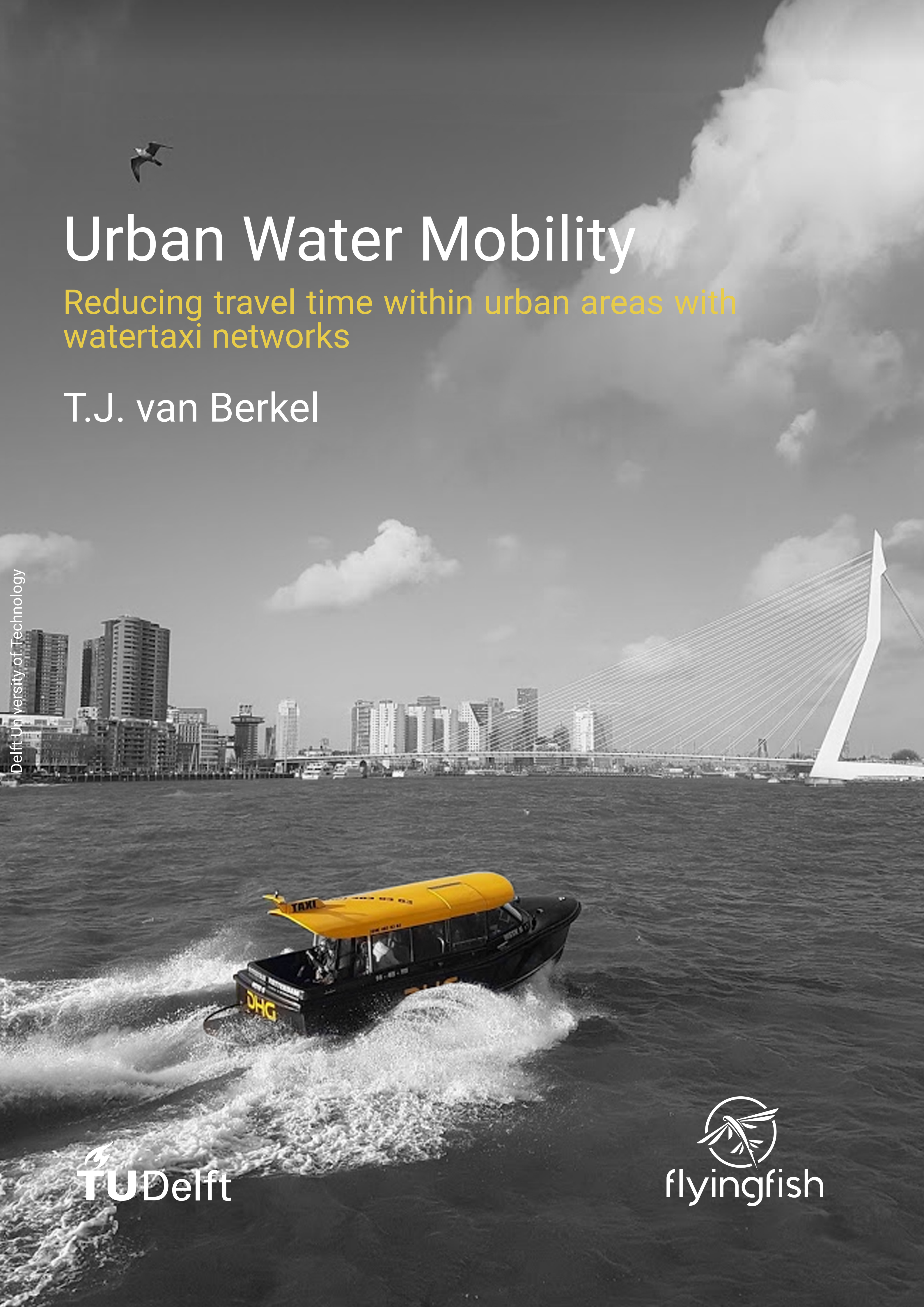
Urban Water Mobility

Reducing travel time within urban areas with watertaxi networks

T.J. van Berkel



Delft University of Technology

TUDelft

flyingfish

Urban Water Mobility

Reducing travel time within urban areas with watertaxi networks

by

T.J. van Berkel

Master Thesis

In partial fulfilment of the requirements for the degree of

Master of Science
in Mechanical Engineering

at the Department Maritime and Transport Technology of Faculty Mechanical, Maritime and Materials Engineering of Delft University of Technology
to be defended publicly on October 15, 2020 at 10:00AM.

Student number:	4225503	
MSc track:	Transport Engineering and Logistics	
Report Number:	2020.TEL.8455	
Supervisor:	Ir. M.B. Duinkerken	
Thesis committee:	Prof. dr. R.R. Negenborn,	TU Delft committee Chair, 3mE
	Dr. ir. J.F.J. Pruijn	TU Delft committee member, 3mE
	Ir. G.P. van Marrewijk	Company Supervisor, Flying Fish Water Mobility
Date:	October 5, 2020	

Preface

This master thesis would have been written about vehicle routing problems and Artificial Intelligence, if it was not for some (un)fortunate events. Just before the start of that project, a contractual matter stopped me from using that project to graduate. This left me in the situation without a graduation project. I had to start over with looking for a project and as the reader may know, this process can take months. Fortunately, I spoke with Johan, my cousin and also Flying Fish co-founder, at our family Christmas brunch. As I explained my situation, he told me to talk with Gijsbert, one of the co-founders, about the possibility to graduate at Flying Fish and with a world record breaking speed, a thesis subject was formulated. Within 2 weeks I was an internal graduate. I am very grateful that I could start within such a short period of time and with all the freedom I got to formulate my own subject.

My time at Flying Fish brought me more skills than I had imagined beforehand. The daily standups and two weekly sprint meetings have given a positive boost to the process of this research. For that and all his effort, I want to thank Gijsbert. With a limitless number of ideas, there was always something to improve the project. I have learned a lot from Gijsbert's experience of his own graduation project. I also want to thank him for the Mapmaker software he provided me to make nice plots of the city of Rotterdam. With all expertise and help from Jaime, Laurent, Casper, Johan, Tim, Gijs, Mohammad, Gerben and Thore, this project turned out better than I expected beforehand. Thank you all for your effort to include me in all (social) acitivities. During my internship, you have learned me lots of lessons on project management, time estimations and software development. Those skills are valuable when I become an Engineer.

I want to thank Mark Duinkerken for his direct feedback every three weeks on work I delivered. The feedback I received was in most cases formulated in such a way that is was directly implementable. I want to thank Rudy Negenborn for his guiding role in this Graduation project.

Last but not least, I want to thank my girlfriend Tanja and my parents for all the limitless support I got in good times but especially at the difficult moments. Mentally, but also with advice on moment I really needed it. Special thanks to my sister Lotte, who made a nice title page design out of the photos I delivered.

I am proud of the work I have delivered and the all the hurdles I have overcome. Now am ready for the next step. I hope you will enjoy reading.

Thijs van Berkel
3rd October 2020

{Front cover: photos made by author}

Summary

From a historical perspective, cities have often been built often around bodies of water [26]. These bodies of water can bring opportunities for transport of people, for example by the usage of watertaxis. Watertaxis work in a similar way as a regular taxi service. A boat can be ordered online or by phone and leaves from a selected jetty. Then the itinerary begins at the agreed time and ends at the agreed arrival jetty. To find the total travel time for passengers, choices have to be made on number of jetties and the locations of jetties. Since jetties require investments and changes in infrastructure, a limited number can be placed. The placement of jetties has to be done strategically to serve the most customers. Therefore it is important to find the relation between the number of placed jetties and the average reduced passenger travel time within an urban area. The main question of this research is: *What is a generic method to find the relationship between the number of placed jetties for watertaxis and the passenger travel time within an urban area?*.

The watertaxi system is modelled as an undirected complete graph, with two types of nodes: city nodes and jetty nodes. City nodes contain the origins and destinations of travellers. The travel time between the city node locations is determined by the duration of taking public transport. The jetty nodes set contains all possible jetty locations. The travel time between jetty nodes depends on the time it takes a watertaxi to sail between the jetty node locations. Between city nodes and jetty nodes, the travel time is equal to the walking time. A selected group of travellers between city nodes is used to find an optimum. The number of travellers is determined before finding a solution and is time invariant. From the jetty nodes, a predetermined number of nodes is chosen. The jetties are chosen in such a way that the sum of the travel time for all travellers is minimized. Due to crowd regulations, the number of people that arrive and leave at a jetty should be limited.

In literature, Hub Location Problems (HLPs) are closely related to problem described. Different variants exist on HLPs. For this research the Median p-HLP design by Campbell [4] is chosen as basic model. This design allows the usage of one type of nodes. This model is combined with the capacity constraint developed by Campbell [6]. To complete the design, a second type of nodes is added and an extra hub is added to serve as direct link. Three performance indicators are used: the percentage of people that take the watertaxi, the average reduction in travel time and the average calculation time. To limit the total time needed to find a optimum, the total calculation time is required to be lower than 90 minutes.

To find the locations of the city nodes and the travellers between the city nodes, the travellers survey OViN 2017 is used. 79 city nodes and 720 unique itineraries are registered in this OViN 2017 dataset. 75 jetties in the city center of Rotterdam are used as jetty nodes. To find the shortest route between city nodes by public transport and the shortes route between jetty nodes by walking, Google Static Maps API is used, and to find the fastest routes on the water, the Flying Fish route planning algorithm is assessed.To stay within the range of the total calculation time, only travellers that are able to reduce their travel time by taking the watertaxi are taken under consideration. After that, only city nodes are selected where people arrive at and leave from. This results in a selection of 21 city nodes and 27 unique itineraries.

21 different numbers of jetties are allocated in the main experiment, starting with allocating 2 jetties and ending with 22 jetties. The average reduced travel time is 105 seconds with allocating two of jetties. With an increasing number of jetties, the reduction in travel time increases. The average reduced travel time goes asymptotically to 517 seconds when 22 jetties are allocated. The percentage of people that takes the watertaxi is 100 percent when 22 jetties are allocated, due to the preselection on city nodes and travellers. The total calculation time for all 21 steps is 57 minutes, which is lower than the allowed maximum of 90 minutes. In a second experiment, the number of city nodes and jetty nodes is varied to make estimations on calculation time. The total calculation time scales exponentially with the number of nodes used.

Further research with the developed method into other cities is recommended. With Google static Maps API, the walking and public transport duration data of other cities can be found. OViN 2017 survey is not useful in other cities due to its low number of travellers between city nodes. The Flying Fish planning algo-

rithm and jetty locations cannot be used in other case studies, since all data is based on Rotterdam. Allowing an optimality gap in the solver of the developed model, could result in a tradeoff between optimality of the solution and the calculation time. This could allow more nodes to be used while still meeting the total calculation time requirement. To analyse the reduced travel time in another manner, walking can be replaced by biking or public transport. With the required travel time data, this method can be used as a first step in the development of watertaxi networks in other cities.

Contents

List of Abbreviations

AP	Australian Post
C	Capacitated
CAB	Civil Aeronautics Board
DSA	Delft System Approach
Fix	Fixed
GIS	Geographic Information System
HLP	Hub Location Problem
KPIs	Key Performance Indicators
MA-p-HLP	Multiple Allocation p-Hub Location Problem
MA-p-HCP	Multiple Allocation p-Hub Maximal Covering Location Problem
MILP	Mixed Integer Linear Programming
OViN	Onderzoek Verplaatsingen in Nederland
UN	United Nations
Var	Variable

List of Symbols

C_n	City node n
Ca_k	Capacity limit at jetty k
$d_{zone,n}$	The distance travelled in zone n
H	Predetermined number of hubs
J_n	Jetty node n
Sa^{km}	Sailing time from jetty node k to jetty node m
$t_{accelerate}$	Time in seconds for a watertaxi to accelerate to the allowed maximum speed
t_{aboard}	Time in seconds to get aboard and go offboard of a watertaxi
$t_{watertaxi}$	Time in seconds it takes a traveller to travel with a watertaxi from its origin jetty to its destination jetty
T_{ij}^{km}	Travel time from i via k and m to j.
$v_{max,n}$	The maximum allowed sailing speed in zone n
W_{ij}	Number of travellers from i to j
Wa_i^k	Walking time from city node i to jetty node k
X_{ij}^{km}	The fraction of the travellers between i to j that travels via k and m
Y_k	Variable that stipulates if jetty node k is a jetty

1 Introduction

For the longest part in history, mankind lived in rural areas. In the last centuries, humans started to live in bigger communities. The United Nations (UN) conducted research on this topic to quantify this trend. With data, gathered by the World Urbanization Prospect of the UN[1], one can plot the percentage of people living in urban areas. As the definitions and criteria differs per country for this percentage, the UN attempted to keep the national criteria consistent over the years to estimate the trend of urbanization. Based on estimations of the UN, the proportion urban from 1950 to 2050 is plotted in figure 1.1. From this figure, one can estimate that after the period between 2005 and 2010, more than half of the world population is living in urban areas. Since that time, urbanization did not stop. The UN estimates that this growth will continue for the next decades. Figure 1.1 shows the increase of percentage of people living in urban areas for the coming 30 years. The percentage of people living in urban areas worldwide will increase to 68 percent. With an estimated world wide population of 10 billion in 2050 [2], the total urban population world wide will continue to grow rapidly. With the growth of the proportion people living in urban areas one can conclude that this will result into one or combination of the following issues:

1. Increase of size of urban areas
2. Increase of population density

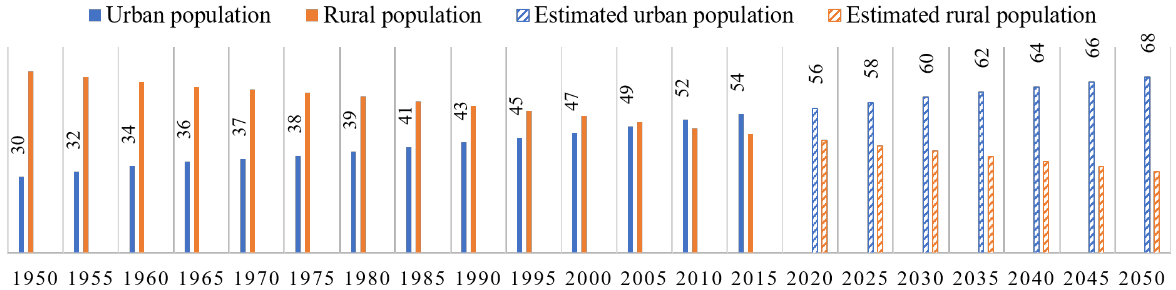
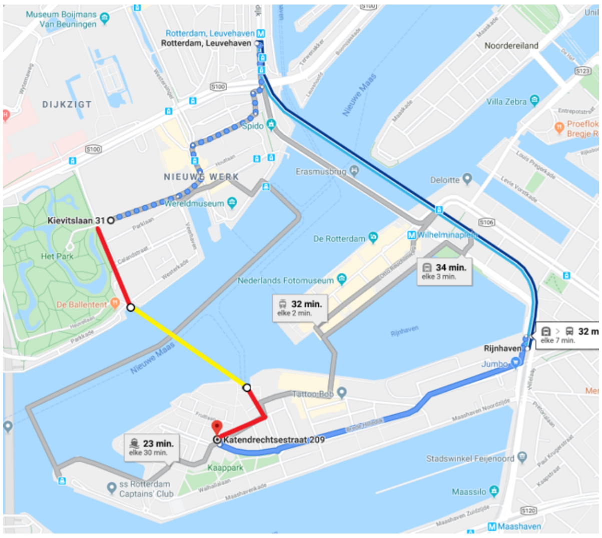


Figure 1.1: Percent of people living in urban areas based on estimations of United Nations

From a historical perspective, cities have been built often around bodies of water [26]. These bodies of water can bring opportunities for transport of people. One way to use water as a transport mode is the usage of watertaxis, which is a taxi service that can be ran on the water. Watertaxis work in a similar way as regular taxi services. A boat can be ordered at a locations where it can moor: a jetty. Then the itinerary begins at the agreed time and ends at the agreed arrival jetty.

[1] United Nations, Department of Economic and Social Affairs, Population Division (2018). World Urbanization Prospects: The 2018 Revision, Online Edition.

[2] United Nations, Department of Economic and Social Affairs, Population Division (2019). Probabilistic Population Projections based on the World Population Prospects 2019



Kievitslaan 31 to Katendrechtstraat 209	
Walking	34 minutes
Public transport	31 minutes
Taxi	11 minutes
Watertaxi + walk	10 minutes

Figure 1.2: Example: route planning from Kievitslaan 31 to Katendrechtstraat 209 in Rotterdam. The red line shows the walking route from and to the jetties, in yellow the route the watertaxi takes is shown. This is an example where using a watertaxi results in a reduction in travel time compared to travel time with public transport, walking and taking a regular taxi. Data recieved from Google maps on: 1 februari 2020. The sailing time is estimated with existing water taxi duration data of Flying Fish.

A watertaxi can provide a reduction in travel time. In Rotterdam, watertaxis are used to transport people over the Nieuwe Maas. In Rotterdam taking a watertaxi can result in a reduction in travel time. Figure 1.2 shows an example where a combination of walking and taking the watertaxi could result in a reduction of 24 minutes of travel time compared to travelling with public transport.

In Rotterdam 600.000 people make use of the watertaxi services every year. According to Watertaxi Rotterdam every year the percentage of passengers grows by 10 percent. They operate with 6 different types of boats that have maximum sailing speeds between 15 and 55 km/h. Figure 1.3 shows the locations of watertaxi jetties within Rotterdam[3].

Figure 1.3: Watertaxi jetties within Rotterdam[4].

[3]Retrieved on 25 january 2020 from: https://www.watertaxirotterdam.nl/over-ons
[4]Retrieved on: 30 january 2020 from https://www.watertaxirotterdam.nl/steigerlocaties

1.1. Objective of this research

The company that runs this watertaxi network, Watertaxi Rotterdam, asked Flying Fish to improve their operations. From Flying Fish, the question arose if watertaxi networks could provide a reduction in passenger travel time in other urban areas. Therefore, Flying Fish wants to have method developed that can be used to estimate the total travel time for a selected group of people when watertaxi network is placed within an urban area.

Develop a method that can be used to determine the locations of the jetties by minimizing the average travel time.

1. A method that can be used to determine the walking times of travellers from and to jetties.
2. The locations of the jetties should be determined considering public transport.
3. Optimize the locations in order to minimize the summation of the total travel time for the passengers.
4. The method should be able to limit the capacity to ensure not too many people gets assigned to a single jetty.
5. The time to determine these jetty locations, the calculation time, should not exceed 90 minutes on a personal computer.

1.2. Research questions

To find the total travel time for passengers, choices have to be made on number of jetties and the locations of jetties. The placement of jetties has to be done strategically to serve the most customers. To provide destination choice for customers, a high number of jetties is preferred. Since jetties require investments and changes in infrastructure, a limited number can be placed. Therefore it is important to find the relation within an urban area between the number of placed jetties and the total reduced passenger travel time. The main question of this research is:

What is a generic method to find the relationship between the number of placed jetties for watertaxis and the passenger travel time within an urban area?

When a generic method is determined, the method is tested in a case study. To be able to produce realistic outcomes for the case study, travel data within the area of interest is required. The following sub questions are answered:

1. What are the characteristics of the watertaxi system?
2. Which related mathematical problem can be used to determine a relationship between the number of jetties and the passenger travel time?
3. How is the watertaxi system going to be modelled?
4. Which suitable data sources are going to be used for a case study to the city of Rotterdam?
5. What are the results of the implemented method for the city of Rotterdam?
6. How generically applicable is the described method?

1.3. Scope

To answer the research questions, the following assumptions are made:

1. Cost is not taken into account.
2. Travellers are modelled as origin and destination points.
3. To determine travel time within an urban area, the time of day and rushhours are not taken into account.
4. To determine the locations of a jetties, no cost of allocation is taken into account.

1.4. Methodology

The research is going to be divided into 5 main subjects. Based on Case Study Methodology [8], the intervention cycle is used. Every main subject below has its own chapter. To visualise this, figure 1.4 shows a roadmap of this research.

1. Problem finding (green in figure 1.4)
2. Problem diagnosis (grey in figure 1.4)
3. Design of intervention (yellow in figure 1.4)
4. Implementation (orange in figure 1.4)
5. Evaluation (blue in figure 1.4)

The problem finding consist of identification and definition of the problem. The problem diagnosis is used to find out why the problem exists. In the design of intervention, an intervention is designed that should tackle the problem. In implementation the intervention is going to be implemented. Finally, in the evaluation phase, is used to reflect on the intervention.

1.4.1. Problem finding

First the current situation is evaluated. For this the Delft System Approach [30] is going to be used. In this chapter, first the Delft System Approach (DSA) is going to be explained. In this chapter, the model used to describe the problem is going to be discussed.

1.4.2. Problem diagnosis

When the model is found, literature is needed. A literature study has to be done on a mathematical description of the model. With a suitable mathematical description, literature is going to be found on Key Performance Indicators (KPIs), benchmark problems and optimization methods.

1.4.3. Design of intervention

The design is going to be explained in the 'design' chapter. First, the basic model found in literature is going to be discussed. Since this model does not meet all requirements, adaptions are going to be made. After that, The KPIs are defined. The KPIs are divided into KPIs to evaluate the case study (logistics KPIs) and KPIs for the performance of the designed method (Research KPIs). Finally, the design is going to be verified by 9 case studies.

1.4.4. Implementation

The implementation phase is covered by two chapters. The 'data case study' chapter, the data is gathered on the city centre of Rotterdam. The data is validated in this chapter. In the 'Case Study' chapter, the found method is experimented with in the city of Rotterdam. With the logistic KPIs, the case study results are discussed.

1.4.5. Evaluation

The evaluation phase is described in the last 'reflection' chapter and in the 'conclusion' chapter. The reflection chapter, the performance of the method is discussed.

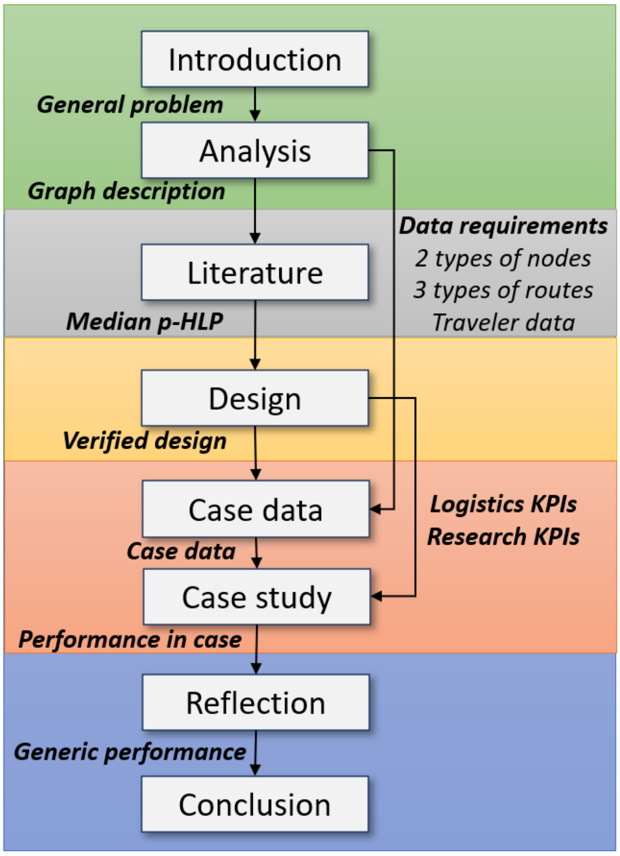


Figure 1.4: Roadmap of research. The research is divided into five phases.

2
Analysis

The watertaxi and public transport have a complex relationship with the travel time through an urban area. Lots of different elements play a role in determining the travel time between two points. To model the total travel time by watertaxis and walking, assumptions have to be made. In order to do so, the sub question *'What are the characteristics of the watertaxi system?'* is answered.

To be able to answer the subquestion, the watertaxi system is analysed on three levels: operational, tactical and strategical. At the operational level, the basic elements of the system are described. Also the basics of the public transport system are introduced. At the tactical level, the watertaxi planning is evaluated. At the strategic level, the jetty placement within an urban area is described. After that, the analogy with network design is explained.

2.1. Operational level: transit operations

The itinerary within an urban area is analysed with the black box approach, as described in Veeke et al. [30]. A black box could be a device, system, or object, which is described in terms of its inputs and outputs, without any knowledge of its actual internal workings [31]. To model the system, two separate systems are distinguished. First, the system is described where a traveller takes the watertaxi. After that, the system is discussed where a traveller takes public transport.

2.1.1. Watertaxi system

Figure 2.1 shows the black box of the trip including a watertaxi. In this research, people enter the system when starting their trip, people leave the system by reaching their destination. In the watertaxi system, only people that travel with a watertaxi are considered. The watertaxi system does not only consist of the itinerary with the watertaxi, but also the itinerary from the origin and the itinerary to the destination.

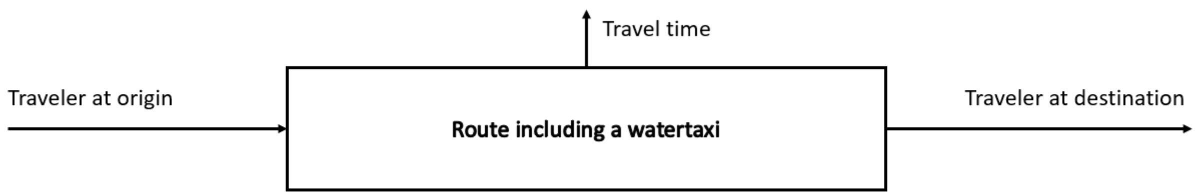


Figure 2.1: Black box approach of the watertaxi system. The travel time is the main performance indicator.

The watertaxi system is subdivided into four main parts: the route to a jetty, the waiting time before a watertaxi arrives, the sailing between two jetties and the route to the destination of the traveller. Figure 2.2 shows the subdivided system. There are many transport modes to travel from and to a jetty. In this research, only walking is taken into consideration. With this assumption, the route to the jetty and the route to the destination are analysed in the same manner.

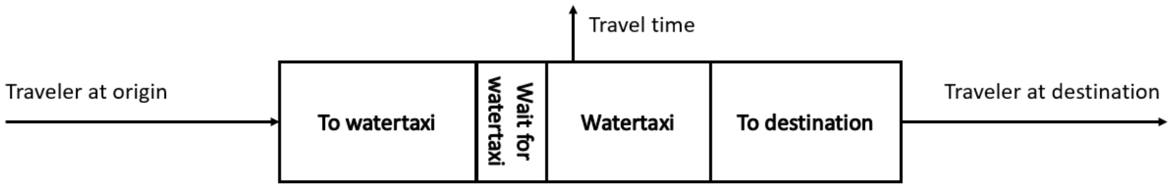


Figure 2.2: The route including the watertaxi through an urban area split up in four parts. The travel time is its performance indicator.

Figure 2.3 shows different possible routes for two travellers and four jetties. With the origins of the travellers are numbered P_1 and P_2, the destinations numbered P_1 and P_2. The jetties are numbered J_1, J_2, J_3 and J_4. People depart from the origins and to arrive at their destinations. Different routes are available. In this research, it is assumed that people attempt to minimize the total travel time. The total travel time of a passenger is the summation of the travel time to the jetty, the total sailing time and the travel time to the destination.

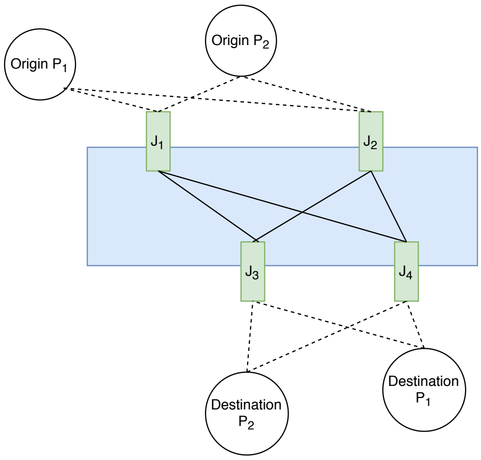


Figure 2.3: The origins of the travellers P_1 and P_2, the destinations numbered P_1 and P_2. The jetties are numbered J_1, J_2, J_3 and J_4. The routes to the jetties and to the destinations are dashed black. The possible sailing routes are shown in black. For simplicity reasons, not all possible sailing routes and walking routes are shown.

2.1.2. Public transport system

Based on the travel time, the traveller can choose to take public transport. Generally speaking, public transport consist of busses, trains, metros and other transport modes that leave at scheduled times. Google Maps defines public transport as follows:"a transportation service that is open to the public, and operates with fixed schedules and routes" [1]. All transport modes within Googles definition of public transport are used, except from ferries and water busses. These types of water transport compete with the watertaxi and are left out of scope.

Modelling the public transport and finding the travel time is complex and not in the scope of this research. Therefore, the black box shown in figure 2.4 is not analysed in more detail. A method is needed to find the travel time from a selected origin to a selected destination within this research.

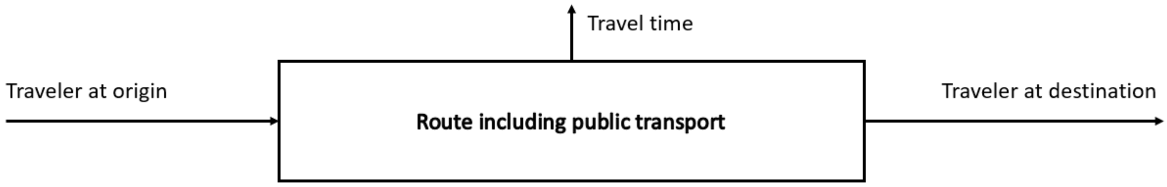


Figure 2.4: The public transport system. The performance is measured by the travel time.

2.1.3. Actors

To further understanding of the watertaxi system, the different actors are described. Three actors are discussed: a watertaxi (the used boat), jetties and travellers through an urban area.

Watertaxi Watertaxi Rotterdam uses different types of boats to transport travellers over the Maas. Three types of watertaxis used by Watertaxi Rotterdam are shown in figure 2.5. The speed of the boats is important to minimize the travel time, only the boats that can reach speed above 50km/h are shown in this figure. The watertaxis can carry up to 12 people and have a water displacement up to 4500kg.

Travellers The number of travellers depends on different factors. Therefore, a predetermined number of travellers is taken into consideration. Data is required on the origins of travellers and their destination to make estimations on travel time.

Jetties Jetties are locations where watertaxis can moor and people can come aboard or step off board. Jetties are build on designated locations at the waterside. Figure 2.6 shows two example of jetties within Rotterdam. With the current limitations on crowd sizes, the method should be capable of limiting the number travellers that arrive and leave from one jetty.

[1] retrieved from: www.google.com/transit on 20-09-2020
[2] Retrieved from: https://www.watertaxirotterdam.nl/steigerlocaties at 20th March 2020

MSTX 1 - 5

Capacity	12 passengers
speed	50 km/h
Water displacement	-

MSTX 6

Capacity	12 passengers
speed	55 km/h
Water displacement	3500kg

MSTX 7 - 14

Capacity	12 passengers
speed	55 km/h
Water displacement	4500kg

Figure 2.5: Examples of three types of watertaxi boats. Retrieved from:https://www.watertaxirotterdam.nl/over-ons on March 25 2020

Figure 2.6: Example of two watertaxi jetties. Jetty at the left is at Katendrechtse hoofd. Jetty at the right is located at the Euromast in Rotterdam.[2]

2.1.4. Travel time estimations

With the black box approach, three different systems are found: sailing, walking and public transport. Travel time has to be determined for the three systems. This data is gathered from external sources. To compare the external sources, first general requirements for the external data sources are set. After that, for specifics for each system are discussed. The main requirements for the data sources:

1. Accurately estimate the travel time between the two points of interest.
2. The gathering of data should not take more than 1.5 hour.
3. Data should be gathered in an automatic manner to cope with large numbers nodes.

Public transport As mentioned in section 2.1.2, public transport within an urban area consists of various options and can be evaluated in different ways and on different aspects. The estimation should include the duration of walking to the destination from the last stop of public transport.

Walking Walking distances to all potential jetty locations have to be found to determine the least time consuming route from an origin to a jetty. To ensure no ferries are used, the designed method should exclude ferries.

Sailing time When a traveller arrives at a jetty, the traveller has to go on board. Then, the waterataxi accelerates to the maximum allowed speed. When the boat arrives at the desired location, the traveller pays for the trip. Ideally, the next passenger is waiting to get into the watertaxi. On the water in Rotterdam, three speed zones are defined. To compensate for the acceleration of the watertaxi, $t_{accelerate}$ is introduced. To compensate for the time it takes to go aboard and go off board, t_{aboard} is introduced. Equation 2.1 is introduced to calculate the sailing time. $d_{zone,1}$, $d_{zone,2}$ and $d_{zone,3}$ are sailing distances in the different speed zones. $v_{max,1}$, $v_{max,2}$ and $v_{max,3}$ are the maximum allowed speeds in the different speed zones. To find the sailing times, a planning algorithm is needed to find the shortest route over water. The designed method should be able to find all routes between the potential jetty locations.

$$t_{watertaxi} = t_{aboard} + t_{acceleration} + \frac{d_{zone,1}}{v_{max,1}} + \frac{d_{zone,2}}{v_{max,2}} + \frac{d_{zone,3}}{v_{max,3}} \tag{2.1}$$

2.2. Tactical level: transit planning

In the previous section, the black box of the watertaxi system is subdivided into four different blocks. The second block, the ‘wait for watertaxi’ block, depends on the planning of the watertaxis.

After a passenger chooses a jetty, the watertaxi central should be informed about the time the passengers wants to be picked up. With thi information a watertaxi is planned to pick up the traveller at the desired time. With a new request for a watertaxi, the request should be fitted into the planning. One of the goals of the watertaxi planning is to fit all requests for watertaxis into the planning in such a manner that all customers of the watertaxi have their watertaxi arriving at the desired time.

It is assumed that the watertaxi succeeds in making the planning for the watertaxi in such a way that all travellers are covered. With that assumption, the planning of the watertaxi does not play a role in calculating the total travel time. This means that the waiting time may be set to zero.

2.3. Strategical level: strategic determination of jetty locations

At last, the strategical level is taken into consideration. In this level, the locations of the jetties are determined. From a simplified perspective, the main constraint of a watertaxi jetty is that it has to lay on the interface between water and land. It is possible to extract the riversides from the map of the selected urban area and make a set with coordinates on the riverside. From a ’real world’ perspective, more aspects are important. The depth of water, the riverbed and whether the government owns the land are important. This makes the determination of a set of possible locations complex. Therefore, a list with coordinates of potential locations is necessary. The set containing all potential jetty locations is named ’jetty nodes’.

2.4. Network analogy

With all assumptions made in this chapter, the problem can be defined as a complete undirected graph. Two types of nodes are described: city nodes and jetty nodes. City nodes are the origins and destination locations of the travellers through the area. The jetty nodes are the initial set of locations where a jetty can be allocated on. Out of all jetty nodes a predetermined number of nodes is chosen in such a way that it minimizes the total travel time for all travellers within the system. A possibility should exist to limit the number of travellers that leaves and arrives a a single jetty due to possible crowd limitation. Between all nodes, connections exist. Between the city nodes, public transport is used. Between the jetty nodes, watertaxis sail. Between jetties and city nodes, travellers walk. For all connections, assumed is that the travel does not change over time. Figure 2.7 shows the complete graph with four city nodes and four jetty nodes. To keep the figure clear, not all possible routes are drawn.

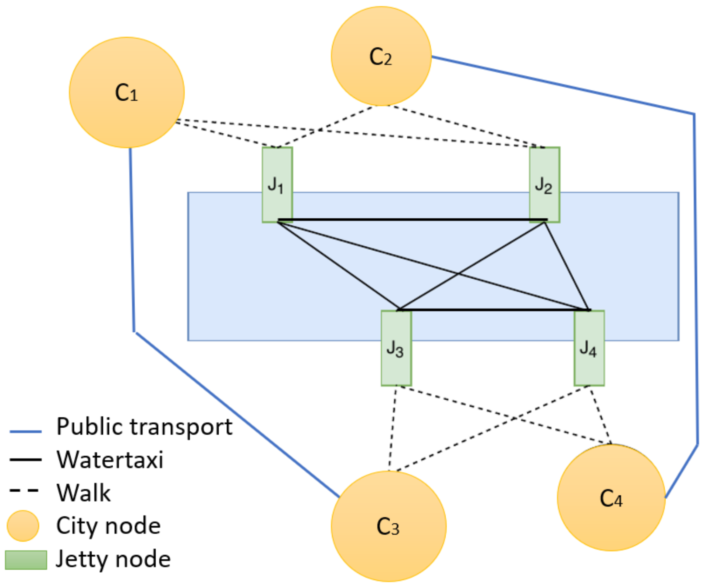


Figure 2.7: A complete graph of the watertaxi system with 4 city nodes and four jetty nodes. Between the city nodes people travel. From the jetty nodes a predetermined number of jetties is chosen. To keep the figure clear, not all routes are drawn.

2.5. Conclusion

In this chapter, the subquestion *'What are the characteristics of the watertaxi system?'* is answered by analysing the system at three levels. The watertaxi system is modelled as a complete graph with travellers between the city nodes. From the jetty nodes, only a predetermined number of jetties is allocated. The second constraint is that arrive and depart from a single jetty can never exceed a certain number. To perform a case study, six different sets of data are required. The six different of dataset consist of the two types of nodes, three types of routes and the demand through the network. Requirements are set to ensure the data gets collected fast and that the collected data is useful.

3

Literature review

Within this chapter the sub question *Which related mathematical problem can be used to determine a relationship between the number of jetties and the passenger travel time?* is answered. In the previous chapter, a description of the problem is given. In this chapter, a mathematical formulation for a complete graph with two types of nodes is found. The objective function is to minimize the travel time. From the jetty nodes, a predetermined number of jetties is chosen. The problem should be capable of limiting the number of arrivals and departures from a single jetty.

Mohri and Akbarzadeh [22] performed research on designing a van-taxi network in Iran. The objective of the research of Mohri and Akbarzadeh [22] is to minimize the driven kilometers by the vans. For the research of Mohri and Akbarzadeh [22], the most suitable locations for the van locations had to be chosen. To find the most suitable location for these hubs, the Hub Location Problem (HLP) is used. This problem is discussed within this chapter.

First, a suitable mathematical formulation for the HLP is elaborated on. Second, a method to solve a HLP implementation is discussed. At last, a benchmark dataset is elaborated on, to verify the implementation of the HLP. To evaluate the performance within the case study, the performance of the solver, and to verify the implementation with the benchmark datasets, performance indicators in literature are described.

3.1. Hub Location Problem

Klincewicz [17], who performed research to Heuristics for hub location problems, gives the following description for hub location problems: *'The hub problem is a facility location problem that can be viewed as a type of network design problem. Each node, within a given set of nodes, sends and receives some type of traffic to and from the other nodes. The hub locations must be chosen from among these nodes to act as switching points for the traffic. Network links are placed between pairs of hubs so that the hubs are fully interconnected; each of the remaining nodes, in turn, is connected to one of the hubs. The traffic may represent telecommunications traffic, data transmissions, airline passengers, express packages, etc. In the HLP, locations of Hubs have to be found out of a set of locations.'*

3.1.1. Criteria

The field of studies to HLPs has been expanded in different directions. Farahani et al. [13] reviewed these different directions. Variants on the HLP can be categorized based on the criteria found by Farahani et al. [13]. In table 3.1 the different models of are shown and compared to the criteria found in the previous chapter.

Objective The objective is the main equation of the model. Normally, this function is either maximized or minimized. In most studies, the cost is the objective that is minimized. In other studies, the coverage is the objective that is maximized. When coverage is maximized, the hubs are placed in such a way to use transport through hubs connections instead of direct transport. In this research the total time is minimized.

Nodes There are two types of nodes described in HLPs: discrete and continuous. When nodes are discrete, the nodes are part of a particular set of a finite number of nodes. When the nodes are continuous, the domain of hub nodes is a plane or a sphere ('cont' in table 3.1). The discrete type of node is subdivided into two subcategories. The nodes in the problem can be part on one set (1 type), but also the nodes can be subdivided into multiple set (2 types). In the problem described in previous chapter, two types of nodes exist: city nodes and jetty nodes. Therefore the type of nodes is '2'.

Source determining the number of hubs to locate The determination of the number of used hubs can be either set beforehand or be a part of the solution. Two options are distinguished: fixed (fix) and variable (var). If the number of hubs is variable, finding the number of hubs is part of the solution. For this research, the number of hubs is fixed. Therefore, 'fix' is chosen.

The number of hubs In literature a difference exists between problems where one hub has to be found and multiple hubs have to be determined. The models where only one hub has to be found are often more simplified than the models where multiple hubs have to be found. In this research, multiple hub locations have to be found. This means that the 'multiple hubs' have to be determined.

Hub capacity In some cases a limit to the capacity of the hubs is useful to increase the accuracy of the problem. Two options are used in literature: uncapacitated (U) (unlimited) and capacitated (C) (limited). As described in the analysis chapter, the number of passengers arriving and departing from jetty should be limited. Therefore, 'capacitated' (C) is chosen.

The cost of locating hub nodes Assigning a location to become a hub can come with costs. In literature, these costs are generally considered in two different manners: fixed cost, and variable cost. For other research, the costs are not taken into account. In this research, the allocation costs are assumed to be zero.

The allocation of a non-hub to hub nodes The allocation of a non-hub to hub nodes describes the amount of hubs where a non- hub node is connected to. Two options are described in literature: the 'multiple' and 'single'. When this characteristic is 'single', all transport from a non hub not has to go to via a single node, even when the transport via another hub could be beneficial. If multiple allocation is allowed, the transport from a node can go through every chosen hub. In this research multiple allocation is needed to meet the requirements.

The cost of connecting non-hub nodes to hub nodes Besides taking cost into account for allocating costs of hubs, cost for connecting hubs to non hubs can be taken into account. Calculation of the cost can be done by giving the links a fixed cost and variable cost. In this research the cost of setting up links is assumed to be zero.

Mandatory transport through hubs As in most models all transport goes through at least one hub. In some cases, transport could benefit by having direct transport instead of transport via the hubs. Since passengers are allowed to take alternative transport instead of the watertaxi, the passengers do not have to travel through the hubs. This characteristic is therefore set to 'no'.

3.2. Model determination

Farahani et al. [13] reviewed different types of HLPs. Based on the work of Farahani et al. [13], table 3.1 is created. In this table eight different types of HLPs are shown. For every model its general characteristics are listed. When comparing the requirements for this research with the characteristics for of the different models, one can conclude that none of the listed models does solely cover the needed requirements. Therefore, a combination between models is needed.

From the listed models, multiple types of models could be required to cover all characteristics. Starting with the 'p-HLP with limited capacity' can be useful for its requirement on limiting the capacity on the jetties. The second model is the p-Hub Maximal Covering Location Problem, which could be used to add direct travel methods that do not require transport hubs. The third option that is taken into consideration is the Median p-HLP.

Table 3.1: Different versions of the p-HLP with their general characteristics. The HLPs that are the closest related to this research are coloured light green. The requirements for this research are shown in dark green.

	Objective	Nodes	Number of hubs to locate	Hub capacity	Cost of locating	Allocation of a non-hub to hub	Mandatory transport through hubs
p-HLP	Cost	1 type	Set	U	No cost	Single	yes
Median p-HLP	Cost	1 type	Set	U	No cost	Multi	yes
p-HLP with limited capacity	Cost	1 type	Var/ Fix	C	No cost	Multi	yes
Continuous p-HLP	Cost	cont	Fix	U	No cost	Single	yes
p-Hub center location problem	Cost	1 type	Fix	U	No cost	Multi	yes
p-Hub covering location problem	Cover	1 type	Var	U	No cost	Multi	yes
Hub set covering location problem	Cover	1 type	Var	U	fixed cost	Multi	no
p-Hub Maximal Covering Location Problem	Cover	1 type	Fix	U	no cost	Multi Single	no
This research	Time	2 types	Fix	C	no cost	Multi	no

3.2.1. p-HLP with Limited capacity

Several studies are done on limiting the capacity of hubs. Campbell [6] describes different types of integer programming models and lays the foundation for the multiple allocation HLP with capacity constraints. Ebery et al. [9] uses the formulation of Campbell and updates it. In the model of Ebery et al. [9] only a capacity restriction is applied on the volume of traffic entering a hub via collection. Rodríguez et al. [27] based their study on the model proposed by Aykin [1]. Every hub has two peak usage periods. The first of these is when the trucks arrive with the cargo from the hubs assigned nodes. Here, the constraints ensure that the allowed nominal capacity of the hub is not exceeded by preventing cargo from entering. The second peak usage period is when trucks arrive from other hubs. Here, the constraints limit the maximum cargo load entering the hub at that moment. This means that Rodríguez et al. [27] uses two separate hub capacities. Meraklı and Yaman [21] hubs in an N-node network with capacity and distance limitations when the service standard offered needs to be fulfilled. Table 3.2 shows an overview of the different models. Most models use the number of hubs as a variable. Since the number of hubs needs to be fixed, only the formulation of the restriction on capacity is used. The formulation of Campbell [6] fits into the work of Campbell [4] and is therefore used in this research.

3.2.2. Multiple Allocation p-Hub Maximal Covering Location Problem

The p-hub maximal covering location problem (MApHCP) maximizes the demands that are covered with a predetermined number of hubs. Different studies on MApHCP are shown in table 3.2. The MApHCP was first introduced by Campbell [6]. After that Kara and Tansel [16] improved the formulation. After reviewing 3.2, the formulation of the MApHCP is too far from the requirements in this research. This means that the general formulation by Kara and Tansel [16] and the MApHCP design of Campbell [6] are not usable for this research.

Table 3.2: Models by different authors. The characteristics are compared with the characteristics needed in this research. In light green the two

	Objective	Node	Number of hubs to locate	Hub capacity	Allocation of a non-hub to hub	Mandatory transport through hubs
Multiple Allocation p-Hub Maximal Covering Location Problem						
Weng et al. [33]			x		x	x
Peker and Kara [25]			x		x	x
Silva and Cunha [28]			x			x
Máximo et al. [20]			x		x	x
Ebrahimi and Ahmad [10]			x			x
Hwang and Lee [15]			x			x
p-HLP with limited capacity						
Campbell [6]				**x**	**x**	
Marín [19]				x	x	
Gelareh and Pisinger [14]				x	x	
Ebery et al. [9]			x	x	x	
Rodríguez et al. [27]			x	x	x	
Meraklı and Yaman [21]				x	x	
Median p-HLP						
Campbell [4]			**x**		**x**	
Skorin-Kapov et al. [29]			x		x	
Ernst and Krishnamoorthy [12]			x		x	
Boland et al. [3]			x		x	
Kratica [18]			x		x	
This research	**Time**	**2 types**	**Fix**	**C**	**Multi**	**No**

3.2.3. Median p-HLP

The median p-HLP has straightforward applications to transportation and telecommunication networks in which the objective is to minimize the total cost of movement [6]. The Median p-HLP is first proposed by Campbell [4]. Since then, minor changes have been made to this model. Skorin-Kapov et al. [29] uses a formulation which uses fewer constraints. Ernst and Krishnamoorthy [12] found a three index variables to model the uncapacitated problem. Boland et al. [3] uses a slightly different formulation that is used in more recent literature. In Kratica [18] an electromagnetism-like method is proposed for solving this Median p-HLP. Özgün-Kibiroğlu et al. [24] describes the hub location problem in which capacity restrictions are introduced is addressed into the objective function as a penalty cost to represent their congestion effects on respective hubs. Table 3.2 shows an overview of these different models.

3.2.4. Conclusion on model

As can be concluded from the previous subsections and table 3.2, no model has all requirements needed. Therefore, choices have to be made on which model is the closest related to this research. To do so, first a basic model is chosen. Extra constraints have to be designed to ensure all requirements are met.

In the literature found on this topic, the objective function is normally based on cost. The case studies that use the benchmark datasets (AP, CAB and TN, found in section 3.4), the costs are build up by the usage of routes. In this research, the amount of time is the value on the links. This means that the objective of minimizing cost and minimizing time are similar. The objective function found in the MApHCP differs on multiple points from the objective function required and is therefore not suitable as basic model.

This leaves papers based on Median p-HLP with and without capacity constraints as options. The median p-HLP with capacity constraints by Ebery et al. [9] and Rodríguez et al. [27] are the closest related to this research but have more complex designs that making adjustments complicated.

For this research the the basic model by Campbell [4] is used. This model is combined with the capacity constraint in Campbell [6]. The model of Campbell [4] is widely used and has a basic architecture, which makes it suitable for adjustments. The work of Campbell [6] does not predefine the number of hubs that has to be chosen. The formulation of the model by [6] is closely related to the work in Campbell [4], which makes it possible to fit the capacity constraint of Campbell [6] on to the work of Campbell [4].

3.3. Solving Algorithms

In literature, different approaches exist to solve Hub Location Problems. Peker and Kara [25], Marín [19], Gelareh and Pisinger [14] use Mixed Integer Linear Programming (MILP) solvers, which are normally based on exact algorithms. Other studies use heuristics. Weng et al. [33], Silva and Cunha [28] use the Tabu search and Rodríguez et al. [27] simulated annealing. One of the advantages of heuristics is normally the shorter computing time in comparison with exact algorithms. However, heuristics can be difficult to implement. Heuristics are therefore kept out the scope of this research. It is assumed that a (commercial) MILP solver is sufficient for this research.

3.4. Datasets used for HLP studies

To verify the implementation of the chosen HLP in the chapter 4, a benchmark dataset is used. Three datasets are widely used in literature. These datasets are explained below. Since the benchmark dataset is only used for verification purposes, the smallest dataset is selected. Three datasets are discussed: Turkish Network, Civil Aeronautic Board and Australian Post.

3.4.1. Turkish Network

The Turkish Network (TN) dataset was introduced by Çetiner et al. [7]. It was used to analyse the Turkish postal services. This TN dataset contains 81 major Turkish cities. This dataset can be considered to be too large, due to the number of nodes [34].

3.4.2. Civil Aeronautic Board

The Civil Aeronautics Board (CAB) data set was first introduced by O'kelly [23]. It was based on airline passenger flows between 25 cities in the USA. The passenger flows are symmetric. Five problem sizes are considered: the first 5 cities, the first 10 cities, the first 15 cities, the first 20 cities and all 25 cities [32].

3.4.3. Australian Post

The Australian Post (AP) data set was first introduced by Ernst and Krishnamoorthy [11]. It contained 200 nodes, each representing a postal district. For small size problems, only a small number of the 200 nodes is selected. The numbers of nodes were set to be 10, 30, 25, 40 and 50. For large-size problems, the numbers of nodes were set to 100, 125, 150, 175 and 200 [32].

3.4.4. Conclusions on dataset

?? Since the benchmark dataset is only used for verification purposes, a small set is preferred. The CAB dataset is the most suitable for this research due to its size. The smallest described option contains 5 different cities. In the work of Boland et al. [3] the optimal solutions for the first 5 cities of the CAB dataset are presented.

3.5. Relevant Performance Indicators in Median p-HLP

Since the Median p-HLP is used in this research, its performance should be evaluated. Performance indicators are split up in three categories: logistics KPIs, research KPIs and verification performance indicators. The logistic KPI is used to indicate the performance of the outcome of the designed model. The research KPI indicates the general performance of the implementation. The verification performance indicators is used to find out if the outcome is similar to expectations.

3.5.1. Logistics KPI

To evaluate the performance of the method in a case study, performance indicators are used. In literature, the objective function outcome for a case study is used by Mohri and Akbarzadeh [22].

Objective outcome The outcome of the objective function. In the Median p-HLP implementation of Mohri and Akbarzadeh [22], the total kilometre driven by taxis is summed up and used to compare different scenarios with different numbers of chosen hubs. The outcome of the objective of this research is used in a similar manner.

3.5.2. Research KPI

The research KPIs are used to determine the performance of the implementation. As stated in the introduction of this thesis, a limitation to the total calculation time is set. In literature the following KPIs have been found:

Calculation time The calculation time is an indicator for the amount of time for the computer to come to a solution. Normally the time is measured in seconds [9], [33].

Gap with optimal solution Ebery et al. [9] defines 'Gap' as the relative deviation of the heuristic solution from the optimal solution, expressed as a percentage. Ebery et al. [9] uses the gap with the optimal solution to compare the outcomes of a heuristic with optimal solution of the CAB dataset. Allowing a gap with the optimal solution can reduce the calculation time.

3.5.3. Performance indicators for verification

The CAB benchmark dataset is chosen to verify the implementation of Campbell [4]. To compare the performance of the literature, performance indicators for verification are needed. Two indicators are used by Boland et al. [3]: the hub Locations and the total cost.

Total cost The cost summation is used as the main verification indicator. The total cost consist of all cost made by using the found routes.

Hub Locations Out of the set of nodes, a number of hubs is chosen. Boland et al. [3] describes the locations of the hubs for the CAB dataset.

3.6. Conclusion

As can be concluded from literature, the Median p-HLP (as found by Campbell [4]) is closely related to the requirements and chosen as basic model. To let the work Campbell [4], meet the requirements, extra constraint is necessary to limit the capacity. The constraint as found by Campbell [6] is implemented in the next chapter. When the work of Campbell [4] is implemented, the a solver is needed. An MILP solver is used. To verify the implementation, the CAB benchmark dataset is used. With the work of Boland et al. [3], the implementation is verified by comparing on the performance indicators 'hub Locations' and 'total cost'. To be able to design a model that covers all the predetermined requirements, the following two points have to be designed in this research:

1. **Non obligatory hub transport** Since passengers are not obliged to take the watertaxi, 'direct' travel route should be added.
2. **two types of nodes** the network within HLP consist of two types of nodes: city nodes and jetty location nodes. Within the Median p-HLP by Campbell [4] no distinction is made.

4
Design

This section answers the sub question *How is the watertaxi system going to be modelled?*. To answer this question, the Median p-HLP as described by Campbell [4] is explained. To meet all requirements set in this research, three adaptions are made to this design. An option to directly travel to a jetty has to be added. Also, the set of possible jetty locations should be changed to ensure that only the designated locations can be chosen as hubs. Thirdly, the capacity constraint found in Campbell [6] is added to the model. To analyse its performance, KPIs are defined.

To ensure the implementation is correct, the designs are implemented in Python. First the model of Campbell [4] is implemented. Its performance is compared with literature to the CAB benchmark dataset in three cases. After that, the implementation of the extended model is verified with six cases.

4.1. Main model

The design consists of an undirected network $N = (V, A)$ with $V = \{v_1, v_2, ..., v_q\}$ set of nodes. Every link is in $(a, b) \in A$. No links have a negative weight $d(a, b) = d(b, a)$. The travel time from origin i to destination j via jetty k and m (T_{ij}^{km}) is the sum of walking from i to k (Wa_i^k), sailing between k and m (Sa^{km}) and walking from m to destination j (Wa_j^m). The mathematical formulation for travel time T is $T_{ij}^{km} = Wa_i^k + Sa^{km} + Wa_j^m$. If $k = m$, no transport takes place between the jetties [4]. The following formulations was first proposed by Campbell [4].

4.1.1. Parameters

T_{ij}^{km}	Total travel time from origin i to destination j via jetty k and m
W_{ij}	Number of passengers travelling from origin i to j
H	Predefined number of hubs

4.1.2. Variables

Y^k	is 1 if jetty node k is chosen as jetty, 0 if k if it is not
X_{ij}^{km}	Flow (fraction) of the travellers that travel from i to j via k and m.

4.1.3. Objective and constraints

The equation 4.1 shows the objective function of the model by Campbell [4]. The total travel time is minimized for all passengers by multiplying the number passengers travelling from i to j (W_{ij}) by the travel time T_{ij}^{km}. X_{ij}^{km} is the flow from i to j via k and m and has a value between 0 and 1.

$$\text{Minimize} \sum_i \sum_j \sum_k \sum_m W_{ij} X_{ij}^{km} T_{ij}^{km} \tag{4.1}$$

Variable Y^k is introduced in equation 4.2. Y^k is set to 1 if Y^k is a hub and 0 if not. Equation 4.3 ensures that the number of hubs is sets to H.

$$Y^k \in \{0,1\} \forall k \in V \tag{4.2}$$

$$\sum_k Y_k = H \tag{4.3}$$

The value of X_{ij}^{km} is the fraction of the travellers that travels from i to j via k and m. Equation 4.4 ensures that the flow has a value greater than or equal to 0. Equation 4.5 and 4.6 stipulate that the flow via hub k and m is only possible if hubs k and m are chosen. Equation 4.7 ensures that the total flow from i to j is equal to 1.

$$X_{km}^{ij} \geq 0 \forall i,j,k,m \,;\, i,j,k,m \in V \tag{4.4}$$

$$X_{km}^{ij} \leq X_m \forall i,j,k,m \,;\, i,j,k,m \in V \tag{4.5}$$

$$X_{km}^{ij} \leq X_k \forall i,j,k,m \,;\, i,j,k,m \in V \tag{4.6}$$

$$\sum_k \sum_m X_{km}^{ij} = 1 \forall i,j \,;\, i,j \in V \tag{4.7}$$

4.2. Extended model

In this section the extensions to the existing model by Campbell [4] are explained. To meet the requirements, three features have to be changed. First, the the set of nodes V is split into 2 the sets C and J. After that, the capacity constraint described by Campbell [6] added to the model. Finally, non obligatory hub transport added to make integration of public transport possible.

4.2.1. Two types of nodes

To speed up the optimization process and ensure that only potential hub location nodes are chosen as jetties, the set V is subdivided into two sets. Two sets of nodes are introduced: first the set of possible jetty locations. This set of all possible jetties is $J = \{1,..,m\}$, containing all possible locations for jetties. The second set is the set of possible origins and destinations of passengers. The set $C = \{1,..,n\}$ contains all possible origins and destinations of travellers, the city nodes. The constraint equations are updated. i and j are chosen out of the subset C. k and m are selected from set J. The constraint formulations are updated: equation 4.2 is changed to 4.8. Equation 4.4 is updated to 4.9. Equation 4.5 becomes 4.10. Equation 4.6 is updated to 4.11. Equation 4.7 is changed to equation 4.12.

$$Y^k \in \{0,1\} \forall k \in C \tag{4.8}$$

$$X^{ij}_{km} \geq 0 \forall i,j,k,m\,;\, i,j \in C\,;\, k,m \in J \tag{4.9}$$

$$X^{ij}_{km} \leq X_m \forall i,j,k,m\,;\, i,j \in C\,;\, k,m \in J \tag{4.10}$$

$$X^{ij}_{km} \leq X_k \forall i,j,k,m\,;\, i,j \in C\,;\, k,m \in J \tag{4.11}$$

$$\sum_k \sum_m X^{ij}_{km} = 1 \forall i,j\,;\, i,j \in C \tag{4.12}$$

4.2.2. Capacity constraint

The capacity of a hub is determined by the flow (X^{km}_{ij}) multiplied by the number of people that travel from i to j within a certain time frame. In this research, the capacity constraint is used to limit the number of people arriving and leaving from a single jetty on daily basis. Equation 4.13 shows the equation that adds capacity constraints to the total flow Campbell [6]. The parameter is introduced, Ca_k, which is the limit on the capacity of jetty k. In equation 4.13 the total flow through a hub is limited by capacity Ca_k.

Ca_k Capacity of jetty at k

$$\sum_j \sum_i W_{ij} * (\sum_m (X^{km}_{ij} + X^{mk}_{ij}) - X^{kk}_{ij}) \leq Ca_k * X_k) \tag{4.13}$$

4.2.3. Non obligatory hub transport

To integrate the possibility to travel with public transport, a link between city nodes added. To make the least amount of changes to the model, an extra hub is added with special properties. The total travel time along this virtual hub is the travel time that it takes to travel with public transport. Figure 4.1 shows a visual representation.

The T_{ij}^{km} is a four dimensional matrix containing all possible travel times between i and j, where k and m are the jetty nodes. To add the travel time with public transport, a virtual jetty node is added to represent the travel time with public transport at jetty node 0. By adding jetty 0, all public transport travel times between i and j can be added to $T_{i,j}^{0,0}$.

When adding an extra jetty node that represent the travel time with public transport, the mathematical setup allows the algorithm to ignore the virtual jetty node. Therefore, the setup is changed in order to ensure the virtual jetty node is always part of the solution. From the variable X^k a predetermined number of nodes H is chosen (constraint equation 4.3). Equation 4.14 updates equation 4.3 by adding 1 to the total of predetermined number of hubs. Equation 4.15 ensures that virtual jetty 0 is chosen as jetty. No transport is allowed from other jetties to the virtual jetty. Therefore, equation 4.16 and equation 4.16 are used set the 'forbidden' combinations to infinite.

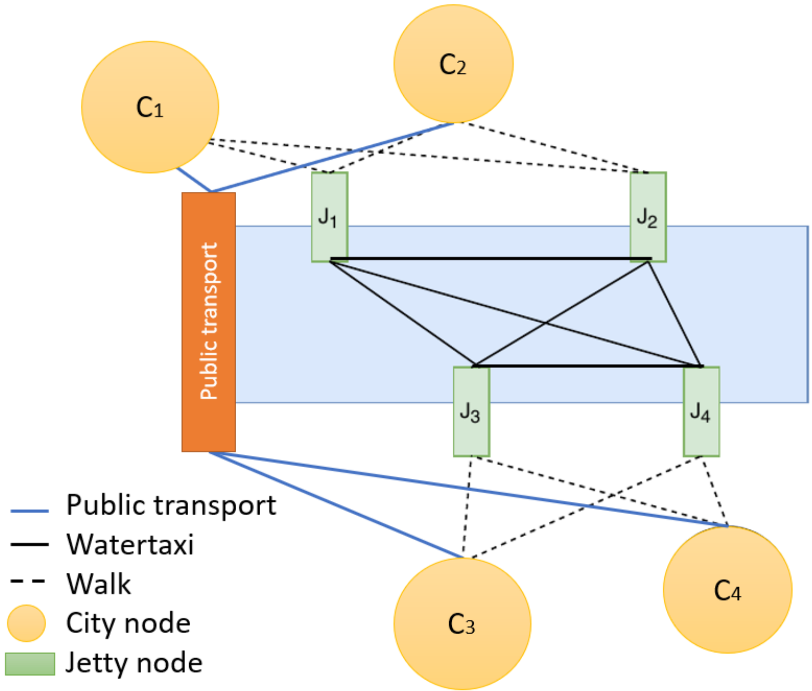


Figure 4.1: Connection of city nodes to public transport. Blue is water, J1 to J4 are jetties C1 to C4 are city nodes.

$$\sum X^k = H + 1 \tag{4.14}$$

$$X^0 = 1 \tag{4.15}$$

$$T_{i,j}^{k,0} = inf i, j \in C; k \in J \tag{4.16}$$

$$T_{i,j}^{0,m} = inf i, j \in C; m \in J \tag{4.17}$$

4.2.4. Key Performance Indicators for case study

In section 3.5, two performance indicators are found: objective outcome and the hub locations. Two of the KPIs of this research are based on the performance indicators in literature, the third KPI is based on the outcome of one of the decision variables. In this subsection, five Performance Indicators are introduced. With the design of the model in place, the mathematical background can be shown of the KPIs.

The performance indicators are divided into three categories: logistics KPIs, research KPIs and verification performance indicator. The logistics KPIs are used to evaluate the performance of the model in a case study and are numbered 1 and 2. The research KPIs are commonly used performance indicators in the field of HLPs in this research and are numbered 3 and 4. The verification performance indicators are used for the verification of the implementation in the coming section.

1. Average reduced travel time(s), ***logistics KPI*** This KPI shows average reduction of travel time for travellers for a configuration of hubs. For this KPI the outcome of the objective function is needed. Equation 4.18 shows the definition of this performance indicator.

$$\mathrm{KPI}_1: \frac{\text{Traveltime public transport} - \text{obj. outcome}}{\text{Total number of travellers}} = \frac{\sum_i \sum_j N_{ij} X_{ij}^{0,0} W_{ij}^{0,0} - \sum_i \sum_j \sum_k \sum_m N_{ij} Z_{ij}^{km} W_{ij}^{km}}{\sum_i \sum_j N_{ij}} \tag{4.18}$$

2. Percentage of people taking the watertaxi (%), ***logistics KPI*** To find out what percentage of the people takes the watertaxi instead of public transport, percentage of people taking the watertaxi is introduced. This KPI is defined in equation 4.19. To find its value, the flow through hub 0, the Public Transport hub, need to be found. Assumed is that all flow that does not take public transport takes the watertaxi. The mathematical formulation allows travellers to walk to a particular jetty and walk from that jetty to the destination. In that case, begin jetty m and end jetty k are the same. It requires a detour to the jetty, when a direct link is available. Therefore, it is assumed that no detours are taken.

$$\mathrm{KPI}_2: \text{percentage watertaxi} = 100 - \frac{\sum_i \sum_j Z_{ij}^{0,0} * W_{ij}}{\sum_i \sum_j W_{ij}} \tag{4.19}$$

3. Average calculation time (s), ***Research KPI*** One of the requirements of this research states that the calculation time of the designed model has to stay below 1.5 hours on a personal computer. To find a relationship between the number of placed jetties and the average reduced travel time, different numbers of placed jetties are tried. For every configuration, the implementation starts with loading all data, calculating the optimal solution and at last print all outputs. The time it takes to calculate all configurations is defined as the total calculation time. The total calculation time divided by the number of different configurations tried is the average calculation time. Equation 4.20 shows mathematical formulation.

$$\mathrm{KPI}_3: \text{Average calculation time} = \frac{\text{Total calculation time}}{\text{Number of configurations}} \tag{4.20}$$

4. Gap with optimal solution (%), ***Research KPI*** As described in the literature chapter, allowing a gap with the optimal solution can reduce the calculation time. In this research, the optimal solution researched. Therefore, no gap is allowed. Ebery et al. [9] defines the gap with the optimal solution as the relative deviation of from the optimal solution, expressed as a percentage.

5. jetty locations (-), ***verification KPI*** The jetty locations can be determined by examining the decision variable Y^k. If Y^k is 1, k is chosen out of of the location set to be jetty. This performance indicator is used in the next chapter to verify the implementation of the model of Campbell [4]

6. Objective outcome (s and \$), ***verification KPI*** To compare the outcome of the implementation with benchmark dataset, the objective outcome is used. The unit in case 1 to 3 is in dollars, in case 4 to 7, the unit is time in seconds.

4.3. Verification of model

To ensure the implementation of the literature is correct, 9 cases are discussed. The implementation of the mathematical model is done in Python 3.7.4. To find an optimal result, the python MIP package (version 1.7.1) is used. According to their website: *Python-MIP is a collection of Python tools for the modeling and solution of Mixed-Integer Linear programs (MIPs). Its syntax was inspired by Pulp, but our package also provides access to advanced solver features like cut generation, lazy constraints, MIP starts and solution pools.*[1]

4.3.1. Verification of main model

To verify the implementation in Python of MA-p-HLP described by Campbell [4], the CAB benchmark datasets (section 3.4.2) is used. The performance of the implementation is compared with the performance indicators found in literature (section 3.5). Two relevant performance indicators were found: the objective outcome and the hub locations. Boland et al. [3] presents the optimal objective outcomes with the Median p-HLP for the CAB benchmark dataset and Campbell et al. [5] presents the hub locations in the optimal solution.

To find out if the implementation is correct, three experiments have been done. In all experiments, an initial set of 10 locations were taken. Out of the set of 10 locations, first 2 hubs are chosen. In the second test, 3 hubs have been chosen. In the final test, 4 hubs have been allocated. In table 4.1 are the expected outcomes compared with the outcomes by this study. As table 4.1 shows, the results of the implemented design of Campbell [4] and the values found in literature are similar for the three cases.

Table 4.1: Comparison of results found in literature with results found by this study for CAB dataset. P the total number of locations, the discount factor for travelling between hubs is 1 ($\alpha = 1$) .

Case	**Test values**	**Expected locations of hubs (-) [5]**	**Expected objective ($) [3]**	**Hub locations (-)**	**Objective value** ($)	**Pass?**
1	*2 hubs, P = 10*	7,9	721	7,9	721	yes
2	*3 hubs, P = 10*	4, 6, 7	654	4, 6, 7	654	yes
3	*4 hubs, P = 10*	2,4,6,7	632	2, 4, 6,7	632	yes

4.3.2. Verification of extended model

To verify the inner working, a case is designed with 2 city nodes and 4 jetty nodes. The solutions of this case are checked by hand in Appendix B. Figure 4.2 shows 2 city nodes: postal code 3016 and postal code 3072. The jetty nodes are numbered by 15, 14, 26 and 260. 2 jetties have been chosen. In this section the designed model is verified by six cases. The six cases are described in table 4.2. The used travel times and other parameter values are shown in appendix B. Table 4.3 shows the outcomes of the different test cases. The number of travellers between the city nodes is 8.

Case 4

In the fourth case, 2 jetties have to be determined out of the set of the jetty nodes 260, 26, 15 and 14. Table 4.3 shows that the implementation of the designed model finds jetty 15 and 26 as chosen jetties with a objective value of 4016 seconds. Both findings are equal to the finding of the hand calculations.

[1]Retrieved from:https://www.python-mip.com/index.html#the-package on: 19th may 2020

Table 4.2: Six cases to verify the performance of the designed method.

Case	**Description**
4	Case with 4 jetties and obvious fastest path and if all variables have correct value.
5	Case where the obvious fastest route cannot be taken due to capacity limitations.
6	Case where flow separates due to capacity constraints.
7	Working of public transport.
8	Capacity constraint where part of people take public transport
9	Second city node pair is added

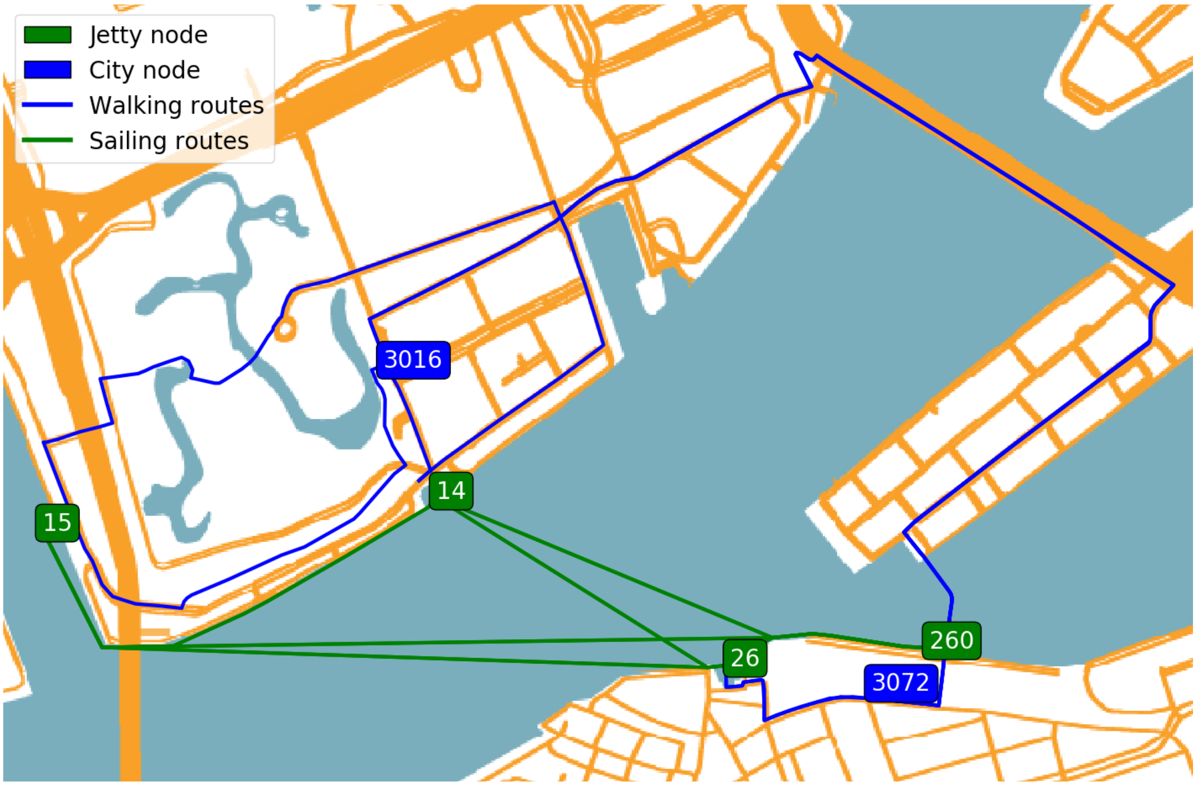


Figure 4.2: Visual representation postal codes 3072, 3016 locations of jetties 15,14 ,26 and 260. Sailing routes between the jetties have been plotted in green. The walking route between all postal areas and jetties have been plotted blue. The background map is originally produced by Google and adjusted for this research.

Case 5
In case five, the capacity constraint is verified. As shown in case four, without capacity constraints jetty 15 and 26 are preferred. In this test, 2 jetties have to be selected. To check the capacity limitation, the allowed maximum capacity is set to 0 for jetty 15 and 26. With the set capacity, one can see in table 4.3 that jetty 260 and jetty 14 are chosen.

Case 6
In case six, the spreading of flow is demonstrated by limiting the capacity. The jetties that result in the lowest travel time are jetty node 26 and 15 (case four). With limiting the capacity to 3 persons for both jetties the number of people travelling should be divided between the preferred jetties and the other jetties. The predetermined number of jetties is 4. As shown in table 4.3, the objective function is equal to the hand calculation. This makes it assumable that the flow of passengers is equal to the predicted values.

Case 7
In this case the public transport is added to the design. The settings of the travel time with public transport hub is 400 seconds. This is lower than the fastest route with the watertaxi. Jetty 5 is introduced as virtual jetty for public transport. The predicted behaviour of the implemented model is that all passengers are going to take the public transport. As all passengers take public transport, the locations of the jetties do not influence the total travel time. Therefore, no predictions can be made on the locations of the jetties.

Case 8
Now, only a percentage of people is travelling with the watertaxi by applying a capacity constraint in order to ensure that a selection of the people takes public transport. The capacity constraints found in B.3 are applied and the travel time with public transport is set to 500. The expected behaviour is that 3 people are going to take public transport (202+120+180)*3 and 5 people public transport (500). 100- 5/8*100 = 37.5 percent takes takes the watertaxi. The average reduced passenger travel time is (600*8-(600*5+3*502))/8 = 36.75 and jetties at 26 and 15.

Table 4.3: Six cases to find out if the implemented design works as expected

Case	Expected			Implementation values			
	Jetty Locations(-)	Watertaxi (%)	objective (s)	Jetty Locations(-)	Watertaxi (%)	objective (s)	pass?
4	J = 26, 15	-	4016	J = 26, 15	-	4016	yes
5	J = 260, 14	-	6168	J = 260, 14	-	6168	yes
6	J = 260, 26,15,14	-	5361	J = 260, 26,15,14	-	930	yes
7	J =0, -,-	0	0	J = 26,15,0	0	0	yes
			Average reduced traveltime (s)			Average reduced traveltime (s)	
8	J =0, 26,15	62.5	36.75	J = 0,26,15	62.5	36.75	yes
9	J =0, 26,15	57.2	56	J =0, 26,15	57.2	56	yes

Case 9
To verify a case with multiple Origin Destination pairs, another city node is added. Table B.4 shows an overview with all travel times. This location is the same as 3016, except from the travel time with public transport. Public transport from 3072 to 3012 is set to 600 seconds. From previous examples, the fastest travel option with the watertaxi is 502. The travel time with public transport is 400 between 3072 and 3016. This is lower than the fastest travel time with the watertaxi. This means that the travelling from 3072 to 3016 goes with public transport. The expected objective outcome should be: 400*6+502*8 = 6416. The average reduction is (400*6+600*8-6416)/14 = 56. Percentage that takes the watertaxi needs to be 100 - 6/14 *100 = 56. Like in the case 4, the expected jetties are 15 and 26.

4.4. Conclusion

In this chapter, the question *how is the watertaxi system going to be modelled* is answered. The capacity constraint of Campbell [6] and the Median p-HLP model by Campbell [4] were explained. To overcome the limitation on direct transport, a virtual jetty is introduced. The virtual jetty node has different properties than other jetty nodes. This jetty node is always part of the chosen jetties and no transport from other jetties is allowed to this jetty. This node is always part of the selected jetties. The travel time via this jetty is always equal to the direct flow between the two city nodes.

To evaluate performance in this research, three types of performance indicators are introduced: the logistics KPI, the research KPI and the verification performance indicator. Logistics KPI consist of the average reduced travel time and the percentage public transport. The research KPIs evaluate the performance of the implementation. And the verification performance indicators are used in this chapter to find out if the model is implemented in a correct way.

5 Data case study

To successfully execute the case study in Rotterdam, accurate data is required. Therefore, the question *which suitable data sources are going to be used for a case study to the city of Rotterdam?* is answered in this chapter. Data is needed on travel times between the nodes, which is shown in 5.1. To ensure the data is accurate, the found data is validated. The following data has to be found:

- **Number of travellers** Number of passengers travelling between begin and end nodes.
- **Jetty nodes** The coordinates of the jetty nodes.
- **City nodes** The coordinates of the locations where people departure from and arrive.
- **Sailing times** All sailing routes and times between jetty nodes.
- **Public transport** travelling time between the begin and end nodes.
- **Walking times** the time it takes to walk from a begin/end node to a potential jetty node.

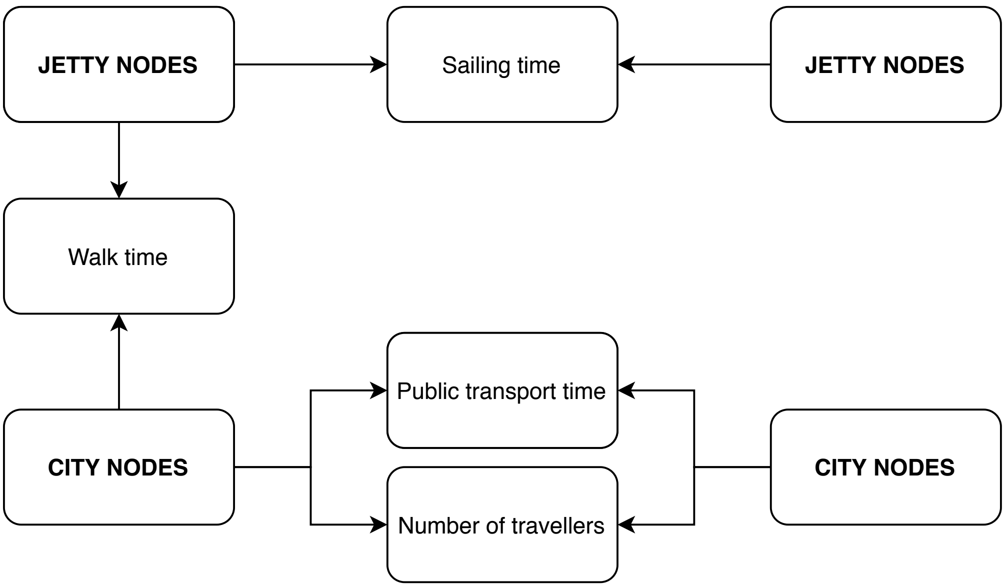


Figure 5.1: Information needed on the legs between the nodes.

5.1. Number of travellers

The travellers data is essential as input for the model that is determined in the design chapter. To answer the question several data sources are evaluated. Different data sources have their own specific advantages and disadvantages. The data sources are evaluated on the following four characteristics:

- **Resolution**: The resolution determines the level of detail of the data. Some data sources can only provide OD data between cities, which makes the OD data irrelevant for this research since its provides no information for the urban area itself. Ideally, the dataset would provide information what the origins and the destinations are of individual people.

- **Accuracy**: To improve the accuracy of the outcome of the method determined in this research, accurate data is required.

- **Accessibility**: As not all data is unlimited and free to used, accessibility needs to be taken into account.

- **User friendliness**: For some data sources, post processing could be necessary to get the required data.

5.1.1. Mezuro

Mezuro, according to their website[1], is a data science company specialized in extracting (urban) mobility information out of network data gathered by telecom providers. Mezuro divides the map of the Netherlands in 1246 cells, where the OD of people is found between. Since the data is based on measurements of cellphone connections, its accuracy is estimated to be fair. To access the data, a fee has to be paid.

5.1.2. Municipality of Rotterdam

The Municipality of Rotterdam owns data of origins and destinations of their citizens for Urban planning purposes. This data could be available Data from the municipality of Rotterdam. A common system to extract data is a Geographic Information System (GIS). This could be a obstacle to since this requires extra research on this type of system.

5.1.3. OViN 2017

The Dutch 'Onderzoek Verplaatsingen in Nederland' (OViN, roughly translated to: Research to trips within the Netherlands) is conducted by the Dutch government. OViN conducts to the travel behaviour of Dutch Citizens. All respondents is asked to monitor one specific day of the year where they travel to, with the mode of transport, with their reason of travelling and their travel time. Besides those questions few general personal and household questions are asked and questions about driving licenses and ownership of vehicles. Based on this research, information was gathered about the travelling of Dutch citizens on Dutch Territory [2]. The OViN 2017 consist out roughly 100.000 monitored itineraries. Origins and destinations are provided on postal code level. To analyse the itineraries within Rotterdam, the itineraries within Rotterdam need to be extracted from the dataset.

To this research, 37016 people responded, which is 0.22 percent of the Dutch population. This means that 1 in the 455 Dutch citizens responded. Due to the lack of better data, assumed is that the sample represents the Dutch citizens perfectly. Therefore, assumed is that every registered trip counts for 455 trips[3].

5.1.4. OD estimation on watertaxi itineraries

Based on data owned by the company 'Watertaxi Rotterdam', estimates can be made on the origins and destinations of people. Watertaxi Rotterdam owns information on the itineraries from origin jetty to destination jetty. To find the origins and destinations of the travellers within the city centre, a separate literature study is needed to done. A way needs to be found to estimate the origins and destinations of the travellers within an urban area on data owned by Watertaxi Rotterdam.

[1] https://www.mezuro.nl/over-mezuro/

[2] Retrieved from: https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:103498/tab/1/rd/1 on 14th of April

[3] https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/onderzoek-verplaatsingen-in-nederland–ovin– retrieved on: 20-08-2020

Table 5.1: OD data sources compared. ++ is very suitable, + is suitable, o is neutral, - inconvenient, – very inconvenient.

	Mezuro	Rotterdam municipality	OD estimation	OViN 2017
Resolution	o	++		+
Accuracy	+		- -	++
Accessibility	–	-	+	+
User friendliness	++	- -	- -	++

5.1.5. Conclusion on traveller data

To find a suitable data source, the different data sources should be evaluated on the four indicators shown in the introduction of this chapter. For every indicator a score is given between inconvenient (–) and very suitable (++). Table 5.1 shows an overview of this comparison. In this table four OD data sources are compared. From this table can concluded that the OViN 2017 data is the most suitable for this research.

5.1.6. Number of travellers (travellers between city nodes)

?? With locations determined, the number of travellers is extracted from the OViN2017 dataset. The following data have been found:

1. **Total number of itineraries: 1899** Within the OViN2017 dataset, roughly 100,000 itineraries are registered between postal codes in the Netherlands. For this research, only itineraries that occur within the postal code area of Rotterdam are considered.

2. **Number of postal code areas: 75** Rotterdam is split up in 75 postal code areas.

3. **Unique itineraries: 720** Within the OViN2017 dataset, the number of unique itineraries within Rotterdam are examined. To find the number of itineraries that occur at least once, the number of unique itineraries is examined. This includes itineraries that have the same origin postal code as destination postal code. With 75 postal code area, there are 75*75 = 5625 possibilities. Only 12 percent of the possible itineraries is registered.

4. **Number unique itineraries: 395 (one direction)** The graph is assumed to be undirected. With this assumption, itineraries from origin to destination have an identical travel time with the travel time in opposite direction. When assuming itineraries only occur in one direction, 395 unique itineraries are identified.

5.2. Coordinates of postal codes (city nodes)

The OD data has a resolution on postal code level. Those postal codes have to be converted to coordinates on the map. Assumed is that people leave from the middle of the postal area. To find the latitude and longitude a database is used[4]. In appendix C, the locations of the postal codes are shown.

To make assumable that this dataset is correct, the postal code of 5 latitude and longitudes is validated. Table 5.2 the longitudes and latitudes of the database are checked on correctness with Google maps. Figure 5.2 shows in red all city nodes based on the dataset that converts postal code areas into city nodes. All postal codes used with coordinates within Rotterdam are shown in appendix C.

Table 5.2: Validation of locations of postal codes match with the postal locations on Google Maps.

Postal code	Coordinates	Adress according to Google	Pass
3012	51.9194046,4.4757692	Karel Doormanstraat 340, 3012 GR Rotterdam	Correct
3027	51.9182359,4.4366362	Da Costastraat 156, 3027 JL Rotterdam	Correct
3072	51.9018612,4.4842992	Veerlaan 13-21, 3072 AN Rotterdam	Correct
3042	51.9342811,4.4327096	Rotterdam Rechter Maasoever, 3042 NL Rotterdam	Correct
3035	51.9324493,4.4811214	Jacob Catsstraat 79-91, 3035 PG Rotterdam	Correct

[4]Retrieved from:https://github.com/bobdenotter/4pp

Table 5.3: Constants used to determine the sailing time between two jetty nodes.

Variable	value
$t_{acceleration}$	92.8 s
v_{slow}	2.4 m/s
$v_{maneuver}$	6.0 m/s
v_{cruise}	12.9 m/s
t_{aboard}	120 s

5.3. Jetty nodes

The jetty nodes are the possible locations where a jetty can be placed. The coordinates of the jetty nodes are chosen on the basis of the existing jetty locations within Rotterdam. A total of 79 jetties are chosen within Rotterdam. Watertaxi Rotterdam does not use all of the jetties within Rotterdam. In figure 5.2 all jetty nodes used for this research within Rotterdam are shown.

5.4. Sailing times (jetty node to jetty node)

With an algorithm that can calculate the distance between two jetties, the total travelling time is calculated based on the three travel zones (equation 5.1, equal to equation 2.1). First, the parameters are filled in equation 2.1. The values of the parameters are shown in table 5.3. After that, a route algorithm developed by Flying Fish is used to find the distance in every speed zone and can the travel time with the watertaxi $t_{watertaxi}$ be calculated.

$$t_{watertaxi} = t_{aboard} + t_{acceleration} + \frac{d_{zone,1}}{v_{max,1}} + \frac{d_{zone,2}}{v_{max,2}} + \frac{d_{zone,3}}{v_{max,3}} \tag{5.1}$$

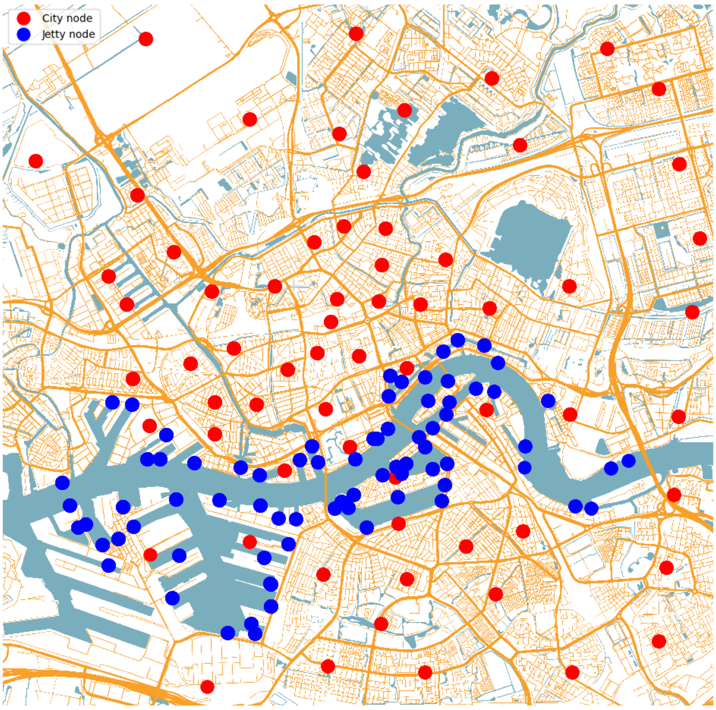


Figure 5.2: City of Rotterdam with all city nodes and jetty nodes plotted. In blue all jetty nodes, in red the city nodes. The background map is originally produced by Google and adjusted for this research.

5.4.1. Validation

To find out if this formulation is accurate, the itineraries are compared with the duration of watertaxi itineraries measured by Flying Fish. Figure 5.3 shows the results based on itineraries that at least occur 10 times in the database. In figure 5.3 the average and error bars are plotted. The error bars are based on the mean plus and minus the Root Mean Square Error. The calculated sailing time is plotted blue when the calculated sailing time falls within the error bars, and red otherwise.

Figure 5.3 shows that the calculated values are correlated to the sailing time. In several cases the calculated value exceeds the error margin. Some itineraries show a major spreading in sailing time. Even when the calculated value deviates significantly from the average, the calculated value does not exceed the error margin. The parameters of the mathematical formulation are tuned on the same sailing duration data as shown in the figure. This could mean that if another set of sailing duration data is used as comparison, the calculated value could show an increased deviation.

5.5. Walking time data (city node to jetty node)

When the locations of the jetties and the postal codes are known, the time it take to walk between locations can be determined. To find the total travel time, an estimation has to made on the travel time to walk from and to the jetties. To find the travel times, the API of Google Maps is used[5]. Google Maps is widely used for estimating travel times.

5.6. Public transport (city node to city node))

To find the travel time with public transport, the departure time is important. It could be possible that some modes of public transport are not available during some hours of the day. Another example can be when travelling during the weekend. It could be possible that some modes of transport are operated less frequently during weekend days. Therefore, the public transport times are taken on a Monday at 8:30AM.

5.6.1. Validation of Public Transport

The time it take to travel between two points within a urban area with public transport has a significant influence on the outcome of the KPIs. The public transport time in most cases consist of walking to public transport, taking public transport and walking to the destination. The average reduced travel time and also the percentage of the number of people in the system that takes public transport can change significantly by small change in travel time.

To find out how accurate the estimated travel times are, three methods of estimating travel time with public transport are compared on their duration. The first method is the travel time that is given back from Google Maps Static API. This method is used to find all travel times by walking and public transport. The second method is using the coordinates of the different city nodes and manually find the travel time. All used coordinates for postal codes are shown in appendix C. The third method consist of manually checking the website 9292.nl. This website provides travel information on public transport within the Netherlands. For all methods, the duration is found for Monday 17 august 2020 8:30AM and are consulted on 14 august 2020. With the duration, only the time it takes to travel is taken into account, the waiting time to start travelling is neglected. With public transport travelling time, also the time to walk to and from public transport is taken into account.

Table 5.4 shows the results of this validation step for the five most used itineraries according to the data used in this research. The table shows that the Google Static maps API shows different results in comparison to the website of Google maps. On average 9292.nl gives higher travel times than the two sources from Google.

The difference in travel times of the different sources can be caused by different reasons. It may be explained by the difference in walking speed or route, but can also be cause by the buffer time between public transport and walking. This could result in for example that the total travel time takes a later metro into account than another travel time finder.

In this research is assumed that the travel time in two directions is similar and therefore taken as equal.

[5]Retrieved from: https://developers.google.com/maps/documentation/maps-static/overview on 4th October 2020

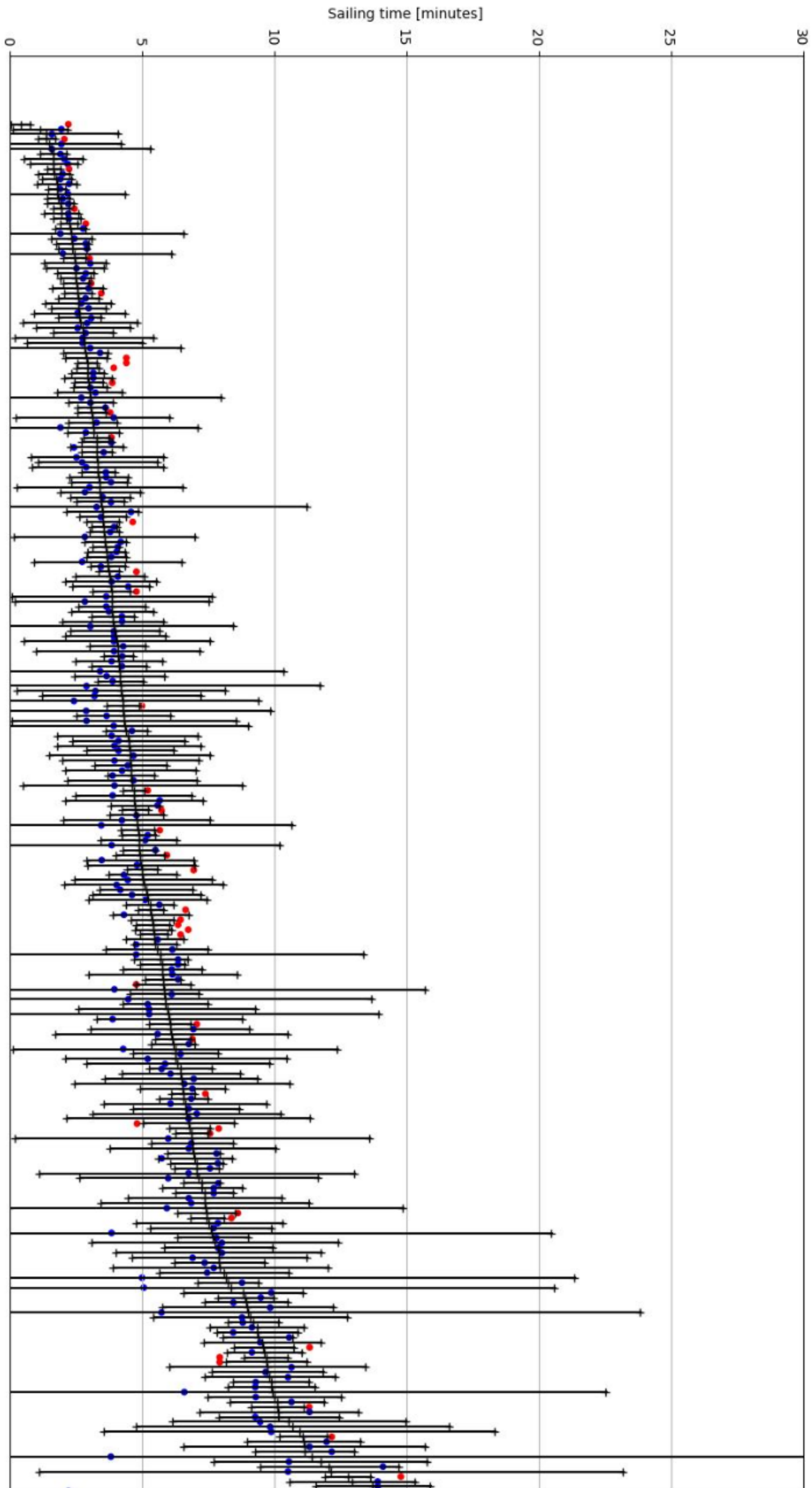


Figure 5.3: Comparison of the duration of the calculated trip duration and the actual trip duration. The colors indicate the difference in duration as deviation.

Table 5.4: Travel times compared for public transport from different sources.

Origin	Destination	Number of itineraries (*455)	Google Static Maps API (minutes)	maps.Google.com (minutes)		9292.nl (minutes)	
					opposite		opposite
3011	3063	19	29.2	26	25	26	26
3011	3072	9	23.0	23	21	24	25
3011	3024	8	23.7	23	22	26	26
3016	3072	6	28.1	28	28	33	32
3071	3082	6	40.9	40	40	36	37

Table 5.5: Number city nodes, registered itineraries, estimation on number of people and unique number of itineraries for every step of the preprocessing steps. The estimation on the number of people is determined by the OViN 2017 survey.

step number	**1**	**2**	**3**	**4**	**5**
Number of City nodes used	75	63	63	63	20
Number of registered itineraries	1899	1899	1899	1283	86
*Estimated number of people (*455)*	864045	864045	864045	583765	39130
Number of unique itineraries	720	720	720	395	27

For the Google maps static API only the trip from origin to destination is taken into account. For the other two methods of finding travel time, for the five compared routes the travel time in the 2 normal and opposite differ maximum of one minute.

5.7. Selection on registered itineraries and city nodes

To decrease the calculation time, the number of city nodes and registered itineraries are reduced. The number of itineraries that is deleted can be found in table 5.5. This is done by carefully filtering on itineraries that are taken into account. In this subsection the following steps are undertaken to limit the number of nodes:

1. **itineraries within a postal code** Within the conducted research on itineraries within the Netherlands, also itineraries within postal codes are registered. Since those itineraries do not have any influence on the outcome of this research, these itineraries are ignored.

2. **Compare travel time** the focus of this research is on the itineraries where a watertaxi can potentially save travel time. Therefore, all itineraries where there is no possibility for a watertaxi to be quicker than public travel time according to the harvested data, the itineraries are ignored.

3. **Mirror routes** Assumed is that routing travel time from Origin to Destination is equal to travel time to Destination to Origin. Therefore, the Destination to Origin trip is added to the Origin Destination trip.

4. **Empty Origins** Only city nodes that are used as origins and destinations have an influence on the outcome. Therefore, all nodes that are not an origin or destination are ignored.

Figure 5.4 show the class diagram of of the four steps described above. In every step, either the number of nodes or the number of travellers decreases. In table 5.5 an overview is given of the number of city nodes and the registered itineraries. The initial number of nodes is 75. After all preprocessing steps, the number of nodes is decreased to 20.

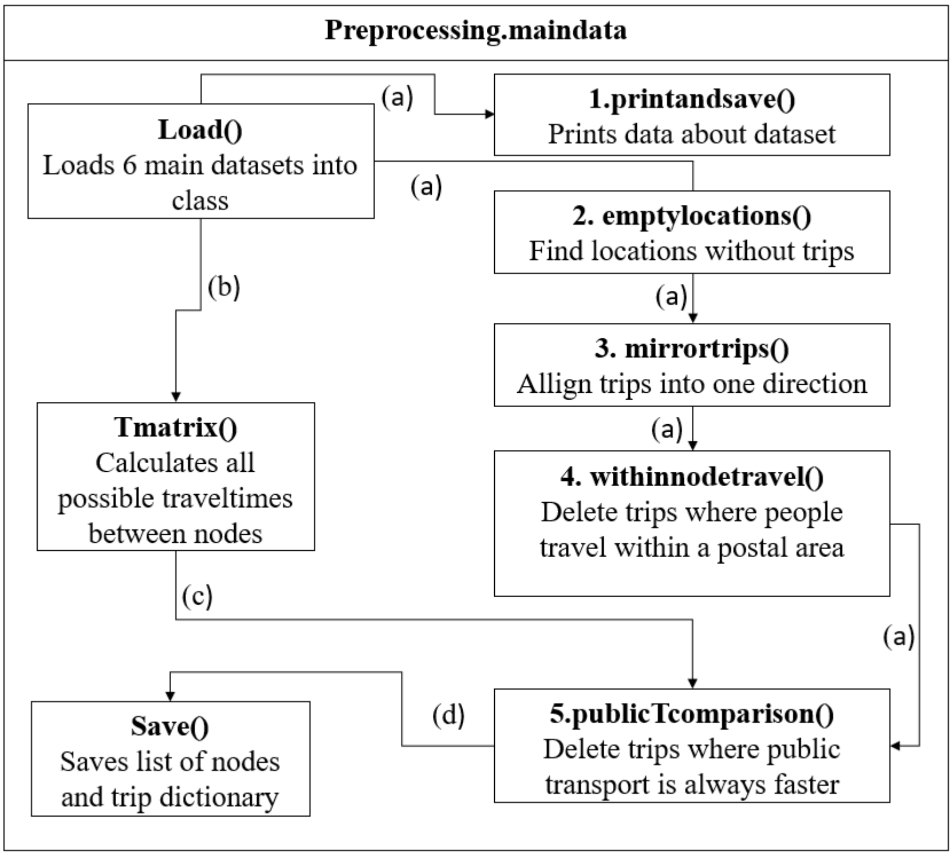


Figure 5.4: Class diagram of preprocessing with the flow of information between different functions. (a) is a dictionary containing all passenger travel. In every step itineraries are deleted that are not relevant. (b) are the 3 dictionaries based on the travel times between Jetty and City nodes. (c) is an dictionary containing the fastest way to travel with an watertaxi from an origin to a destination. (d) contains the list with the selected jetties and the dictionary with all selected itineraries.

5.8. Conclusion

The question *which suitable data sources are going to be used for a case study to the city of Rotterdam?* is answered by finding all six types of data that are needed for the case study. First, the number of travellers and the city nodes have been described. OViN2017 provides itineraries on postal code detail. For the sail times between the jetty nodes, the routing algorithm of Flying Fish is used. To find the coordinates of the jetty nodes, a dataset of Flying Fish is used. Those estimations are calculated for the different speed zones. For walking routes and public transport routes, the API of Google Maps is used. In order to only select travellers that can reduce their travel time by taking a watertaxi, the travellers are selected during a preprocessing process. Table 5.6 gives an overview of all the validation steps, data sources and the number of nodes and routes.

Table 5.6: Six data sources compared on their source, validation step and number of used routes and nodes. (1) is the number of routes or nodes before the preprocessing step. (2) is the number of routes or nodes after preprocessing.

	Source	**Validation**	(1)	(2)
Walking	Google Maps API	Visual	5925	1500
Sailing (route)	Flying Fish Algorithm	Compared with historical data	2775	2775
Public transport (route)	Postal code to coordinate dataset	Compared with 9292ov and Maps	3081	190
Jetty node	Flying Fish dataset	Locations checked on map	75	75
City node	Postal code to coordinate dataset	Sample compared with Google Maps	79	20
Number of travellers	OViN 2017 Traveller survey		720	27

6 Case study

In this chapter the question *what are the results of the implemented method for the city of Rotterdam?* is answered. A total of five experiments is conducted, divided over three sections. The first experiment shows the result of the designed method (general case study experiment). Its results are presented in one map, two plots and two tables. After that, an experiment is conducted with different numbers of people. Thirdly, the number of persons that is allowed to arrive and leave from a single jetty is limited. To find out if this method is capable of predicting the average number of itineraries through the city of Rotterdam, historical sailing data is compared with the outcomes of this research. In the final experiment, the number of nodes is varied to find a relationship between average calculation time and the number of nodes.

All experiments have been run on a Zbook G5 studio with 16GB DDR4 RAM, Windows 10 Home, Gurobi, Python MIP 1.7.2 and Gurobi Optimizer 9.0.2. In all experiments, no gap with the optimal solution is allowed. To keep the calculation time as stable as possible, all background programs were shut down during experiments.

6.1. General case study experiment

For the general case study experiment, the algorithm allocates of different numbers of jetties. A total of 21 different configurations are tried. The experiment starts with the allocation of 2 jetties and ends with 22 jetties. The following data is used: the number of people and the city nodes are based on the OViN 2017 dataset, the duration of public transport and walking is found by Google Static Map API. Table 6.1 shows an overview of all itineraries from origin city node to destination city node. In this experiment the relationship between the average reduced travel time and the number of placed jetties is investigated.

In the following section the data is shown. First, the average reduction of travel time is plotted against the number of jetties used in figure 6.2. After that, a plot is made in the spreading of the results in a boxplot in figure 6.3. A map is plotted with the locations of the city nodes and the chosen jetty nodes in figure 6.1.

6.1.1. Performance visualisation and average calculation time

Figure 6.1 shows the city center of Rotterdam with city nodes, jetty nodes and the selected jetty nodes. For this particular plot 12 jetties were used. Between all city nodes, routes have been plotted. All city nodes are connected to at least one jetty, except from city node 3073 and 3063. These two nodes are connected with public transport, due to the shorter shorter travel time with public transport. A remarkable result is that jetty 126 is not connected with a sailing route. An explanation could be that this jetty is only used to connect two walking routes. This would mean that the assumption made in the design phase is not correct. In section 4.2.4 is assumed that all flow that does not go with public transport would use the watertaxi. Table 6.1 shows an overview of all itineraries from origin to destination. With table 6.1 itineraries are checked on correctness.

Table 6.2 shows the added jetties for every configuration. In this table the first configuration is two jetties. The second configuration uses three jetties. Displayed is which jetties are added in the second configuration. In general, for every step a new jetty is added and no jetty is deleted. There are a few exceptions where two jetties are added and one deleted. Besides the added and deleted jetties, the calculation time for every step is shown. Mean is 161, maximum is 174. Minimum is 151. There is not a clear correlation between number of predetermined jetties and calculation time. The total calculation time is 3440 seconds, less than the required 1.5 hours, which is one of the requirements stated in the introduction.

Table 6.1: itineraries between city nodes from origin to destination. The number of itineraries is multiplied by the factor 455, which compensates for the number of people that took part in the OViN 2017 survey.

Origin	Destination	Number of itineraries (*455)	Origin	Destination	Number of itineraries (*455)	Origin	Destination	Number of itineraries (*455)
3011	3016	3	3024	3063	1	3063	3072	2
3011	3024	8	3024	3071	1	3063	3073	1
3011	3063	19	3024	3072	2	3063	3075	1
3011	3072	9	3025	3087	2	3063	3082	1
3011	3088	3	3027	3072	1	3071	3072	2
3016	3024	1	3028	3071	2	3071	3082	6
3016	3072	6	3032	3071	2	3072	3081	2
3021	3072	1	3032	3088	1	3072	3082	2
3024	3029	4	3036	3088	1	3077	3087	2

6.1.2. Result on logistic KPIs

Results in figure 6.2 show the average reduction of travel time versus the number of jetties used. The percentage of people that take the watertaxi is shown with the blue dots. In blue a line is plotted where the average reduced travel time converges to. From 20 jetties, extra added jetties do not further reduce the average travel time. Between 13 jetties and 20 jetties, the average reduced travel time slowly increases. From 12 jetties, the percentage of people taking the watertaxi is starting to decrease to the value of 7 percent when taking only 2 jetties into account. When looking at the change from 4 to 5 jetties, the result is remarkable. The percentage of people taking the watertaxi decreases. This result can be explained by that another configuration of jetties is chosen that increases the average reduction in travel time.

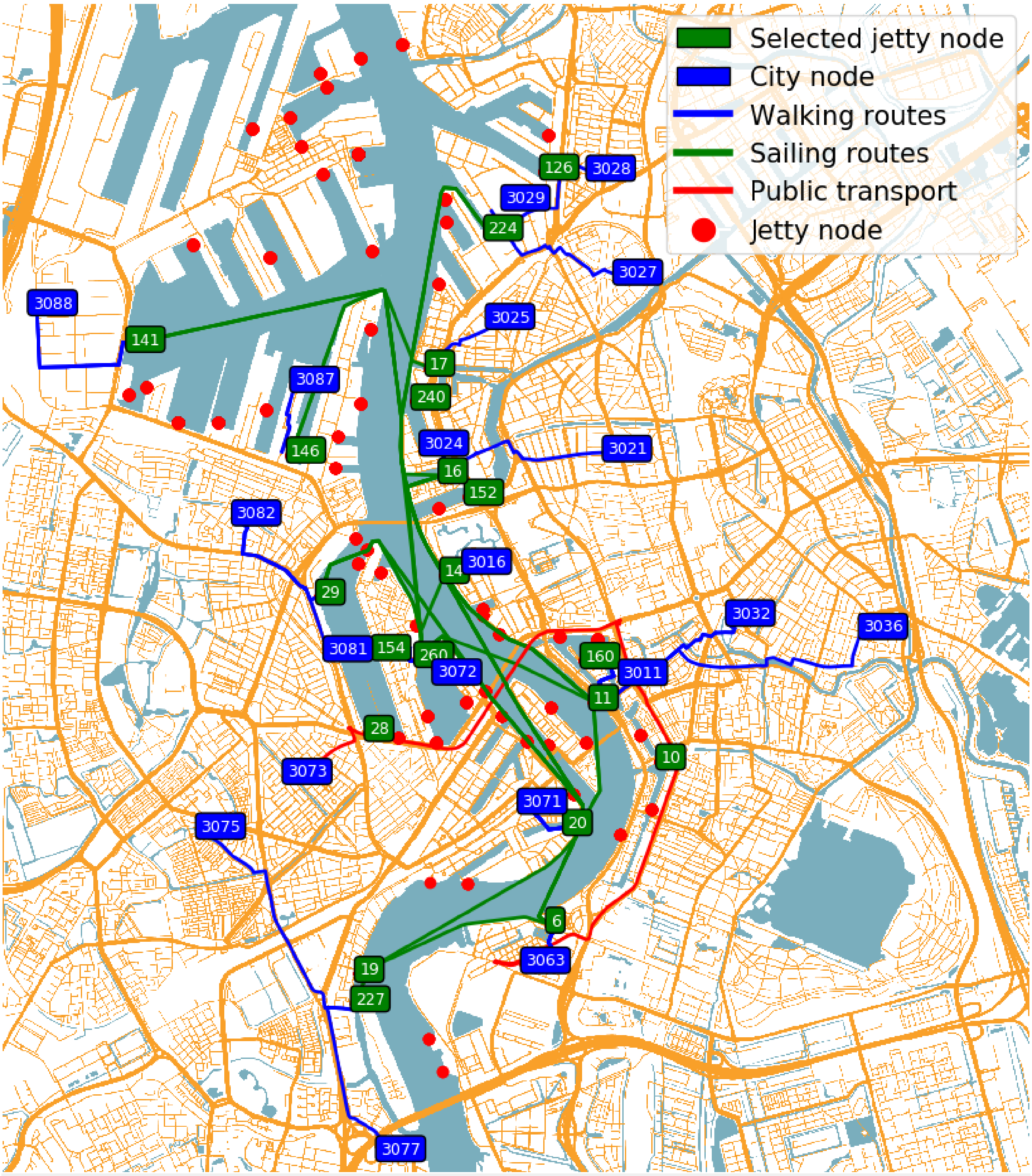


Figure 6.1: Map of Rotterdam (turned 90 degrees) with an overview of the locations of the jetties nodes, the selected jetties and used city nodes. In green the allocated jetty locations with ID and in blue used city nodes with postal code. In red all jetty nodes.The background map is originally produced by Google and adjusted for this research.

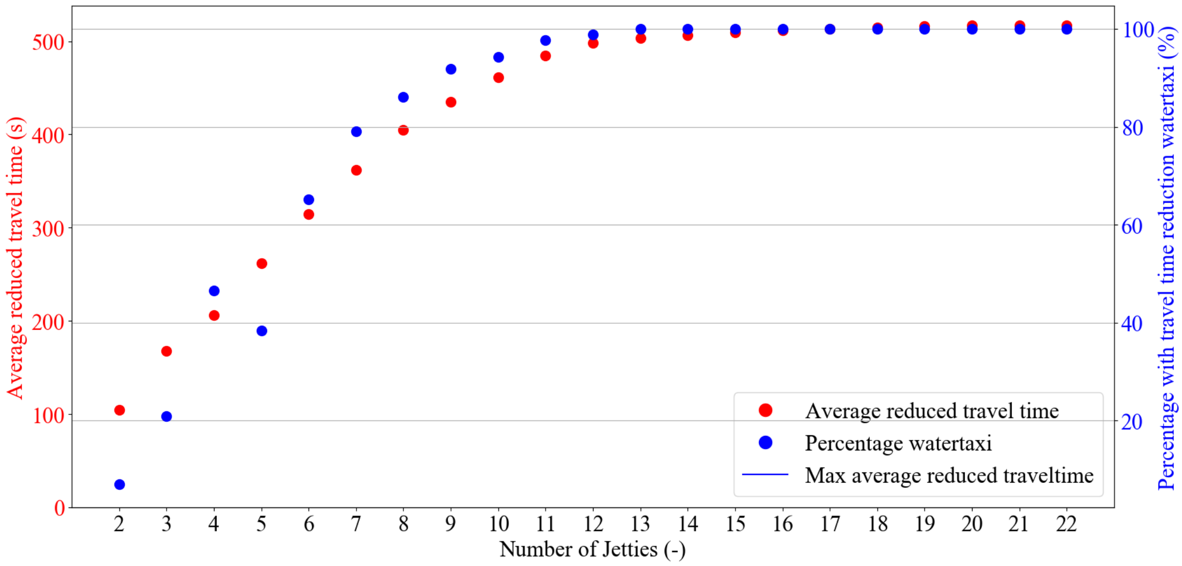


Figure 6.2: The average reduced passenger travel time and the percentage of all travellers that takes the watertaxi.

Spreading in results

The average does not give a perception of the spreading of different reductions in passenger travel time. To find this, the reduction in travel time for every registered trip is extracted. The itineraries without any reduction in travel time are deleted for the overview. 6.3 shows a boxplot for the outcome of 2 to 22 predetermined jetties. In orange the median value is depicted. The first quartile is the median for the lower half of the data points and is shown by the lower bar of the box. The median of the higher half of the data points is depicted by the top bar of the box. The extreme values are visualised by the top and bottom line that are connected with the box. The circle shows an outlier. The number of registered itineraries that is taken into account for every step can be found at the x-axis after the letter 'P' in the figure. The number of jetties that is been chosen can be found after the 'J'.

With 2 jetties chosen, there is no spreading in the reduction in travel time. For all six passengers, the reduced travel time is 1500 seconds. When 3 jetties are chosen, 18 registered itineraries have a reduced travel time by taking the watertaxi instead of public transport. For all registered 18 itineraries, 250 seconds can be reduced. When comparing the plots of 4 and 5 jetties, the number of people that use the watertaxi drops. This is a similar result that can be found in figure 6.2. This graph shows a increase in people that take public transport when comparing the results for 4 and 5 jetties.The highest found reduced travel time is 1800 seconds, the lowest found reduced travel time is 14 seconds.

Table 6.2: Jetties added and deleted for every step, the calculation time per step. In the Jetty ID column, the number of added jetties is shown and in bold and italic the IDs that are deleted for the solution.

Selected Number of jetties	Jetty added ***Jetty deleted***	**Calculation time (s)**	**Selected Number of jetties**	Jetty added ***Jetty deleted***	**Calculation time (s)**	**Selected Number of jetties**	Jetty added ***Jetty deleted***	**Calculation time (s)**
2	260,14	151	9	141	165	16	126	162
3	11	156	10	146,29 ***30***	165	17	152	166
4	6	159	11	227	174	18	154	167
5	29,20 ***6***	160	12	17,16 ***240***	166	19	19	165
6	6	163	13	28	162	20	160	166
7	16	164	14	240	166	21	122	168
8	224,240,30 ***29,16***	166	15	10	164	22	123	169
Total calculation time		**3444s**	**Average calculation time: 164s**					

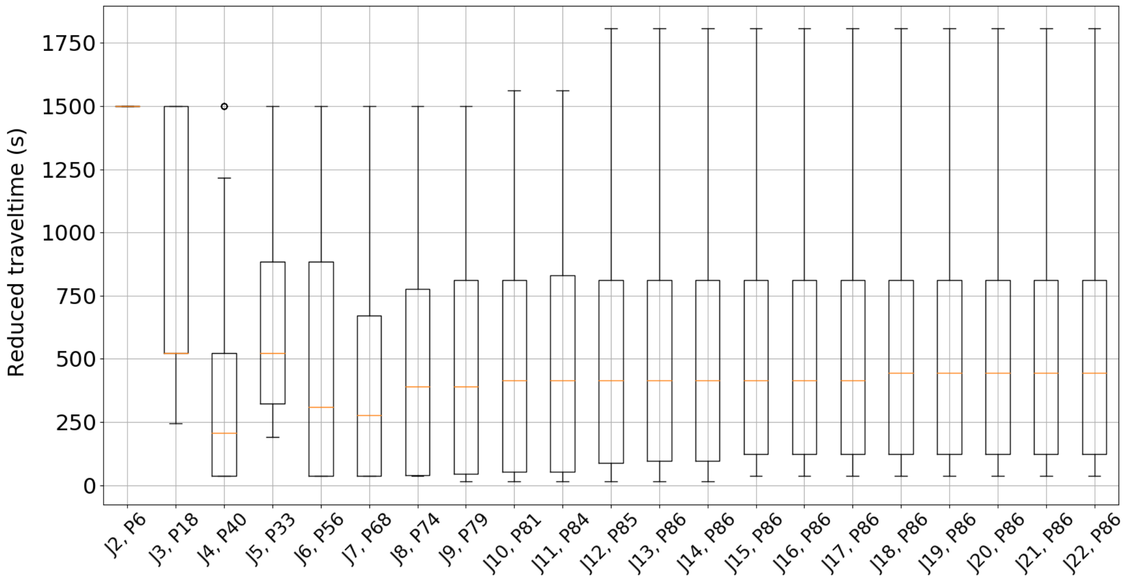


Figure 6.3: Boxplot with the difference in travel time between travelling with public transport and travelling with the watertaxi. Only results where a watertaxi is quicker is taken into account. P is the number of itineraries that are used. J is the number of jetties that is chosen.

6.2. Variations on case study

Other experiments can be done with the case study data. The following experiments are conducted:

1. **Sensitivity on number of travellers** In this experiment, extra travellers are added to the case study data to find out what the influence is on the KPIs. The average calculation time (KPI) is 210 seconds.

2. **Capacity limitations** The method is capable of limiting the capacity on the jetties. This is investigated in this experiment. The average calculation time (KPI) is 152 seconds.

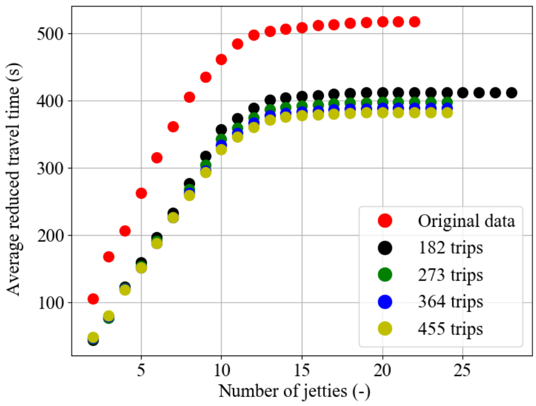
3. **Estimated number of people with travel time reduction** In this subsection, estimation are made on the travellers that can reduce their travel time by using a watertaxi. This data is comparison with historical data of Watertaxi Rotterdam. The average calculation time (KPI) is 62 seconds.

6.2.1. Sensitivity on number of travellers

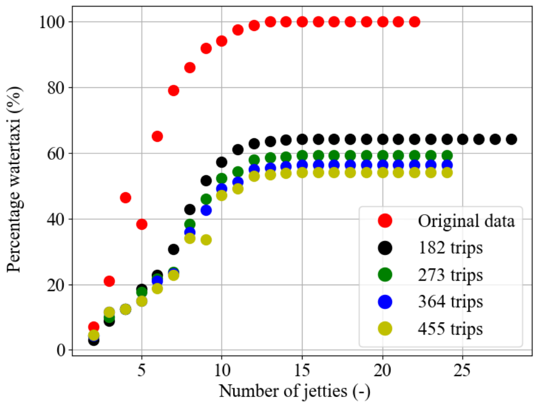
In this experiment the number of registered itineraries between the city nodes is changed. The preprocessed dataset contains, 27 out of the 190 possible different itineraries are used. In this experiment, all of the 190 possibilities is used. Those itineraries are added to the original set of origins and destinations. City nodes are shown in figure 6.1. The original set of itineraries is shown in table 6.1. Four scenarios are investigated and compared with the situation described in the first experiment of this chapter. Since 1 in 455 Dutch citizens submitted the OViN 2017 survey, the number of itineraries is multiplied by 455. To every trip, a factor of 0, 0.4, 0.6, 0.8 and 1 is multiplied by 455. This means that a total of 0, 182, 273, 364 and 455 are added to every trip respectively.

Figure 6.4 shows the result of this experiment compared with the outcome of the general experiment, conducted with similar data. For every scenario, the average reduced travel time follows a similar curvature. The average passenger travel time increases barely after 15 jetties. The percentage that travels with the watertaxi also follows a similar curvature to the number of jetties. For both the percentage and the average reduced passenger travel time, when the number of passengers between city nodes increase, the reduced travel time seems to convert to a limit.

The stop in increase in average travel time reduction at roughly the same number of jetties could be explained by the difference travel speed. Travelling on land is according to the parameters used significant slower than sailing. This could mean that the jetty that is closest to the city node results in a bigger reduction in travel time than a overall shorter route with a shorter path on water. Figure 6.4b shows that for the four added scenarios asymptotically go to values between 64% and 57%. An explanation for this behaviour could be that itineraries are not preprocessed as described in chapter 5.7, which could result in itineraries that are in no scenario faster than public transport. With adding more people to the link, the graph seems to reach a asymptomatic limit. This could be caused by the fact that the influence of the original data decreases in comparison with the added itineraries. With adding even more people to the model, the number of people averages over all itineraries becomes evenly divided.



(a) The average reduction in travel time



(b) Percentage of people taking the public transport

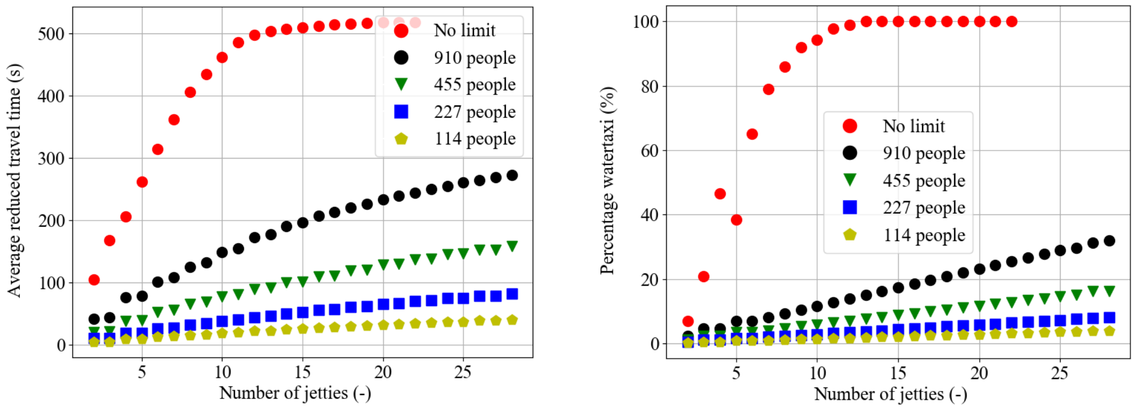
Figure 6.4: The general results compared to results with more itineraries added.

6.2.2. Capacity limitations

At the times this research was conducted, COVID-19 was spread around and the big crowds on jetties should be avoided. The method is capable of limiting the number of people that arrives and leaves from a jetty. The experiments consist of the data found in chapter 5. On every jetty, the daily capacity is limited. Four scenarios are compared to the experiment without capacity limitations. The maximum allowed people on daily basis on a jetty is 910, 455,299 and 114 people. The locations of the city nodes is shown in figure 6.1.

Figure 6.5 shows the results of the four scenarios in comparison with the general experiment. The average reduced travel time steadily increases and the percentage that uses public transport decreases with the number of jetties. With a limit of 910 people, the average reduced travel time is with 28 jetties is 280 seconds. The graph of 227 people and 114 people does not exceed the 100 seconds.

With limitations, more jetties are needed to meet demand for quicker routing. This could explain the steadily increase of reduced travel time. With a limitation of 114 people, only for a small number of people the watertaxi reduces their travel time (around %).



(a) The average reduction in travel time

(b) Percentage of people taking the watertaxi

Figure 6.5: Capacity settings for the jetties of 910, 455, 227 and 114 people compared with the results without capacity limitations.

6.2.3. Estimated number of people with travel time reduction

With the estimations made on the numbers of people that travel through Rotterdam, estimation can be made on the number of people that can reduce their travel time by taking a watertaxi. With finding the jetties, it is possible to find the flow between the jetties with looking at the variable outcomes of the solver. The data found is compared with itineraries made by the watertaxi in 2017. The data of the watertaxi averaged over the year. By comparing the predictions with real data, one can conclude if this method is capable of predicting the number of itineraries. Ideally the result should show a correlation. Historical trip data from the year 2017 is used as comparison with the results from the implemented method. All the estimations are on daily basis

To estimate how many people would reduce their traveltime by taking the watertaxi, first the list of initial jetties is changed to only jetties Watertaxi Rotterdam uses. Watertaxi Rotterdam uses 42 jetties within Rotterdam. By setting the predetermined number of jetties to 42, all jetties nodes are forced to be allocated as jetties. To find out if the method is capable of showing a correlation between the number of predicted itineraries and the real number of itineraries, two graphs are made. The flow is defined as the number that arrive or leave from a jetty. The second graph shows the predicted itineraries versus the real itineraries.

Figure 6.6 shows the estimated flow through several jetty IDs. On the x-axis, the predicted flow is shown. On the y-axis, the real flow is shown relative to the biggest real flow. Figure 6.7 shows the estimated itineraries versus the actual itineraries.

Figure 6.6 and 6.7 show little to no correlation between the simulated flow and the flow based on historical data. The difference between both axis is at least a factor 10.000. In this research travel time reduction is calculated with the number of people that would benefit from a travel time reduction. This model is not capable of predicting number of travellers since that requires research to travellers motivation. This could explain the factor of 10.000 between the two axis.

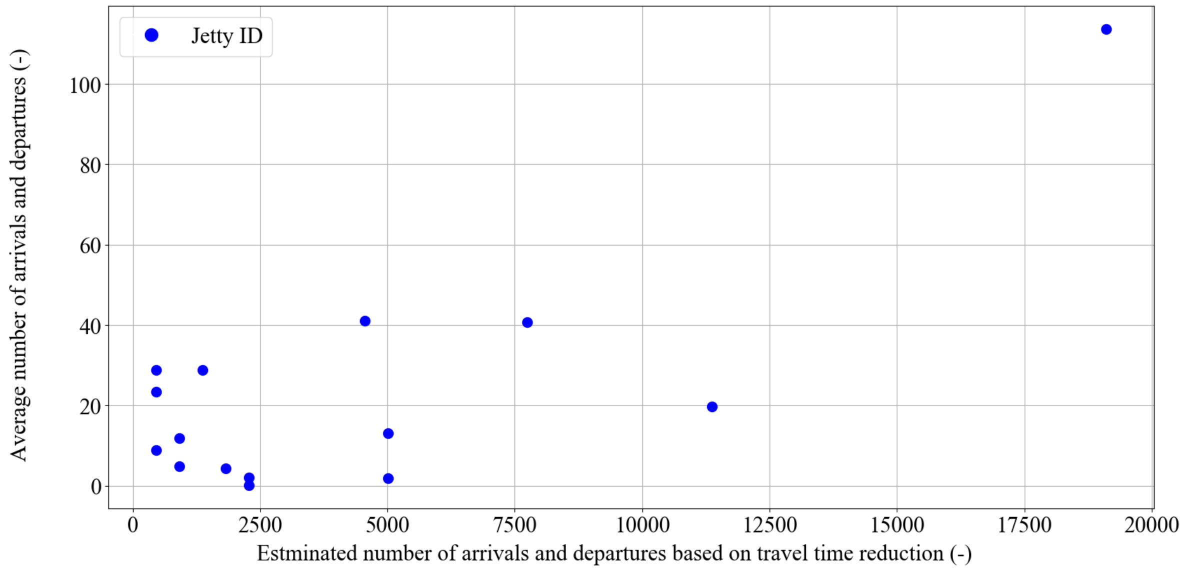


Figure 6.6: Daily number of people that arrive and leave from a jetty based on Watertaxi Rotterdam data in 2017 versus estimations on the number of people that could reduce their travel time with the watertaxi. The estmations are based on the experiment described in section 6.2.3. Every blue dot represents a single jetty.

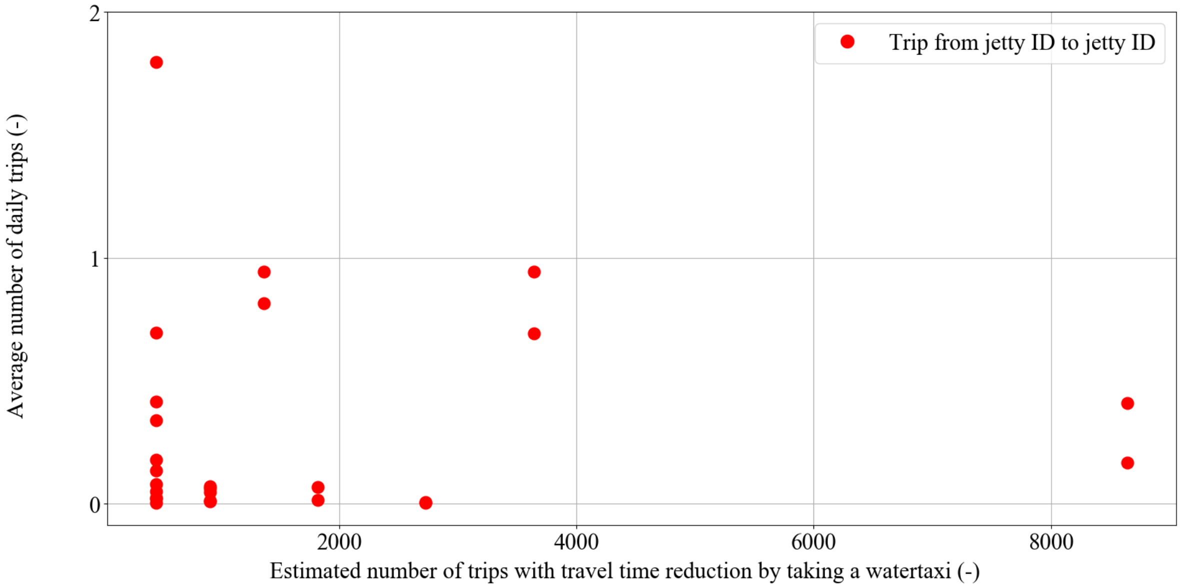


Figure 6.7: Daily number itineraries based on Watertaxi Rotterdam data in 2017 data versus estimations on the number of people that could reduce their travel time with the watertaxi. The estmations are based on the experiment described in section 6.2.3. Every red dot represents an itinerary between two jetties.

Table 6.3: calculation time in seconds of different for different number of city nodes. * indicates that test could not be performed due to a shortage in RAM.

Jetty nodes	City Nodes					
	6	**11**	**21**	**41**	**61**	**80**
75	8.5	35.6	197.7	823.9	*	*
40	2.1	4.17	42.7	162	431	803

6.3. Experiment with average calculation time

Different numbers of nodes are used. By performing several experiments with different numbers of nodes, estimations on runtime can be made for future usage. The number of city nodes is increased at every step. Not every flow variable has a influence on the outcome of the model. The objective function is a product of the number of registered itineraries, the flow between city nodes via jetties and the travel time of that route (in chapter 4, described by 4.1). When no itineraries occur, the variable of flow between city nodes do not have an influence on the objective outcome. To ensure every variable is used, the number of itineraries made between two point is set to '1'.

Table 6.3 shows for different numbers of city nodes the preparation time and the calculation time. In figure 6.8, the average calculation time for the different numbers of city nodes and jetty nodes is plotted. The average calculation time seems to rise exponentially with the number of city nodes. An explanation for this can be that the number of variables rises to the power of two as the number of city nodes doubles. The flow variable described by equation 4.5, has four dimensions. Two of the dimensions are linked to the number of city nodes. With the non linear relation, the calculation time for every step becomes significant. Due to memory limitations on the configuration, the solution to 61 and 80 nodes could not be found.

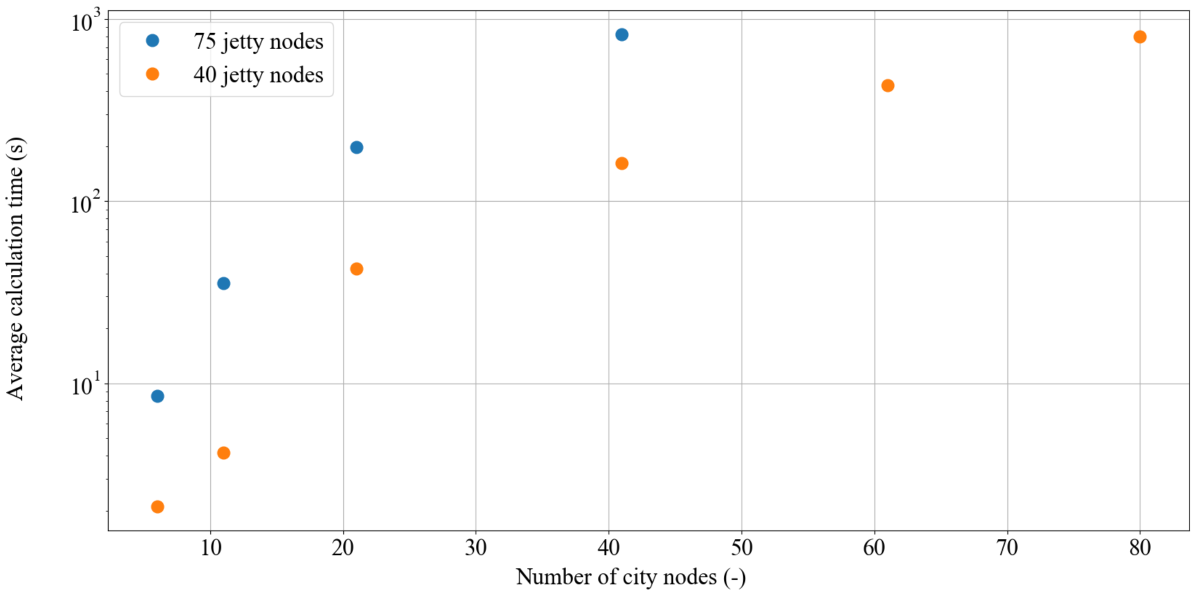


Figure 6.8: Average calculation time versus the number of nodes used.

Table 6.4: Requirements of method for the case study within Rotterdam. The generally applicability (6) is discussed in chapter 5.7.

Requirement	**solution**	**pass?**
1.Determine walking times	Usage of Google Maps API	pass
2.Determining locations considering public transport	Direct links between nodes	pass
3.Optimize for travel time	total travel time objective of optimization	pass
4.Capacity should be considered	Capacity constraint verified	pass
5.Calculation time within 5400 seconds	Calculation time is 3440s	pass
6.The method is generically applicable		

6.4. Conclusion

In this chapter, the question *what are the results of the implemented method for the city of Rotterdam?* is answered by executing a case study. A total of five experiments have been done. First, a general experiment with the designed method is executed. The designed method shows that a average reduction 517 seconds when 22 jetties are allocated. In second experiment, extra travellers are added. The results were compared with the results from the general experiment. In the second experiment, the average reduced travel time stopped increasing with a similar number of jetties. The third experiment consisted of using different capacity settings for the jetties. This experiment showed that the capacity constraint has a negative influence on the average reduced travel time. In the fourth experiment, the number of people which can reduce their travel time is compared with with itineraries from the Watertaxi Rotterdam in 2017. In this experiment, no correlation is found between two datasets. In the final experiment, the number of jetty and city nodes is varied to estimate the relation between the number of nodes and the calculation time. The calculation time seems to rise roughly exponentially with the number of city nodes.

In the introduction, section 1.1, requirements are set. In the table 6.4 the requirements presented. Requirement 1 to 5 are based on the requirements of the research goal. Requirement 6 is based on the main research question. The table shows all requirements are met, except requirement 6. Requirement 6 is discussed in the next chapter.

7
Reflection

In this research, the objective is to find a generic method to find the average reduced travel time. In this chapter, the implemented method is reflected on by answering the following question: *How generically applicable is the described method?* First, the requirements for further research are investigated. After that, the data sources used to find travel time within Rotterdam are evaluated and explored if usage in other urban areas is recommended.

7.1. Requirements for further research

For further usage of the developed model, other data is required. To ensure that the implemented method works with other data, requirements are set. Three tables are shown (table 7.2, 7.3 and 7.1). The first table shows the requirements for the jetty nodes, city nodes and travellers. As both data sources can not be chosen independently and depend on each other, the data is discussed together. The second table (table 7.2) shows the routing requirements. The third shows the format of the data structures needed to ensure the method works. As mentioned before, the method is implemented in python. The data has to be stored in python dictionaries in the format displayed in table 7.3. All recommendations are based on the requirements of the research, which includes the total of 1.5 hour calculation total time and the setup used by the author.

In table 7.1, item 2.2 gives advice on the number of nodes to use. This is an advice based on the calculation time. With another setup or no calculation time requirements, higher number of nodes are possible. In item 1.4 is stated that the predetermined of jetties must be equal or smaller than the number of jetty nodes. To increase the accuracy of the average reduced travel time, high numbers of nodes are preferable. There are no limitations on the number of travellers between the city nodes.

In table 7.2 the route requirements are shown. Item 4.1 states that accurate routing is needed from the regular street to the waterside. From experience, the map makers do not always draw routes to the water correctly, which can result in detours. In item 5.2 is stated that collecting data from multiple points in time would require lots of data, only one moment on the day is taken into account. To make the implementation of the method of this thesis work, between every pair of nodes, a travel time in seconds has to be found. If a connection does not exist, the travel time should be set to infinite (item 7.1). In 7.3 is advised to automatically generate routes. The number of possible itineraries scales exponentially with the number of nodes. A script that can automatically based on coordinates find a travel time is advised to find all travel options.

Table 7.3 shows the data format required. This is nessesary to let the data serve as an input for the implemented model. The structure is based on a python dictionary. The keys of the dictionary are the node IDs.

Table 7.1: Routing Requirements with general requirements and route type specific requirements.

Node and traveler dataset requirements	
1. Select jetty nodes	*2. General requirements for nodes*
1.1 Points on waterside that are reachable by a watertaxi boat.	2.1 Discrete points within area of interest
1.2 Minimum: 2 nodes.	2.2 Advised maximum of total number nodes is 94 nodes.
1.3 All nodes must be reachable by water.	2.3 Output format: table 'Data structure'
1.4 Number of nodes must by higher than number of chosen jetties.	2.4 Possible to reach all nodes by walking
3. Select traveler and city nodes	2.5 Coordinates and unique ID
3.1 Location of city nodes must be reachable by public transport.	
3.2 Minimum: 2 nodes.	
3.3 At least 1 traveler, no limitations.	

Table 7.2: Requirements for data on routes.

Routing requirements	
4. Walking routes	*7. General requirements for routes*
4.1 Accurate routing from regular street to waterside.	7.1 Between pair of nodes, a travel time must be determined.
5 Public transport	7.2 The travel needs to be an accurate representation of the real travel time.
5.1 Include walking into time estimation.	7.3 Automatically generate routes based on coordinates.
5.2 All data is gathered from a single date and point in time.	7.4 Output in format of table 'Data structure'.
6. Sailing routes	7.5 Duration in seconds between nodes.
6.1 Zones where no watertaxi is allowed are avoided.	
6.2 Nautical regulations are respected.	

7.2. Usability of data sources

To find out which of the sources used in the case study is useful in further research, the main data and its sources are discussed. The main sources used for the case study are the Google Maps API for the public transport and walking itineraries, OViN 2017 travellers survey for the traveller data and the Flying Fish route planner. In the three subsections, these data sources are discussed. Table 7.4 gives an overview of the different data sources. To evaluate the usefulness of the data soures, the advantages, disadvantages and the scale where the datasource can be used on are shown.

7.2.1. Flying Fish: routing algorithm and locations jetty nodes

When analysing the operations of Watertaxi Rotterdam, Flying Fish gathered data on the jetties in the Maas in Rotterdam. To estimate the sailing times between the jetties, arouting algorithm was developed. With this routing algorithm, travel times between all jetty locations was calculated. This travel data is not publicly available. This planning algorithm is designed on basis of sailing routes within Rotterdam and is therefore not directly usable in other urban areas.

The equation used to determine the sailing time, (equation 2.1), can been used in other urban areas. Also the constants to determine to make up for acceleration, getting on board and getting of board($t_{acceleration}$) and (t_{aboard}) can be used in other studies as a first estimation.

7.2.2. Routing within an urban area

For routing within an urban area, Google Maps static API is used. With a standard HTTP request data on routing can be gathered from Google. The possibilities include finding route for different ways of travelling within the city. The data that is sent back includes the route in coordinate points, time in seconds and distance in kilometres.

Table 7.3: Requirements for data format for six sources of data. * the routes are not required, but are useful to make plots.

Data structure	
Name	*Structure (python dictionary)*
Number of travellers	{CityNodeID:{CityNodeID:Number of Travellers}}
Locations of city nodes	{CityNodeID:{'long':longitude,'lat':latitude}
Locations of jetty nodes	{JettyNodeID:{'long':longitude,'lat':latitude}
Walking route	{CityNodeID:{JettyNodeID:{duration':int,'routing':list}}*
Public route	{CityNodeID:{CityNodeID: {'duration':int,'routing':list}}*
Sailing route	{JettyNodeID:{JettyNodeID:{'duration':int,'routing':list}}*

For this research Google Static Maps API is convenient to use. It is able to sent back all data needed to find all characteristics needed to find the necessary travel time, the length of the route and the coordinates to plot it. For the public transport, it is able to provide alternative routes. For research to other urban areas, Google Maps can be a easy method to gather data on travel times within Urban area. It has a high route accuracy, but some routes are plotted wrong. Walking routes can lack accuracy. Cases were found in this study with walking routes besides highways. Some direct routing to water was not found, which resulted in detours.

Another disadvantage can be that it is free of charge until a certain number of requests. The number of requests without charge depends on Google policy. On 22 september 2020 this number is 40000[1]. To do this research, around 10000 requests were used.

7.2.3. Travellers data

To make estimations on the routes that people take through Rotterdam and the locations they depart from and arrive at, OViN 2017 is used. In this research postal code areas are used to estimate the origins and destinations of the travellers. This method has two main disadvantages: lack of data and the placement of locations.

The Combination of the travellers survey and the dataset to covert postal codes works best in 'round' postal areas. In irregular postal code areas, the location where the postal code is plotted can be far off from the corners of the postal code. This could result in a inaccurate reduced travel time.The survey used is based on a small sample of the Dutch population within the Netherlands. Only a small sample size out of this research can be used. This means that within Rotterdam only 27 registered itineraries were taken into account. This means that 1 trip can have a major significance on the outcome. Therefore, when using another urban area a check has to be done on the number of measured itineraries. Rotterdam is one of the major cities in the Netherlands. If Rotterdam does not have sufficient datapoints, a fair assumption is that this dataset does not contain enough data points for other urban areas.

Table 7.4: Three sources of data distinguished on their source, scale of usefulness, advantages and disadvantages.

	Sail and jetty data	**routes within Urban area**	**Travellers data**
Source	Flying Fish internally	Google Static Maps API	OViN 2017
Scale of usefulness	Only Rotterdam	Worldwide	Netherlands
Advantages	Based on data of watertaxis	Easy to use worldwide	-
Disadvantages	Not useful for data outside Rotterdam	Free of charge until a certain number of queries	Just a few areas within the Netherlands have sufficient datapoints

7.3. Conclusion

The question *How generically applicable is the described method?* is answered by examining the requirements for future research and looking at the usability of the data sources in further research. The requirements are divided into three categories: node requirements, route requirements and format requirement. Within the three categories, specific requirements are given for each of the six data sources.

When looking at the usability of the data sources in future research, only Google Static Maps API is recommended to be used in a similar way. The OViN 2017 in combination with the dataset to convert postal codes to coordinates gave inaccurate results. The routing algorithm used for the sailing routes can not directly be implemented in case studies to other cities. Since the duration of sailing itineraries is validated, lessons can be learned from the implementation of this algorithm and the parameters used to calculate the duration.

[1]Retrieved from: https://cloud.google.com/maps-platform/pricing/sheet?hl=nl on: 22-09-2020

8

Conclusion and recommendations

The main question, *What is an generic method to find the relationship between the number of jetties and the passenger travel time* is answered in the conclusion section by dividing the main question into sub questions. After that, recommendations for future research are given.

8.1. Conclusion

The question *What are the characteristics of the watertaxi system?* is answered in the second chapter. In the analysis of the watertaxi and other relevant systems, assumptions have been made that allows the problem to become a complete undirected graph consisting of three types of legs. The nodes are divided in two types of nodes: city nodes and jetty nodes. The city nodes are the origins and destinations of the travellers. The summation of the values of the legs used for an individual to travel from origin to the destination is the travel time for each individual. Out of the set of jetty nodes a predetermined number of jetties has to be chosen. With selecting different predetermined numbers of jetties, a relation between passenger travel time and the number of jetties has been found. The number of passengers leaving and arriving at a jetty (the capacity of the jetty) should be limited to avoid congestion.

The question *Which related mathematical problem can be used to determine a relationship between the number of jetties and the passenger travel time?* is answered in the third chapter. To find a solution for the described graph in the analysis of the problem, a literature study has been done. HLPs are closely related to the problem description given. The Median p-HLP described by Campbell [4] is used as basic formulation. The formulation of the capacity constraints of Campbell [6] is used to extend the work of Campbell [4]. This combination of work is able to find a predetermined number of jetties out of one node type and does not allow direct transport between nodes. The second type of node and direct transport have to be added to meet the requirements of this research.

The question *How is the watertaxi system going to be modelled?* is answered in the fourth chapter. The design of Campbell [4] is further extended with direct transport between city nodes and the ability to add the second node type. The direct transport between city node is fixed by introducing an extra jetty node with special properties. The sum of the legs to this virtual jetty is equal to the travel time of the direct connection.

The question *Which suitable data sources are going to be used for a case study to the city of Rotterdam* is answered in the fifth chapter. 79 jetty nodes and 75 city nodes have been found, where 1899 registered itineraries are registered between. Assuming that this sample represents the number of people perfectly, the number of travellers through Rotterdam on a daily basis is 864045 people. With Google Static Maps API, walking distances and public transport times are found. With internal Flying Fish data, the sailing times are gathered. Since this results in a calculation time that exceeds the allowed maximum time, choices have been made on the nodes and itineraries that are taken into account. With filtering and preprocessing, a selection of 20 nodes and 39130 people has been made.

The question *What are the results of the implemented method for the city of Rotterdam?* is answered in the sixth chapter. To validate the travel time, the travel time with public transport and the travel time with the watertaxi are validated. The watertaxi is validated by comparing predicted travel times with historical data. The travel time by walking and taking public transport is validated by comparing the outcomes with 9292ov and Google Maps. The average reduced travel time converges to a value of 517 seconds, when using the described data sources. The average reduction in travel time increases with the number of jetties used. The method is used to compare the found number of people that travel between jetties with historical Watertaxi Rotterdam of itineraries in 2017. When comparing the number of people with historical Watertaxi Rotterdam data, no correlation has been found. The calculation time is important to estimate the scalability of the problem. With varying the number of used city nodes and jetty nodes, the average calculation time is approximated. With the gathered data, an exponential relation is estimated between the average calculation time and the number of used jetty and city nodes is predicted.

To answer the question, *How generically applicable is the described method?*, requirements for generic usage are set. Also, the data sources used are evaluated. The Google Static API seems to work for cities around the globe. The study used for the number of registered itineraries, OViN 2017, contains data from itineraries within the Netherlands. The study is conducted all around the Netherlands, which results a small number of data points in Rotterdam. This could mean that in other areas of interest, the number of data points could be too low for reliable results. The locations of jetties and the sailing times have to be determined in another manner in another case study.

8.2. Recommendations

With executing this research, opportunities for other research arose. The recommendations are divided into research to improve the method (section 8.2.1 to 8.2.3) and suggestions to use the method in other situations.

8.2.1. Calculation time

With roughly an exponential relation between the number of hubs and the runtime, clever choices have to be made on the number of hubs used. In this research, with preprocessing only itineraries are chosen that result in a reduction in travel time. To not compromise on the solution, no gap with the optimal solution was allowed in this research. This could be a cause of the high calculation times. When allowing a gap with the ideal solutions, the calculation time may be lower. In literature, (Weng et al. [33]) Hub Location Problems were solved with heuristics. With heuristics, a small deviation from the optimal solution is allowed, but can decrease the runtime significantly.

8.2.2. Traveller motivation

In this research only the average reduced travel time is taken into account. However, in the real world travellers motives do not only depend on the reduced travel time. Examples of other factors are cost, fun, or the number of times a person has to change from modality. With studying the motives of travellers, one can find out if a correlation exist between the reduced travel time and the number of travellers.

8.2.3. Watertaxi planning

In this research, assumed is that the planning of the watertaxi is ideal. To take the planning of the watertaxi into account has two main advantages. The accuracy of the average reduced travel time can be increased and more information can be gathered about the watertaxi. With taking the planning of the watertaxi into account, estimations can be made on the number of boats needed to serve all costumers. A method developed by Bakker [2] can be used to make estimations on the number of watertaxis that are necessary.

8.2.4. Repeat case study with other transport modes

The method is developed to easily change transport modes. More research can be done on the average reduced travel time by comparing other transport modalities. An example is to replacing the transport of walking by taking a bike. It is also possible to change public transport in another transport mode and compare it with the watertaxi. Repeating the case study by comparing the combination watertaxi and biking does not require changes in the method. Biking data can be required by using the Google Maps Static API and can replace walking, when it is converted to the correct format.

8.2.5. Use method in other logistics

The design is based on the median p-HLP. The adjusted HLP design allows direct transport between hubs, which is not allowed in median p-HLP. With the changes made in design, likewise situations as described in this research can be analysed. An example is the optimization of the locations where shared bikes are stored. With this design, the option of walking to a shared bike storage can be compared with public transport. A likewise situation occurs.

8.2.6. Use for other cities

Chapter 7 shows an overview of the requirements set, the method can be used in cities worldwide. With this method, a first step to the development of watertaxi systems all over the world is made.

Bibliography

[1] Turgut Aykin. Networking policies for hub-and-spoke systems with application to the air transportation system. *Transp. Sci.*, 29(3):201–221, 1995. ISSN 00411655. doi: 10.1287/trsc.29.3.201.

[2] Lenny Bakker. The impact of routing optimization on the service time of the watertaxi of Rotterdam. 2020.

[3] Natashia Boland, Mohan Krishnamoorthy, Andreas T. Ernst, and Jamie Ebery. Preprocessing and cutting for multiple allocation hub location problems. *Eur. J. Oper. Res.*, 155(3):638–653, 2004. ISSN 03772217. doi: 10.1016/S0377-2217(03)00072-9.

[4] J. F. Campbell. Hub location problems and the p-hub median problem. 1991.

[5] J. F. Campbell, A. T. Ernst, and M. Krishnamoorthy. Hub arc location problems: Part I - Introduction and results. *Manage. Sci.*, 51(10):1540–1555, 2005. ISSN 00251909. doi: 10.1287/mnsc.1050.0406.

[6] James F. Campbell. Integer programming formulations of discrete hub location problems. *Eur. J. Oper. Res.*, 72(2):387–405, 1994. ISSN 03772217. doi: 10.1016/0377-2217(94)90318-2.

[7] Selim Çetiner, Canan Sepil, and Haldun Süral. Hubbing and routing in postal delivery systems. *Ann. Oper. Res.*, 181(1):109–124, 2010. ISSN 02545330. doi: 10.1007/s10479-010-0705-2.

[8] Jan Dul and Tony Hak. *Case Study Methodology in Business Research.* Elsevier Ltd, 2008. ISBN 9780750681964.

[9] Jamie Ebery, Mohan Krishnamoorthy, Andreas Ernst, and Natashia Boland. Capacitated multiple allocation hub location problem: Formulations and algorithms. *Eur. J. Oper. Res.*, 120(3):614–631, 2000. ISSN 03772217. doi: 10.1016/S0377-2217(98)00395-6.

[10] Amir Ebrahimi and Zade Ahmad. A modified NSGA-II solution for a new multi-objective hub maximal covering problem under uncertain shipments. pages 185–197, 2014. doi: 10.1007/s40092-014-0076-4.

[11] Andreas T. Ernst and Mohan Krishnamoorthy. Efficient algorithms for the uncapacitated single allocations-Hub median problem. *Rech. - Transp. - Secur.*, 62(3):139–154, 1996. ISSN 07618980.

[12] Andreas T. Ernst and Mohan Krishnamoorthy. Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem. *Eur. J. Oper. Res.*, 104(1):100–112, 1998. ISSN 03772217. doi: 10.1016/S0377-2217(96)00340-2.

[13] Reza Zanjirani Farahani, Masoud Hekmatfar, Alireza Boloori Arabani, and Ehsan Nikbakhsh. Hub location problems: A review of models, classification, solution techniques, and applications, 2013. ISSN 03608352.

[14] Shahin Gelareh and David Pisinger. Fleet deployment, network design and hub location of liner shipping companies. *Transp. Res. Part E Logist. Transp. Rev.*, 47(6):947–964, 2011. ISSN 13665545. doi: 10.1016/j.tre.2011.03.002. URL `http://dx.doi.org/10.1016/j.tre.2011.03.002`.

[15] Young Ha Hwang and Young Hoon Lee. Uncapacitated single allocation p-hub maximal covering problem. *Comput. Ind. Eng.*, 2012. ISSN 03608352. doi: 10.1016/j.cie.2012.03.014.

[16] Bahar Y. Kara and Barbaros Ç Tansel. The latest arrival hub location problem. *Manage. Sci.*, 47(10): 1408–1420, 2001. ISSN 00251909. doi: 10.1287/mnsc.47.10.1408.10258.

[17] J. G. Klincewicz. Heuristics for the p-hub location problem. *Eur. J. Oper. Res.*, 1991. ISSN 03772217. doi: 10.1016/0377-2217(91)90090-I.

[18] Jozef Kratica. An electromagnetism-like metaheuristic for the uncapacitated multiple allocation p-hub median problem. *Comput. Ind. Eng.*, 66(4):1015–1024, 2013. ISSN 03608352. doi: 10.1016/j.cie.2013.08.014. URL `http://dx.doi.org/10.1016/j.cie.2013.08.014`.

[19] Alfredo Marín. Formulating and solving splittable capacitated multiple allocation hub location problems. *Comput. Oper. Res.*, 32(12):3093–3109, 2005. ISSN 03050548. doi: 10.1016/j.cor.2004.04.008.

[20] Vinícius R. Máximo, Mariá C.V. Nascimento, and André C.P.L.F. Carvalho. Intelligent-guided adaptive search for the maximum covering location problem. *Comput. Oper. Res.*, 2017. ISSN 03050548. doi: 10.1016/j.cor.2016.08.018.

[21] Merve Meraklı and Hande Yaman. A capacitated hub location problem under hose demand uncertainty. *Comput. Oper. Res.*, 88:58–70, 2017. ISSN 03050548. doi: 10.1016/j.cor.2017.06.011.

[22] Seyed Sina Mohri and Meisam Akbarzadeh. Incomplete Hub Location Model for Designing Van-Taxi Networks. *Transp. Res. Rec.*, 2672(8):619–628, 2018. ISSN 21694052. doi: 10.1177/0361198118783597.

[23] Morton E. O'kelly. A quadratic integer program for the location of interacting hub facilities. *Eur. J. Oper. Res.*, 46(3):409–411, 1987. ISSN 03772217. doi: 10.1016/0377-2217(90)90018-7.

[24] Çağrı Özgün-Kibiroğlu, Mehmet Nahit Serarslan, and Yusuf İlker Topcu. Particle Swarm Optimization for Uncapacitated Multiple Allocation Hub Location Problem under Congestion. *Expert Syst. Appl.*, 119: 1–19, 2019. ISSN 09574174. doi: 10.1016/j.eswa.2018.10.019.

[25] Meltem Peker and Bahar Y. Kara. The P-Hub maximal covering problem and extensions for gradual decay functions. *Omega (United Kingdom)*, 2015. ISSN 03050483. doi: 10.1016/j.omega.2015.01.009.

[26] L. H. Phong. The relationship between rivers and cities: influences of urbanization on the riverine zones – a case study of Red River zones in Hanoi, Vietnam. *Sustain. Dev. Plan. VII*, 1:27–43, 2015. ISSN 1746-448X. doi: 10.2495/sdp150031.

[27] V. Rodríguez, M. J. Alvarez, and L. Barcos. Hub location under capacity constraints. *Transp. Res. Part E Logist. Transp. Rev.*, 43(5):495–505, 2007. ISSN 13665545. doi: 10.1016/j.tre.2006.01.005.

[28] Marcos Roberto Silva and Claudio B. Cunha. A tabu search heuristic for the uncapacitated single allocation p-hub maximal covering problem. *Eur. J. Oper. Res.*, 2017. ISSN 03772217. doi: 10.1016/j.ejor.2017.03.066.

[29] Darko Skorin-Kapov, Jadranka Skorin-Kapov, and Morton O'Kelly. Tight linear programming relaxations of uncapacitated p-hub median problems. *Eur. J. Oper. Res.*, 94(3):582–593, 1996. ISSN 03772217. doi: 10.1016/0377-2217(95)00100-X.

[30] Hans P.M. Veeke, Jaap A. Ottjes, and Gabriel Lodewijks. *The Delft Systems Approach*. 2008. ISBN 9788578110796. doi: 10.1017/CBO9781107415324.004.

[31] Peter Vos. Optimizing the use of Autonomous Surface Vehicles in the Port of Rotterdam. 2018.

[32] John Wang. *Encyclopedia of Business Analytics and Optimizationitle*. 2014. ISBN 978-1-4666-5203-3.

[33] Ke Rui Weng, Chao Yang, and Yun Feng Ma. Two artificial intelligence heuristics in solving multiple allocation hub maximal covering problem. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 4113 LNCS:737–744, 2006. ISSN 16113349. doi: 10.1007/11816157_90.

[34] T G Wijnsma. Designing Express Networks with Multi-Agent Modelling. (December), 2011.

A

Article

Reducing travel time within urban areas with watertaxi networks

Thijs van Berkel, Mark Duinkerken, Gijsbert van Marrewijk, Rudy Negenborn

Department of Maritime and Transport Technology, Delft University of Technology, The Netherlands

Abstract

A watertaxi can reduce the travel time through the city of Rotterdam in comparison with public transport. To make estimations on the average reduced travel time within Rotterdam, a generic method is developed. This method is capable of finding locations for watertaxi jetties. To achieve this, the watertaxi, public transport and the walking from and to the jetties is modelled as a complete undirected graph. The Median p- Hub Location Problem described by Campbell [1] is changed by adding a second type of node and allow direct links. With a case study to Rotterdam, the method shows to be useful to be used find the average reduced travel time in urban areas.

Keywords: , Watertaxi, Median p-HLP with direct links, Graph, Public transport, Optimization

1. Introduction

From a historical perspective, cities have been build often around bodies of water [2]. These bodies of water can bring opportunities for transport of people. One way to use water as a transport mode is the usage of watertaxis, which is a taxi service that can be ran on the water. Watertaxis work in a similar way as regular taxi services. A boat can be ordered at one of the jetties, locations where watertaxis can moor. The itinerary begins at the agreed time and ends at the agreed arrival jetty. A watertaxi can provide a reduction in travel time. In Rotterdam, watertaxis are used to transport people over the Nieuwe Maas. In Rotterdam taking a watertaxi can result in a reduction in travel time.

The objective of this research is to develop a method that can be used to determine the locations of the jetties by minimizing the average travel time. When developing this method, the following requirements have been considered:

1. A method that can be used to determine the walking times to and from jetties for travellers.
2. By determining the locations of the jetties, the travel time with public transport should be considered.
3. Only passengers where watertaxi transport reduces their travel time should be taken into account.
4. The method should be able to limit the capacity to ensure not too many people gets assigned to a single jetty.
5. The time to calculate the outcome of this research should not exceed 1.5 hours.

1.1. Scope

The following assumption are made in this research:

1. Travellers take the fastest travel option and do not take cost into account.
2. Passengers are modelled as origin and destination points.
3. To determine travel time within an urban area, the time of day and rush hours are not considered.
4. To determine the locations of a jetties, no cost of allocation is taken into account.

In the Analysis section, the modelling assumption are described. In literature, research to Hub Location Problems is elaborated on and one of the described models is chosen as basic model for this research. In the design section, the basic chosen model is further extended and Key Performance Indicators are defined. The data collection section describes the data used in a case study to the city of Rotterdam. The case study chapter, experiments with the gathered data are presented. This paper ends with a conclusion and recommendations for further research.

2. Analysis

The travel time through an urban area depends on a number of factors. When looking at an urban area, not every location at the water side is available for a jetty. Therefore, the jetties are a finite (discrete) set, where a predetermined number of jetties is chosen from. This research keeps the watertaxi planning out of scope. Assumed is that the watertaxi planning is ideal: the duration between two points is time invariant. To simplify the travel time within the city, assumed is that walking and public transport are time invariant and people take the route that takes the least amount of time.

With all assumptions in place, the system can be modelled as an undirected complete graph, with three types of link between the different types of nodes. Origins and destinations of people are city nodes and have a travel time with public transport. The original set of nodes containing all possible jetties are jetty nodes and have a travel time determined by the sailing time.

Between city nodes and jetty nodes, the travel time is determined by the walking time. A selected itinerary between city nodes is known and is time invariant. From the city nodes the predetermined number of nodes is chosen which results in the lowest total travel time. Due to crowd regulations, the number of people that arrive and leave at a allocated jetty should be limited.

3. Literature

Hub Location Problems (HLP) fulfils the requirements of the described problem. Three directions of the Hub Location Problem are investigated with the criteria found by Farahani [3]: Median p-HLP, Multiple Allocation p-Hub Maximal Covering Location Problem and Median p-HLP with capacity Constraints.

3.1. Median p-HLP

The median p-HLP has applications to transportation and telecommunication networks. Often, the objective is to minimize the total cost of movement [4]. The Median p-HLP is first proposed by Campbell [1]. Since then, minor changes have been made to this model. Skorin-Kapov [5] uses a formulation which uses fewer constraints. Ernst [6] found a three index variables to model the uncapacitated problem. Boland [7] uses a slightly different formulation that is used in more recent literature. In Kratica [8] an electromagnetism-like (EM) method is proposed for solving this NP-hard problem.

3.2. Multiple Allocation p-Hub Maximal Covering Location Problem

The p-hub maximal covering location problem (MApHCP) maximizes the demand that is covered with a predetermined number of hubs. The MApHCP was first introduced by Campbell [4]. After that Kara [9] improved the formulation. Maximo [10] and Silva and Cumba [11] explored solution methods. The MApHCP formulation of [9] allows direct links between the non hubs. However, the formulation of the objective function is not usable for the purpose of this research, which makes the formulation not usable.

3.3. Median p-HLP with Capacity constraints

Several studies are done on limiting the capacity of hubs. Campbell [4] describes different types of integer programming models and lays the foundation for the multiple allocation HLP with capacity constraints. Ebery [12] uses the formulation of Campbell and updates it with increased compact formulation. In the model of Ebery [12], a capacity restriction is applied on the volume of traffic entering a hub via collection. Rodriguez [13] based their study on the model proposed by Aykin [14]. Rodriguez [13] uses two separate hub capacities. Merakl [15] hubs in an N-node network with capacity and distance limitations when the service standard offered needs to be fulfilled. The models of [4] and [15] use a variable number of chosen hubs. The models of [12] and [16] do use a predetermined number of hubs.

3.4. Conclusion on model

As can be concluded, no model has all requirements needed. None of the models make use of two types of nodes. Besides that, none of the models use time as their main objective. Therefore choices have to be made on which model is the closest related to this research. To do so, a first a basic model is chosen. Extra constraints have to be designed to ensure all requirements are met.

In the literature found on this topic, the objective function is normally based on cost. The costs are build up by the usage of links. In this research, the amount of time is the value on the links. This means that the objective of minimizing cost and minimizing time are similar. The objective function found in the MA p-HCP differs on multiple points from the objective function required and is therefore not suitable as basic model.

This leaves papers based on Median p-HLP with and without capacity constraints as options. The median p-HLP with capacity constraints by Ebery [12] and Rodriguez [13] are the closest related to this research but have more complex designs that not easily adapted.

For this research the basic model by Campbell 1991 [1] is used. This basic model is combined with the capacity constraint in Campbell 1994 [4]. The model of [1] is widely used and has a basic architecture, which makes it suitable for adjustments. The work of Campbell 1994 [4] does not predefine the number of hubs that has to be chosen. The formulation of the model by Campbell 1994 [4] is closely related to the work in Campbell 1991 [1], which makes it possible to fit the capacity constraint of Campbell 1994 [4] into the work of Campbell 1991 [1].

4. Design

The design consists of an undirected network $N = (V, A)$ with $V = \{v_1, v_2, ..., v_q\}$ set of nodes. Every link is in $(a, b) \in A$. No links have a negative weight $d(a, b) = d(b, a)$. The travel time from origin i to destination j via jetty node k and m (T_{ij}^{km}) is the sum of walking from i to k (Wa_i^k), sailing between k and m (Sa^{km}) and walking from m to destination j (Wa_j^m). The mathematical formulation for travel time T is $T_{ij}^{km} = Wa_i^k + Sa^{km} + Wa_j^m$. Set V is split into set C and J. The set of all jetty nodes is $J = \{1, .., m\}$, containing all jetty nodes. The set $C = \{1, .., n\}$ contains all possible origins and destinations of travellers, the city nodes. If $k = m$ no inter jetty transport takes place [1]. The following formulations was first proposed by [1] and adjusted in this research to fit the requirements:

4.1. Parameters

T_{ij}^{km}	Travel time from origin i to destination j via jetty node k, m
W_{ij}	Number of passengers travelling from origin i to j
H	Predefined number of hubs

4.2. Variables

The variables used in this research:

Y^k is 1 if jetty node k is allocated as jetty, 0 if k if it is not
X_{ij}^{km} Fraction of travellers that travel from i to j via k and m.

4.3. Objective and constraints

The objective function 1 the total travel time is minimized for all passengers by multiplying the number passengers travelling from i to j (W_{ij}) by the travel time T_{ij}^{km}. X_{ij}^{km} is the flow from jetty node i to j via city node k and m and has a value between 0 and 1.

$$\text{Minimize} \sum_i \sum_j \sum_k \sum_m W_{ij} X_{ij}^{km} T_{ij}^{km} \tag{1}$$

Variable Y^k is introduced in equation 2. Y^k is set to 1 if Y^k is a hub and 0 if not. Equation 3 ensures that the number of hubs is sets to H.

$$Y^k \in \{0,1\} \forall k \in J \tag{2}$$

$$\sum_k Y_k = H \tag{3}$$

The value of X_{ij}^{km} is the flow between city node i to j via jetty node k and m. Equation 4 ensure that the flow has a value is greater than or equal to 0. Equation 5 and 6 stipulate that the flow via hub k and m is only possible if hubs k and m are chosen. Equation 7 ensures that the total flow from i to j is equal to 1.

$$X_{km}^{ij} \geq 0 \forall i,j,k,m; i,j \in C; k,m \in J \tag{4}$$

$$X_{km}^{ij} \leq X_m \forall i,j,k,m; i,j \in C; k,m \in J \tag{5}$$

$$X_{km}^{ij} \leq X_k \forall i,j,k,m; i,j \in C; k,m \in J \tag{6}$$

$$\sum_k \sum_m X_{km}^{ij} = 1 \forall i,j; i,j \in C \tag{7}$$

4.4. Capacity constraint

Equation 8 shows the equation that adds capacity constraints to the total flow (Campbell [4]). A new parameter is introduced, Ca^k. This is the capacity of jetty node k. In equation 8 the total flow through a hub is limited by capacity Ca^K. The capacity is determined by the flow through the hub times the number of people travelling through the hub.

Ca^k Capacity of jetty node at k

$$\sum_j \sum_i W_{ij} * (\sum_m (X_{ij}^{km} + X_{ij}^{mk}) - X_{ij}^{kk}) \leq Ca^k * X_k \tag{8}$$

4.5. Public transport

To integrate the possibility to travel with public transport, one virtual jetty node is added to the system. The total travel time along this route is the duration of a watertaxi itinerary. To integrate public transport, three constraints are added. Jetty node 0 is chosen to represent public transport. Y^0 is be set to 1. The total of Y^k has to sum up to the total number of jetties that have to be chosen.

$$\sum Y^k = H + 1 \tag{9}$$

$$Y^0 = 1 \tag{10}$$

4.6. Key Performance Indicators for case study

Three of the Key Performance Indicators (KPIs) of this research are based on the performance indicators in literature. The Average reduced travel time (s) KPI, shows average reduction of travel time for travellers for a number of placed jetties. For this KPI, the outcome of the objective function is needed. The percentage of people taking the watertaxi (%) is used to find out which percentage of the people takes the watertaxi instead of public transport. The flow through hub 0, the Public Transport hub, is used to determine the percentage of people that travel the watertaxi by assuming that all travellers that do not use public transport, takes the watertaxi.

The third used KPI is the average calculation time (s). One of the requirement of this research states that the calculation time of the model has to stay below 1.5 hours. To find a relationship between the number of placed jetties and the average reduced travel time, different numbers of placed jetties are tried. For every configuration, the implementation starts with loading all data, calculating the optimal solution and at last print all outputs. The time it takes to calculate all configurations is defined as the total calculation time. The total calculation time divided by the number of different configurations tried is the average calculation time.

5. Data collection

For the case study 6 datasets are needed: city nodes and jetty nodes, three types of routes and data on the travellers between the city nodes.

The city nodes and travellers are based on the Dutch 'Onderzoek Verplaatsingen in nederland' (OViN, roughly translated to: Research to itineraries within the Netherlands). The OViN 2017 is conducted by the Dutch government. OViN conducts to the travel behaviour of Dutch Citizens. All respondents is asked to monitor one specific day of the year where they travel to, with the mode of transport, with their reason of travelling and their travel time [1]. The OViN 2017 consist out roughly 100.000 monitored itineraries. Origins and destinations are provided on postal code level. To analyse the itineraries within Rotterdam,

[1]Retrieved from: https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:103498/tab/1/rd/1 on 14th of April

the itineraries within Rotterdam should be extracted from the dataset. To this research, 37016 people responded, which is 0.22 percent of the Dutch population. Assuming that the sample represents Dutch citizens correctly, does every registered itinerary count for 455 itineraries.[2]. With a dataset [3] the postal areas are translated into coordinates. A sample of 5 random postal coordinates has been validated with Google Maps. The city nodes within Rotterdam are shown in figure 1.

The locations of the jetty nodes are based on the locations of jetties in Rotterdam. Within Rotterdam, a selection of jetties is used by the company 'Watertaxi Rotterdam'. With a planning algorithm provided by Flying Fish, estimations have been made on the sailing time between jetty nodes. The jetty nodes within Rotterdam are shown in figure 1. A sample of the sailing times has been validated by comparing the estimated sailing times with historical sailing times by Watertaxi Rotterdam.

The Google Static Maps API is used to determine the travel time between the city nodes with public transport and the walking time between the city nodes and the jetty nodes. The travel time with public transport validated by comparing the travel time found with the Google Static Maps API with the travel time of 9292ov, a commonly used public transport planner. Five routes have been compared. The difference in travel time between the two sources is on average 3 minutes, with a maximum value of 5 minutes.

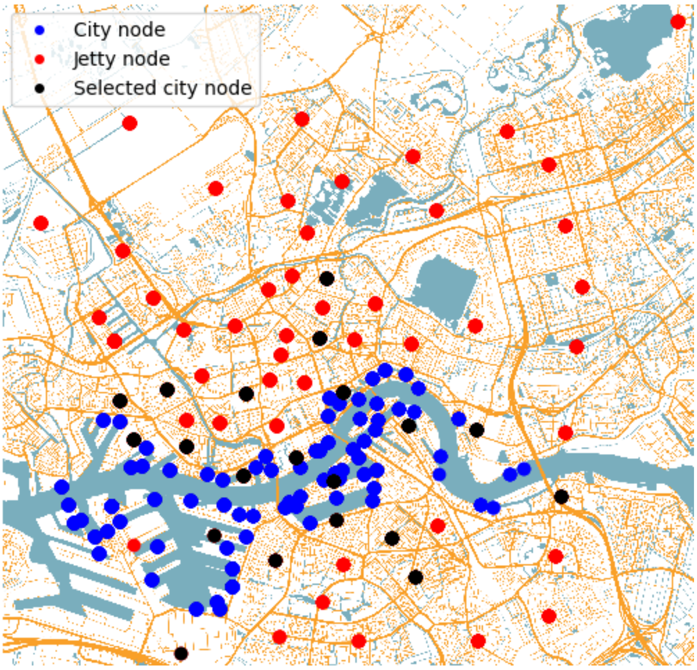


Figure 1: The locations of the city nodes, the jetty nodes and the selected jetty nodes.

To decrease the calculation time, the number of city nodes and registered itineraries is reduced. First, all itineraries are deleted where the travel time with public transport is in all cases lower than the combination of walking and taking a watertaxi. After that, all nodes are deleted where no travellers are arrive or depart from. The number of city nodes is reduced from 75 to 20.

6. Case study

To evaluate the performance of the design, experiments have been conducted. The experiments have been run on a Zbook G5 studio with 16GB DDR4 RAM, Windows 10 Home, Gurobi, Python MIP 1.7.2 and Gurobi Optimizer 9.0.2. The following experiments have been conducted:

1. General outcome on average reduced travel time and percentage watertaxi
2. Capacity limitations
3. Sensitivity on number of travellers
4. Number of nodes versus calculation time

6.1. General outcome

Results in figure 2 show the average reduction of travel time versus the number of jetties used. The percentage of people that take the watertaxi is plotted blue. From 20 allocated jetties, extra added jetties do not further reduce the average travel time. Between 13 allocated jetties and 20 allocated jetties, the average reduced travel time slowly increases. From 12 allocated jetties, the percentage of people taking public transport is starting to decrease to the value of 7 percent when taking only 2 allocated jetties into account. When looking at the change from 4 to 5 allocated jetties, the result is remarkable. The percentage of people taking the watertaxi decreases. This result can be explained by that another configuration of allocated jetties is chosen that increases the average reduction in travel time.

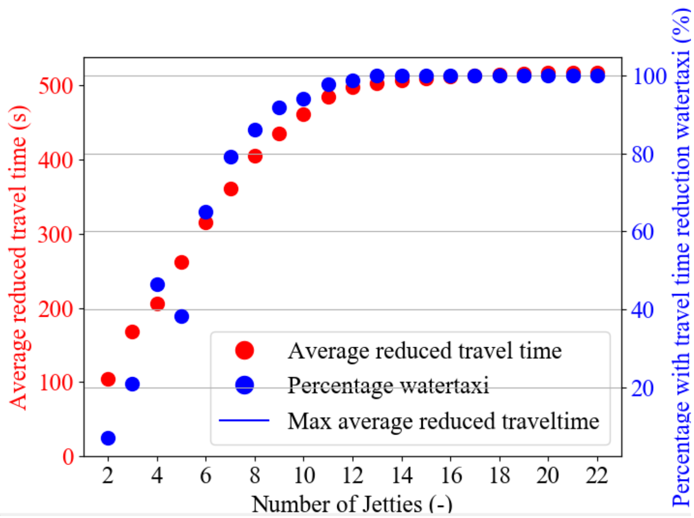


Figure 2: The predetermined number of jetties versus the average reduced travel time and the percentage of people that take the watertaxi.

[2]Retrieved from: https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:103498/tab/1/rd/1 on 14th of April

[3]Retrieved from:https://github.com/bobdenotter/4pp

6.2. Capacity

In current time with the Corona virus, governments look for methods to decrease crowd sizes. The method is capable of limiting the number of people that arrive and leave from a jetty node. On every jetty node, the daily capacity is set. Four scenarios are compared to the experiment without capacity limitations. The maximum allowed people on daily basis on a jetty node is set to 910, 455, 299 and 114 people respectively.

Figure 3 shows the results of the four scenarios in comparison with the general experiment. The average reduced travel time steadily increases and the percentage that uses public transport decreases with the number of allocated jetties. With a limit of 910 people, the average reduced travel time is with 28 allocated jetties is 280 seconds. The graph of 227 people and 114 people does not exceed the 100 seconds.

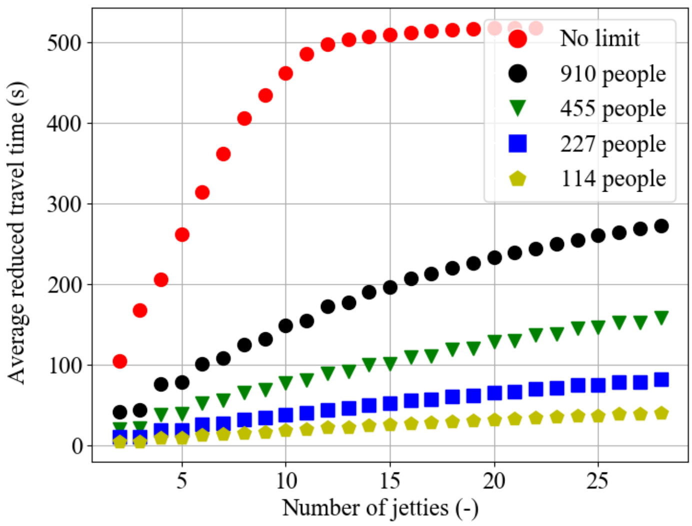


Figure 3: Experiments with a capacity constraint of 114, 227, 455 and 910 people compared with the general experiment without a capacity constraint.

6.3. Added itineraries

In this experiment the number of registered itineraries between the city nodes is changed. In the preprocessed dataset, 27 out of the 190 possible different itineraries are used. In this experiment, all of the 190 possibilities is used. Those itineraries are added to the original set of origins and destinations. Four scenarios are investigated and compared with the situation described in the general experiment. Conversion rate between number of itineraries 455 times is a factor of 0, 0.4, 0.6, 0.8 and 1. This means that a total of 0, 182, 273, 364 and 455 are added to every possible itinerary. Figure 4 shows the result of this experiment compared with the outcome of the general experiment.

The stop in increase in average travel time reduction at roughly the same number of allocated jetties could be explained by the difference travel speed. Travelling on land is according to the parameters used significant slower than sailing. This could mean that the allocated jetty that is closest to the city node results in a bigger reduction in travel time than a overall shorter route with a shorter path on water. With adding more people to the link, the graph seems to reach a asymptomatic limit. This could be caused by the fact that the influence of the original data decreases in comparison with the added itineraries. With adding even more people to the model, the number of people average over all itineraries becomes evenly divided.

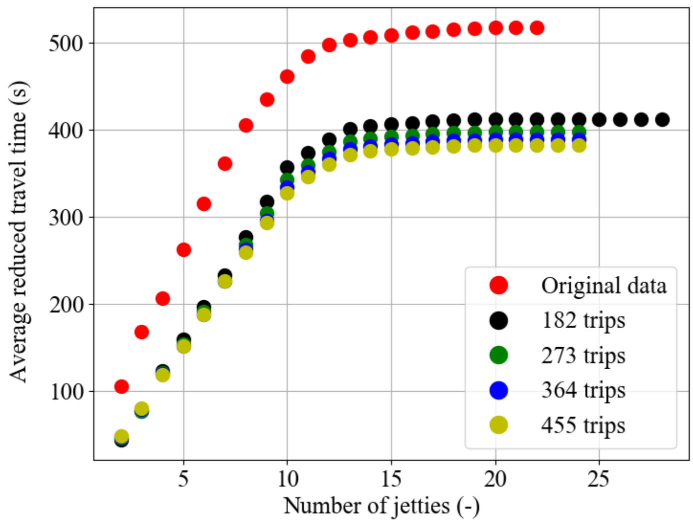


Figure 4: Experiments with 182, 273, 364 and 455 travellers added to every possible itinerary. The data is compared with general experiment.

6.4. Number of nodes

To make estimations on the average calculation time in future case studies, the number of city and jetty nodes used is increased at every step. Not every flow variable has a influence on the outcome of the model. The objective function is a product of the number of registered itineraries, the flow between city nodes via jetty nodes and the travel time of that route. When no itineraries occur, the variable of flow between city nodes do not have an influence on the objective outcome. To ensure every flow variable is used, the number of itineraries made between city nodes is set to 1.

Table 1 shows the average calculation time for different numbers of city and jetty nodes. The calculation time seems to rise exponentially with the number of city nodes. An explanation for this can be that the number of variables rises to the power of two as the number of city nodes doubles. The flow variable described by equation 5, has four dimensions. Two of the dimensions are linked to the number of city nodes. With the non linear relation, the calculation time for every step becomes significant. Due to memory limitations on the configuration, the solution to 61 and 80 nodes could not be found.

7. Reflection

With roughly an exponential relation between the number of hubs and the runtime, clever choices have to be made on the number of hubs used. Other research can be done in urban areas to find the average reduced travel time. An example is

Table 1: calculation time in seconds of different for different number of city nodes. * indicates that test could not be performed due to a shortage in RAM memory

Jetty nodes	City Nodes					
	6	**11**	**21**	**41**	**61**	**80**
75	8.5	35.6	197.7	823.9	*	*
40	2.1	4.17	42.7	162	431	803

to change duration of walk time in to the time to ride a bike. It is also possible to change public transport in another type of transport and compare it with the watertaxi.

The Google Maps API can be used to gather data on the duration of walking and public transport on different places in the world. When using the API for this purpose, first the availability of public transport has to be checked. Also, the possibilities on walking to the waterside has to be checked.

8. Conclusion and recommendations

The mathematical design in this research is an extension on the median p-HLP with capacity constraints and a predetermined number of hubs. The main drawback is the separation in two types of nodes, which does not allow a hub location on every node. The separation of nodes, however, can reduce the calculation time of the Hub Location Problems. The direct links can bring other opportunities for Hub Location Problems. The direct links give possibilities to select the demand that has a positive influence on the outcome of the objective function of the model.

To assume that people only travel on time as objective is a assumption that simplifies the model. However, in the real world travellers motives are also heavily dependent things such as cost and fun, or the number of times a person has to change from modality. Therefore, a study has to be done to the motives travellers have to travel with the watertaxi. This requires another type of research. Placing jetties depends numerous factors, such as cost.

With roughly an exponential relation between the number of hubs and the runtime, clever choices have to be made on the number of hubs used. In this research, with preprocessing only itineraries are chosen that result in a reduction in travel time. To not compromise on the solution, no gap with the optimal solution was allowed in this research. This could be a cause of the significant calculation times. In literature, Hub Location Problems were solved with heuristics. With heuristics, a small deviation from the optimal solution is allowed, but can decrease the runtime significantly.

Other research can be done in urban areas to find the average reduced travel time. An example is to change duration of walk time in to the time to ride a bike. It is also possible to change public transport in another type of transport and compare it with the watertaxi.

The direct links give possibilities to select the demand that has a positive influence on the outcome of the objective function of the model. As the literature study of this research shows, the median p-HLP is used in cases where cost is the leading objective.

This method gives insights in the average reduced travel time in urban areas. With the requirements found in the reflection chapter, the method can be used in cities worldwide and can help the research to further implementation of watertaxis all over the world.

References

[1] J. F. Campbell, Hub location problems and the p-hub median problem (1991).

[2] L. H. Phong, The relationship between rivers and cities: influences of urbanization on the riverine zones – a case study of Red River zones in Hanoi, Vietnam, Sustain. Dev. Plan. VII 1 (2015) 27–43. doi:10.2495/sdp150031.

[3] R. Z. Farahani, M. Hekmatfar, A. B. Arabani, E. Nikbakhsh, Hub location problems: A review of models, classification, solution techniques, and applications (2013). doi:10.1016/j.cie.2013.01.012.

[4] J. F. Campbell, Integer programming formulations of discrete hub location problems, Eur. J. Oper. Res. 72 (2) (1994) 387–405. doi:10.1016/0377-2217(94)90318-2.

[5] D. Skorin-Kapov, J. Skorin-Kapov, M. O'Kelly, Tight linear programming relaxations of uncapacitated p-hub median problems, Eur. J. Oper. Res. 94 (3) (1996) 582–593. doi:10.1016/0377-2217(95)00100-X.

[6] A. T. Ernst, M. Krishnamoorthy, Exact and heuristic algorithms for the uncapacitated multiple allocation p-hub median problem, Eur. J. Oper. Res. 104 (1) (1998) 100–112. doi:10.1016/S0377-2217(96)00340-2.

[7] N. Boland, M. Krishnamoorthy, A. T. Ernst, J. Ebery, Preprocessing and cutting for multiple allocation hub location problems, Eur. J. Oper. Res. 155 (3) (2004) 638–653. doi:10.1016/S0377-2217(03)00072-9.

[8] J. Kratica, An electromagnetism-like metaheuristic for the uncapacitated multiple allocation p-hub median problem, Comput. Ind. Eng. 66 (4) (2013) 1015–1024. doi:10.1016/j.cie.2013.08.014. URL `http://dx.doi.org/10.1016/j.cie.2013.08.014`

[9] B. Y. Kara, B. Ç. Tansel, The latest arrival hub location problem, Manage. Sci. 47 (10) (2001) 1408–1420. doi:10.1287/mnsc.47.10.1408.10258.

[10] V. R. Máximo, M. C. Nascimento, A. C. Carvalho, Intelligent-guided adaptive search for the maximum covering location problem, Comput. Oper. Res. (2017). doi:10.1016/j.cor.2016.08.018.

[11] M. R. Silva, C. B. Cunha, A tabu search heuristic for the uncapacitated single allocation p-hub maximal covering problem, Eur. J. Oper. Res. (2017). doi:10.1016/j.ejor.2017.03.066.

[12] J. Ebery, M. Krishnamoorthy, A. Ernst, N. Boland, Capacitated multiple allocation hub location problem: Formulations and algorithms, Eur. J. Oper. Res. 120 (3) (2000) 614–631. doi:10.1016/S0377-2217(98)00395-6.

[13] V. Rodríguez, M. J. Alvarez, L. Barcos, Hub location under capacity constraints, Transp. Res. Part E Logist. Transp. Rev. 43 (5) (2007) 495–505. doi:10.1016/j.tre.2006.01.005.

[14] T. Aykin, Networking policies for hub-and-spoke systems with application to the air transportation system, Transp. Sci. 29 (3) (1995) 201–221. doi:10.1287/trsc.29.3.201.

[15] M. Meraklı, H. Yaman, A capacitated hub location problem under hose demand uncertainty, Comput. Oper. Res. 88 (2017) 58–70. doi:10.1016/j.cor.2017.06.011.

[16] I. Rodríguez-Martín, J. J. Salazar-González, Solving a capacitated hub location problem, Eur. J. Oper. Res. 184 (2) (2008) 468–479. doi:10.1016/j.ejor.2006.11.026.

B

Hand calculated cases

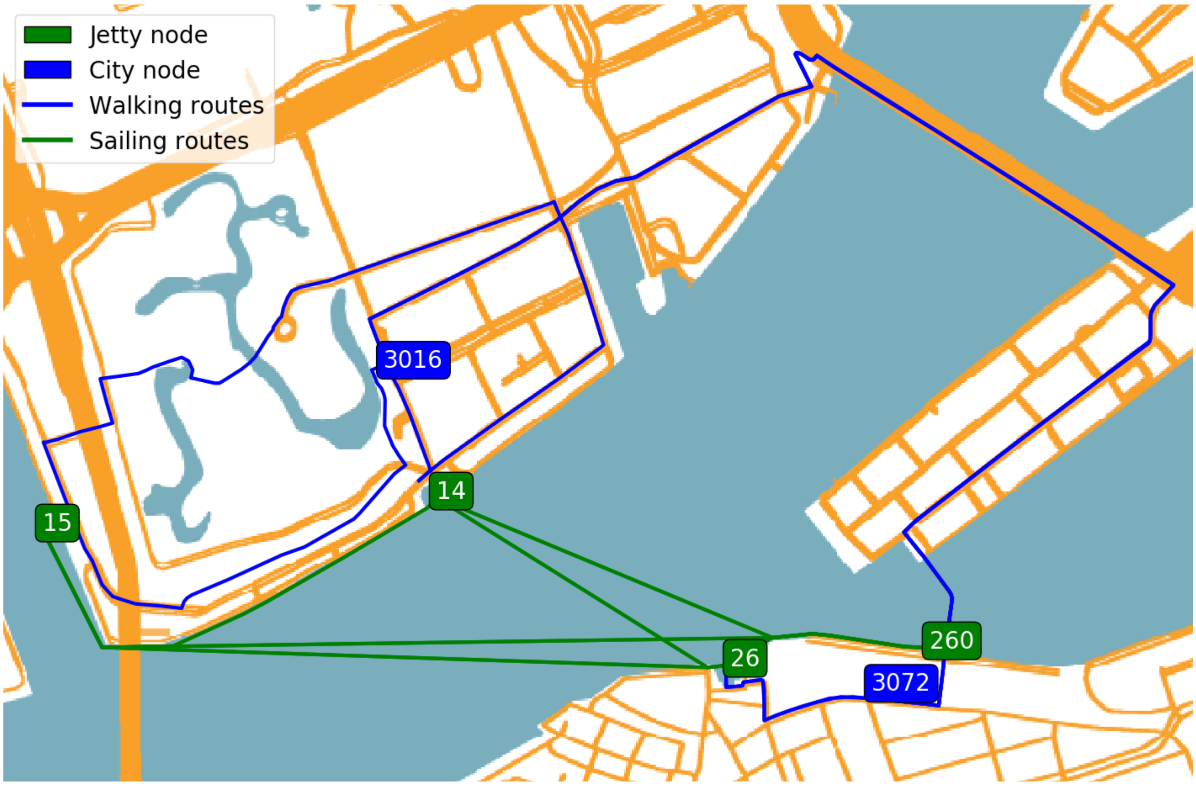


Figure B.1: Visual representation postal codes 3072, 3016 locations of jetties 15,14 ,26 and 260. All sailing routes have been drawn

To verify the inner working, a simple example is taken. The solution is possible to check by hand. Figure B.1 shows 2 points where people travel between: a point in postal code 3016 and a point in postal code 3072. Out of the selection of the jetties 15, 14, 26 and 260, 2 jetties have to be chosen. Travel times between origin, destination and between jetties are shown in B.1.

In the coming examples the following datasets are going to be used. First the travel time between the different jetties. The travel times in seconds are shown in table B.1. The travel time in seconds between the city node and the jetty node is shown in table B.1. Between the postal areas 8 people travel: 4 in every direction.

B.1. Case 1

In the first case no restrictions are set. The number of allowed jetties is set to p = 2. With 8 people travelling between two points: (202+120+180)*8 = 4016. Therefore the selected jetties are 15 and 26.

Table B.1: travel times between jetties 260,26,15,14 and to postal areas 3072 and 3016

	260	26	15	14	3072 (51.9018612,4.4842992)	3016 (51.9062395,4.4736472)
260	0	136	239	171	180	1800
26	136	0	202	136	180	2280
15	239	202	0	169	1760	120
14	171	136	169	0	1920	420

B.2. Case 2

In case two the capacity constraint is going to be verified. As shown in case 2, without capacity constraints jetty 15 and 26 are preferred. To check the capacity limitation, the capacity is going to be set to 0 for jetty 15 and 26. The other capacities are shown in table B.2. The total objective time should be therefore (171+180+420)*8 = 6168 and chosen jetties are 14 and 260.

Table B.2: Capacity settings case 2

260	100
26	0
15	0
14	100

B.3. Case 3

In chase two the spreading of flow is going to be shown. The preferred jetties are 26 and 15. With limiting the flow to 3 persons for both jetties the number of people travelling should be divided between the preferred jetties and the other jetties. The capacities are shown in B.3 The number of chosen jetties p = 4. 3*(202+120+180)+5*(171+180+420)= 5361.

Table B.3: Capacity settings case 3

260	100
26	3
15	3
14	100

B.4. Case 4

In this case the public transport is going to be added. The settings of the public transport hub are 400. this is lower than the fastest route with the watertaxi In that case the expected behaviour is that only the public transport is used: (202+120+180) = 404. The expected percentage of taking public transport instead of watertaxi is 100%

B.5. Case 5

Now, only a percentage of people is travelling with public transport by applying a capacity constraint in order to ensure that a selection of the people takes public transport. The capacity constraints found in B.3 are going to be applied and the traveltime with public transport is going to be set to 500. The expected behaviour is that 3 people are going to take public transport (202+120+180)*3 and 5 people public transport (500). 5/8*100 = 62.5 percent takes takes public transport and the average reduced passenger travel time is going to be (600*8-(600*5+3*502))/8 = 36.75 and jetties at 26 and 15.

B.6. Case 6

To verify a case with multiple Origin Destination pairs, another city node is added. Table B.4 shows an overview with all travel times. This location is going to be the same as 3016, except from the traveltime with public transport. Public transport from 3072 to 3012 is set to '600'. From previous examples, the fastest travel option with the watertaxi is 502. The travel time with public transport is 400 between 3072 and 3016. This is lower than the fastest traveltime with the watertaxi. This means that the travelling from 3072 to 3016 goes with public transport. The expected objective outcome should be: 400*6+502*8 = 6416. The average reduction is (400*6+600*8-6416)/14 = 56. Percentage that takes the public transport should be 6/14 *100 = 43. Expected jetties: 15,26.

Table B.4: travel times between jetties 260,26,15,14 and to postal areas 3072, 3012 and 3016

	260	**26**	**15**	**14**	**3072**	**3016**	**3012**
260	0	136	239	171	180	1800	1800
26	136	0	202	136	180	2280	2280
15	239	202	0	169	1760	120	120
14	171	136	169	0	1920	420	420
Public transport							
3072					0	600	400
3016					600	0	200
3012					400	200	0
Number of people travelling							
3072					0	8	6
3016					0	0	0
3012					0	0	0

C

Coordinates of postal codes

Postal code	*latitude*	*longitude*	Postal code	*latitude*	*longitude*	Postal code	*latitude*	*longitude*
3000-3009	51.92679	4.421901	*3037*	51.93809	4.472285	*3066*	51.93625	4.554884
3011	51.91766	4.486852	*3038*	51.9357	4.465404	*3067*	51.94703	4.550146
3012	51.9194	4.475769	*3039*	51.92942	4.456277	*3068*	51.95782	4.545404
3013	51.92429	4.469251	*3041*	51.92865	4.441546	*3069*	51.96358	4.533328
3014	51.91977	4.466089	*3042*	51.93428	4.43271	*3071*	51.91165	4.505364
3015	51.91175	4.468044	*3043*	51.94255	4.424291	*3072*	51.90186	4.484299
3016	51.90624	4.473647	*3044*	51.93088	4.417564	*3073*	51.89197	4.500548
3021	51.91738	4.459214	*3045*	51.95349	4.450397	*3074*	51.89411	4.513878
3022	51.9205	4.446733	*3046*	51.965	4.426248	*3075*	51.88508	4.507544
3023	51.91236	4.451917	*3047*	51.94741	4.400721	*3076*	51.87381	4.52518
3024	51.90284	4.4585	*3051*	51.94587	4.476788	*3077*	51.89939	4.54884
3025	51.90809	4.442312	*3052*	51.95138	4.471178	*3078*	51.88886	4.547115
3026	51.91273	4.442243	*3053*	51.96581	4.475129	*3079*	51.87833	4.545392
3027	51.91824	4.436636	*3054*	51.95478	4.486354	*3081*	51.89522	4.484849
3028	51.91609	4.423319	*3055*	51.95931	4.5066	*3082*	51.88794	4.467456
3029	51.90932	4.427105	*3056*	51.94978	4.513182	*3083*	51.8872	4.486798
3031	51.92681	4.489956	*3059*	51.98291	4.582107	*3084*	51.8808	4.4809
3032	51.92718	4.480271	*3061*	51.9262	4.5061	*3085*	51.8738	4.491117
3033	5.192755	4.470588	*3062*	51.92947	4.524507	*3086*	51.87465	4.468561
3034	51.93321	4.495865	*3063*	51.91091	4.524732	*3087*	51.89255	4.450364
3035	51.93245	4.481121	*3064*	51.9106	4.549866	*3088*	51.87174	4.44055
3036	51.93772	4.481972	*3065*	51.92571	4.553156	*3089*	51.89077	4.427389