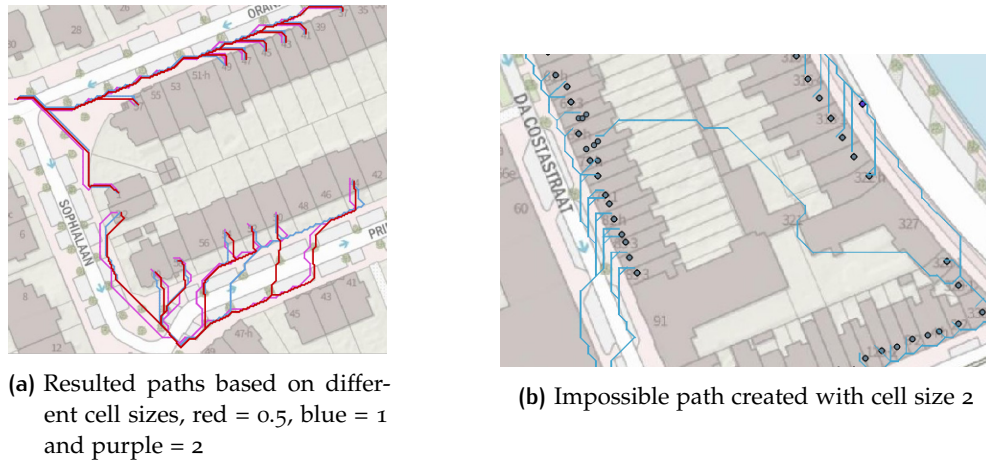
Figure 4.13: Different cell sizes and their effect on the LCPA

of sidewalks, paths, trails, or feature that provide perceived paths on which to walk'. Therefore the pedestrian path is given value 1, it is the second need to be satisfied.

Another statement of Alfonzo (2005) is: 'For destination walking, the perception of distance to a particular destination may affect the person's level of satisfaction with accessibility'. Thus, pedestrians are willing to cross certain objects to make their path shorter. Therefore alternative paths are given value 2, since they are relatively easy to cross and can make a path significantly shorter. Road paths are more difficult to cross and therefore given value 4. Private areas are preferably not crossed, but if it makes the path ten times shorter, it can be crossed, accordingly assigned with value 10. (see table 4.2).

BGT object	description	walkability value
Pedestrian path	Sidewalks and pedestrian area	1
Alternative path	Green space, parking spaces, cycling paths	2
Road path	Roads, rails, bus lanes	4
Private area	Private gardens and terraces	10
Buildings	Buildings and other structures	999
Water	Canals and open water	999

Table 4.2: BGT objects and their walkability value (z-value)

Transformation 4: rasterize

Now that the z-value is added to all the objects, the objects can be rasterized. By rasterizing vector objects details get lost, because the raster is containing cells of a certain size. The best raster representation of the vector objects is the smallest cell size, although a bigger cell size is contributing to a noticeable speed increase for the LCPA. The biggest cell size needs to be identified, which is still able to represent the objects of the environment in a realistic way. In figure 4.13a three different sizes are compared, some of the paths cross the street at different crossovers, but all the sizes find the same closest trashcan. Following this argumentation, cell size 2 meters is the ultimate, big and still finding the same trashcan as smaller cell sizes. Nevertheless, when size 2 is tested on the other test location an error occurred. A path is found through the building block, this is possible since the algorithm is able to make diagonal connections, but due to the resolution the buildings are not properly closed (figure 4.13b). This path is not desirable and that is why one cell size smaller is picked to be the optimum.

Transformation 5: sort
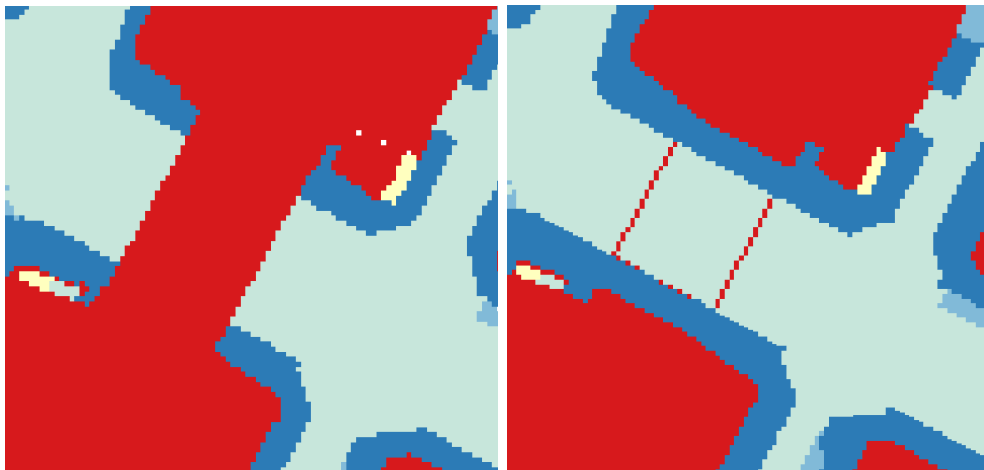
(a) Order of the objects when water is situated above the road objects

(b) Order when the objects are correctly situated

Figure 4.14: Visualizing correct and incorrect order possibilities

Size the objects are rasterized individually, an order is added to specify the sequence of the objects when they overlap. Without sorting, the raster visualized in figure 4.14a is obtained, the *WaterBody* object placed above the road object and therefore the bridge disappeared. By creating an order to make sure the road stays above the water, the raster in figure 4.14b is created. The figure is an example of the problems that appeared. By creating an order, a logical raster is produced.

Transformation 6: raster mosaicker

Now that the order of the individual objects is obtained, the output raster is created by combining all the individual rasterized objects. By adding a sorter in FME the control of the insert order is taken. The raster mosaicker has an option to indicate to take the last insert value as the top value, therefore the order is inserted in that order. The result is visualized in figure 4.14b.

Transformation 7: focal statistics

After the raster mosaicker is preformed, the raster is not yet complete. Due to the overlap of the different BGT objects, some objects do not perfectly connect to their neighbouring objects (figure 4.15a). Cells with no value appear in the raster. Multiple Geographical Information System (GIS) packages offer different approaches to fill no data cells in rasters. Because all of these approaches take averages (or other combined values) of the cells around the missing cell, these techniques could not be used. Since this is creating values which are not specified. To make a focal statistics that is capable of finding the best value to fill in the no data values, Python is used [1].

Another error, already visualized in figure 4.14b, is a line over the road on a bridge. This line consists of values 999 while the road is value 4, this causes a congestion. This error is solved in the same Python script [1].

In pseudocode 4.1 the general approach is written. First is discussed how the no data values are filled, afterwards the errors on the bridges are removed. For the first approach, first, the raster is opened, secondly the raster is converted into an array. Thirdly a *for* loop inserted to go over every value in the raster. When the cell holds a no data value, the cells around it are checked. Because the background of the raster is likewise consisting of no data values, the *if* statement makes sure these are not filled. When the values around the cell are not no data, this means a hole

---

[1] Script for focal statistics:
https://github.com/DaveyOldenburg/Reversehouseholdwasteflowmapping/blob/master/Focalstatistics.py

(a) Objects not connected to each other   (b) Objects connected without no value cells
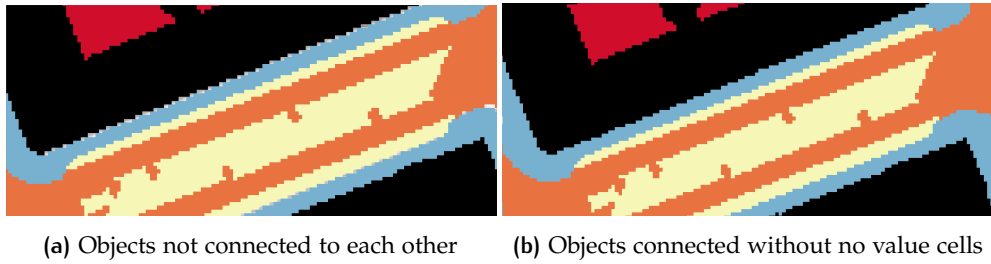
Figure 4.15: Comparison of before and after the focal statistics

is found. All the eight values of the neighbouring cells are inserted in a list, where after the new value of the cell is the median of this list. In this way the cell holds a value that is not rounded or strange for the cell location. The result of this part of the algorithm can be seen in figure 4.15b.

The errors on the bridges require a different approach. Since this error only occurs on bridges and in the middle of roads, this is taken into account while developing the algorithm. The *for* loop is the same as the previous approach because it saves time. When the cell equals 999 the cell is considered, because this also counts for buildings or water, an extra check is implemented. When the 999 cell is not having multiple 999 values diagonal of it, it means the cell is not part of a surface and can be considered. By inserting the neighbouring values into a list, the count of the value 4 (since 4 represents the road) is found, when the count is higher or equal to 4 the cell is considered an error and replaced by the value 4.

Algorithm 4.1: Focal statistics (original on Github [1])

    Input: The outcome of transformation 6
    Output: An updated cost raster without holes and without congestion in
          road due to error

1 **create** array from input raster
2 **for** *value* **to** *array* **do**
3   **if** *cell == NoData* **then**
4   **if** *cell top left =! NoData or cell top right =! NoData or cell bottom left =! NoData or cell bottom right =! NoData* **then**
5   **add** values around cell to list **when not** nodata
6   **update** value of cell to **median** of list

7   **if** *cell == 999* **then**
8   **if** *cell top left =! 999 or cell top right =! 999 or cell bottom left =! 999 or cell bottom right =! 999* **then**
9   **add** values around cell to list
10   **if** *count 4 in list $>=$ 4* **then**
11   **update** value of cell to 4

12 **transfrom** array to raster
13 **save** raster

4.2.2 Household to retailer

To make the implementation of the method household-to-retailer possible, the inputs from figure 4.6 are necessary. The inputs one until five are treated respectively. Due to some overlap with the other method, some inputs are already prepocessed. This section is describing the preprocessing of input 2&3 and 5. The inputs: households and hotspots are already satisfied.

Input 2&3: preprocess companies food waste

To be able to find the food retailers that are reachable within the isochrone areas, the retailers need to be properly located. Since the retailers are missing most of the revenue, the floor size of the retailers is to be added. The floor size is giving an indication of the relevance of this retailer (as discussed in section 4.1.2). Figure 4.16 shows the flowchart for this preprocessing, the numbering in the figure is the same as used below.
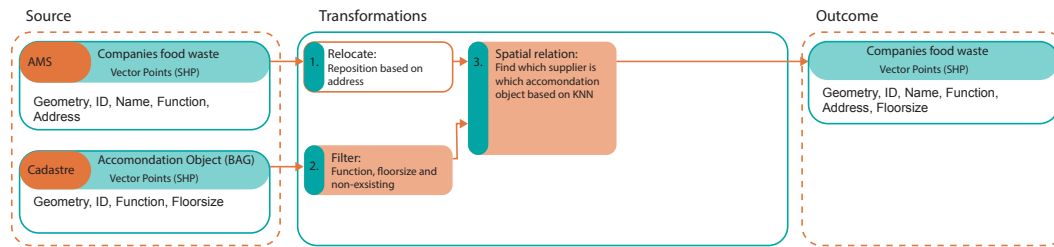
Figure 4.16: Overview of the needed steps to achieve input 2&3.

Source

Both of the sources used for this preprocessing are used before. The companies dataset is containing all the food waste contributing actors. The accommodation object is used to represent the households before, but in this preprocessing it is representing the objects classified as store.

Transformation 1: relocate

Since the food waste retailers dataset is an Orbis dataset, the precision of the points is not high (Bureau Van Dijk, 2019). The Orbis dataset is containing details about millions of companies around the globe, but the location of the objects is not always accurate. Some of the points lie in the middle of the street. To achieve a better precision, the points need to be relocated. Since one of the accommodation objects is representing this specific store, an approach to link the object with the store is tested. Multiple approaches are tried, the addresses are used to find a connection to the accommodation objects of the BAG, due to too many exceptions this is not possible. Connecting on postal code is also not possible. Therefore finding the neighboring accommodation objects is the only option, so the attributes can be shared with the point. To make sure the retailer geometry is as precise as possible, they are relocated using BatchGeo (BatchGeo, 2019). This tool is able to read addresses and visualize them in Google maps, assuming that the retailers are within the right building after this processing.

Transformation 2: filter

To make sure the spatial relation is succeeding, the accommodation object is filtered on three different variables. First the object needs to exists, secondly the object should have a 'shop' function and finally the shop needs at least a size of $15\,m^2$. After filtering these objects, a better result is expected. Since most of the accommodation objects are households, the objects total decrease from 840.00 before and 11.000 after filtering.

Some of the floor sizes in the accommodation object dataset are more than 55.000 square meters. Since this research is implemented in an urban area, supermarkets or food retailers with these dimensions are not expected. To give a comparison, the Albert Heijn XL supermarket which is known for their big floor size, is never bigger than 4.500 square meter in the city of Amsterdam. Therefore accommodation objects with a bigger surface than 4.500 m^2 are neglected. This filter results in a decrease of 4 percent of the usable stores.

Transformation 3: spatial relation
The retailer of food is finding the nearest accommodation object in 2D euclidean distance. Where after the attributes are merged, the result is the retailer having information about the floor size of the store. The uncertainty is significant, but restrained to a minimum by filtering on decent store sizes and the function *store*.

Outcome
The outcome of this preprocessing is formatted in the same way as the dataset of the companies food waste which is used as input. However, now the floor sizes retrieved from the accommodation objects is added.

Input 2&3: preprocess network

To be able to create the isochrone area mentioned in the section 4.1.2, the road network needs to be obtained. A reliable and easily achievable source for road networks is OSM.
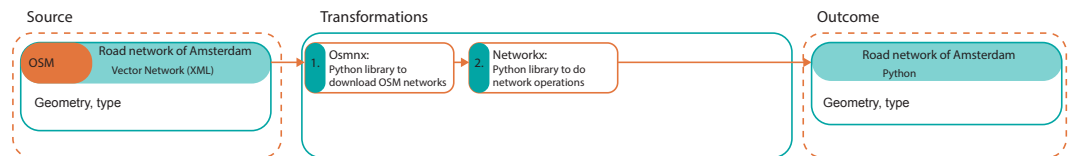
Figure 4.17: Overview of the needed steps to achieve input 5.

Source
The source of this preprocessing is the OSM street network. Since the implementation of method household-to-retailer, is executed in Python, loading the street network directly into a Python environment is the most efficient method. Thus, the *.xml* OSM is not downloaded as a file, but obtained in the python environment.

Transformation 1: osmnx
The osmnx python package is able to download a OSM network based on a name of the city. "OSMnx is a Python package that lets you download spatial geometries and construct, project, visualize, and analyze street networks from OpenStreetMap's APIs. Users can download and construct walkable, drivable, or bikable urban networks with a single line of Python code, and then easily analyze and visualize them." (Boeing, 2017). Since the isochrone area is calculated based on trips on bike trips, the bike network is downloaded. The network is plotted in figure 4.18, here is visualized how the network is transformed to a graph, containing vertices and edges. Due to the graph being present, network operations can be preformed.

Transformation 2: Networkx
The retrieved network and graph from the previous transformation is the base of this transformation. Since routes, travel times and shortest paths are calculated, a library enabling these functions is inserted. The Networkx library is; "a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks" (Developers NetworkX, 2019).

4.3 IMPLEMENTATION

This section emphasizes on the implementation of the two main methods. The outcomes of section 4.2 are composing the input for these methods. Subsection 4.3.1 is

Figure 4.18: Plot of the OSM network in Python

concentrated on the method household-to-trashcan. Subsection 4.3.2 is focusing on the method household-to-retailer. Both of the methods are treated in the following way, first a visual representation of the method is used to specify the fundamental steps, thereafter a pseudo code of the applied approach is used to give technical details. Since multiple approaches are tested, additionally the not implemented approaches are described and explained why they are not used.

4.3.1 Household to trashcan

This subsection is following the method described in section 3.3. For the summary figure 4.19 is created. Since the required inputs are preprocessed, the process can be executed. First the accommodation objects from section 4.2.1 are inserted, they represent the households (1). Secondly (2) all the trashcans from the municipality dataset are inserted (section 4.2.1), thereafter the trashcans within a bounding box are considered. This distance is set on 100 meters, approximately 2-4 trashcans lie within this box for every household. The 100 meters is used because it finds multiple trashcans, but not too many. This means the building box is 200 by 200 meters. Where after, all the considerable trashcans are found. A bounding box is chosen above a normal buffer since a *within* function for buffer is a relatively time consuming geo-processing operation. The building box is created by mathematically checking if the coordinates are within the extents of the bounding box. Which is resulting in a faster execution.

After the candidate trashcans are found, there are two options. First (3), the trashcan is so close, that regardless of the environmental factors, this disposal is unconditionally used. This distance is set on 20 meters. 20 meters is the distance from the front door of the household to the other side of the street (in the test area). Therefore this distance is set as the closest trashcan. When the trashcan falls within this distance, the trashcan is so remarkably close, that disposing your trash in this trashcan is not making the walking distance much longer, not even when the final destination is the opposite direction.

Second (4), when there is no remarkably close trashcan, the method proceeds. To implement the expected preference direction, the closest hotspot for the considered household is found. To speed up the algorithm, to closest hotspot is initiated in 2D euclidean space. The preference direction is created, by adjusting the raster.

After creating the new adjusted raster, the LCPA can be executed on the considerable trashcans (5). The results are the paths to all these trashcans, to determine which trashcan is the used disposal, the cost values for every path are compared. The cost value is the sum of all the cell values used to create the path. The length of
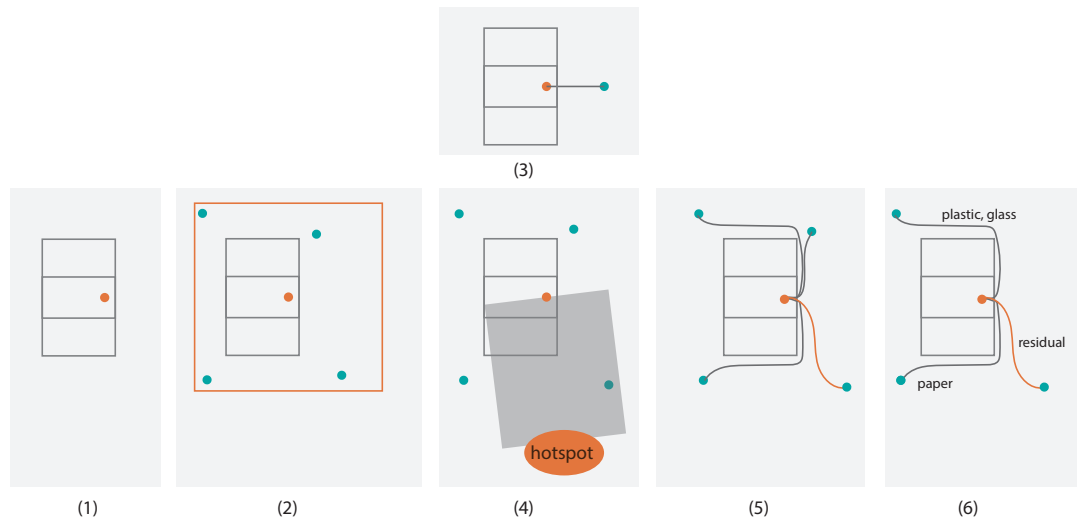
Figure 4.19: Overview of the steps in method household-to-trashcan

the path is on purpose not used. The adjustment of the raster is lowering the cost value of the path, therefore the influence of the raster adjustment is only effective when comparing the cost values. The path with the lowest cost value is picked and written as the used trashcan.

The last step (6) is an additional execution. Amsterdam is collecting four different types of waste in public space. From experience of the author of this thesis is assumed that when a resident is disposing another type of waste, the resident is willing to walk further and to make the trip as efficient as possible, takes the residual waste together with the other type of waste. The resident walks further with their residual waste, but only due to the disposal of another type. By checking which types of waste are available at the used trashcan, a list of missing types is generated. Then the locations containing these types are found to add the missing waste types. The outcome is a minimum of one path (all the types are present at the used trashcan) or a maximum of four paths (the different types are available at different locations).

To highlight the difficulties of this algorithm, the sections below elaborate on the solutions for these complications. References to pseudo code 4.4 are made to address specific executions.

No trashcan in bounding box

In line 15 of algorithm 4.4 the search bounding box is specified. Since this is an *if* statement after a *for* loop, all the trashcans are considered and only the considerable objects go into the condition. When an object is considerable, the LCPA is preformed and the result is added to a list, afterwards the list is filtered on the lowest cost value and this result is written.

In rare occasions, it happens that there is no trashcan in the bounding box. This is not expected, since the test areas are dense urban environments, but in case it appears that the list with considerable trashcans is empty, another execution is preformed. By enlarging the bounding box and iterating through the trashcans again, the closest trashcan is found. Since the distance towards this trashcan is considered distant, EM is used to find the closest of the newly find disposals. When found, the LCPA is executed and the path written.

This analysis takes three inputs: starting point, end point and the least cost raster. This operation is performed on a raster, since Python is faster to preform operations on arrays, the raster is transformed into a array. This is a 2D array, where every cell is represented by an element in the array.

To be able to create a path through the array, the indices of the start- and stop-point are found. Now the function *route_through_array* from the library *skimage.graph* can be preformed. This function returns the cost value of the found path and the indices of the used cells. A 0s array with the same dimensions as the input is created and the indices are written as 1s in the path-array.

This path-array needs to be transformed into a linestring, by finding the coordinates of the elements in the array with value 1, a dictionary is created. The shape is as follows: pointDict[count] = (Xcoord, Ycoord). The count is important because the order of the points representing the line are critical to create the linestring. When all the points are added to the dictionary, a loop through the dictionary is executed. The line is created by adding the first point and the second point to a linestring, then the second point and the third point and so on. The results are multiple linestrings between all the centre points of the considered cells. By adding all the linestrings to a multilinestring, a path is created. This path is a Well Known Text (WKT) formatted multilinestring and hereby converted from raster to vector.

Shape of preference direction

The LCPA is only taking into account the values given in the preprocessing of the least cost raster. To validate the hypothesis, another factor needs to be implemented. The residents are willing to walk further when the trashcan is on the route towards a city hotspot. Figure 4.20(1) shows the point where the preference is the most important, both trashcans in this figure are within the same distance, however following the hypothesis the trashcan in the direction of the hotspot is the used trashcan. When the resident leaves his/her house, the choice if he/she goes either left, right or straight, has a big impact on the trashcan he/she uses. In figure 4.20(2) the hotspot in inserted and a polygon is created in this direction. The polygon is drawn by using the coordinates of the house and the hotspot, creating a line between them and buffering the result (without round edges). The result is a path that lies completely within this direction. Therefore it is expected that the trashcan within the path is chosen above the other trashcans. The usage of this simplistic shape puts the focus on the first choice after leaving the house.
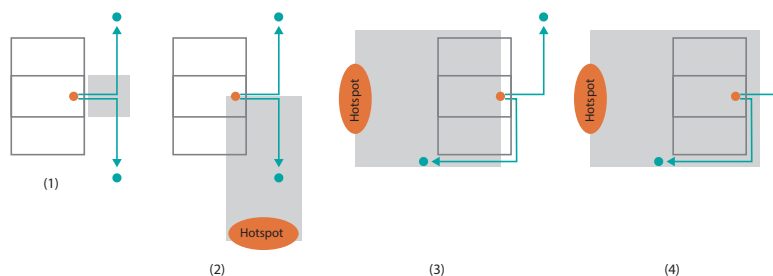
Figure 4.20: (1) where the direction choice is the most important. (2) when the walking direction falls within the hotspots path. (3) when the hotspot lies behind the considered household. (4) solution for the occasion that the hotspot is behind the household

Figure 4.20(3) shows an example when this shape is not ideal. Since the polygon is derived from a line with household and hotspot as start- and end-point. The polygon is falling within the building footprint and therefore cannot be used right after the resident leaves his/her house. The solution is to extend the line in the

opposite direction of the hotspot. The result is visualized in figure 4.20(4), because the polygon is wide enough to cover the area around the building block and the line is extended in the opposite direction from the hostspot, the path to the trashcan falls within the polygon.

Algorithm 4.2 shows the pseudo code to enlarge the line between the household and the closest hotspot. A geometrical approach is executed to calculate the new x and y coordinate. The enlargement is 5 meters, this value is based on the placement of the accommodation objects in the BAG dataset. Averagely the points are placed 3 meters away from the boundary of the building footprint. Since the polygon should also cover an area on the pavement, 5 meters is chosen. This approach is a part of the final code of the method household-to-trashcan [2].

Algorithm 4.2: Preference direction shape improvement (original on Github [2])

Input: Hostspot coordinates, considered household and the buffer size
Output: Shape representing the first choice

1 **if** *household x > hostspot x* **then**
2   **angle** = $(household_y - hotspot_y)/(household_x - hotspot_x)$
3   **calculate** distance between household and hotspot
4   **new distance** = $distance + 5$
5   **change in x direction** = $cos(angle) * newdistance$
6   **change in y direction** = $sin(angle) * newdistance$
7   **new house coordinates** = $(hotspot_x + changein_x, hotspot_y + changein_y)$
8 **if** *household x < hostspot x* **then**
9   **angle** = $(hotspot_y - household_y)/(hotspot_x - household_x)$
10   **calculate** distance between household and hotspot
11   **new distance** = $distance + 5$
12   **change in x direction** = $cos(angle) * newdistance$
13   **change in y direction** = $sin(angle) * newdistance$
14   **new house coordinates** = $(hotspot_x - changein_x, hotspot_y - changein_y)$

15 **create** linestring from new x & y
16 **buffer** linestring without rounding
17 **return** shape

Alternative shape of preference direction

An alternative approach for the shape of the previous section is explained within this section. After elaborating both approaches the previous execution is considered more accurate, but the alternative is discussed here.

Since the direction polygon is a rather simplistic shape, a more specified shape is tested. In the method household-to-retailer, a network based analysis is used to create isochrone areas (discussed in section 4.3.2). The implementation of this method raised a new possibility for the direction shape. By using the network to create a path from the household towards the hotspot, a shape following the road network can be created. Since the original shape discussed in the previous section is not taking into account any obstacles (like road network pattern or building polygons) the newly proposed shape is providing a configuration which acknowledges these.

Algorithm 4.3 shows the applied approach to create a shape based on the street network. The full approach can be found as unused function in the method household-

---

2 Part of method household-to-trashcan
https://github.com/DaveyOldenburg/Reversehouseholdwasteflowmapping/blob/master/
MethodHouseholdtotrashcan.py

Algorithm 4.3: Alternative preference direction shape improvement (original on Github [3])

Input: Hostspot coordinates, considered household, OSM network and the buffer size

Output: An alternative shape representing the direction towards the hotspot

1 **transform** coordinates of hotspot and household from RD to WGS84
2 **find** the vertex in the network closest to household
3 **find** the vertex in the network closest to hotspot
4 **create** shortest path weighted on distance
5 **find** coordinates of found vertex in shortest path
6 **transform** vertex coordinates from WGS84 to RD
7 **create** linestrings from the found points
8 **append** linestring from household coordinates to closest vertex coordinate
9 **buffer** linestrings to create polygon
10 **return** shape

---

to-trashcan [3]. The result is a shape going beyond the important area from figure 4.20(1) and is therefore more representative for longer and further routes to hotspots. Because the OSM network holds CRS in WGS84, line 1 transforms the coordinates from the households and hotspot into WGS84 coordinates. After the network operations, the coordinates are transformed back to RD coordinates. Line 2 finds the closest vertex to both inputs. Since network operations can only be performed from the nodes within the graph, the shortest path sometimes does not come close to the household (when the first vertex in the graph is far away from the object). When this occurs, the important area from figure 4.20(1) falls not within the shape. To prevent this, the linestring from the household to the closest vertex is inserted manually (see line 8). This works in most of the cases, but as already mentioned in the comparison of distance measure tools in section 4.1.1, the closest vertex to the building, can be behind it. Therefore this approach is not always working, an example is shown in figure 4.21, where orange represents the household and the closest vertex in green. The connection is inserted manually, but the vertex is behind the household. Since the main choice of the direction is right after exiting the house, the other method is always correct, where this approach lacks consistency.
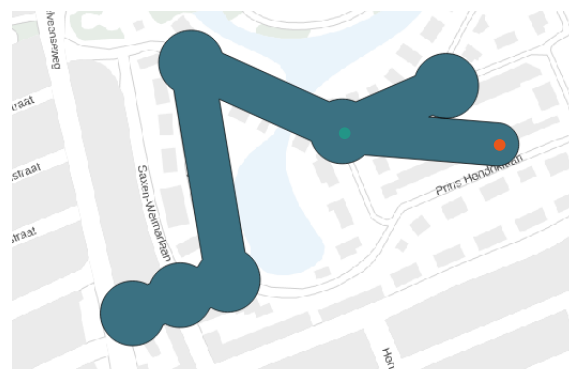
Figure 4.21: An example of a vertex behind a household in the alternative direction method.

Rasterize preference direction

After the shape of the preference direction is found, a new question rises. How can the LCPA be manipulated, to force a certain direction? The only inputs of the LCPA

---

3 Part of method household-to-trashcan
https://github.com/DaveyOldenburg/Reversehouseholdwasteflowmapping/blob/master/MethodHouseholdtotrashcan.py

are the begin point, end point and the values of the raster. Therefore is concluded that the values of the raster should be manipulated to force this direction. Since the cost values of the paths are compared to find the used trashcan, these are manipulated. By lowering the cost values that fall within the preference shape, trashcans within this shape are more likely to have lower cost values for the paths and thereby more feasible to be the used trashcan.

To adjust the pixels that fall within the shape, the shape needs to be rasterized and combined into the raster. For rasterizing a vector shape, the algorithm of Bresenham (2010) is used. The library *rasterio* is imported because it contains a rasterize function that uses the Bresenham approach. The vector shape is inserted in a *GeoJSON* format, together with the original raster, where after the function returns a new array with 0s and 1s in the same dimensions. The 0s represent the cells which are not overlapped by the vector shape and the 1s represent the cells that are overlapped. Since the goal is to multiply the original raster with the adjustment raster, the 0s are eliminated so percentages can be used. 1s represent 100%, which means the values stay the same. When the values within the shape should be dived by 2, the following approach is executed (see figure 4.22). (1) represents a part of the original raster and (2) the adjustment raster with the preference shape. As shown in the adjustment raster (2) the cells which are not touched by the shape are represented by 1s, therefore the orginial cells stay the same after the multiplication. By multiplying the two rasters the needed result is obtained (3). Multiplying by 1 results in the same value, while multiplying by 0.5 divides the value by 2.

```
1 1 1 1 1 1 1 2 2          1 1 1 1 1 1 1 1 1          1 1 1 1 1 1 1 2 2
1 1 1 1 1 1 2 2 2          1 1 1 1 1 1 1 1 1          1 1 1 1 1 1 2 2 2
1 1 1 1 1 2 2 2 2    *     1 1 0.5 0.5 0.5 0.5 1 1 1    =     1 1 0.5 0.5 0.5 1 2 2 2
1 1 1 1 2 2 2 2 2          1 1 0.5 0.5 0.5 0.5 1 1 1          1 1 0.5 0.5 1 1 2 2 2
1 1 1 2 2 2 2 2 2          1 1 0.5 0.5 0.5 0.5 1 1 1          1 1 0.5 1 1 1 2 2 2
1 1 2 2 2 2 2 2 2          1 1 1 1 1 1 1 1 1          1 1 2 2 2 2 2 2 2
       (1)                          (2)                          (3)
```
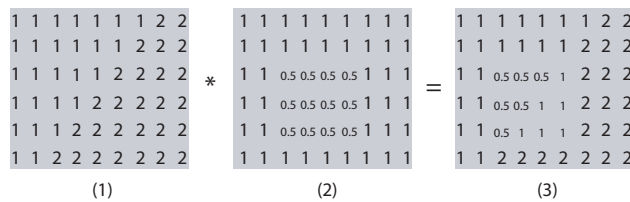
Figure 4.22: Example of an adjustment of 0.5 on the raster

The result is an adjusted raster, useful for a household on a specific location. The adjusted values within the direction of the hotspot, cause a preference because the values are lower and therefore more likely to create a lower cost value of the path. In the example above 0.5 is used as value to adjust the raster, this adjustment has much influence on the results of the least cost value comparison. Therefore this value is optimized in the validation of this method. An adjustment of 1 is also used, which is testing if the preferential routes approach can be validated.

Performance improvement

Because finding a path through an array is a rather slow operation, the execution of the algorithm is slow. To improve the computing time the following technique is applied. Since the households in one of the two testing areas are multiple family homes, these homes are on different floors, but use the same front door. Due to the usage of the same door, the environmental factors of these households are equal. They have the same least cost raster and are close to the same city hotspot. Therefore can be concluded, that accommodation objects that are within the same building footprint, contain the same paths. A dictionary is kept having the *buildingID* as key and the values of the paths as values. In line 11 of algorithm 4.4 is checked if the *buildingID* of the considered household is already in the dictionary. When this is the case, the results are written without preforming multiple LCPAs. After implementation of these improvements 720 objects can be executed in 50 minutes.

Algorithm 4.4: method household-to-trashcan (original on Github [3])

Input: Inputs visualized in figure 4.5: hotspot shape, household shape, trashcan shape and least cost raster.

Output: Walk path between household to the used trashcan, including additional trashcans

1 **open** textfile
2 **for** *hotspot* **to** *hotspot shape* **do**
3   **add** hotspot to list

4 **create** dictionary;
5 **for** *household* **to** *household shape* **do**
6   **for** *hotspot* **to** *hotspot list* **do**
7   **find** closest hotspot
8   **create** direction polygon in direction to closest hotspot
9   **rasterize** polygon
10   **adjust** cost raster with rasterized polygon
11   **if** *footprint of household in dictionary* **then**
12   **write** paths to **textfile**
13   **for** *trashcan* **to** *trashcan shape* **do**
14   **if** *residual in trashcan* **then**
15   **if** *coordinate trashcan in bounding box* **then**
16   **if** *coordinate trashcan = very close to household* **then**
17   **write** path to **textfile**
18   **add** path to **dictionary**
19   **break**
20   **else**
21   **preform** least cost path analysis in adjusted raster
22   **add** path to **dictionary**

23   **compare** cost values of every trashcan
24   **find** smallest cost value
25   **write** path of closest trashcan to **textfile**
26   **if** *closest trashcan does not contain all waste types* **then**
27   **find** closest paper container
28   **preform** least cost path analysis in original raster
29   **find** closest glass container
30   **preform** least cost path analysis in original raster
31   **find** closest plastic container
32   **preform** least cost path analysis in original raster
33   **write** additional paths
34   **add** all paths to dictionary **as** footprint

4.3.2 Household to retailer

This subsection is elaborating the method described in section 3.4. For this summary figure 4.23 is created. Since the required inputs are preprocessed, the process can be executed. First the accommodation objects from section 4.2.1 are inserted, they represent the households (1). Secondly (2), all the food retailers are inserted (section 4.2.2), the final needed data (3) is the OSM network subtracted within Python. Then all the input data required is inserted to be able to execute this method. The final step (4) visualizes the three different isochrone areas suggested by Benoit and Clarke (1997), based on the network and the determined travel time for the specific mode.

(1)   (2)   (3)   (4)

Figure 4.23: Overview of the steps in method household-to-retailer

To specify which mode is most likely to be used to travel to the retailers, the modal split is used. The modal split is the percentage of travellers using a particular type of transportation or number of trips using a type (Eurostat, 2014). To find out what is the most used mode in Amsterdam, the modal split just gives an indication. The difficulty with the split is that most of the data is based on the distance made with the specific mode, therefore the car always seems very present. This can be caused due to the fact that the car mostly overpasses more kilometers per trip. Still, Eurostat (2014) shows a modal share of the bike in Amsterdam of 40% against 27% for private motor vehicles, therefore the bicycle is the most likely to be used to travel to the retailers of food in Amsterdam. To be able to create the isochrone area, the average speed of the bicycle is needed. How far can a bicycle go in a certain time frame? The Dutch bicycle organization did a national research where they calculated the average bicycle speed in Amsterdam of $14.4 kmph$ (Fietsersbond, 2017). This speed can be used as the first input for the isochrone area, the second needed input is the travel time.

According to Dunkley et al. (2004) in Great Britain, 31 percent of the supermarket trips made once a week require less than five minutes of travel time and 55 percent require less than ten minutes. After a GIS analysis of the supermarkets and households in Amsterdam, the average distance to the nearest supermarket is 383 meter (for the municipality of Amsterdam). Which makes that, taken the average travel time in Britain, inhabitants not always opt for the closest supermarket. This conclusion matches the study by Dunkley et al. (2004), which states that people are willing to overcome more distance for better quality or cheaper food. Since this method is based on literature & assumptions and has no possibility to be validated. The travel time for this method is set on 10 minutes. Due to the on average closely located supermarkets, as well as the dense urban area where the test areas are situated. It is expected that taking more than 10 minutes is involving more retailer than needed. As discussed in section 4.1.1 closer stores are more likely to be visited, but the further ones should not be ignored. To achieve this necessary result, the travel time is divided into three equal parts. This results in three different isochrone areas.

Doing groceries by foot is automatically integrated in the isochrone area for the bicycles, going by foot is slower and is less fast. Therefore the area to be covered by 10 minutes of walking falls completely within the bicycle area.

Shape of isochrone area

As mentioned before, an isochrone area is an area of which the boundary contains points that are reachable within the same travel time. The approached followed in this section is form Boeing (2017), by combining the *OSMnx* and *networkx* Python packages, examples of isochrones are generated. The first possibility is creating a isochrone based on a convex hull, Boeing (2017) states that this is not the most accurate way to go (see figure 4.24). After downloading the network, the starting vertex is chosen, in the case of this method this is the household. By reprojecting the household to *WGS*84, the closest vertex is found. Since the length of the edges are in meters, the determined speed is transformed from *kilometerperhour* to *meterpersecond*. Where after the attribute *time* is added to the edges in the subgraph for this starting vertex. *time* is calculated by dividing the length of the edge by the speed. By inserting the maximum travel time, the vertices representing the end of the isochrone area can be found and connected. As seen in figure 4.24 only the end vertices are found and the representation does not take into account any boundaries (like water). If this area is used to find the retailers, unreachable actors are found. Therefore this is not the most suitable approach for this method.
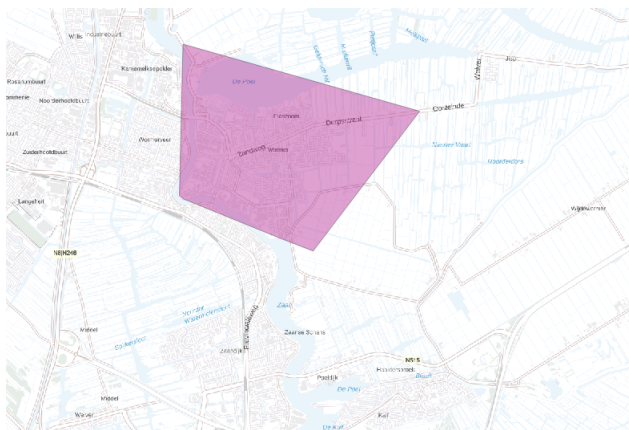
Figure 4.24: Example of an isochrone where only the points are connected

Improvement of the shape of the isochrone area

A more realistic isochrone area is created when the edges reachable within the specified time are buffered. The result of this approach is visualized in figure 4.25, this isochrone area takes into account boundaries and does not reach into places where no road is available. The implementation is a bit different than the one in the previous section. First, a sub-graph is created with the trip time as radius. Second, the coordinates of the vertices are reinserted (they are necessary for the buffering). Thirdly, the line strings can be created by finding the coordinates of the begin- and end-point of all the vertices that fall within the determined travel time. The line strings have geometries and can be buffered with a simple *shapely* library buffer operation. The line strings are separate geometries and can be combined. The combined shape is a polygon and can be filled (to prevent holes within the shape).

Addition of hotspot

The method household-to-trashcan is partly based on the hotspots in the city. Since the approach of this method is validated within section 4.4 the importance of the
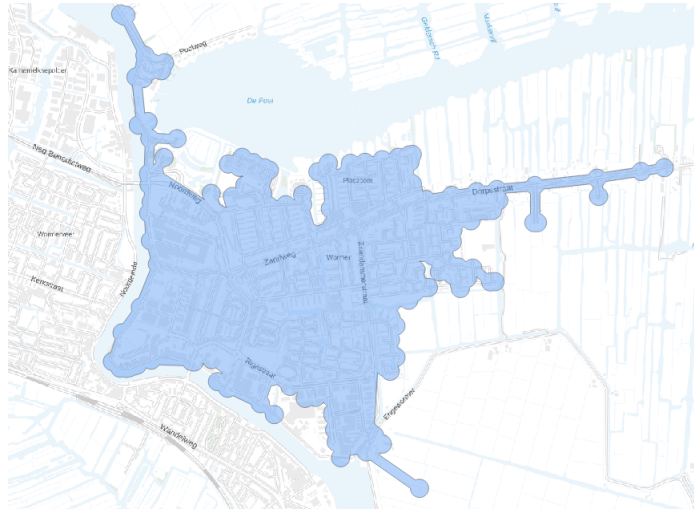
Figure 4.25: Example of an isochrone where the edges are buffered

hotspot can not be neglected. Therefore the shortest path from household to hotspot from algorithm 4.3 is implemented. There is a high possibility the closest hotspot is within one of the generated isochrone areas, but if this is not the case, the path towards the hotspot is inserted. This results is the addition of the retailers on the route to the closest hotspot.

Importance value

After the isochrone areas are found, the retailers which are within the areas can be determined. As shown in figure 4.23(4) three different isochrone areas are created, following the suggestion of Benoit and Clarke (1997). Where the closer isochrone area has more influence on the household than the further ones. The retailers with the added floor size from the accommodation objects from the BAG are found within the three different isochrone areas. The formula to calculate the importance value of a retailer is written in the formula below, first the total square meters of all the retailers within the isochrone areas is calculated, taken into account the values visualized in table 4.3. x represents the proximity value from the same table. When there is one store of the same size in all three of the isochrone areas, the store in the closest area get a value of 50% importance. In algorithm 4.5 starting from line 21 is written how this is implemented.

$$\text{importance value} = \frac{\text{floor size of retailer} * x}{\text{total square meters}} \times 100$$

isochrone number	close proximity value
1	3x
2	2x
3	1x

Table 4.3: Close proximity values for retailers in different isochrone areas (see figure 4.26)

Computation time improvement

The creating of the isochrone is a rather time consuming operation. New sub-graphs are created for every household. It takes 300 seconds to find the food stores within the isochrone areas for 10 households. The starting point of the created sub-graphs is the vertex closest to the considered household. Since this starting vertex can
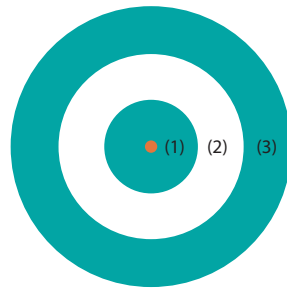
Figure 4.26: Isochrone numbers

be equal for multiple households, this results in the same outcome for all these households. Therefore the following approach is implemented (line 6 of algorithm 4.5), if the closest vertex of the household is existing in the updated dictionary, the results can be directly written towards the text file. If it does not exist in the dictionary the isochrone areas and retailers are found, where after they are added to the dictionary.

To decrease the computation time even more, the same approach as the computation time improvement of method household-to-trashcan can be used. If the household is within the same footprint as an earlier calculated households, they share an exit of the building and therefore have the same retailers. This approach is faster than the improvement based on the closest vertex, since the vertex does not have to be found. Therefore this improvement is executed first (line 5 in algorithm 4.5). The orginial algorithm can be found on Github [4].

4.4 VALIDATION

This section describes on the validation of the two main methods. In section 4.4.1 the correlation between the mass of the trash in the trashcan and the number of households is elaborated. As already mentioned in the methodology, there is no validation for the method household-to-retailer (section 4.4.2), therefore a methodology is discussed, but not implemented.

4.4.1 Household to trashcan

The section is elaborating the validation of the method household-to-trashcan. Different adjustment values for the raster are executed in different loops of the same test trashcans. The value for no adjustment is executed, but also different values. Therefore this validation is as well a validation as an optimization. When the method can be validated the most optimal adjustment is found.

Method

Due to the computation speed of the algorithm for the method household-to-trashcan, a validation method containing test trashcans is created. To test twenty test trashcans the following method is applied. For all of these trashcans all the households that belong to this trashcan should be found, algorithm 4.4 should pick this trashcan, without being limited by the selection of households. In figure 4.27 is shown how the trashcan is buffered to find the households that are inserted in algorithm 4.4. The buffer size is very important in this implementation, when the buffer is too small, not all the households belong to this trashcan are found, which might affect

---

4   `https://github.com/DaveyOldenburg/Reversehouseholdwasteflowmapping/blob/master/MethodHouseholdtoretailer.py`

Algorithm 4.5: Method household-to-retailer (original on Github [4])

Input: Inputs visualized in figure 4.6: hotspot shape, household shape, retailer shape, accommodation object shape and network.
Output: Retailers possibly used by household with their importance value

---

1   **download** osm network for mode 'bike'
2   **open** textfile
3   **for** *hotspot* **to** *hotspot shape* **do**
4   **add** hotspot to list

5   **create** dictionary for foot print;
6   **create** dictionary for household vertex;
7   **for** *household* **to** *household shape* **do**
8   **for** *hotspot* **to** *hotspot list* **do**
9   **find** closest hotspot

10   **if** *footprint of of household in dictionary* **then**
11   **write** importance value for every retailer per household to **textfile**

12   **if** *vertex of household in dictionary* **then**
13   **write** importance value for every retailer per household to **textfile**

14   **else**
15   **find** nearest vertex to the household
16   **create** sub-graphs containing travel times from nearest vertex
17   **create** isochrone areas for three different travel times
18   **buffer** path to closest hotspot
19   **union** buffer to closest hotspot together with isochrone area
20   **for** *retailer* **to** *retailer shape* **do**
21   **if** *first isochrone area* ***contains*** *retailer* **then**
22   **add** floor size *3 to total

23   **if** *second isochrone area* ***contains*** *retailer* **then**
24   **add** floor size *2 to total

25   **if** *third isochrone area (plus hotspot route)* ***contains*** *retailer* **then**
26   **add** floor size *1 to total

27   **calculate** percentage of contribution for every retailer
28   **write** importance value for every retailer per household to **textfile**

the correlation. The buffer distance is set on 130 meters, this dimension is more than the maximum distance from a household to the closest trashcan. In this way, all the considered households for a specific trashcan are found.

Algorithm 4.4 is adjusted to fit this validation. Instead of writing the household *id* and the path to the used trashcan. It writes the trashcan and counts the number of houses that have this trashcan as outcome. A high correlation between the sum of the households and the mass within the trashcan is expected. The mass is derived from the preprocessing of the municipality trashcan data.
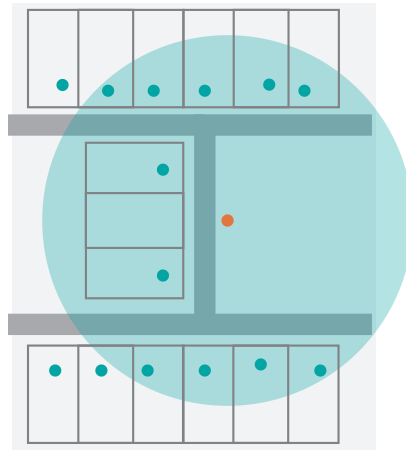
Figure 4.27: Validation method: households are picked around the trashcan

Test trascans

The test trashcans are the trashcans within the test area defined in the methodology of this thesis, to create a higher creditably a few more trashcans are picked. These additional trashcans are surrounded by different hotspots, in this way the influence of the direction to hotspot is as high as possible. In figure 4.28 the trashcans are mapped. The pink circles represent the centre points of the hotspots.

Correlation

To find the correlation between the mass in the waste of the trashcan and the number of households using this trashcan, the programming language R is used (code can be found on Github [5]). When the outcome of algorithm 4.5 is inserted, the mass information of the trashcans can be added. Now the correlation can be calculated between the weight in kg per year for every trashcan.

The first positive correlation in figure 4.29a is representing the connection between the household count and the kg of residual trash per year, when no hotspot is taken into account. This means that only the LCPA is executed, therefore only the minimization of the routes. The R^2 is 0.64, which shows a correlation between an increase in the count and increase in weight of the trashcan.

When the hotspot analysis is added as influence factor, three different values of the adjustment are tested, 0.8,0.4 and 0.1 of the original cost values. Figure 4.29b shows the adjustment of 0.8, the difference in the R^2 is minimal (increase of 0.02), but an improvement is visible.

When the adjustment is made smaller (which is having a bigger influence of the cost value of the routes) the following results are possible. 0.4 adjustment creates a R^2 value of 0.79, a significant increase relative to the 0.8 and the no adjustment

---

5 https://github.com/DaveyOldenburg/Reversehouseholdwasteflowmapping/blob/master/
Correlations.R
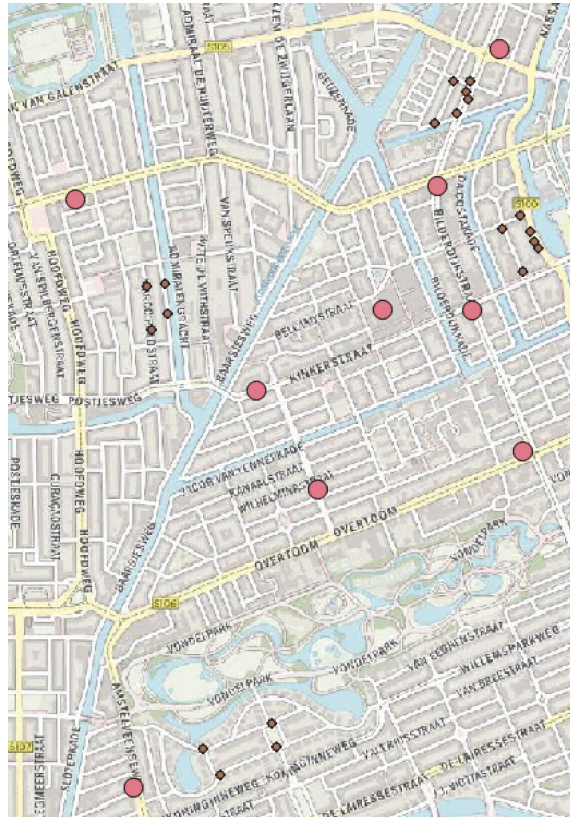
Figure 4.28: Trashcans picked for the validation, the big pink circles represent the centre points of the hotspots

values. When 0.1 adjustment is used, a R^2 of 0.84 is found which is an even bigger increase.

Concluding from the different R^2 values, the adjustment of 0.1 is predicting the increase of trash weight in a trashcan the most. However, an adjustment of 0.1 means that only 10% of the cost values of the raster are used to calculate the LCPA. This is creating a path that almost always goes into the direction of the hotspot.

$y = 935.22\,x + 3078.3$
$R^2 = 0.6414$, $p < 0.0001$
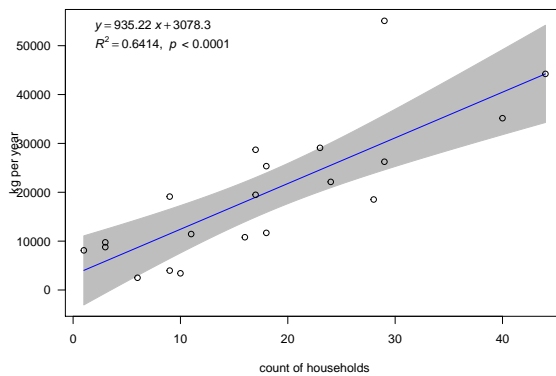
(a) Correlation when only minimizing the distance

$y = 926.55\,x + 3371.3$
$R^2 = 0.65936$, $p < 0.0001$
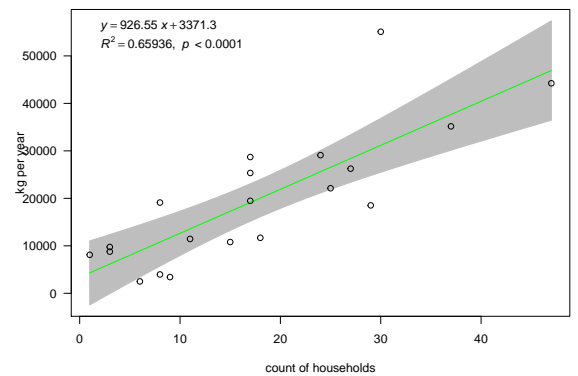
(b) Correlation with adjustment of 0.8 of cell value

Figure 4.29: Correlation plots for first two of the tested adjustments
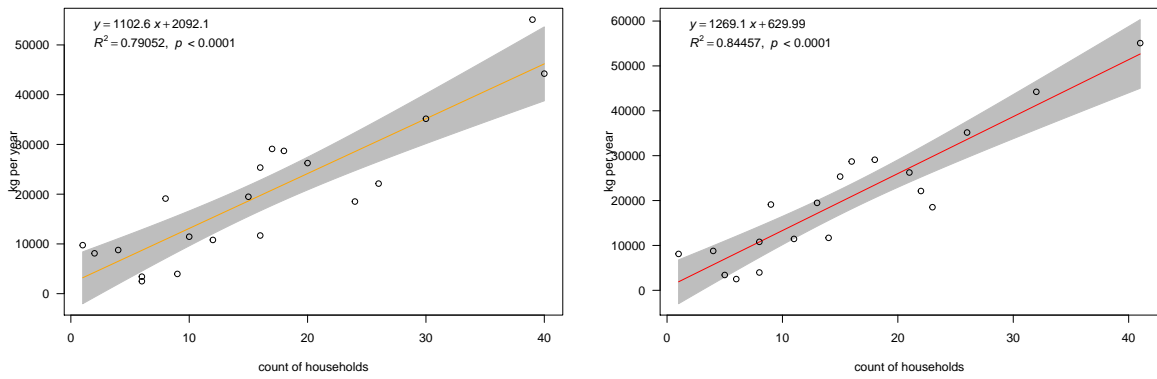
(a) Correlation with adjustment of 0.4 of cell value   **(b)** Correlation with adjustment of 0.1 of cell value

Figure 4.30: Correlation plots for last two of the tested adjustments

Bigger buffer size

For the sake of validation method verification, an extra test is executed. The buffer size to find the households around the trashcan are chosen so that all the households which belong to a trashcan are found. Since a bigger raster adjustment might result in larger paths to these trashcans (the adjustment makes the cost value of the path lower, causing longer paths). The extra test is performed with a buffer size of 200 meters instead of 130 meter, on the adjustment values 0.4 and 0.1. Only the smaller adjustment sizes are chosen because these are the ones likely to create longer paths which might fall out of the buffer.
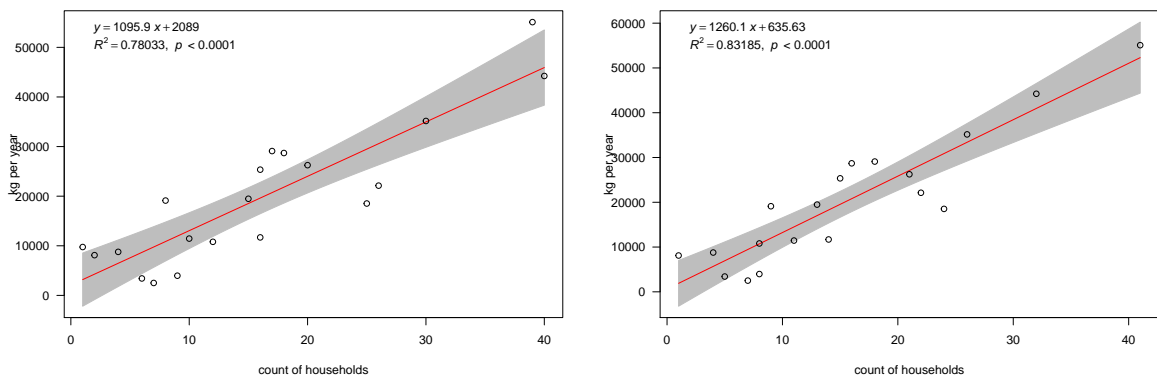
(a) Correlation with adjustment of 0.4 of cell value in a bigger buffer   **(b)** Correlation with adjustment of 0.1 of cell value in a bigger buffer

Figure 4.31: Correlation plots for two raster adjustments with a bigger buffer size

Figure 4.31 is showing the correlations between the bigger buffer sizes and therefore more households. While expecting a difference in the R^2 values, the result is similar to figure 4.30. When comparing the tables behind the plots, only one of the 20 counts for the trashcans increased, The rest stayed equal. The high correlation is therefore not caused by a mistake in buffer size.

Extra adjustment

Since the smallest adjustment value among the tested values has the highest R^2 value, the counterpoint of the increase in R^2 has not been found. To test if more

adjustment always results in a higher positive correlation, an extra adjustment value is implemented. By running the algorithm with an adjustment value of 0.01, the influence of the LCP raster is made much lower and the influence of the hotspot much higher, than the previously tested values.
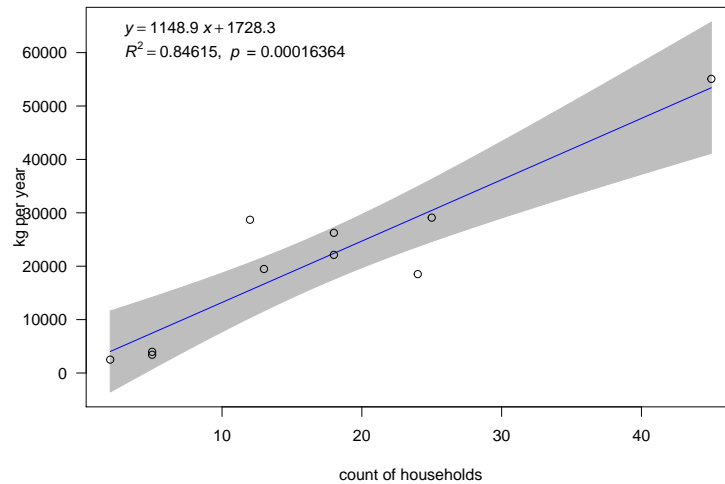
Figure 4.32: Correlation with adjustment of 0.01 of cell value

Figure 4.32 shows the result of the testing of the extra adjustment. The predicting capabilities are similar to the previously mentioned adjustment values. Something important to notice is the decrease of points in figure 4.32, only the trashcans inside the test dataset are considered, therefore the method assigns households to trashcans which are not in the test dataset. This is resulting in a higher p-value and less values to base the R^2 and p on. The increase in the p-value is so high, that no conclusions can be drawn from the R^2 value. However, since the R^2 is similar to the R^2 of 0.1 adjustment, there can be mentioned that lowering the adjustment value lower than 0.1, does not improve the prediction capabilities of this method more. Therefore 0.1 is still the adjustment with the most promising R^2 value, considering that the p-value in 0.1 is valid.

Mapped result

To be able to determine which of the adjustments is the best to apply in the method household-to-trashcan, the results of adjustment 0.1 and 0.4 are mapped to visualize the differences.

Due to lower adjustment values, longer paths are expected. But since Amsterdam is quite dense and the trashcans are equally spread over the city, all of the directions have a trashcan present. Therefore the path is not necessarily much longer, but the trashcan in the direction of the hotspot can be considered the more often used disposal.

In figure 4.33 the results of 0.1 adjustment and 0.4 adjustment are mapped. As expected, the paths are not longer in the 0.1 approach, but some households go to another trash location. The correlations suggest that the 0.1 method is better to predict the weight of the trash. Something to remark is the small difference between the two, only five footprints go to another trashcan.

Conclusion

This validation is a combination of an optimization and a validation. The hotspot direction is improving the result, higher R^2 values are visible when the direction

Figure 4.33: Overlay of the result of 0.1 (in orange) and 0.4 (in green) adjustment

preference is made stronger. When taken an extra adjustment of 0.01, the R^2 increases a little, but is similar to 0.1. Thus the hypothesis from section 3.3 can be considered approved. The most suitable adjustment value for the hotspot direction is 0.1, since 0.1 is having the highest R^2 value (with a valid p-value). This is the adjustment used to create the results elaborated in the result section 4.5. The comparison is visualized in table 4.4.

Adjustment value	Buffersize 130 meter	Buffersize 200 meter
No adjustment	0.64	-
0.8	0.66	-
0.4	0.79	0.78
0.1	0.84	0.83
0.01	0.85	-

Table 4.4: R^2 values of the different adjustment values compared

4.4.2 Household to retailer

As mentioned in section 3.4 there is no validation method for the method household-to-retailer. The validation method described in this section is not implemented and a suggestion for an approach when certain data would be available.

The same approach as the method household-to-trashcan can be followed to validate the results. When a particular store is picked, all the households inside a significant buffer should be found. Since the travel time is set on 10 minutes, the buffer should be at least 10 minutes times 14,4 $kmph$, considering that the method is using the same approach. The importance values of the found households for this specific store can be summed and correlated with a store depended value that represents the attraction of the considered store. This value can be the revenue or the

mass of sold products in a certain time frame. The correlation should be a positive relation.

4.5 RESULTS

In this section the result of the method household-to-trashcan and the method household-to-retailer are discussed. First a summary of the method is given, following with a mapping result and concluding with the file structure of the output. The results elaborated here are end products of the two methods, the testing before is discussed in the section 4.3. However the process is iterative, the intermediate steps are discussed before.

4.5.1 Household to trashcan

The method household-to-trashcan is created to find the used trashcan of a specific household. By using a LCPA of multiple considerable trashcans the used trashcan is found. An additional influence factor is inserted, the direction towards the hotspot of the city. After the validation can be concluded that this approach is convenient. The optimization shows that the more influence this factor has, the better the result can predict the mass of the trashcan waste.

Results on the map

After the used trashcan is found, the result can be mapped for the two test building blocks, as shown in figure 4.34. The walking path between every accommodation object (household) in the test area towards the used trashcan is created. For households that lie within the same building footprint, the line is drawn multiple times (every household has a geometry for the path). In figure 4.34b the paths are easier to distinguish than in figure 4.34a, mainly due to the increase in the number of households in the building block and the smaller foot prints of the buildings in figure 4.34b.

   Something else to remark is the clear structure of the raster visible in the paths. Since only the paths are compared and connecting the trashcan to the households is the main purpose of this method, the raster structure of the paths can be neglected. A more realistic smooth walking path would have represented a more human like path, but the paths give a realistic indication of what the actual path could have been.

   In this method not only the closest residual waste trashcan is found, but also the paths and locations of the other types of waste. Therefore figure 4.35 shows the mapping of the paths to paper waste (figure 4.35a) and the paths to plastic waste (figure 4.35b). Something to remark is the missing paths in a few households in figure 4.35b, this is due to the households which are connected to a residual trashcan that is containing a plastic trashcan at the same location. These households do not need an extra connection to plastic waste, because it is located at their used trashcan. In figure 4.35a no missing households are present, apparently non of the households are connected to a residual trashcan with a paper waste container already present.

Result file structure

The algorithm of method household-to-trashcan, writes a file in *.tsv* format in the structure visualized in table 4.5. The main geometry is the path to the used residual waste trashcan, followed by the accommodation object *id*. This *id* corresponds with the accommodation object in the BAG and therefore the results of the method is easy to connect to the BAG dataset. The *length* represents the actual length of the line

(a) Results of the first method on the Nassaukade

(b) Results of the first method on the Prinshendriklaan

Figure 4.34: Results on the test area

(a) Walking paths to paper waste container Prinshendriklaan

(b) Walking paths to plastic waste container Nassaukade

Figure 4.35: Results of additional trashcan walking paths

string in RD coordinates, thus not the cost value (which is only used to compare the paths). The location *id* is the location of the used trashcan, corresponding with the municipality dataset containing the measurements of the trashcans. The other attributes contain the geometries of the different routes to additional trashcans. These can be empty when the waste type is already covered in the used trashcan.

Attribute Name	Example value
geometry	MULTILINESTRING
accommodation object id	0363010000781096
length	34.24
location id	318136
geometry paper	MULTILINESTRING
location id paper	318136
geometry glass	MULTILINESTRING
location id glass	318136
geometry plastic	MULTILINESTRING
location id plastic	318136

Table 4.5: File structure of the result of method household-to-trashcan

4.5.2 Household to retailer

The method household-to-retailer is created to find the used retailers by a particular household, to give insight in the possible retailers of waste in the food waste chain. By creating an isochrone area, the retailers within a specified travel time of the households can be found and enriched with an importance value. Since no validation method is possible for this method, the results discussed in this section remains a demonstration.

Results on the map

Since this method is developed to map a connections of waste flow nodes, this section is giving an indication of how the isochrone areas are shaped. The outcome of this method is a list of retailers and their importance value. In figure 4.36 an isochrone area is mapped for one of the households. Important to notice is that this isochrone area is one of the three rings, the colored area is reachable in 3.3 minutes, while the size of the total isochrone is 10 minutes. Therefore the list of retailers which are reachable in 10 minutes is quite long.

Figure 4.37 shows the result for one of the households. The green dot represents this accommodation object in one of the test building blocks. The other dots represent the food retailers in the surrounding of the household which are covered by one of the three created isochrone areas. The color of the retailers visualizes the importance value of this retailer, the more red the color, the more influence this store has on the waste production of the household. Looking back in the dataset, two of the red dots represent big Albert Heijn supermarkets. These are expected to have a big influence and the method returns this as well, while not neglecting the smaller food stores, but by giving them less importance.

Result file structure

The algorithm of method household-to-retailer, writes two files in *.tsv* format in the structure visualized in table 4.6 & 4.7. The file represented with table 4.6 is to give an impression of which retailers are found within which of the isochrone areas. The lists of retailers *id*s can be extensive. The file represented with table 4.7 has a new line for every retailer for every accommodation object. The geometry can be added
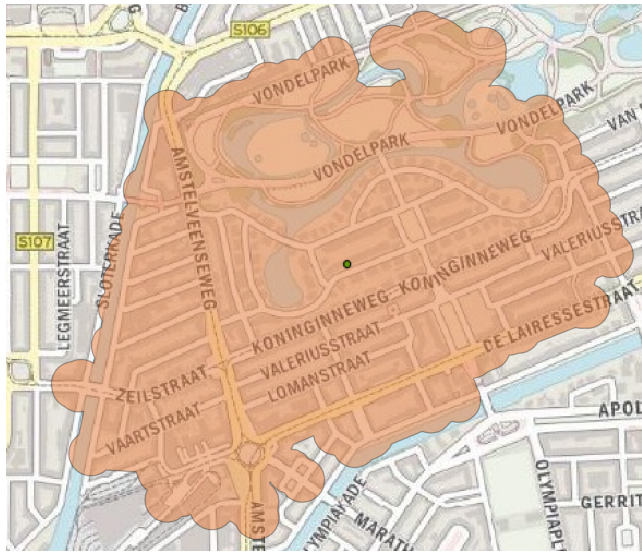
Figure 4.36: Isochrone area of 3.3 minutes biking around one of the accommodation objects
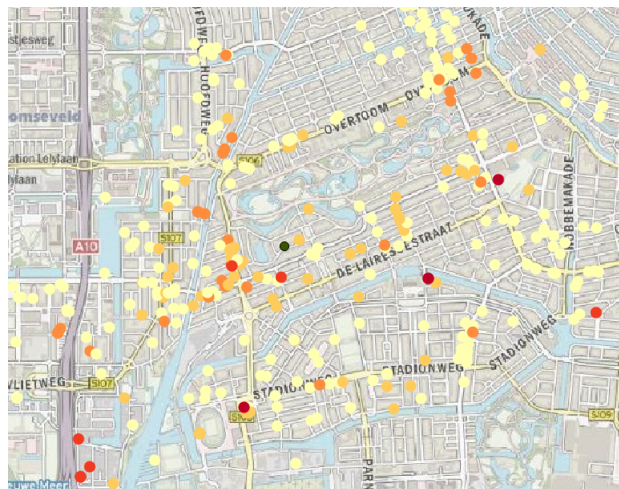

Figure 4.37: All retailers which can be considered usable by the specified household, the more red the color, the more influence the food retailer has according to the method.

easily by connecting on the key *retailer id* from the retailer data set, to produce a result such as 4.37.

Attribute Name	Example value
geometry	POLYGON
accommodation object id	0363010000781096
retailer id (isochrone area 1)	14414,14364,14340
retailer id (isochrone area 2)	14414,14364,14340
retailer id (isochrone area 3)	14414,14364,14340

Table 4.6: First file structure of the result of method 2: household to retailer

Attribute Name	Example value
accommodation object id	0363010000781096
retailer id	14414
influence	2.33

Table 4.7: Second file structure of the result of method 2: household to retailer

4.6 RELATION BETWEEN THE TWO METHODS

This section is discussing the relation between method household-to-retailer and method household-to-trashcan. Since the household is used as starting point for both of the methods, it is also the connection between the nodes in the waste chain. The used trashcan is found for a specific household, as well as the retailers that are potentially used by this specific household. Therefore the possible retailers of waste are found and these can be followed via the households to the trashcan.

As described in the methodology (chapter 3), both of the methods have different approaches based on the mind set of residences. While method household-to-trashcan is based on a time minimization, because it is a mandatory trip. Method household-to-retailer is mandatory, but residences are willing to travel further for better quality food (Dunkley et al., 2004), this results in higher uncertainty. Where in the first method one trashcan can be designated as the used disposal. The latter method does not have this possibility, because of the willingness to travel further and personal retailer preference.

Now that the mass of the trashcan is available and connected to the households that use this trashcan, a calculation of the produced waste per household can be implemented (see section 4.7). It seems logical to do the same calculation towards the retailers used by the household, to find where this waste is coming from. However, the material in the flows are different. Where in the method household-to-trashcan connection the material is domestic waste. In the connection household to retailer the material can be anything consumable. Another reason why the mass of the trash of a household is not directly related to the retailers is because it is not known what happens within the households. A big part of the groceries are eaten, stay in the house or are disposed on other ways (see figure 2.1). Basically, the behaviour of a specific household cannot be fully predicted. However, this thesis is giving a start to predict more of this behaviour.

CE is stated as the main motive for this research. Now that the connections are created between the households that use a specific trashcan, as well as which food retailer this household is using, new implementations are possible. For example, the composition of the trash in the trashcan can be related to the causing households, but also the connection to the food retailers is made, therefore the food retailers selling the products in the trashcan be identified. When a significant mass of plastic waste is found, the retailer selling a lot of plastic can be targeted. To take this a

step further, trash tax is mostly payed by households, but it is highly depending on the offer of the food retailers. If more nodes were connected, even further in the chain. If the created connection can be made more accurate, taxing division can be questioned and re-assessed. Leading to more aware behavior and handling of non-circular materials both in residual en commercial environments.

Another possibility is monitoring the composition of the trash and the relation to the different retailers used. For example, if a retailer packs its vegetables in plastic, can a relation be found between more plastic and less food waste? In this way, tests can be executed to see the value of different packaging and different handling of food products. Barlow and Morgan (2013) states that beef has a CO^2 equivalent of 14.8 per kilogram, only 0.5 percent of this value is coming from the plastic packaging. Wasting the product has more impact on the environment than the packaging. The plastic packaging can extend the shelf life of meat from 7 days to 21 days (Karabagias et al., 2011).

4.7 IMPLEMENTATION: MASS OF FLOW

This section emphasizes on the implementation of the mass of the flow. Since all the households used by specific trashcan are known as outcome of the method household-to-trashcan, the waste inside this trashcan is coming from the households related to this trashcan. As stated by Parizeau et al. (2015), household size is positively correlated with total waste weight. Therefore, the average waste volume per m^2 is calculated. Where after the total mass of waste coming from the household can be determined.

Following the total mass of waste for the household. The importance values of the retailers for this specific household can be used to divide the generated waste over the retailers. Important to realize is the temporal dimension of this mass. The waste mass of the trashcan is calculated for a full year, thereby is the waste coming for the household likewise based on a year. The food bought at the retailers is consumed in the households before being disposed, spreading the generated waste by the households over all the retailers is not the most accurate way to predict their contribution. However, it can give a suggestion of the retailers which have major impact of the trash composition of a specific trashcan.

To visualize the implementation, figure 4.38 & 4.39 are created. The work-flows show the processing of the outcomes of the method household-to-trashcan and method household-to-retailer and their transformations to derive the outcome of the waste mass per household. First figure 4.38 discusses to calculate the mass of waste per year per household, where after figure 4.39 discusses to calculate which retailers are the source for the household waste.
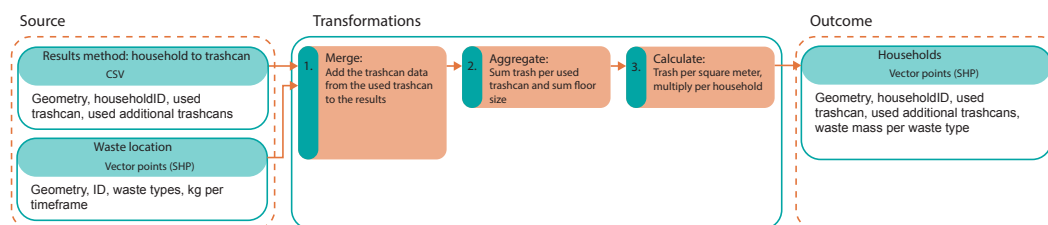
Figure 4.38: Overview of the needed steps to achieve the waste mass produced by a household, derived from method household-to-trashcan.

Mass of flow, household to trashcan

Source

The first source to achieve the required result is the result file of method household-to-trashcan, which contains information about which households use which trash-can. The second source are the trashcans from the municipality dataset, containing information about the mass of the different waste types in all trash locations.

Transformation 1: merge

First, the attributes of the trashcan (e.g.mass per year per waste type) are merged with the results file. Hereby, the mass per year for the used trashcan per waste type is known.

Transformation 2: aggregate

Second, by aggregating on a trashcan id, the sum of the floor size of all the households using this trashcan is calculated. The total sum of waste in this trashcan is calculated for the different waste types at this location.

Transformation 3: calculate

Finally, after calculating the sum of the floor size of all the households, the waste per m^2 is calculated. By dividing the total waste per type per trashcan by the total floor size of the households using this trashcan, the waste per m^2 is achieved. Afterwards, the waste per m^2 can be multiplied by the floor size of the households to achieve their contribution to the specific trashcan.

Outcome

The transformations result in the following outcome: the households, enriched with the mass of the trash for this specific household.

Mass of flow, household to retailer

By calculating the mass of waste coming from a household in the previous section, the next step is to use this mass to calculate which retailers contribute how much to this mass. This can be done based on the previously calculated importance values. For this thesis is assumed that 100 percent of the waste coming from a household is derived from the retailers assigned to this household in method household-to-retailer. Since the importance value is a percentage, the calculation of the contribution of a specific retailer can be calculated as written below.
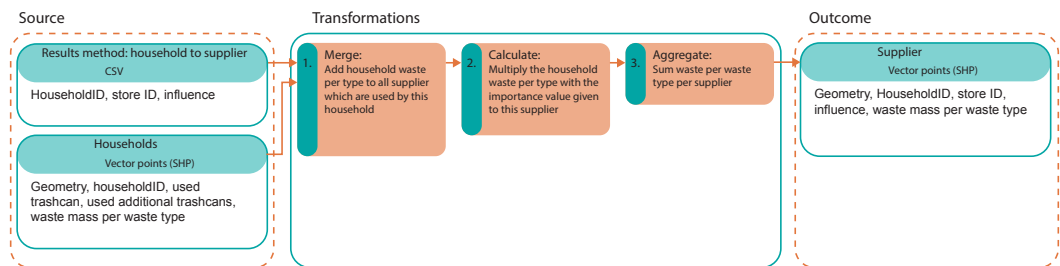
Figure 4.39: Overview of the needed steps to achieve the possible contribution of household waste by a specific retailer, derived from method household-to-retailer.

Source

The first source needed to achieve the desired result is the outcome of the method household-to-retailer, this dataset contains all the retailers and their importance values. The second source is the outcome of the previously described mass of the flow calculation.

Transformation 1: merge

First, the mass of the different waste types is added to every line in the results of household to retailer dataset which holds the same household id.

Transformation 2: calculate

Secondly, the importance value of the retailer is multiplied with the mass of the different waste types. Since the total importance value for every household is 100%, all the household waste coming from this household is distributed over all the considered retailers.

Transformation 3: aggregate

Finally, all the retailers have a waste mass coming from all the different households. By summing all the values belonging to one retailer, the contributed waste mass per type is found for this retailer.

Outcome

The outcome is all the retailers with their total contribution to household waste separated in the different waste types, according to the method used in this thesis.

4.8 RESULTS: MASS OF FLOW

In this section the result of the method: mass of the flow is presented. First a mapping result is shown, concluding with the file structure of the output. Second, the household waste mass is discussed, where after the retailer contribution is elaborated.

Results on the map

To map the results, the connections made with the method household-to-trashcan and method household-to-retailer are used to distribute the known mass in the used trashcan. Figure 4.40 shows all the households, the more red the color, the more waste is coming from this household. The average kg per household per year according to the generated dataset is 180kg of residual waste. According to Municipality Amsterdam (2015a), there is an average of 690kg per year per household for all waste types together.

The contribution of the retailers is mapped in figure 4.41, the more red the color, the bigger the contribution. According to the generated data, the Albert Heijn supermarket is contributing for 24000 kg of residual waste per year. Please note that due to the smaller testing area, not all households using this retailer are found, therefore the mass can be lower than when the methods were applied on the entire municipality.

Result file structure

The method discussed in this section consists of two parts, the mass of the trashcan to the household and the mass of the household to the retailer. The input is derived from the method household-to-trashcan and household to retailer. The mass of the flow method writes two *.shp* files. The first based on the first connection and the latter on the second. Table 4.8 visualizes the output of the first connection. The trash in the trashcans is distributed over the households which use this trashcan.

Table 4.9 visualizes the output of the second connection. The waste coming from the households, summed for a specific retailer.

Figure 4.40: Result of the distributing of mass of the trash over the households. More red means more waste generated by this household
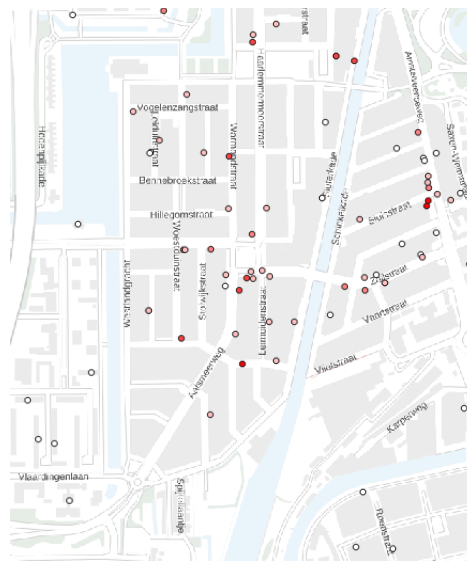
Figure 4.41: Result of the distributing of mass of the trash over the retailer. More red means more waste coming from this retailer.

Attribute Name	Example value
geometry	POINT
accommodation object id	0363010000781096
location id trashcan	318136
location ids additional trashcans	318136
waste mass per location id	1200kg

Table 4.8: File structure of the results the mass method of method household-to-trashcan.

Attribute Name	Example value
geometry	POINT
accommodation object id	0363010000781096
store id	21538
influence	4%
waste mass per waste type	400kg

Table 4.9: File structure of the results the mass method of method:household to retailer.

4.9 IMPLEMENTATION: MAPPING

To create an overview of the data gathered with the methods described above, this section is describing and mapping the previous methods. By creating a spatial Sankey diagram, showing the different paths to different waste types, showing the waste contribution of retailers and creating areas which belong to a certain trashcan, mapping of the achieved results is presented.

Method household-to-trashcan

The method household-to-trashcan enriched with the information derived from the method mass-of-the-flow, are implemented here. By picking a household, the data available is presented in different manners. First, all the paths to the different waste types for a household are mapped. Second, a spatial Sankey diagram is created visualizing the kg of waste per year from this household to the trashcans. Finally, a different approached is used, the creation of convex hulls of the households which belong to a specific trashcan. 'The convex hull of a finite point set S in the plane is the smallest convex polygon containing the set' (Kirkpatrick and Seidel, 2005). Meaning that if all the households belonging to a trashcan are visualized as points, the convex hull is the minimum polygon containing all the households. By mapping these polygons, the area using a specific trashcan is mapped.

Method household-to-retailer

The method household-to-retailer enriched with the information derived from the method mass-of-the-flow, are implemented here. The contribution of the retailers are mapped by using the mass of the flow method.

4.10 RESULTS: MAPPING

In this section the result of the implementation of mapping is presented. First, the maps belonging to the method household-to-trashcan are elaborated. Finally, the maps derived from the method household-to-retailer are presented.

Method household-to-trashcan

The first map generated from this method is mapping the different paths to all different waste types available in Amsterdam. Figure B.1 shows a household and its paths. A minimum of 1 and a maximum of 4 paths are created. This example is containing 3 paths, since glass and paper waste are available at the same location.

The second map generated from this method is a differentiation of the first map. By adding information from the method mass-of-the-flow, a spatial Sankey diagram is generated. Thicker lines means more waste going towards this waste type location. Figure B.2 shows this map.

Finally the map with the convex hull belong to a specific trashcan is created. Figure B.3 shows areas in which the households are that are using this trashcan. In this way the scope of a trashcan is shown.

Method household–to–retailer

The map generated from the this method is showing the contribution of a retailer to residual household waste in kg per year (Figure B.4). These kg per year are derived from the method mass-of-the-flow.

5 | CONCLUSION

The main aim of this thesis is to connect nodes in the waste chain. To draw the conclusion of this research, first the sub-questions introduced in the introduction (section 1.3) are answered, where after the main question is answered. In section 'contribution to research field' the problem statement is elaborated with the knowledge gathered in this thesis.

5.1 ANSWER RESEARCH QUESTIONS

In the introduction the main question is stated as follows:

How can household waste flows be mapped in reverse based on their environment?

Deriving from this main question, multiple sub-questions are generated. The following section answers those sub-questions.

5.1.1 Sub-questions

Theory: Why is it necessary to map waste flows? How has the mapping of (waste) flows done in the past? How are waste flows visualized/mapped?

This first set of sub-questions focuses on the related work done within the field of waste flows and flows in general. The need for the mapping of waste flows derives from the missing of potential behaviour data of households. A wider elaboration of flows is necessary. By giving insight in the flow of waste through a city, while also identifying sinks and underlying flows that might have stayed hidden, a MFA can expose steps for prevention or recycling policies. Since this thesis is adding a more detailed approach to mapping waste flows, new possibilities arise. Household level adds the possibility for a more detailed recycling policy approach can be created. The connected nodes in this thesis establish better insight in the result of retail behaviour on trash composition. due to the introduction of the level of the household, specific households and specific retailers can be addressed.

MFAs allows one to track specific material flows through a socio-economic system and provide a method to identify sources of generation and account for hidden flows and sinks that could remain unexplained, when only looking at the end of the flow (Brunner and Rechberger, 2004).

A Sankey diagram visualizes the flow that can represent the movement of a material or energy. The nodes in the Sankey diagram are representing intermediate steps of the flow. The endpoint of the diagram is either representing a transformation, transportation and/or storage (Brunner and Rechberger, 2004).

The two methods above do not necessarily have a spatial dimension. They are visualizing or analyzing the flows, while this thesis is focusing on the spatial dimension. Multiple papers add spatial dimension to existing flow visualizations. Roy et al. (2015) developed a method to create a SAMFA, this model is able to 'draw the concepts of urban metabolism and links these with an ability to visually represent the implications of spatially and temporally differentiated scenarios related to the accumulation of construction material flows and associated energy use in the housing sector'. Font Vivanco et al. (2012) developed a model for SMFA. The spatial dimension of the method was added by tracking all transport actors.

The WFM model from Kurdve et al. (2015) is using seven steps (figure 2.3). By mapping collection points and map internal waste logistics on site, the collection points can be connected to their out of site directions.

For this research, the previously mentioned visualization and mapping methods are mainly to indicate the possibilities of the potential outcome. The possibility of MFA to identify sinks is an example of the possible usage of the method of this thesis, but this thesis is focused on specific households. Most of the literature relates to this topic, but is not applicable due to the human behaviour influence of the method in this thesis to achieve the needed level of detail.

Design: How can methods be developed to add a spatial dimension to the household waste flows?

Since this question is rather broad, this sub-question is again separated in three questions. The first question is answered in relation to the method household-to-trashcan and the latter for the method household-to-retailer.

Which datasets are needed to develop such methods?

As described in the methodology of this thesis, multiple inputs are needed to derive to the goal of this research. The inputs are listed below and are discussed separately.

1. *Households*

   The households are needed to represent the consumption in private household node, visualized in the system diagram by REPAiR (figure 2.4). This can also been seen as the start point of the path towards the used collection of non hazardous waste. The households are derived from the BAG dataset; the accommodation objects represent every household in the Netherlands. The preprocessing is implemented in section 4.2. The table below shows the necessary attributes for this input.

Necessary attributes
Point geometry
Object ID
Function
Floorsize
ID of footprint

Table 5.1: Needed attributes in households input

2. *Trashcans*

   The trashcans are needed to represent the collection of non hazardous waste. The municipality provided this research with their dataset, which is also containing measurement data for every trashcan, as well as the location of these containers. The attributes listed below are necessary the execute the algorithm.

Necessary attributes
Point geometry
Object ID
Available waste types
Kg per timeframe

Table 5.2: Needed attributes in trashcans input

3. *LCP raster*

   The connection between the two nodes described above is shaped by the environment of these connections. The environment is represented by the BGT

dataset. The BGT is a detailed map of The Netherlands, containing all objects like: roads, houses, pedestrian paths, water and more. This dataset is resulting in a LCP raster based on walkability values for pedestrians. With this data input, pedestrian behaviour is predicted. The raster is a continuous field where every cell has a value. To obtain these values, a dataset (like the BGT) is necessary.

Necessary attributes
Polygon geometry
Classification of polygons

Table 5.3: Needed attributes as input to create the LCP raster

4. *Hotspots*

   The hotspots of the city are simulated in such a way that they are similar to the *area of interest* created by Google Maps (2019). By using density based clustering from the actors of food waste dataset achieved from the REPAiR project (REPAiR, 2018), clusters similar to the Google approach are created. Other datasets representing the interesting areas of a city can be used to create similar hotspots as the ones presented in this research.

Necessary attributes
Point geometry

Table 5.4: Needed attributes as input to create the city hotspots

For the method household-to-retailer, additional datasets and inputs are necessary.

5. *Network*

   Since isochrone areas are created, a road network is needed to produce the desired output. The road network is downloaded from OSM via the OSMnx Python library (Boeing, 2017). The network is specified for a specific travel time and a specific travel mode. This approach is applicable on every city where OSM is available.

6. *Food retail*

   This dataset is obtained from the REPAiR project (REPAiR, 2018). Appendix A is showing the usage details, since the project purchased this dataset. The retailers of food dataset represents all the actors in the food waste chain. Therefore all the supermarkets, restaurants, hotels and more. For this research, the retailers of food are the main component of this information.

Necessary attributes
Point geometry
Object ID
Retailer Name
Function
Address
Floorsize

Table 5.5: Needed attributes in retailers of food retail

Which trashcan is used by a household based on factors of their environment?

This sub-question is answered by the method household-to-trashcan. By using a LCPA starting at a specific household and creating paths to trashcans around it,

the path with the smallest cost value is found. This specific cost value for every path is influenced by factors of the environment. These factors are: objects from the BGT dataset (see table 5.6) and a preference walking direction towards a city hotspot. This approach is also taking into account the direct environmental factors (such as a pedestrian path) and factors from further away such as the influence of a city hotspot. In this way the trashcan for a specific household is found, including the walking path.

BGT object	description
Pedestrian path	Sidewalks and pedestrian area
Alternative path	Green space, parking spaces, cycling paths
Road path	Roads, rails, bus lanes
Private area	Private gardens and terraces
Buildings	Buildings and other structures
Water	Canals and open water

Table 5.6: BGT objects

Which food retailers are used by a specific household based on factors of their environment and characteristics of the retailers

This sub-question is answered by the method household-to-retailer. By using multiple isochrone areas, retailers of food are found as a point in polygon algorithm approach. The uncertainty of this method is quite large, due to the personal preferences of residents of the households. Therefore this method is including all retailers within the isochrone areas and giving them importance values based on distance and floor size of the retailer. Floor size can be seen as variables predicting the revenue of a store, which therefore relates to the probability a household choosing this store.

The methods described in the last two sub-questions are elaborating on how a spatial dimension is added to waste flows. By using the coordinates of the households, the node *the consumption in private households*, is directly spatially enriched. By inserting the used trashcans, they present the node *collection of non hazardous waste*. The connection between these nodes is added by using environmental factors like distance, hotspots and objects from the BGT. The node *retail and market* is added by finding considerable retailers and determining their importance. The results are the connection from retailers to households to used trashcan.

Validation: How can the municipality dataset on trashcan weight be used to validate the spatial dimensions of the waste flows?

To validate and optimize the method household-to-trashcan, the existence of a positive correlation is tested between the total number of households which are assigned to a trashcan and the mass of this trashcan per year. A high positive correlation between the two inputs is expected. After adjusting multiple manual parameters the highest positive correlation among the tested values is found and in this way the method is validated and optimized.

The method household-to-trashcan could be validated in a similar way, but due to missing data on (for example) the revenue of the stores, this method remains merely a demonstration and cannot be validated.

Mass of flow: How can the connections in the waste chain nodes be used to follow waste mass through the flow, based on the municipality dataset on trashcan weight?

The trashcan mass is known, therefore the reverse connection of nodes can pinpoint the households and retailers responsible for the composition of this trashcan. The household contribution is dependent on the floor size of the house, as found

in literature. The retailer contribution to this household waste is dependent on the importance value of the retailer (determined in this research).

The household contribution to the trashcan composition is relatively accurate, considering that the household is close to the trashcan and the floor size is taken into account. However, relating the household waste contribution to the retailers which sold the materials, is creating some uncertainty. The residences of the households eat, transport or store products before they (eventually) end up in the trash. Using the importance values of the retailers to see how much household waste is generated by them, is giving only an indication. Improvement of the two main methods results in a more reliable connection and therefore a better mass indication.

5.1.2 Main question

Based on these answers to these sub-questions, the main question can be answered:

How can household waste flows be mapped in reverse based on their environment?

Whereas REPAiR connected *collection of non hazardous waste* to *the treatment and disposal of non hazardous waste*, this thesis focuses on the reverse mapping of the nodes that happens before the phases that REPAiR focuses on. The nodes *collection of non hazardous waste* is connected to *consumption in private households* and in its turn *retail and markets*. This direction in the waste chain is named reverse.

To connect the households with their used trashcans, the BGT is used to simulate the environment and the city hotspots to predict a direction preference. To connect the household with their food retailer, literature about travel time for groceries is used to create isochrone areas which are suggesting the retailers used by this household. By including characteristics of those retailers a prediction is presented. Since the connection household to their possible retailers is not validated, this method is a demonstration of an approach.

As mentioned in the introduction, the application of the method is to be able to create policies to prevent waste instead of a reaction program. The method would give a possibility to implement the policies closer to the source and check the results further in the chain. The connections created within this thesis are a step in the direction of this possible approach. The validation of the method shows promising results, but to address specific households and retailers, extra research is necessary.

The methods are likely to work in other relatively densely built environments, however, additional adjustments for the methods to work with other data sources or less density of hotspots/trashcans need to be implemented.

The methods are applicable on other locations outside of the Netherlands, provided that the input data is containing all the necessary attributes. Trashcan used in Amsterdam are free to choose by inhabitants of the city. If another trash system is used, the method is not applicable.

5.2 DISCUSSION

For the first method household-to-trashcan, path minimization is the main predictor of this approach. Some bigger apartment buildings have a trashcan inside. Since this cannot be deduced from the used data in this thesis, this is neglected. The exit of a building is partly neglected, the algorithm finds the shortest path from the accommodation object point towards the boundary of the building footprint polygon. Since the exit location is neglected, the path can exit the building on the wrong location. For smaller buildings the differences in the path are negligible, but for bigger buildings containing multiple households, a different exit changes the starting point of the path and thereby the result of the method. Since the test area does not contain these types of buildings, these building-types are not tested.

Should this method be developed further, more data is required, such as the location of the building exits.

To create the paths the LCP raster is developed. The values in the raster are based on literature and tests. In theory, the paths generated with the raster predict pedestrian movement in a reliable way, but temporal aspects are not taken into account. When a car is parked on a parking spot, a pedestrian cannot walk over it, despite that the raster has a relatively low value at this location. The same applies for the crossing of the street. Pedestrian crossings are not present in the raster, however this is probably the place were pedestrians would cross. Include, most of the pedestrian crossings are located on the place were the pedestrian paths are closest together, therefore the LCPA creates the path there automatically, but it is not implemented in the method. To give a more reliable result of the generated paths, the a preference should be implemented on road crossings.

The method household-to-trashcan is validated within section 4.4.1. The correlation is determined between the sum of all the households which are assigned to a certain trashcan and the sum of waste collected from this trashcan. No control variables are used in this analysis. Therefore the effectiveness of the method can be influenced by unknown factors. To give a more reliable result, control variables with known influence on the choice of the trashcans should be implemented.

Since the method household-to-retailer cannot be validated, the implemented ideas in this method remain not more than an hypothesis. The store influence is based on the floor size and distance to the household. In this method three isochrone areas are created to find the different considerable food retailers of a household. An approach in which the areas are not limited to three areas would give a more accurate result. By calculating the distance to every individual retailer, the distance factor of the value can be calculated more accurately. The distance and floor size of the retailer were described by literature and were implemented in this thesis. If the revenue of the food retailer was known, a more reliable result could be produced, because insight in the attractiveness of a store is raised. A survey would give insight in the behaviour of the households, the outcome could be used to implement more factors that are influencing the choice of a retailer.

As mentioned in section 2.1, food retail is not the only source of household food waste. Since this thesis is focused on food retail, only considering physical store locations, other possible sources are neglected. Other possibilities such as, home grown food, takeaway food or temporary food markets should be implemented to give a more accurate explanation of the household waste composition.

The testing areas are specially chosen to represent the biggest group household types in Amsterdam and the rest of the Netherlands. Since these are only the biggest groups, there are more household types present in Amsterdam. After validating the method household-to-trashcan for the test area, the method should be applied on households beyond the test area.

5.3 CONTRIBUTION TO RESEARCH FIELD

As written in the introduction, the problem statement elaborates on a problem of missing information to tackle waste problems closer to the source. By geographically connecting nodes in the household waste chain, this thesis gives the first attempt to give a solution to this problem. Thereby, more insights in household waste behaviour are created and waste can be followed in reverse order towards the retailers of the material. Together with the REPAiR project, four nodes of the waste chain diagram are mapped (figure 2.4).

An indication of companies used by a specific household is generated. By approaching this method from the other direction, an indication of the households which are using a specific store is created. In this way, companies can be made aware of the range of their store and thus follow their product from store to house-

hold, from household to trashcan, from trashcan to waste treatment plant. When this approach is executed for multiple retailers, a policy change can be tested and afterwards implemented.

As mentioned in section 4.6, the established connection between the food retailer, households and used trashcan create an opportunity to analyze impacts on waste composition. When a retailers packs all its food in plastic wraps, can a differentiation of the waste composition be noticed? The established connections create a possibility to pinpoint the retailers which have an influence on the composition of this specific trashcan.

CE is 'a regenerative system in which resource input and waste, emission, and energy leakage are minimized by slowing, closing, and narrowing energy and material loops' (Geissdoerfer et al., 2017). This thesis gives insight in the location of the nodes in the material loop of waste. Therefore, the possibility of pinpointing the specific locations of improvement to slow, close and narrow the material loop of waste rises as possibility. By geographically connecting a part of the food waste material flow, the decisions in a node can be related to the impacts on other nodes, like the example in the previous paragraph. In this way, the results of the applied policies or ideas can be tested.

Finally, multiple weekly (or more than weekly) walking paths are simulated within this thesis. These paths give insight in frequently used routes by a pedestrians. The LCPA shows where shortcuts are used to shorten the path. Urban designers can use this information to optimize their walking paths and prevent deduction of public space by walking.

5.4 RECOMMENDATIONS ON FUTURE WORK

This thesis researches how households spatially relate to their trashcans and demonstrates how households spatially relate to their retailer. However, some of the methods still have their shortcomings. Multiple directions for future researches are presented.

- Since the method household-to-retailer cannot be validate, more research on this topic should be done. This thesis provides a direction and some approaches, but when additional data is available this method could be improved and eventually validated.

- The method household-to-retailer is developed for the most used travel mode in Amsterdam, the bike. Since the car is not considered as possibility, potential retailers are ignored. The possible use of the car should not be neglected in further research. Due to the probable expansion of the isochrone area and increase of potential retailers the uncertainty rises.

- The processing of the first method household-to-trashcan is done in the Python programming language. Since large arrays are adjusted and paths are calculated through this array, a more efficient programming language can reduce the processing time. Therefore implementing the idea of this thesis in for example C++ might result in faster response of the method.

- The scope of this thesis was determined on three nodes of the waste chain. The REPAiR H2020 project mapped two nodes and this thesis is adding another three. Future research can focus on the remaining nodes to map the full cycle of the material flow chain.

- Despite the promising results of the validation of the method using the municipality trashcan data, adding a survey in the test areas to get background information about the residents' waste behaviour would strengthen the method.

The outcome of the survey adds new input for a more promising outcome of both of the methods.

- As already described in section 5.2, the exit of a building is highly important for the method to find the used trashcan. The exit of the building is the starting point of the LCPA and therefore influencing the path that is chosen. No specific method is developed to find the exit of the building based on the position of the point representing the accommodation object. Further research or contact with the Dutch Cadastre about their placement of the points would increase the credibility of this method.

- Finally, this research has been conducted to give insight in human waste behaviour. Since households are one of the biggest contributors for food waste (Parfitt et al., 2010). The outcome of this thesis can help create new waste prevention policies based on specific locations

BIBLIOGRAPHY

Accorsi, R., Manzini, R., Pini, C., and Penazzi, S. (2015). On the design of closed-loop networks for product life cycle management: Economic, environmental and geography considerations. *Journal of Transport Geography*.

Alfonzo, M. A. (2005). To walk or not to walk? The hierarchy of walking needs. *Environment and Behavior*.

Ammerlaan, R. (2016). Prediciting microscale pedestrian route choice for optimising household waste container locations. Technical report, Wageningen University.

Barlow, C. Y. and Morgan, D. C. (2013). Polymer film packaging for food: An environmental assessment.

Bastein, T., Roelofs, E., Rietveld, E., and Hoogendoorn, A. (2013). Kansen voor de circulaire economie in Nederland. *TNO-rapport*.

BatchGeo (2019). BatchGeo.

Benoit, D. and Clarke, G. P. (1997). Assessing GIS for retail location planning. *Journal of Retailing and Consumer Services*.

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*.

Borgers, A. and Timmermans, H. (1986). A Model of Pedestrian Route Choice and Demand for Retail Facilities within Inner-City Shopping Areas. *Geographical Analysis*.

Bresenham, J. E. (2010). Algorithm for computer control of a digital plotter. *IBM Systems Journal*.

Brunner, P. H. and Rechberger, H. (2004). Practical handbook of material flow analysis. *The International Journal of Life Cycle Assessment*.

Bulkeley, H. and Gregson, N. (2009). Crossing the threshold: Municipal waste policy and household waste generation. *Environment and Planning A*.

Bureau Van Dijk (2019). Orbis.

CBS (2019). Woningvoorraad.

Chen, L. F. and Tsai, C. T. (2016). Data mining framework based on rough set theory to improve location selection decisions: A case study of a restaurant chain. *Tourism Management*.

Collischonn, W. and Pilar, J. V. (2000). A direction dependent least-cost-path algorithm for roads and canals. *International Journal of Geographical Information Science*.

CREM Waste Management (2016). Bepaling voedselverspilling in huishoudelijk afval Nederland. Technical report, CREM Waste Management.

Daamen, W. and Hoogendoorn, S. P. (2007). Experimental Research of Pedestrian Walking Behavior. *Transportation Research Record: Journal of the Transportation Research Board*.

Developers NetworkX (2019). NetworkX: Software for complex networks.

Dunkley, B., Helling, A., and Sawicki, D. S. (2004). Accessibility Versus Scale. *Journal of Planning Education and Research*, 23(4):387–401.

Eurostat (2014). Modal Share.

Fietsersbond (2017). Bicycle speed in Amsterdam.

Font Vivanco, D., Puig Ventosa, I., and Gabarrell Durany, X. (2012). Building waste management core indicators through Spatial Material Flow Analysis: Net recovery and transport intensity indexes. *Waste Management*.

Geissdoerfer, M., Savaget, P., Bocken, N. M., and Hultink, E. J. (2017). The Circular Economy – A new sustainability paradigm?

Geldermans, B., Wandl, A., Steenmeijer, M., and Furlan, C. (2016). D3.3 Process model for the two pilot cases: Amsterdam, the Netherlands & Naples, Italy. Technical report, REPAiR.

Google Maps (2019). Google street view.

Herzog, I. (2013). Theory and Practice of Cost Functions. *Proceedings of the 38th Annual Conference on Computer Applications and Quantitative Methods in Archaeology in Fusion of Cultures*.

Hoogendoorn, S. P. and Bovy, P. H. (2004). Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological*.

Hoogendoorn, S. P. and Bovy, P. H. (2017). Normative Pedestrian Behaviour Theory and Modelling. In *Transportation and Traffic Theory in the 21 st Century*.

Hoogendoorn, S. P. and Daamen, W. (2005). Pedestrian Behavior at Bottlenecks. *Transportation Science*.

Karabagias, I., Badeka, A., and Kontominas, M. G. (2011). Shelf life extension of lamb meat using thyme or oregano essential oils and modified atmosphere packaging. *Meat Science*.

Kirkpatrick, D. G. and Seidel, R. (2005). The Ultimate Planar Convex Hull Algorithm? *SIAM Journal on Computing*.

Kurdve, M., Shahbazi, S., Wendin, M., Bengtsson, C., and Wiktorsson, M. (2015). Waste flow mapping to improve sustainability of waste management: A case study approach. *Journal of Cleaner Production*.

Li, M. and Bailiang, Z. (2019). Discover the action around you with the updated Google Maps.

MacArthur E. (2013). *Towards a Circular Economy - Economic and Business Rationale for an Accelerated Transition*.

Municipality Amsterdam (2015a). Afvalketen in beeld: Grondstoffen uit Amsterdam. Technical report, Municipality Amsterdam, Amsterdam.

Municipality Amsterdam (2015b). Duurzaam Amsterdam. Technical report, Municipality Amsterdam.

Municipality Amsterdam (2016). Circular Amsterdam. Technical report, Circle Economy, Amsterdam.

Municipality Amsterdam (2019). Road Travel Intesity.

Murray, A., Skene, K., and Haynes, K. (2017). The Circular Economy: An Interdisciplinary Exploration of the Concept and Application in a Global Context. *Journal of Business Ethics*.

O'Sullivan, D., Morrison, A., and Shearer, J. (2000). Using desktop GIS for the investigation of accessibility by public transport: An isochrone approach. *International Journal of Geographical Information Science*.

Parfitt, J., Barthel, M., and MacNaughton, S. (2010). Food waste within food supply chains: Quantification and potential for change to 2050.

Parizeau, K., von Massow, M., and Martin, R. (2015). Household-level dynamics of food waste production and related beliefs, attitudes, and behaviours in Guelph, Ontario. *Waste Management*.

Quested, T. E., Marsh, E., Stunell, D., and Parry, A. D. (2013). Spaghetti soup: The complex world of food waste behaviours. *Resources, Conservation and Recycling*.

R.C. Lupton, J. M. A. (2017). Hybrid Sankey diagrams: Visual analysis of multidimensional data for understanding resrouce use. *Elsevier*.

REPAiR (2018). AMS Repair: Resource management in peri-urban areas: going beyond urban metabolism.

Roy, M., Curry, R., and Ellis, G. (2015). Spatial allocation of material flow analysis in residential developments: A case study of Kildare County, Ireland. *Journal of Environmental Planning and Management*.

Schmidt, M. (2008). The Sankey diagram in energy and material flow management: Part I: History. *Journal of Industrial Ecology*.

Searchinger, T., Hanson, C., Ranganathan, J., Lipinski, B., Waite, R., Winterbottom, R., Dinshaw, A., and Heimlich, R. (2013). Creating a Sustainable Food Future : A menu of solutions to sustainably feed more than 9 billion people by 2050. *World Resources Report 2013-14*.

Vignieri, S. N. (2005). Streams over mountains: Influence of riparian connectivity on gene flow in the Pacific jumping mouse (Zapus trinotatus). *Molecular Ecology*.

Vreugdenhil, C., Bregt, A., and Steenhuis, F. (2016). Stimulus Project: Smart Wasting in Amsterdam. Technical report, Wageningen University & Research (WUR), Wageningen.

Wallsten, B. (2015). Toward social material flow analysis on the usefulness of boundary objects in urban: Mining research. *Journal of Industrial Ecology*.

Wang, I. J., Savage, W. K., and Bradley Shaffer, H. (2009). Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander (Ambystoma californiense). *Molecular Ecology*.

Product name	Orbis Europe		
Update number	169		
Software version	129.00		
Data update	22/03/2018 (n° 16903)		
Username	a.wandl		
Export date	28/03/2018		
		Step result	Search result
1. All active companies and companies with unknown situation		69,573,928	69,573,928
2. World region/Country/Region in country: Aalsmeer (Netherlands), Almere (Netherlands), Amstelveen (Netherlands), Amsterdam (Netherlands), Beemster (Netherlands), Beverwijk (Netherlands), Blaricum (Netherlands), Bloemendaal (Netherlands), Diemen (Netherlands), Edam-Volendam (Netherlands), Gooise Meren (Netherlands), Haarlem (Netherlands), Haarlemmerliede Ca (Netherlands), Haarlemmermeer (Netherlands), Heemskerk (Netherlands), Heemstede (Netherlands), Hilversum (Netherlands), Huizen (Netherlands), Landsmeer (Netherlands), Laren (Netherlands), Lelystad (Netherlands), Oostzaan (Netherlands), Ouder-Amstel (Netherlands), Purmerend (Netherlands), Uitgeest (Netherlands), Uithoorn (Netherlands), Velsen (Netherlands), Waterland (Netherlands), Weesp (Netherlands), Wijdemeren (Netherlands), Wormerland (Netherlands), Zaanstad (Netherlands), Zandvoort (Netherlands)		953,452	538,120
3. NACE Rev. 2 (Primary codes only): 471 - Retail sale in non-specialised stores, 4711 - Retail sale in non-specialised stores with food, beverages or tobacco predominating, 4719 - Other retail sale in non-specialised stores, 4721 - Retail sale of fruit and vegetables in specialised stores, 4722 - Retail sale of meat and meat products in specialised stores, 4723 - Retail sale of fish, crustaceans and molluscs in specialised stores, 4724 - Retail sale of bread, cakes, flour confectionery and sugar confectionery in specialised stores, 4725 - Retail sale of beverages in specialised stores, 4729 - Other retail sale of food in specialised stores, 4781 - Retail sale via stalls and markets of food, beverages and tobacco products, 4789 - Retail sale via stalls and markets of other goods, 479 - Retail trade not in stores, stalls or markets, 4791 - Retail sale via mail order houses or via Internet, 4799 - Other retail sale not in stores, stalls or markets		4,429,202	14,667
Boolean search : 1 And 2 And 3			
		TOTAL	14,667

*Manually removed: 4719, 4729, 4789, 4791, 4799

Fiscal year end:31/03

Current search settings:
- priority given to the most recent accounts available
- exclusion of companies with no recent financial data and Public authorities/States/Governments

Definition of the Ultimate Owner:
- minimum percentage that must characterise the path from a subject Company up to its Ultimate owner: 50.01%
- UO can have no shareholder identified or all its shareholders have an unknown percentage

Figure A.1: Usage restrictions of food supplier dataset

B | MAPS

Figure B.1: Household and its paths to the different waste types

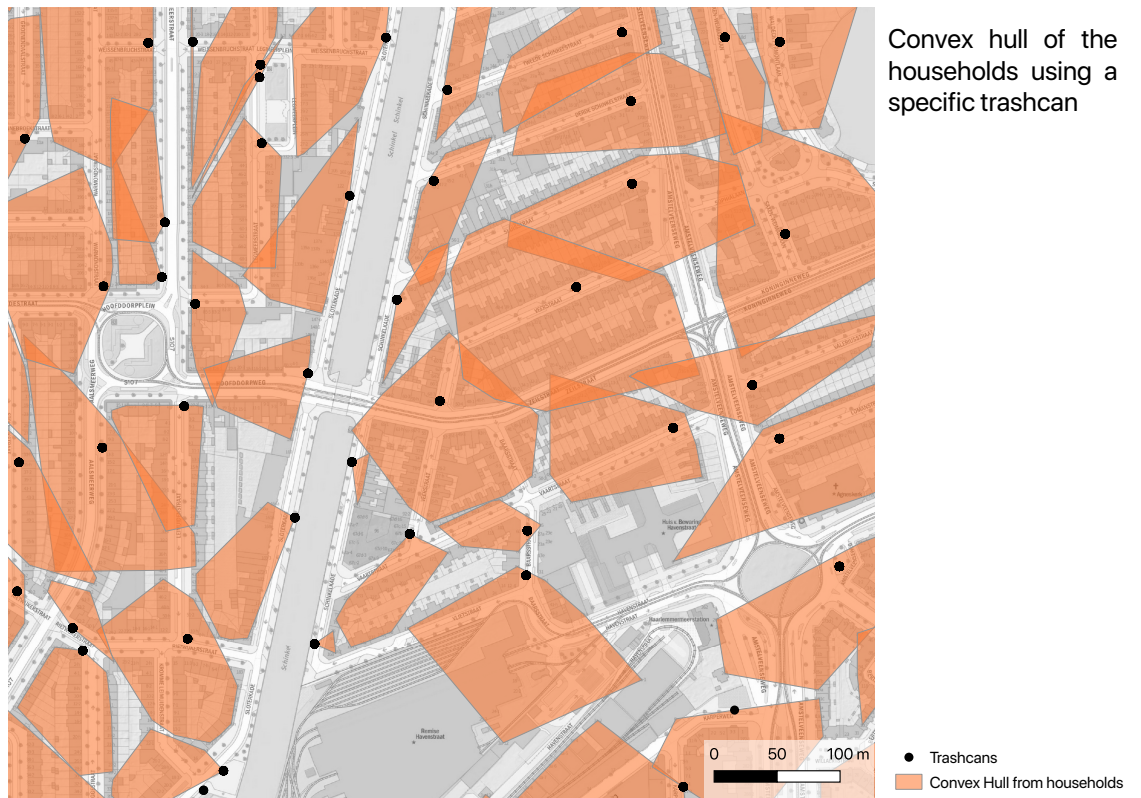**Figure B.2:** Household and its paths to the different waste types as spatial Sankey Diagram

Figure B.3: Convex hull of the households using a specific trashcan

Figure B.4: Retailer contribution to residual household waste in kg per year