A study into the joint activity - route choice behaviour of pedestrians during large scale events based on revealed preference data.

SAIL EVENT case study

Anastasia Iliadi

Master Thesis
August 2016
A study into the joint activity-route choice behaviour of pedestrians during large scale events, based on revealed preference data

SAIL event, case study

Anastasia Iliadi 4301935

Transport and Planning
Delft University of Technology

Graduation committee:

Prof.dr.ir. S.P. Hoogendoorn – Chair
Delft University of Technology Faculty of Civil Engineering – Transport & Planning
S.P.Hoogendoorn@tudelft.nl

Dr.ir. W. Daamen – Daily Supervisor – TU Delft
Delft University of Technology Faculty of Civil Engineering – Transport & Planning
W.Daamen@tudelft.nl

Dr.ir. Stefan van der Spek – External Supervisor – TU Delft
Delft University of Technology Faculty of Civil Engineering – Transport & Planning
S.C.vanderSpek@tudelft.nl

Ir. P.B.L. Wiggenraad – Graduation Coordinator
Delft University of Technology Faculty of Civil Engineering – Transport & Planning
P.B.L.Wiggenraad@tudelft.nl
Preface
This thesis is submitted as part of the requirements for obtaining the degree of Master of Science of the track Transport & Planning, in the faculty of Civil Engineering and Geosciences, at University of Delft. The thesis was completed at Delft University of Technology.

I am sincerely grateful to all people and circumstances that contributed to my arrival at the current point and have made my thesis possible. Therefore, I would like to devote this preface into expressing my gratitude towards the people without whom I would not be able to arrive here.

First of all I would like to thank my daily supervisor at TU Delft, Winnie Daamen for the fruitful meetings and her support throughout the entire project. Furthermore, I would to thank the other members of my graduation committee: Serge Hoogendoorn, Stefan van der Spek and Paul Wiggenraad – for their encouragement and feedback. A special thank goes to Danique Ton who was interested in my research and taking the time to help me when I needed it.

I am grateful to my family, Athina, Charalampos, Ioanna for their understanding, patience and support. Without them this master degree will not be reality. A special gratitude to my friend Vasilis who stood by me and supported my effort to accomplish my thesis. Moreover, I would like to thank all my friends during my graduation project for their patience. I can promise them that from now on I can devote more time to them. Last but not least, I would like to thank the other students from Afstudeerhok for fun times and fruitful discussions during coffee and lunch breaks.

Sincerely,

Anastasia Iliadi
Summary

Nowadays, there is an increasing number of large scale events and a growing number of people tend to participate to these events. In general, during large scale events the combination of the lack of sufficient information to support crowd management with the movements of the crowd can create not only problems on the undisturbed movements of pedestrians but also potentially dangerous situations i.e serious incidents or even to fatalities (Love Parade in Duisburg in 2010, Hindu-Fest Maha Kumbh Mela in India in 2013). Despite the severity of these problematic situations, still there is limited knowledge about the behaviour of pedestrians during mass events which could support the development of a sufficient crowd management strategy. Therefore the objective of this thesis is to analyse pedestrians’ activity and route choice behaviour at mass events and examine how pedestrians make these choices using SAIL event as a case study. The analysis of pedestrians’ activity and route choice is based on Revealed Preference (RP) data.

This objective is reflected in the main research question:

‘Which factors influence the activity and route choice behaviour of pedestrians during the SAIL mass event and to which extent?

In order to address to this objective literature review is conducted in order to derive information about pedestrians’ behaviour. Literature regarding the choice behaviour of pedestrians during large scale events hardly exists. However, it can be said that all the type of choices (activity choice, route choice etc.) that were identified in literature can be recognized during large events. Two of the most fundamental choices for the crowd management of large scale events are the activity and route choice. Therefore, the factors that influence these choice aspects are recognized. There is limited literature related to the factors that influence activity and route choices during large scale events. Hence, based on the state-of-the art on pedestrian choice behaviour in combination with observations and experience derived from large scale events, the factors that influence the choices of pedestrians during large scale events are identified. These factors were divided into four groups: personal, system, event and external factors. One factor which is related to the impulsive behaviour on the decision making leads to the distinction of pedestrians in two types; the ones that have a predetermined way on making choice and the others that have an intuitively way. These two groups behave in a significantly different way on the decision process and therefore different frameworks should be identified for each group. However, this thesis focuses on the theoretical framework of pedestrians with a pre-determined decision on preference for activity.

Several theoretical frameworks derived from pedestrian behaviour studies described the relationships between pedestrians’ choices and the influential factors. Elements of these frameworks were used for the creation of the theoretical framework for this study.

The last part of the literature refers to discrete choice models. These models can be used for the quantification of the relationships between the influencing factors and the decisions like route choice and activity choice.

The Revealed Preference (RP) data are being assessed based on their capability to capture the factors that influence the activity and route choice behaviour during large scale events. The RP research took place during the SAIL event. Real-time data have been collected using different kinds of sensors: counting cameras (equipped with Wi-Fi sensors), Wi-Fi sensors and GPS trackers. The data collection methods were pre-determined before the start of this thesis. The advantages and the limitations of the available RP data derived in SAIL event are described. The result of this assessment is a selection of the list of the potential influencing factors. The GPS data collection method is more suitable to extract information on route choices by capturing the spatiotemporal movements of pedestrians in comparison with the other available collection methods. In addition, travel time information which is related to the time spent on specific areas (area presence time) can be derived by the analysis of the GPS data. Thus the process of the
raw GPS data aims in extracting information regarding the routes taken and the choice of performing an activity or not. During the SAIL event 110 GPS trackers were distributed 322 times. This implies that a total of 322 trips over the five days of the event were collected. Since the focus area of this study is the Orange route, the trips that were not appeared at this part were excluded from the database. Next, general remarks on pedestrians’ movement patterns and behaviour retrieved from GPS data, along the Orange route are described. The analysis of the GPS data showed that during the SAIL event pedestrians tend to choose routes which are indicated by the SAIL organisation and provide with the opportunity of enjoying the view of the attractions of the SAIL event.

Another interesting finding, is the existence of relationship between the initial route taken and route for the way back. In most cases the initial route is found to be different than the way-back route. A Pearson chi-square test was conducted and it was found that the relationship is considered to be significant with a 95% confidence level. Furthermore, a remark that should not be omitted is related to the percentage of pedestrians that choose to cross the Orange route from station to Verbindingsdam and the way back. It was noticed that a higher percentage of pedestrians cross the area with direction to Verbindingsdam in comparison with the percentage of pedestrians with direction to the station. On the way back (direction to the station) there is a variability; pedestrians choose either to return on food or take a ferry or use PT. This information is valuable for the next SAIL event and the regulatory bodies should take into consideration the aforementioned options (for example increasing the frequency of PT).

Onward, the construction of a joint activity-route model for pedestrians’ choices during the SAIL event is attempted. The analysis is based on the predetermined preference of activities for pedestrians. In this way, a visitor knows beforehand whether he is willing to perform an activity but is uncertain about the activity location. The quantitative model is performed for a small part of the event area. According to the model estimation, the activity time and the walking time are shown to be significant factors for the joint activity-route choice. The activity time and the walking time while performing an activity have positive influence on the joint activity-route utility, while when there is no preference for activity, the walking time has a negative impact.

The method used for the quantification of the relationships between the factors and the choices is discrete choice using the concept of utility maximization. The underlying assumption is that pedestrian chooses the alternative that provides the highest subjective utility.

The descriptive statistical analysis results to additional factors which are not captured by the model. Apparently, there is a high preference for the areas of the main activity of the SAIL event, which simultaneously determines the route and the activity choice. Additionally, the performance of an activity seems to affect the next activity, as the time spent for the latter is relatively lower. Finally, it is concluded that the sociodemographic characteristics of pedestrians do not influence the joint route-activity choices.

The final part of the present work highlights the main conclusions during the investigation process, while the main research question is answered. Several points are raised, regarding the suitability of RP collection methods and the generalization or the boundary boxes approach. Finally, the relevance of the results to other large scale events cases studies is discussed.
# Table of Contents

- **Preface** .................................................................................................................................................. i
- **Summary** .............................................................................................................................................. ii
- **Table of Contents** ................................................................................................................................. iv
- **List of Figures** .................................................................................................................................... iv
- **List of Tables** ..................................................................................................................................... vii

1. **Introduction** ........................................................................................................................................ 1
   1.1 Problem statement .............................................................................................................................. 1
   1.2 Research objective and research questions ..................................................................................... 2
   1.3 Research Scope ................................................................................................................................. 2
   1.4 Contribution of this research ........................................................................................................... 4
   1.5 Research methodology .................................................................................................................... 4
   1.6 Report structure ............................................................................................................................... 5

2. **Literature review on choice behaviour of pedestrians** .................................................................... 6
   2.1 Choices of pedestrians ..................................................................................................................... 6
   2.2 Choice behaviour of pedestrians in large scale events .................................................................. 7
   2.3 Influences on the choice behaviour of pedestrians in large scale event ...................................... 8
       2.3.1 Personal factors ....................................................................................................................... 10
       2.3.2 System factors ....................................................................................................................... 11
       2.3.3 Event factors .......................................................................................................................... 13
       2.3.4 External factors ..................................................................................................................... 14
   2.4 Relationships between pedestrians ‘choices’ ................................................................................... 14
       2.4.1 State of the art of the relationships between pedestrians’ choices ...................................... 14
       2.4.2 Proposed Theoretical framework .......................................................................................... 17
   2.5 Discrete choice modelling .............................................................................................................. 19
   2.6 Conclusions ....................................................................................................................................... 21

3. **Selection of under-study choice behaviour using revealed preference data** .............................. 22
   3.1 Overview of revealed preference data collection methods ............................................................... 22
   3.2 Revealed preference data collection methods used during SAIL .................................................. 24
   3.3 Usability of data on selecting the choice behaviour for this study ............................................... 27
   3.4 Conclusions ....................................................................................................................................... 29

4. **Introduction to the case study: Sail, Amsterdam 2015** ................................................................. 31
   4.1 Description of the SAIL area ........................................................................................................... 31
   4.2 Overview of SAIL event .................................................................................................................. 32
   4.3 Expected outcome of the case study ............................................................................................... 35
   4.4 Conclusions ....................................................................................................................................... 35

5. **Process and analysis of the GPS data** ............................................................................................... 37
List of Figures

Figure 1.1 Selected part of the Orange route for analysis ........................................................................................................... 3
Figure 1.2 Research approach ....................................................................................................................................................... 5
Figure 2.1 Theoretical framework on activity choice and sequences (Root & Recker, 1981) ..................................................... 15
Figure 2.2 Levels in pedestrian behaviour based on Hoogendoorn et al. (2001) ................................................................. 16
Figure 2.3 Route choice process for the individual traveller (Hoogendoorn & Bovy, 2005) ............................................... 16
Figure 2.4 Proposed framework ................................................................................................................................................ 19
Figure 3.1 Overview of sensor locations in Google Earth. Orange numbers were given to the cameras for the need of the example of walking between two cameras ........................................................................ 26
Figure 3.2 Flowchart with selection criteria of factors ............................................................................................................. 27
Figure 4.1 Coloured oceans of SAIL area (SAIL organisation, 2015) ....................................................................................... 31
Figure 4.2 Part of the Orange Ocean (Orange and Purple route) ............................................................................................... 32
Figure 4.3 Veemkade (August 20, 2015) ......................................................................................................................... 33
Figure 4.4 Veemkade (August 19, 2015) canteens and tall ships ................................................................................................. 33
Figure 4.5 Pawns divide multiple modalities (August, 2015) ................................................................................................. 33
Figure 4.6 Behind the Amsterdam train Station at Ruijterkade (August, 2015) ................................................................. 34
Figure 4.7 Overview of the different types of information signs at SAIL (August, 2015) (Galama, 2015) ................................. 34
Figure 4.8 Information point before the divergent point of the Orange and Purple route ............................................................ 35
Figure 4.9 Flow diagram of processing the GPS data ................................................................................................................. 38
Figure 5.1 Trajectories of pedestrian with trip ID 254 ................................................................................................................ 39
Figure 5.2 Spatial bounding boxes ......................................................................................................................................... 40
Figure 5.3 Data selection algorithm ........................................................................................................................................ 40
Figure 5.4 Gender, age, time of day, group size, day of the week and trip duration distributions over the 230 trips sample ................................................................. 44
Figure 5.5 Javakade (22/08) ........................................................................................................................................ 45
Figure 5.6 Veemkade (22/08) ........................................................................................................................................ 45
Figure 5.7 Focus study area .................................................................................................................................................. 46
Figure 5.8 Trip patterns (direction from Amsterdam Central Station to Verbindingsdam) .................................................... 46
Figure 5.9 Visualized common pedestrian patterns (direction from Amsterdam Central Station to Verbindingsdam)-Examples ................................................................. 47
Figure 5.10 Chosen route alternatives (direction from the Amsterdam Central Station to Verbindingsdam) ................................................................. 48
Figure 5.11 Distribution of 172 trips over the 4 alternative routes- Intermediate stops ............................................................... 49
Figure 5.12 Distribution of trips over the reasons why they did not arrive at Verbindingsdam ....................................................... 49
Figure 5.13 Share of trips at intermediate destinations ............................................................................................................ 49
Figure 5.14 Trip distribution over intermediate stops ................................................................................................................ 50
Figure 5.15 Trip patterns (direction from Verbindingsdam to Amsterdam Central Station) ...................................................... 51
Figure 5.16 Origins of the trips that crossed the area .................................................................................................................. 51
Figure 5.17 Trip patterns (direction from Verbindingsdam to Amsterdam Central Station) ...................................................... 52
Figure 5.18 Route alternatives for the direction from Verbindingsdam to the station (Origin: Box 21, Destination: Box 58) ......................................................................................... 53
Figure 5.19 Trips share over the 4 alternatives (direction from Verbindingsdam to the station) ...................................................... 54
Figure 5.20 Trips share over the shortcuts connecting Veemkade with Piet Heinkade ......................................................................... 54
Figure 5.21 Average time spent for combinations of routes for the way forth (from station to Verbindingsdam) and back (from Verbindingsdam to station) ................................................................. 57
Figure 5.22 Trip patterns (direction from Verbindingsdam to Amsterdam Central Station) ...................................................... 58
Figure 5.23 Trip patterns (direction from Verbindingsdam to Amsterdam Central Station) ...................................................... 59
Figure 5.24 .................................................................................................................................................................................. 60
Figure 5.25 .................................................................................................................................................................................. 60
Figure 5.26 .................................................................................................................................................................................. 60
The estimation of the joint activity route choice. The available activities are activity times for each area separately (AB, BC) when Route1a_2a is chosen.

Table 2.1 Factors which could influence the activity and walking time.

Figure 6.9 Visualization of joint activity route choices in area AC. The choice 06 is illustrated in case that there is an activity in the shortcut.

Figure 6.10 Distribution of pedestrians that do not perform an activity.

Figure 6.11 Relationship between socio-demographic characteristics of pedestrians (gender, age, group size) and joint activity-route choice model.

Figure 6.12 Relationship between the group-size and the performance of activity or not.

Figure 6.13 Distribution of the area presence time over the group-size.

Figure 6.14 Distribution of joint activity route choices along the days.

Figure 6.15 Time of day distribution over the joint activity-route choice.

Figure 6.16 Distribution of the joint activity route choices according to the time that a pedestrian has spent before arrives at the understudy area.

Figure 6.17 Area presence time distribution over the joint activity route choices.

Figure 6.18 Walking time distribution over the joint activity route choices.

Figure 6.19 Distribution of average walking time according to the time of day for each alternative.

Figure 6.20 Activity time distribution over the joint activity-route choices.

Figure 6.21 Distribution of activity times for each area separately (AB, BC) when Route1a_2a is chosen.

Figure 6.22 Individual activity times for every pedestrian that chose to perform activities in both parts of the network.

Figure 6.23 Locations of Wi-Fi sensors used for the estimation of the activity and walking time.

Figure A.1 Theoretical framework on activity choice and sequences (Root & Recker, 1981).

Figure A.2 Levels in pedestrian behaviour based on Hoogendoorn et al. (2001).

Figure A.3 Route choice process for the individual traveller (Hoogendoorn & Bovy, 2005).

Figure A.4 Framework for familiar travellers (Ton, 2014).

Figure A.5 Framework for unfamiliar travellers (Ton, 2014).

List of Tables

Table 2.1 Factors which could influence the activity-route choice behaviour of pedestrians - Relationships between influencing factors and choice aspects.

Table 3.1 Factors that can be captured using the collected data (in grey are the factors that cannot be captured by RP data collection methods or other available tools).

Table 3.2 Advantages and drawbacks of the revealed data collection methods used during the SAIL event.

Table 5.1 Overview of the processing of the GPS data.
Table 5.2 Descriptive Statistics for the examined dataset ............................................................. 43
Table 5.3 Box sequences of alternative routes (direction from station to Verbindingsdam) .......... 48
Table 5.4 Trip share over the 4 routes (direction from station to Verbindingsdam) ..................... 49
Table 5.5 Trip shares over the route alternatives (direction from Verbindingsdam to train station) ....... 53
Table 5.6 Cross table of the trips that walked from the station to Verbindingsdam and the way back ... 55
Table 5.7 Chi-Square Tests ......................................................................................................... 55
Table 5.8 Average total time spent (in minutes) for the way forth and back .................................. 59
Table 5.9 Activities that can be performed at the boxes of Veemkade ........................................ 59
Table 5.10 Kolmogorov-Smirnov Test .......................................................................................... 64
Table 6.1 Descriptive Statistics for the examined dataset ............................................................. 77
Table 6.2 Physical attributes of the route alternatives from A to C ................................................ 78
Table 6.3 Correlations of socio demographic characteristics versus joint activity-route choice ........ 79
Table 6.4 Distribution of sociodemographic characteristics over the joint activity route choices ....... 80
Table 6.5 Distribution of joint activity route choices along the days ............................................. 84
Table 6.6 Relationship between time of day and joint activity-route choice ................................. 85
Table 6.7 Distribution of joint activity route choices along the days ............................................. 86
Table 6.8 Variable description ..................................................................................................... 93
Table 6.9 MNL model - ASC ....................................................................................................... 97
Table 6.10 Overview of the individual models of each parameters/attributes ................................. 99
Table 6.11 Correlation table of the walking time versus the distance from the activity performed ...... 101
Table 6.12 General information on the combined MNL model ................................................... 102
Table 6.13 Utility parameters for combined MNL model .............................................................. 102
Table 6.14 Overview of the individual models of each parameters/attributes without ASC .............. 105
1 Introduction

Large scale events or more specifically mass events in urban areas where the available infrastructure might be limited are regularly organized and attract large crowds. In The Netherlands, a great number of mass events are taking place in and around the area of city centres. Thousands of visitors are participating to these events. For example, around 700,000 people visited the city centre of Amsterdam during the King’s Day and the succession of Willem-Alexander in 2013 (COT, 2012), (COT, 2013), while approximately 130,000-300,000 visitors participated in various events in Nijmegen’s city centre during Vierdaagsefeesten in 2013 (Respons, 2013) to name a few. A similar mass event is organized in the city centre of Amsterdam: SAIL Amsterdam. SAIL event is Europe’s biggest free accessible nautical event, organised every 5 years on Amsterdam which attracts thousands of visitors. In 2015, SAIL event took place during the period of 19th-23rd of August and around 2.3 million national and international visitors (distributed over the 5 days of the event) gathered in the area around the IJhaven in Amsterdam.

The municipality of Amsterdam is interested in running a project on the crowd behaviour throughout Amsterdam’s network during mass events. SAIL event has been used as a driving example to illustrate various concepts in crowd monitoring and crowd management which can be applied to future events. For that purpose, during the SAIL event the municipality of Amsterdam in collaboration with Amsterdam Institute of Advanced Metropolitan Solutions (AMS), the organisation of SAIL and DAT Mobility performed a pilot study which uses a combination of various methods of real time data collection in order to give an optimum picture of pedestrian flows along the SAIL route. Thus the main focus of the pilot study is on how to gain reliable information on pedestrian flows during large-scale public events such as SAIL and use it effectively for crowd management. Real-time data have been collected using different kinds of sensors: counting cameras, Wi-Fi sensors and GPS trackers.

Real time measurement techniques offer insight into the flow and the behaviour of visitors to SAIL, the routes people have taken and the amount of time spent at a given location. This information is of important value for the organisation of future events. This thesis uses the data which were collected during SAIL event in order to analyse the route and activity choices of pedestrians.

1.1 Problem statement

In general, during large scale events the movements of the crowd can develop critical dynamics leading not only to problems on the undisturbed movements of pedestrians but also to potentially dangerous situations i.e serious incidents or even fatalities (Love Parade in Duisburg in 2010, Hindu-Fest Maha Kumbh Mela in India in 2013).

During all these problematic situations, not only large crowds were involved but also there was lack of accurate and sufficient information to support crowd management. Sufficient information could be obtained by developing a system for measuring and managing pedestrian flows, such as the real-time crowd monitoring system which was tested at SAIL. However, this monitoring system could be more accurate and useful as prediction and decision support tool for crowd management measure deployment if knowledge of choice behaviour of the crowd (pedestrians) is available. The resulting information from the monitoring phase can be used to gain insight into the choice behaviour of visitors which is important for prediction and thus management. The link between the data derived from the monitoring phase and their use in predicting pedestrians’ behaviour during mass events needs to be thoroughly assessed. The high complexity of pedestrians’ behaviour dictates the requirement of its simplification in order to be comprehended. A first step towards this direction is studying the factors that influence the choice aspects of pedestrians.

The influence of different factors on the route and activity choice behaviour of pedestrians has been investigated in previous studies. Borgers and Timmermans (1986) examined the route choices of pedestrians in inner-city shopping centres, Zomer (2014) investigated the influence of information
measures on the activity choice behaviour during the mass event of Vierdaagsefeesten in Nijmegen. Ton (2014) addressed the activity and route choice behaviour of pedestrians in train stations. Each of these studies uses data derived from different collection methods (revealed or stated preference) to analyse the pedestrians’ choices and predict the influencing factors on their choice behaviour. However, up to date, there is little knowledge in choice behaviour of pedestrians during large scale events and more specifically knowledge of how pedestrians choose their routes and the activity areas is yet limited. Moreover, a joint model for pedestrian’s activity and route behaviour during large scale events estimated by revealed data has not been structured and tested yet.

During mass events, various aspects of pedestrians’ choice behaviour (route and activity choices, choices related to walking behaviour etc.) are observed. This thesis is focusing on the choices which are related to the routes taken and the activity choices.

1.2 Research objective and research questions
The objective of this research can be summarised as follows:

The objective of this thesis is to analyse pedestrians’ activity and route choice behaviour at mass events based on revealed preference data and examine how pedestrians make these choices using SAIL event as a case study.

In order to structure the objective, several research questions are defined. The main research question is:

‘Which factors influence the activity and route choice behaviour of pedestrians during the SAIL mass event and to which extent?’

Besides the main research question a number of side research questions are proposed to compose the objective of this thesis. These side research questions are the following:

1. Which factors influence the choice behaviour (activity and route choice) of pedestrians in large scale events according to literature and field observations?
2. Which relationships exist between activity and route choice behaviour of pedestrians according to literature?
3. Which modelling approaches that describe discretionary activities such as route choice behaviour can be applied to estimate model(s) on activity and route choice behaviour of pedestrians?
4. To what extent is the Revealed Preference data a suitable data collection method for collecting quantitative data on route and activity choice behaviour of pedestrians in large scale events, like SAIL?
5. What information is extracted on the activity and route choice during the SAIL event from the collected data?
6. Which factors have significant influence on the activity and route choice behaviour of pedestrians during SAIL event?

1.3 Research Scope
The scope of this research is determined in this section. The following aspects are addressed in this thesis: the use of revealed preference data collection method, the location of the data collection (SAIL event, Orange route) and the level of detail in choice process.
Use of Revealed preference data collection method

A pilot has been performed to develop a crowd monitoring dashboard for the municipality of Amsterdam. For this reason, real-time data have been collected using different kinds of sensors: counting cameras (equipped with Wi-Fi sensors), Wi-Fi sensors and GPS trackers. These data are being examined on their capability to capture the activity and route choices that pedestrians made during large scale events, as well as the factors that influence these choices. Chapter 3 gives information on the data collection methods used during the event.

The location of the data collection (SAIL area)

The SAIL event was chosen as a case study in order to investigate the activity and route choice behaviour of pedestrians at mass events. SAIL event is an interesting research case study given that SAIL Amsterdam is the largest public event in the Netherlands and attracts a great number of visitors. Next to that, the SAIL area offers a range of possible activities to perform as well as routes that can be taken. Also, during the event, real-time data have been collected using different kind of sensors. The analysis of these data can give more information regarding the choice behaviour of the visitors. However, due to the extent of the SAIL area and the time constraints of this thesis, a part of this area is selected to study which factors influence the activity and route choice behaviour of visitors. The selected area for further analysis is a part of the Orange route and is depicted in Figure 1.1, while for the estimation of a joint activity-route choice model a small part of this area is used (Figure 1.1, blue circle). More information about the selected study areas is given in Chapter 5 and Chapter 6. Orange route was the main area at SAIL event according to the SAIL organization. Other parts of the SAIL area are not considered for this thesis.

Level of detail in decision making

Three different levels of pedestrian's behaviour can be distinguished (Hoogendoorn, et al (2001)). These levels are the following:

- Strategic level: activity set choice
- Tactical level: activity scheduling, activity area choice, route choice
- Operational: walking, waiting, performing activity, trajectory choice

The data collection methods used (counting cameras, Wi-Fi, GPS) for analysing the choice behaviour of pedestrians might set a constraint on the level of detail. According to their capability to capture the choice behaviour of pedestrian's choice behaviour during mass events, the level of decision making is determined. For example, the available data cannot capture the exact activity locations but the activity areas (where activities are performed). More discussion about it is done in Chapter 3.

Figure 1.1 Selected part of the Orange route for analysis
This thesis focuses on the tactical level which includes activity area choice and route choice. The interactions between different decisions levels are not considered in this thesis. For example measures taken on the strategic level and influence pedestrians’ behaviour are not taken into account to this thesis.

1.4 Contribution of this research
This thesis contributes both scientifically and in practice. On a scientific level this research contributes in gaining insight into the activity and route choice behaviour of pedestrians at mass events. The methodology that is developed in this thesis contributes to understand the possibilities of using Revealed Preference (RP) data collection methods in order to research pedestrian’s choice behaviour. Thus the limitations of this approach are stated and discussed. Furthermore, the quantification of the factors that influence the activity and route choice by means of RP data provides an added value to science. Last but not least, an evaluation of potential modelling jointly the activity and route choice behaviour of pedestrians during large scale events is discussed.

On a practical level, the findings and insights related to the factors that influence the activity and route choice behaviour of pedestrians during a large scale event can help on prediction and thus on the development of a crowd management strategy. In other words, the results of this thesis regarding the analysis of the route and activity choices of visitors could be considered for crowd management measures for the next SAIL event (2020). In addition, the methodology that is developed on this thesis regarding the extraction of information about the route and activity choices from revealed preference data can be used by other mass events. This implies that the resulting information could be a source of crowd management measures at other mass events.

1.5 Research methodology
The research approach of this thesis is visualized in Figure 1.2.

The research starts with the problem recognition which formulates the objective and the research questions. The literature review aims in providing information related to the objective of this research topic. It gives insight into the factors that influence the route and activity choice behaviour of pedestrians in large scale events. In the identification process of the possible influencing factors a selection is made, based on the extent in which these factors may influence the activity and route choices and their relevance with large scale events. Based on the state-of-the art on pedestrian choice behaviour in combination with observations and experience derived from large scale events, the factors that could influence the choices of pedestrians during large scale events are identified.

The relationships between the influencing factors and choice aspects are recognized. This leads to the creation of a theoretical framework on choice behaviour. This framework provides a qualitative conceptualization of how people make their choices. Based on the collected revealed preference data during the SAIL event, the relationships between the influencing factors and the choice aspects are tested on their ability of being estimated. Once the collected data are processed and analysed the influencing factors need to be quantified in order to define their influence on the activity and route choice of pedestrians. This results in a better understanding of the route and activity choice behaviour of pedestrians during large scale events.
1.6 Report structure

The report structure follows the research approach sequence presented in Figure 1.2. In Chapter 1 the problem is identified and the objective and research questions are formulated. Chapter 2 focuses on the literature which is related to activity and route choice behaviour of pedestrians during large scale events, the factors that influence these choices as well as the relationships between the pedestrian choices (by presenting several theoretical frameworks from the state-of-the-art). Chapter 3 discusses some of the collection methods that were applied in other studies to derive revealed preference data while the method used for the data collection during SAIL event is assessed based on the capability to capture activity and route choices. Chapter 4 gives an overview of the case study (SAIL event). In Chapter 5 the collected data are processed and analysed while in Chapter 6 a joint activity route choice model is estimated. Finally, in Chapter 7 the final conclusions, discussions and recommendations are presented.
2 Literature review on choice behaviour of pedestrians

Pedestrians’ behaviour is a complex phenomenon. Capturing and forecasting require (among other) analysis on various behavioural dimensions in terms of choice. The concept of choice is present in many aspects of pedestrians’ behaviour. The aim of this chapter is to recognize the choices made during large scale events as well as the factors that influence these choices.

In Section 2.1, it is clarified that from the major choices that pedestrians are facing, this thesis will focus on the activity and route choice behaviour. In addition, the ways that these choices have been addressed in the literature are described shortly. Section 2.2 focuses on investigating the activity and route choices made in large scale events. However, in order to better understand the choice behaviour of pedestrians in large events, the following research question is discussed in Section 2.3.

1. Which factors influence the choice behaviour (activity and route choice) of pedestrians in large scale events according to literature and field observations?

In Section 2.4 the relationships between the choices made are reviewed in respect to the research question below:

2. Which relationships exist between activity and route choice behaviour of pedestrians according to literature?

When these relationships are identified, the theoretical framework can be created. This is discussed in Section 2.4.2.

For researching choice behaviour of pedestrians discrete choice models are widely used. Thus, in Section 2.5 an overview of discrete choice models regarding the choices of pedestrians and more specifically the route and activity choice are being described. Accordingly, the following research question is addressed:

3. Which modelling approaches that describe discretionary activities such as route choice behaviour can be applied to estimate model(s) on activity and route choice behaviour of pedestrians?

Finally, in Section 2.6 the conclusions are given.

2.1 Choices of pedestrians

The concept of choice is present in many dimensions of pedestrians’ behaviour. Most of the choices are highly interrelated and usually considered jointly in the literature. The most fundamental choices that are met in the literature are the following: activity choice, destination choice (is related to the choice of the activity location), mode choice and route choice. (Bierlaire & Robin, 2009). This thesis will focus on the activity and route choice of pedestrians, while the mode choice will be left aside. Thus analysis on how each of these choices has been addressed in the literature is required.

Hoogendoorn and Bovy (2001) described the choice behaviour of pedestrians at three different levels: strategic level, tactical level and operational level. At the strategic level, pedestrians choose which activities they want to perform (activity set). The order in which the activities will be performed is chosen. This is the activity scheduling and is determined on the tactical level. Next, the activity area choice is performed which concerns the choice of location to perform an activity by individual pedestrians. The last process at the tactical level is the route choice between origins, possible intermediate destinations and pedestrians’ final destinations. It was found that route choice has been considered in a variety of research fields (such as psychology, geography, and traffic engineering) (Daamen, 2004).

Finally, at the operational level, pedestrians take instantaneous decisions for the immediate next time period, in line with the choices made at the tactical level. Most of the pedestrians’ decisions on the
operational level are related to walking, waiting, performing an activity and interaction with Public Transport.

The choice of activity triggers the need for travelling. This means that the activities are usually used as a purpose for travel. Only a few authors have analysed the activity choice in the specific case of pedestrians. Among the limited literature, references are made on Borgers & Timmermans (1986), Hoogendoorn & Bovy (2004), Zhu (2008), Daamen (2004), Ton (2014). Ton (2014) states that activity choice happens on different levels. On the higher level the main activity or purpose for travel is chosen while on the lower level the activity choice while travelling towards the main chosen activity is done. Related to activity choice three more choices are identified. Activities need to have a relative importance compared to other activities (activity hierarchy). This is necessary because the set of potential activities may demand more than the available time of a person. Therefore, the importance of each activity needs to be recognised. Next to that, the set of activities needs to have a sequence in order to be performed (activity sequence). Finally, the activity often can be done at different locations (activity location choice).

The choice of the route is a critical dimension of the pedestrian behaviour, hence a great number of studies refer to this choice (Seneviratne & Morrall (1985), Bovy & Stern, (1990), Daamen, et al (2005), Hoogendoorn & Bovy (2004), Daamen & Hoogendoorn (2003), Antonini, et al (2006), to name a few). A route choice refers to a trip from a given origin to a given destination between which multiple route alternatives exist. Three types of choice processes may be distinguished (Bovy & Stern, 1990), namely simultaneous choice (choice of entire routes), sequential routes (choice of sub-routes at decision points in the network), and hierarchical choice (similar to sequential choice, but choice behaviour depends on previous choices). Quite distinct from any specification of the choice process is the phenomenon of adaptive route choice (Bovy & Stern, 1990), (Stern & Sinuani-Stern 1989), where travellers make choices depending on changing conditions they encounter while they are on their way. A choice set is a group of alternatives, out of all possible routes between a given origin and a given destination from which travellers will make their choice (Daamen 2004).

Most of the available literature deals with pedestrians’ choices in facilities like public transit station, urban walkway and shopping mall, while literature on pedestrians’ choices in large scale events hardly exists. Next section (Section 2.2) focuses on investigating the choices that are made during large scale events based on the literature found regarding the pedestrians’ choices.

2.2 Choice behaviour of pedestrians in large scale events

Literature regarding the choice behaviour of pedestrians during large scale events hardly exists. However, it can be said that all the type of choices (activity choice, route choice etc.) that were identified in literature can be recognized during large events.

Several choices are made by pedestrians in large scale events. The most obvious choice made in an outdoor large scale event like SAIL event is activity choice. As mentioned in Section 2.1, the activity choice occurs on different levels. On the higher level the main activity is chosen. This can be e.g. joining SAIL event (decision to go to the event). On the lower level is the activity choice that can be performed during the SAIL event (watching and visiting tall ships, eating/drinking, participating to a concert). Depending on the type of the mass event, the variability of the available activities and the size of the activity choice set differ. Furthermore, every mass event may have different character and configuration which influence the dynamic and the flexibility of pedestrians’ choice behaviour as well as the assigned priority to the activities. Activities need to have a relative importance compared to other activities (activity hierarchy) since the available time of each person may be not enough to cover the set of potential activities during the event or maybe the activity itself has a time restriction (opening hours for example). Thus, it needs to be decided which activities are important and which are less important. This means that a priority needs to be allocated/assigned to each activity.
Next to the activity hierarchy, the activity set needs to have a sequence in order to be performed. For all the activities in the activity set, it needs to be determined at which locations these activities will be performed. Each activity can often be performed at different locations. Usually, tools like maps, signs or real time information can assist on finding where to perform an activity.

Apart from the choices related to the activity choice a pedestrian also chooses a route. Route choice is possible when multiple routes (choice set) are available. The attributes of the available alternatives describe the different characteristics of the routes in terms of preference and attractiveness. In this research the route definition is the path that connects the decision points used in the network. At these points decisions are made regarding the performance of an activity or not. For both practical and scientific purposes, it is interesting to know the distribution of pedestrians over the routes and in which way the choice of each route is made during large scale (mass) events, like SAIL.

Despite the fact that pedestrians have to face similar choices either during an event or at a train station, there is a key difference. The main difference lies on the process of decision making (behaviour) regarding the choices as well as the relationships between the choices (activity area choice- route choice for instance).

Regarding the decision process during large scale (mass) events, to the best of our knowledge, there is no model to estimate this decision process. Hoogendoorn’s theory on normative behaviour which assumes that route choice, activity area choice and activity scheduling are simultaneously optimized, is mostly applicable on commuters (Hoogendoorn & Bovy, 2004). It is expected that behaviour of visitors during large scale events is different. Pedestrians in mass events do not only move from one point A to point B, towards a goal with the most direct route but also engage more personal motivated, multipurpose behaviour during the trip. According to Zomer (2014), in large scale events a variety of purposes exists, which can be divided into two categories:

- Obligatory or functional purposes such as eating, searching for a toilet, shopping etc.
- Enriching purposes such as sightseeing, socializing or entertainment.

Borgers and Timmermans (1986) consider impulse stops, where the activity choice is not planned, but triggered by stimuli in the pedestrian’s environment. This is particularly relevant not only for shopping (Borgers & Timmermans, 1986) and tourism (Stewart & Vogt, 1997), where individuals can easily be diverted from their original plans, but also for large scale events where pedestrians with impulsive behaviour can be met.

In general, the choices that pedestrians make in events are the following: activity choices, activity hierarchy, activity sequence, activity location choice and route choice. After the identification of pedestrians’ choices, it is necessary the attributes that are of most influence on the choice behaviour of pedestrians during large scale events to be investigated in order to get better insight into the activity and route choice. Next section (Section 2.3) elaborates on the factors that exert influence on activity and route choice behaviour of pedestrians.

### 2.3 Influences on the choice behaviour of pedestrians in large scale event

In this section literature is studied extensively in order to determine the factors which influence the activity and route choice of pedestrians during large scale events. However, there is limited literature related to the factors that influence pedestrians’ choices during large scale events. Hence, based on the state-of-the art on pedestrian choice behaviour in combination with observations and experience derived from large scale events, the factors that could be relevant to pedestrians’ choices during large scale events are identified. This means that in this identification process a selection is made based on the extent in which these factors may influence the activity and route choices and their relevance with large scale events.
According to Daamen (2004), the factors that influence pedestrian choices can be divided into four groups: network characteristics, route characteristics, personal characteristics and trip characteristics. Ton (2014) used a different grouping structure for her research. The groups that she identified are the following: personal factors, system factors, Public Transport factors and external factors. The group of system factors was created by combining the network and route characteristics, since as she stated network, and route characteristics are not as important for activity choice as they are for route choice. Similar categorization can be applied to this thesis. The personal factors identified by Daamen (2004) can also be used in this research. Moreover, since this thesis is focusing on pedestrians’ choice behaviour during events, the event factors are also included. Another group that affects both activity location and route choice is the external factors such as natural environmental factors. To sum up, the groups identified for this research are personal factors, system factors, event factors and external factors. An overview of the factors that influence the activity and route choice behaviour of pedestrians in large scale events is given in Table 2.1.

Table 2.1 Factors which could influence the activity-route choice behaviour of pedestrians- Relationships between influencing factors and choice aspects

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Choice aspects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Personal factors</strong></td>
<td></td>
</tr>
<tr>
<td><strong>General factors</strong></td>
<td>Route</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Route</td>
</tr>
<tr>
<td>Familiarity</td>
<td>Activity sequence, activity location, route</td>
</tr>
<tr>
<td>Emotional State</td>
<td>Activity hierarchy, activity sequence, route</td>
</tr>
<tr>
<td>Decision style (Herding behaviour)</td>
<td>Activity location, route</td>
</tr>
<tr>
<td><strong>Trip factors</strong></td>
<td></td>
</tr>
<tr>
<td>Trip purpose</td>
<td>Choice of activities, activity hierarchy, activity sequence, route</td>
</tr>
<tr>
<td>Time spent in the event area</td>
<td>Activity sequence, activity location, route</td>
</tr>
<tr>
<td>Group composition</td>
<td>Choice of activities, activity sequence, activity location, route</td>
</tr>
<tr>
<td>Time of day or week</td>
<td>Choice of activities, activity location, route</td>
</tr>
<tr>
<td>Impulse behaviour</td>
<td>Activity sequence, activity location, route</td>
</tr>
<tr>
<td><strong>Location factors</strong></td>
<td></td>
</tr>
<tr>
<td>Dimensions of the location</td>
<td>Activity location, route</td>
</tr>
<tr>
<td>Visibility</td>
<td>Activity location, route</td>
</tr>
<tr>
<td>Amount of shops and other activities available</td>
<td>Activity sequence, activity location, route</td>
</tr>
<tr>
<td><strong>System factors</strong></td>
<td></td>
</tr>
<tr>
<td>Travel time</td>
<td>Activity location, route</td>
</tr>
<tr>
<td>Distance</td>
<td>Activity location, route</td>
</tr>
<tr>
<td>Crowdedness</td>
<td>Activity location, route</td>
</tr>
<tr>
<td>Number of attractions</td>
<td>Route</td>
</tr>
<tr>
<td>Weather conditions and weather protection</td>
<td>Activity location, route</td>
</tr>
<tr>
<td>Directness</td>
<td>Activity location, route</td>
</tr>
<tr>
<td>Environment</td>
<td>Activity location, route</td>
</tr>
<tr>
<td><strong>Event factors</strong></td>
<td></td>
</tr>
<tr>
<td>Event characteristics</td>
<td>Choice of activities, activity hierarchy, activity sequence, activity location, route</td>
</tr>
<tr>
<td>Information and crowd management measure</td>
<td>Activity sequence, activity location, route</td>
</tr>
<tr>
<td><strong>External factors</strong></td>
<td></td>
</tr>
<tr>
<td>Natural environmental factors</td>
<td>Activity sequence, activity location, route</td>
</tr>
<tr>
<td>Land use of the event area</td>
<td>Route</td>
</tr>
</tbody>
</table>
These factors will be analysed extensively in the following Sections (2.3.1-2.3.4).

2.3.1 **Personal factors**

The personal factors that influence choice behaviour can be divided into the following two categories: general factors of each pedestrian and factors related to the trip

**General personal factors**

The general personal factors refer to age, gender, familiarity, emotional state, and decision style.

- **Age**
  
  Hill (1982) and Seneviratne & Morrall (1985) concluded that people older than 60 years old choose routes based on different reasons than other age groups. For example, seniors place a higher value to the safety of the route (Seneviratne & Morrall, 1985). On the other hand, children due to their explorative nature, make choices on a different and unpredictable way. According to Hill (1982) it was found that they prefer routes with a higher complexity in comparison with other age groups. The complexity relates to the number of turns one needs to make. It is plausible that in a large scale event people behave similar to these findings.

- **Gender**
  
  Seneviratne & Morrall (1985) and Verlander & Heydecker (1997) found that there is no difference on choosing a route between men and women. However, according to Hill (1982) women tend to choose more complex routes than men. This difference may stem from the research method used (Golledge, 1999). Seneviratne & Morrall (1985) used stated preference data, while Hill (1982) observed the routes of people (revealed preference). Thus different findings are seen in literature regarding the gender.

- **Familiarity**
  
  Familiarity was addressed by e.g. Dijkstra, et al (2009) and Bovy & Stern (1990). People tend to choose routes and activities they know and are hesitating or avoiding to choose new ones (Golledge, 1999). A number of studies found that pedestrians familiar with the understudy environment (in every case) behave significantly different from unfamiliar pedestrians. Unfamiliar pedestrians show more searching and exploring behaviour and they do not always make the logical choices because they have limited knowledge regarding the available choice set in comparison with familiar pedestrians. For the case of mass events, familiar pedestrians can be defined as the ones that they have knowledge of the event area and the activity locations since it is not the first time that they visited the event area. On the other hand, unfamiliar pedestrians do not have any experience regarding the event and the event area. This factor is important to take into account in this research.

- **Emotional state**
  
  In a crowded environment like the one during mass events, feelings like panic, frustration, unsafety may appear (Lee, et al., 2001). This is because, people may feel restricted in their movement and the unavailability of space may lead to stress and uncertainty. Furthermore, it is possible for pedestrians to get tired of the crowdedness so they will seek for another route less congested or will avoid performing activities. Emotional state however is a subjective factor that is hard to be quantified. Hence, it is only described in literature in a qualitative way.

- **Decision style**
  
  According to Bovy & Stern (1990), Seneviratne & Morrall, (1985) each pedestrian makes decisions based on a decision style, which may be different for the various types of decisions. Decision styles are based on herding behaviour, utility maximisation, and habitual behaviour (Avineri & Prashker, 2003).
Trip factors
Trip factors refer to trip purpose, time spent in the event area, group composition and impulsive behaviour.

- Trip purpose
  Several trip purposes are distinguished into compulsory/mandatory trips (working trips, educational trips) and discretionary/optional trips (social trips, recreational trips). It was found that a significant difference arises in choice behaviour when looking at trip purpose (e.g. van Hagen (2011)). For instance, people heading for work or school pay less attention to the attractiveness of the route (Seneviratne & Morall, 1985). In this thesis the trip purpose is leisure/social recreational, thus more attention will be paid in the characteristics of the route (attractiveness for instance) and the opportunity of performing an activity while walking.

- Time spent
  Time spent in the event area can be related to the activities performed during this period of time. More available time increases the probability of performing (more) activities. This was never quantified in research so it would be interesting to examine in this research the influence of the time spent in the event area on pedestrians’ choices during large scale events.

- Group size and composition
  Individual choices can differ from choices made by a group (Hill, 1982) (Bierlaire & Robin, 2009). Individuals’ decisions are influenced by other members of the group (Goldstone & Janssen, 2005). Regarding the group composition, Bierlaire & Robin (2009) found the groups consisting solely of females visit more shops than groups consisting of men.

- Impulse behaviour
  Impulse behaviour was researched by Borgers and Timmermans (1986). They consider impulse stops, where the choice of the activity is not planned, but is triggered by stimuli in the pedestrian’s environment. This means that individuals can easily be diverted from their original plans during a large scale event. In mass events, 2 main types of pedestrians can be identified. The ones that know beforehand if they want to perform an activity and the others that decide whether they want to perform an activity or not while walking in the event area (intuitive choices). More information about this factor is given in Chapter 6. Impulse behaviour is also addressed in the research of Dijkstra, et al (2009).

- Time of day or week
  Golledge (1999) and Seneviratne & Morall (1985) discussed this factor. The choices and reasons for choosing a route can change during the day or week (Golledge, 1999). In the case of large scale events, it is possible that the time of day may influence the type of the activities that pedestrians want to perform. For instance, during afternoon people usually have lunch, so they will opt for finding locations to satisfy this need. Not much research considering this factor has been carried out so this offers opportunities for this research.

2.3.2 System factors
The system factors that influence choice behaviour of pedestrians are focusing on large scale events. Within this group a distinction is made between location factors and route attributes.

Location factors
The location factors concerned are the following: dimension of the location, visibility, amount of shops and other activities available.
• Dimensions of the location
Timmermans (2009), Schadschneider (2009), Bierlaire and Robin, (2009) were addressed to this factor. The dimensions of an area and more specifically the width, influence the movement of people and have an impact on their route and location choice.

• Visibility
Routes that are partially invisible, they may be skipped when a person is not aware of their existence. The location of the invisible route is also often underestimated (Montello, 1991). This happens, because one can first see the visible locations and routes thus he/she is more tempted to choose the visible ones. This factor therefore relates to the familiarity of pedestrians in the event area.

• Amount and type of activities
The availability of multiple activities or different types of activities influence the choice behaviour (especially related to activity choice). It was observed that the presence of some activities at a certain location along the event area triggered different behaviour than when those activities were not present. An example is the presence of the café/ restaurant in the Bimhuis at the SAIL area. Eating or having a coffee (activities) at this location triggers people to stay longer time, while the opposite happens when someone buys a take away- coffee from a canteen.

Route attributes
The route attributes concerned are: travel time, distance, crowdedness, number of attractions, comfort, weather protection and directness.

• Travel time
Many studies, (Cheung & Lam, 1998), (Daamen & Hoogendoorn, 2003) investigated the influence of travel time on choosing a route. Along the routes there are activities (visiting/watching tall ships, eating, buying coffee) where queuing can occur locally. This significantly influences the travel time in some cases. In this thesis, the travel time is related to the walking time. Walking time can provide information about the prevailing conditions and the potential delays due to crowding. This can be done by comparing the real walking time with the free flow walking time. Pedestrians can therefore switch routes if there is a long queue (Voskamp, 2012). It would be interesting to investigate how pedestrians weight their travel time during large scale events and whether they choose route where a long queue or crowdedness occur.

• Distance
According to Seneviratne & Morrall (1985) routes that are shortest in distance versus shortest in time are ambiguous for pedestrians. Travellers do not know whether they optimise for time or distance. However, some studies (Borgers & Timmermans, 1986) use distance in their research. The distance is often addressed in literature, but only related to route choice. However, it was observed that people also take distance into account when choosing a location for their activity. If a choice between two equal services has to be made the closest is often selected, unless there are specific preferences for the other service. The influence of distance in activity choice location during mass events has not been researched before and provides an interesting research opportunity.

• Crowdedness
Lee, et al (2001), Helbing (1997) and Daamen & Hoogendoorn (2003) to name a few, refer to crowdedness. Crowdedness occurs when the available space does not satisfy the demand. When there is crowdedness the level of comfort is low (Lee, et al., 2001) and high travel time is expected. Therefore a pedestrian may consider to change the route that he/she followed (Voskamp, 2012). However Daamen (2004) stated that even if the progress on a direct route is relatively slow the
choice of a longer route (in distance) is seldom made. Galama (2015) found that the crowdedness influence the route choice behaviour of pedestrians during the SAIL event. According to revealed preference data it was retrieved that crowdedness had a highly attractive value. On the contrary, based on the stated preference data she found that crowdedness is a repellent influential factor on the route choice.

To the best of author’s knowledge the influence of crowdedness on activity choice during mass events has not been addressed in literature so it offers a research opportunity for this study.

- **Number of attractions**
  Hoogendoorn and Bovy (2004), Seneviratne and Morrall (1985) denote that the importance of this attribute is related to the trip purpose. A person visiting a city is more triggered by attractions on the route than a person going to work. During a large scale event, like SAIL, there are multiple attractions (ships, stages, performances) which attract a great number of visitors. The existence and the variety of these attractions can influence the activity and route choice behaviour of pedestrians.

- **Weather**
  According to a study by SBB in Switzerland (Thurau, 2013), weather conditions influence the behaviour of passengers waiting on stations. More specifically, temperature has higher influence than rain. The influence of the weather conditions depends on the existence of protection against the weather conditions. Weather protection can be related to the level of comfort that travellers experience. The level of comfort influence the route and activity choice. If a route or the location of activity offers protection against weather conditions, it is expected that this route or this location will be more often chosen than a route or activity which is exposed to weather condition. It would be interesting to investigate if this statement can be applied for the case of large scale events.

- **Directness**
  A route or an activity location can be classified as direct when a pedestrian can reach to his/her destination walking on a straight line without being hindered by obstacles or by the interaction with other pedestrians. Helbing (1997) was addressed to this attribute stating that the order in which activities are performed generally depends on the directness, implying that some activities can be performed only once others are completed. Directness is could be related to visibility and number of turns in a route.

### 2.3.3 Event factors

The event factors are related to: the type of the event, the provided information during the event and crowd management measures.

- **Event characteristics**
  Type of event (indoor, outdoor, free entrance) and the type of activities that can be performed during the event are the main characteristics of the event. The majority of the activities taking place on an event start and end on a specific time. Thus, in case that someone wants to attend one of these activities should set a priority on the activity choice which is restricted by time. However, during the SAIL event, with some exceptions, the main activities (visiting/watching tall ships, drinking /eating) were open to the public during the whole day (09:00-23:00).

- **Existence of information and crowd management measures**
  During mass events, multiple measures are implemented in order to give information about the main attractions, routes crowdedness, etc. Zomer (2014) found that these information measures can have high influence on route choices of pedestrians and that they could even control the crowd by steering or spreading it over the network. Next to that, most of the times, during events there is an organization or a crowd management team that gives certain directions to the crowd.
Therefore these directions have major influence on the route and activity location choices of pedestrians.

2.3.4 External factors
The external factors that influence choice behaviour of pedestrians are related to environmental influences, land use of the event area

- Natural environmental factors
  According to Bafatakis (2014), and Hill (1982) the existence of grass, trees as well as the presence of canals, rivers can influence the route choice. This is related to the pleasantness that the aesthetic of the route’s environment can offer to pedestrians. In CROW (1998), it is stated that as routes become more attractive and pleasant, walking time becomes a less important factor. This statement would be interesting to be explored in this master thesis. Furthermore, noise and air pollution, day or night, weather conditions (Bovy & Stern, 1990) could also influence pedestrians’ choices in a large outdoor event. In the case of SAIL event, no noise or air pollution were present to such an extent that could exert influence on the choice behaviour of pedestrians.

- Land use of the event area
  In most of the cases, events take place in areas where the land use is not only dedicated for the organization of the event. This means that the event area has multiple land uses during the event period. Along the event area, there could be shops, offices, houses etc. For example in the case of the SAIL area, along the part of the event area there are houses. This implies that during the event, there would be pedestrians with different trip purposes depending on the land use. In other words there would be pedestrians that are crossing the SAIL area in order to arrive at their homes (inhabitants), while at the same time there are pedestrians that are visitors of the event. Thus, this factor is related to the interaction of pedestrians with different trip purposes.

In this section, the factors that influence the choice behaviour of pedestrians are identified. The relationships between the influential factors and the choice aspects (activity, route choice) during events are mentioned (Table 2.1). Due to the long list of the potential influencing factors a selection of these factors will be done in Chapter 3.

2.4 Relationships between pedestrians’ choices
After identifying the factors that influence the choices related to the activity and route choice (Section 2.3) during mass events, it is necessary to investigate the relationship between the choices that pedestrians make during events, focussing mostly on the activity and route choices. No theoretical framework focusing on pedestrians’ choices during events and capturing the relationships between the choices was found in literature.

Nevertheless, several theoretical frameworks derived from pedestrian behaviour studies describe the relationships between the choices are reviewed by means of the state of the art and are presented in Section 2.4.1. Furthermore, the adaptation of these frameworks to the case of pedestrians’ choices during events will be discussed. The investigation of the relationships derived from literature would set the stimuli for a preliminary design of a theoretical framework for pedestrians visiting large scale events (Section 2.4.2).

2.4.1 State of the art of the relationships between pedestrians’ choices
Some of the theoretical frameworks found in literature that introduce a relationship between the choices of pedestrians are elaborated below. More detailed information about these frameworks are given in Appendix A.
Theoretical framework of Root & Recker (1981)

Root & Recker (1981) introduced a theoretical framework of decision taken which consists of two phases: the pre-travel phase and the travel phase (Figure 2.1). In the pre-travel phase the activities are chosen and an activity program is planned. The decisions in this phase influence the choices made during the travel phase. The travel phase starts with the execution of the first activity included in the activity program.

This framework suggested by Root & Recker (1981) provides useful insights into the relationship between the activity choices and the planned activity program before the trip. The planning of the activity sequence before travelling can be applied to the case of large event with the assumption that a pedestrian has knowledge on the spatial and temporal availability of activity sites for each activity. This means that the pedestrian is well informed about the event area, the available activities, the locations where each activity can be performed, the activity duration. The adaptation of the activity schedule during the trip due to unforeseen situations, it can be suitable for the event case. For example crowded activity sites can cause delays resulting in unexpected increases in activity duration or sometimes in reality the activity is more entertaining than expected so pedestrian may prefer to spend more time to this activity.


Hoogendoorn, et al (2001) introduce a framework that shows different levels of pedestrian behaviour (Figure 2.2). These levels are comparable to the levels introduced by Root & Recker (1981). Hoogendoorn, et al (2001) only distinguish three levels instead of two: strategic (pre-travel phase), tactical (travel phase) and operational (travel phase). Daamen (2004) also adapted this framework for her research.

The framework of Hoogendoorn, et al (2001) offers interesting insights for this research. This thesis is focuses on the tactical level. According to this framework, the location choice of an activity is decided while walking at the event area, together with the route choice. Considering the main assumption of normative theory that pedestrians maximize the utility (simultaneous choice of optimal activity pattern and path by optimizing the utility of the activities and minimizing the walking cost of the activities), it can be said that normative choice theory will not fully cover real-life human choice behaviour and more specifically the behaviour of pedestrians during large scale events. It does however provide a very
convenient framework for modelling human decision making. The elements of the described theoretical framework could be adapted for the case of pedestrians’ behaviour during event and could serve as ‘building blocks’ for the design of a theoretical framework for pedestrians during events.

**Theoretical framework of Hoogendoorn & Bovy (2005)**

Hoogendoorn & Bovy (2005) also introduce a simultaneous choice process between route choice and activity location choice (Figure 2.3). The main difference in this framework compared to the framework of Hoogendoorn, et al (2001) is that the strategic level decisions (activity choice) influence the decisions made on a tactical level but not the other way around. In Hoogendoorn, et al (2001) the strategic level decisions were also influenced by decisions on a tactical level (feedback relation). This could mean that the experience gained during that trip is taken into account for the next trip.

This feedback loop for the case of pedestrians that are visiting an event means that next time that they will visit the same event they will have experience. However, this means that the layout and the activities related to the event will remain the same, which is not always the case.
**Theoretical framework of Ton (2014)**

Ton (2014) suggests two theoretical frameworks; one for familiar departing passengers in train station and another one for unfamiliar passengers. The structure of this theoretical framework is based on the theoretical frameworks introduced by Root and Recker (1981) (pre-travel, travel phase) and Hoogendoorn et al. (2001) (levels of decision: strategic, tactical). Thus, these frameworks include decisions made on two levels: before trip and during trip. The familiar passenger knows the station and can make plans while for the unfamiliar passenger is not the case. Therefore the familiar passenger plans all aspects of his trip beforehand. Then he adapts the sequence of activities, route and locations of activities when needed. Furthermore, the familiar passengers can make a simultaneous choice for route and location of activity. On the other hand, the unfamiliar passenger is only able to plan the activities, assign importance to these activities and organize a basic activity sequence before the trip. The route choice and the location choice are not simultaneously determined. During the trip the actual activity sequence, route and locations are determined.

The theoretical frameworks for unfamiliar and familiar departing passengers in train stations suggested by Ton (2014), could be adapted for the needs of this thesis. This can be done by considering a possible correspondence of the two different groups of passengers in stations with the two types of visitors in large scale events; the ones that have a predetermined way on making choice and the others that have an intuitively way (Section 2.3).

Taking all these frameworks into consideration, it can be said that each of them provides interesting insights into pedestrian’s choice behaviour and can be used for the conception of the theoretical framework for this thesis.

**2.4.2 Proposed Theoretical framework**

In Section 2.3, two types of pedestrian are distinguished regarding their choice behaviour during an event; the ones that have a predetermined way on making choices and the others that have an intuitively way.

It is evident that these two groups of pedestrians behave in a significantly different way on the decision process. Hence, a different approach should be identified for each group.

The pedestrians that have a predetermined decision are able to plan their choices more precisely than a pedestrian with an impulsive behaviour. To be more specific, they know beforehand whether they have preference to perform an activity or not. Conversely, pedestrians with impulsive behaviour cannot do this at the same extent because they might not have sufficient knowledge or they have a different state of mind and they are not interested in planning their potential choices. Basically, for pedestrians with an impulsive behaviour the performance of activity is decided while travelling (intuitive choices). That decision process would lead to a completely different framework from the ones with a pre-determined decision.

In the remainder of this thesis, the research is based on revealed preference data, which makes difficult to distinguish whether a decision regarding the performance of activity was predetermined or intuitive. However, in this research it is assumed that pedestrians know beforehand if they want to perform an activity or not. Consequently, the theoretical framework which will be presented, focuses on the pedestrians that have predetermined to perform an activity.

Although pedestrians have a predetermined decision regarding the preference of performing an activity, they cannot create a route or determine the location of the activity before the trip. Therefore, the route and activity location choice are determined during the trip. In Figure 2.4 the theoretical framework for pedestrians with predetermined decision on the performance of activity, is shown.
This framework is split into two parts; before the trip and during the trip.

Before the trip, a decision is made on which activities will be performed in the event. These decisions are determined according to trip, personal and event factors (inputs). The output of the activity choice is the set of the activities that can be performed during the event. The derived activities choice set is being processed based on personal, trip and event factors and as a result the relative importance of each activity in the activity set is determined (priority). The next step is to make a sequence of the activities planned. Usually, pedestrians that participate in large scale events may be uncertain due to their limited knowledge on the area topology. Hence, by knowing the activity choice set and by assigning a weight of importance to each activity, the process of determining the sequence of the activities starts. In parallel, the activity sequence is influenced by personal, trip, location and event factors. The output is the sequence of activities in the activity set with the relative importance of activities in mind. This output serves as input for the decision made during the trip.

The activity sequence is updated during the trip if the situation changes or if new information is gathered by visiting the event. This implies that there is the possibility of adding activities which were not known before the start of the trip. However, due to uncertainty, this will only be done if there are no constraints (time, budget etc.). The activity sequence gets input from the planned activity sequence and the available information during the event (maps). Pedestrians during the event carry a map of the SAIL event where the location of activities are illustrated. The location of activities influences the next activities to be performed. This can for example be expressed as skipping an activity because the location is too far from the final destination and there is not enough time available. The factors that influence the activity sequence are personal, trip, location, external and event factors. The output produced by this decision is the adapted activity sequence.

The next decision is a joint decision regarding their preference to perform an activity and route choice. By choosing an activity area consequently means that there is a preference to perform an activity. The factors that influence this decision belong to the following groups: personal, trip, location, route, external and event factors. Thus it can be said that the output of the decision on route choice is the route taken when executing the activities in the activity sequence on the chosen locations.

The joint activity- route choice provides a dotted feedback line into the activity choice of the next trip. However, it is not known how and whether the visitors of large scale events will use this feedback. This means that the information and knowledge gathered from the time a pedestrian visited the event may be useless till the next time that he/she visits the event due to changes in the configuration and the layout of the event area. At this point, it should be considered the fact that some large scale events take place once a year or even less often (SAIL event takes place every 5 year) implying that the configuration of the network and the event area may have changed over this period of time.

The relationships identified in the theoretical framework form the basis for the quantitative study that will be tested to which extent each factor influences the choices made during an event. However the theoretical framework covers a lot of relationships all of which cannot be tested in this research. This is due to time constraints and limitation from the data collection method (revealed preference data) which cannot capture all the described relationships of the theoretical framework. Therefore a selection of the relationships concerning activity and route choice should be made. This selection is done in the next Chapter (Chapter 3).
2.5 Discrete choice modelling

This thesis focuses on discretionary pedestrian choices, thus the literature review on pedestrian choices serves this direction. Several pedestrian modelling approaches which describe discretionary choices such as route choice behaviour have been proposed and tested in literature.

Discrete choice models (Ben-Akiva & Lerman, 1985)) have been widely applied in the context of travel decisions (Ben-Akiva & Bierlaire, 1999). These models are disaggregate in nature and are based on random utility theory. According to literature random utility modelling appears to be the common way to model pedestrian choice behaviour.

More specifically, a decision maker $n$ is considered the one who chooses among a set $C_n$ of $J_n$ alternatives. It is assumed that the decision maker $n$ attaches a utility $U_{in}$ to each alternative $i$ within $C_n$ and selects the alternative corresponding to the highest utility (utility maximization). Another example of the use of utility is regret minimization (e.g Chorus, et al., (2008)). The utility is modelled as a random variable to account for uncertainty due to various issues, including unobserved variables and measurement errors. The utility is decomposed into a deterministic part $V_{in}$ and an error term $\varepsilon_{in}$ so that

$$ U_{in} = V_{in} + \varepsilon_{in} $$

And the probability that individual $n$ is selecting alternative $i$ is

$$ P_n(i|C_n) = P_r(U_{in} > U_{jn} \forall j \in C_n) $$
Based on the explicit specification of $V_i$ and the distribution assumptions of the error term $\varepsilon_{in}$ choice models can be derived. Starting with the specification of $V_i$, it depends on the selection of the explanatory variables, i.e. the attributes of each alternative $i$. The complexity of the model can be determined by the distribution assumptions of the $\varepsilon_{in}$. The most widely used model is the logit model (the basic model is the binary model), which assumes that $\varepsilon_{in}$ are independent across both $i$ and $n$, and identically distributed with an extreme value distribution leading to a simple/easy formulation. Thus the model imposes constraints related to the covariance structure (Ben-Akiva & Bierlaire, 1999). The result is that many other more complex logit models (multinomial logit, nested logit, cross-nested logit, mixed logit) have been developed in order to relax these restrictions that may be unrealistic in some contexts.

The models that can be used for the estimation of route and activity location choice behaviour are briefly introduced below:

**Multinomial logit (MNL):** The multinomial logit model is the extension of the binary choice model, because it considers more than two alternatives. It assumes a linear and additive utility function. The model also assumes that the random error term is independent and identically Gumbel-distributed. In case of correlation between the alternatives, the model will not provide realistic probabilities. The best example to illustrate this is the red/ blue bus paradox (Ben-Akiva & Lerman, 1985).

The probability that alternative $i$ is chosen within choice set $C$ is given by:

$$P(i|C) = \frac{e^{Vi}}{\sum e^{V_j}}$$

**Nested logit (NL):** The nested logit model is a common solution for the problem of correlated alternatives. The alternatives are divided into nests. Within a nest correlated alternatives can be present, because the utility of a nest is based on the utility of each member of the nest. For the red/ blue bus paradox would mean that these two buses are in the same nest. The random error term is assumed to be independent and identically Gumbel-distributed. The model does not allow for correlation between the nests.

The probability that alternative $i$ is chosen within a choice set $C$ is given by

$$P(i|C) = P(C_n|C)P(i|C_n)$$

Where $P(C_n|C) = \frac{e^{\mu VC_n}}{\sum e^{\mu VC}}$ and $P(i|C_n) = \frac{e^{\mu Vi}}{\sum e^{\mu V_j}}$

**Cross-nested logit (CNL):** The cross-nested logit model is a direct extension of the nested logit model, which resolves the issue of allowing only one nest and thus allows alternatives to be members of multiple nests.

**Mixed logit (ML):** The mixed logit model can be derived for many different behavioural specifications. Each derivation provides a different interpretation of the model. Basically, the mixed logit probability is a weighted average of the multinomial logit function evaluated on different values of the parameters. More information about the mixed logit model is referred to Hensher & Greene (2003).

To sum up, this section provides a brief introduction to discrete choice modelling. In this research, discrete choice modelling aim to provide insight into the route and activity location choice behaviour of pedestrians in SAIL event. Decisions like route choice and activity choice are mutually exclusive. This means that only one option can be chosen (Ben-Akiva & Bierlaire, 1999) which makes discrete choice analysis a suitable tool for the analysis and the prediction of the decision making in events.
2.6 Conclusions
This chapter addressed the literature review on the choice behaviour of pedestrians during large scale events. Most of the available literature deals with pedestrians’ choices in facilities like public transit station, urban walkway and shopping mall, while literature on pedestrians’ choices in large scale events hardly exists.

Hence, the factors that influence the choices of pedestrians during large scale events were identified based on the state-of-the art on pedestrian choice behaviour in combination with observations and experience derived from large scale events. The factors were categorized in the following four groups: personal, system, event and external factors.

Several theoretical frameworks derived from pedestrian behaviour studies described the relationships between pedestrians’ choices and the influential factors. Each theoretical framework gives an element for the creation of the theoretical framework of this thesis. The pre-travel phase and the travel phase of Root and Recker (1981) as well as the adaptation of the activity schedule during the trip due to unforeseen situations can be considered for setting the theoretical framework of pedestrians during events. The normative theory of Hoogendoorn et al. (2001) gives the idea of modelling the activity and route choice jointly since these choices are interrelated. In addition, the theoretical frameworks for unfamiliar and familiar departing passengers in train stations suggested by Ton (2014) presents a lot of similarities with the 2 types of pedestrians during the event, the ones with a predetermined decision and the others with an intuitive behaviour.

Lastly, discrete choice models can be used for the quantification of the relationships between the influencing factors and the decisions like route choice and activity choice. Among the discrete choice models that are identified (multinomial logit (MNL), nested logit, cross-nested logit and mixed logit model), the MNL is used in this thesis due to its simplicity. The estimated MNL models are presented in Chapter 6. It can be said that discrete choice modelling is a suitable tool for the decision making in events.

The factors that are identified in Section 2.3 constitute a long list which not all of them will be tested on the extent to which they can influence the activity and route decisions during large scale events. The selection criteria will be presented in Chapter 3.
3 Selection of under-study choice behaviour using revealed preference data

In the previous Chapter (Chapter 2) a theoretical framework for visitors (pedestrians) during large scale events is proposed. In this theoretical framework 5 types of choices can be identified: choice of activities, activity hierarchy, activity sequence, activity location choice and route choice. It is seen that some of these choices (activity choice, activity hierarchy, planned activity sequence) are made before trip while others (updated activity sequence, activity location and route choice) during trip. Various factors (personal, trip, system, external, event factors) exert influence on the aforementioned choices. The data collection method was determined before the start of this thesis, in the context of developing a crowd monitoring system. Thus, the available data for this research are revealed preference data that were collected during the SAIL event.

The aim of this chapter is to assess the revealed data based on their capability to capture the factors that influence the activity and route choice behaviour during large scale events. On that way, a selection of the list of the potential influencing factors that were identified in Section 2.3 is done.

In literature several revealed data collection methods are used to capture real life information (what people actually did, chose or experienced). An overview of these findings is described in Section 3.1. Next, Section 3.2 addresses in the revealed data collection methods used for this study. These methods are assessed regarding their suitability for the research objective of this study. Thus, the following research question will be discussed:

4. To what extent is the Revealed Preference data a suitable data collection method for collecting quantitative data on route and activity choice behaviour of pedestrians in large scale events, like SAIL?

In Section 3.3, the collected data are being examined on their capability to capture the choices that are identified in the theoretical framework and the influential factors. Thus, a selection of the factors that influence the activity and route choice behaviour of pedestrians for this study is made depending on the usability of the collected revealed data. Finally conclusions are derived in Section 3.4.

3.1 Overview of revealed preference data collection methods

Revealed preference data are used to capture the actual choices of people (real life information). Revealed preference data (RP) collection method focuses on the investigation and interpretation of the visible motion behaviour.

Several revealed preference (RP) datasets have also been used in the literature to analyse choices. Some of the collection methods that were applied to derive revealed preference data are discussed below.

Direct observation

A common method used in pedestrian behaviour surveys is direct observation, also known as behavioural mapping or “tracking”. Direct observation is distinguished into participatory and non-participatory. Participatory observation involves the observer taking part in the participant’s activities, where the researcher can gain insight to the purposes influencing the subject’s decisions. Nonetheless, this technique faces issues of reliability; knowing that they are under observation, participants may adapt their behaviour to researcher’s expectations. Non-participatory observations avoid the risk of such “observer effects”. These unobtrusive observations have been used for studies concerning the movement behaviour of visitors in museums and exhibitions (Hill, 1984). Tracking involves following the pedestrian at a distance and recording his/her movements by drawing a line corresponding to his/her activities on a map of the investigation area. This method yields information concerning the exact routes and activities...
of pedestrian in urban environments in time and space, while avoiding the risk of influencing the “free” behaviour of participants (Hill (1984), Keul & Kühberger (1996), Keul & Kühberger (1997)). A major drawback of this technique is that it is very time-consuming and labour intensive, and findings are limited to pedestrians’ activities that can be observed since motives and intentions cannot be revealed.

**Survey**

Apart from stated preference data¹, a survey can also provide revealed preference data when the researcher asks people what they did, chose or experienced. Seneviratne & Morrall (1985) asked people to indicate why they chose the route they used (revealed preference). Hill (1982) also used this method to collect data on route choice. A survey can capture much more than only the route choice of a person, for example also the socio-demographic information of each person. However, errors arise when asking people about the route or activities they performed. People tend to adapt their answers – consciously or subconsciously – to what they expect to be socially desired behaviour (Esser, 1985). Another issue could be the fact that pedestrians could not remember/recall the exact route or activities. In other words, it is strongly dependent on individual’s memory. To face the problem of memory, real time trip diaries can be used which can provide detailed information. It helps investigating whether travellers have a good perception of what actually happened during the trip and how expectations are updated based on past/recent experiences. Nevertheless, self-administered diaries (which are written in real-time) require the individual’s cooperation to a very large extent.

**Video cameras**

Technologically sophisticated methods use video analysis to monitor and interpret pedestrian behaviour. Especially the development of agent-based simulation models requires calibration and validation to confirm the accuracy of simulated human behaviour (Daamen & Hoogendoorn (2003a), O’Connor, et al., (2005)). Most studies using video captured data are limited to a very small observation field. However, there are approaches observing a larger area by a network of several surveillance cameras (Millonig & Schechtner, 2007b).

Lee et al. (2001) collected data on the route choice of a pedestrian by means of video cameras. They observed movements of pedestrians through a station to deduce which route pedestrians took through the station. The same method was applied by Antonini et al. (2006). This method captures (nearly) everyone that passes the camera(s) and provides a visual image of the route taken. The camera footage of multiple cameras needs to be combined and the same person needs to be identified during the entire route. This configuration is very time consuming. Next to that, mistakes can be made when combining the cameras. Human error increases with the use of multiple cameras. Also, cameras on different angles capture different images. The method needs a translation to data that can be easily interpreted. The configuration and translation of the camera footage into usable data is very time consuming and prone to human errors. Moreover, only visible behaviour can be investigated, leaving the individuals’ intentions and motives as well as most other personal characteristics unknown.

**Photos and Video time lapse movies**

Helbing et al. (2002) also made use of photos of multiple locations. Photographs of an area can be used to identify the density in that area at several moments in time. A photo shows the actual density or even the self-organization effects occurring in pedestrian crowds, but it needs a lot of time for the processing of the collected data. Due to the level of details that a photo can offer, it is almost impossible to identify the route and activity choices of each individual.

¹ Stated preference data are based on responses to hypothetical travel situations in a survey context. Stated preference data are mainly used to capture information and data about future or possible situations, which often cannot be captured in real life. This can therefore be defined as laboratory setting.
Digitally based localization technologies (GPS, Wi-Fi, Bluetooth)

Spatiotemporal behaviour of pedestrians can be covered by digitally based localization technologies (Shoval & Isaacson, 2007), (Spek, 2007), (Svetsuk, 2007). These include for example GPS (Global Positioning System), Bluetooth tracking and Wi-Fi sensors. All these technologies track the movements of pedestrians. Depending on the accuracy of each technology, the level of detail may vary. The type and location of activities, the sequence of the performed activities and the routes that pedestrians follow can be captured by these technologies. Another advantage of using tracking technologies lies in the possibility to gather data within a very large study area. On the other hand, these tracking technologies can be of a rather invasive nature, if the participants have to be equipped with tracking devices, and therefore, again, researcher effects may be suspected. In addition, apart from (GPS, Bluetooth) devices, a smartphone can produce all these sources of data (GPS, Wi-Fi, Bluetooth tracking) when these functions are enabled. However, the use of data gathered from private mobile phones without knowledge of their owners may pose various ethical questions. Apart from these issues, there is also the weakness of capturing the underlying reasons of individuals’ choices.

Social media (Facebook, Twitter, Instagram, Foursquare)

Several studies (Quercia & Sáez-Trumper (2014), Cranshaw, et al. (2012), Del Bimbo, et al. (2014), Noulas, et al. (2011)) share in common the potential of easily accessible and geo-localized social media data (Twitter, Foursquare) in providing scalable solutions for exploring varied city dynamics. They all argue for the limitations traditional urban data gathering methods have in providing relevant insights at very large scales. These studies focus entirely on content stemming from a single data source. In the study of Psyllidis, et al (2015), insights from several – municipal, sensor- and web-based – sources were combined. The inherent diversities in social (media) data were considered, by taking into account the various geographic, demographic, and contextual biases, instead of following a rather uniform approach.

According to Psyllidis, et al. (2015), understanding the complexity of urban dynamics requires the combination of information from multiple city data sources. Besides traditional urban data, geo-localized social media provide human-generated content, which may reflect in (near) real-time the activities people undertake in cities. The integration and analysis of such heterogeneous sources of information gives the opportunity to explore, monitor, and visualize urban dynamics. It further shows how the inherent spatiotemporal, demographic, and contextual diversities of social data influence the interpretations of (dynamic) urban phenomena.

Taking all these collection methods that were applied to derive revealed preference data into account, it is seen that decisions taken during the trip (activity sequence, activity location, route choice) can be (partly) identified by all of these methods. Nevertheless, the underlying reasons of each choice, the motivations or the intentions of each individual is not possible to be captured. Furthermore, with the use of revealed datasets, it cannot be tested whether the activity hierarchy and the planned activity sequence before the trip is the same as during the trip and how the activity sequence is adapted to the prevailing conditions during the trip. Capturing these choices would need a different data collection approach. Another limitation of this method is that it is not possible to examine how changes in the conditions could affect the choice behaviour. This means that the type of RP data that has been usually collected are of limited use for the assessment of behavioural changes over time (Moraes Ramos, et al., 2012). These limitations should be taken into consideration during the data analysis part.

3.2 Revealed preference data collection methods used during SAIL

After discussing the capabilities and the limitations of the data collection methods that can be applied to derive revealed preference data, in this Section a description of the collection methods that were applied to this study is presented. In this study, as it was mentioned before, a pilot has been performed to develop a crowd monitoring dashboard for the municipality of Amsterdam, so the data collection was
predetermined before the start of this thesis. For this reason, real-time data have been collected using different kinds of sensors: counting cameras (equipped with Wi-Fi sensors), Wi-Fi sensors and GPS trackers.

**Wi-Fi and camera sensors**

Wi-Fi sensors and counting-cameras (which include Wi-Fi sensors) were placed at the area were the SAIL event took place. The position of each sensor was determined by DAT. Mobility and TU Delft based on the location of important diverging/merging points in the network and identified or potential bottlenecks. Apart from the aforementioned requirements for the position of counting-cameras and Wi-Fi sensors, some other requirements were considered. To be more specific, the positions of counting-cameras should enable the distinguishing between different modes (car, bicycle, pedestrians) and the derivation of pedestrian flows for both directions. Wi-Fi sensors should be able to cover the cross section of the street and track the pedestrians that have their Wi-Fi function enabled. In Figure 3.1 the locations of each sensor (8 counting-cameras, 15 Wi-Fi sensors) are depicted. Due to long distances between the sensors and the existence of shortcuts in the network between the positions of consecutive sensors, analysing the choice behaviour of pedestrians implies a constraint on the level of detail for the analysis. Walking between two consecutive sensors does not indicate the route that a pedestrian followed from the one sensor to the next. This means that for some parts of the SAIL area is difficult or even impossible to derive the exact route that a pedestrian follows as well as the intermediate stops where a pedestrian could perform an activity (activity location areas). Below the potentials of the Wi-Fi sensors and counting-cameras used in the area are shortly-described separately.

**Wi-Fi sensors**

Regarding the Wi-Fi sensors, as it was aforementioned, they can only track the pedestrians that have their Wi-Fi function enabled. However, the reconstruction of the possible routes taken by the Wi-Fi users is disputable due to long distances between the Wi-Fi sensors. It can only be ascertained that the location of an individual is within a range of positions. Using several closely-located Wi-Fi sensors could help triangulate a more precise location. This implies that many sensors would be needed in order to track a person over a long path, which makes the process costly. However, Wi-Fi data could be used for the estimation of crowd density (when no cameras are available) as well as for the acquisition of travel time information. A pair of Wi-Fi sensors collects information about the unique ID of a Wi-Fi device \(i\) (the MAC address) and a timestamp \(t\). By matching the IDs of the devices with distinct sensors, the time it takes the device to move from the one sensor to the other can be easily determined. These times are referred to as area presence times of device \(i\). It should be denoted that the set of area presence times was collected during some time intervals in which the traffic conditions were more or less similar. An observation of a Wi-Fi device then corresponds to the observation of a pedestrian.

**Cameras**

The cameras, are able to count the number of pedestrians at given location and time. Except for the density which can be derived from camera data, information on the route choice of pedestrians can be given. Considering the long distances between the cameras (Figure 3.1) in some parts of the network, it is obvious that it is difficult to deduce the routes that pedestrians took, by observing their movements. Furthermore, the type of activities and the activity locations that a pedestrian choose can be captured only for some specific parts of the network (only locally). For example if someone is first located at camera 1 and then again at camera 2, it is not known what he/she did in between and which route he/she followed (Figure 3.1). In case that camera data collection will be used, the camera footage of multiple cameras needs to be combined and the same person needs to be identified during the entire route. This configuration is very time-consuming and prone to human errors, especially for the case of mass events.
Figure 3.1 Overview of sensor locations in Google Earth. Orange numbers were given to the cameras for the need of the example of walking between two cameras

GPS

The GPS devices that were used in this study are manufactured by My GPS Tracker B.V. The manufacturer states that position measurements are taken every 5 seconds and uploaded in real time to a server, outputting location estimates with a 5 meter horizontal accuracy. However, little it is said about the impact of the surroundings (building height, population density, amongst other factors) on the accuracy of the devices. Furthermore, the detection rate of stops and movements on very low speeds cannot be captured.

110 GPS trackers were available to be handed out to the visitors of the SAIL event. The standing spot for the GPS distribution was Amsterdam Central Train Station. During the event, a team of graduate students and staff of TU Delft and the AMS Institute was responsible not only for the distribution of the devices but also for checking their function and ensuring a connection between the computer and the GPS signal. The visitors that participated in the GPS data collection had to fill a form with some personal details (age, gender, nationality, if they are alone at the event or part of a group) in order to extract some socioeconomic characteristics of the sample. Thus, during these days, a survey file was created including the GPS ID given to each pedestrian-participant, some personal details of the participants and their departure and arrival time at the standing spot.

With the exception of the first day of the event, where the GPS trackers did not work due to some technical errors at the server, the next 4 days a total number of 322 devices were handed out. This means that there are GPS measurements from 322 pedestrians which capture pedestrian tracks and can be used to identify movements made by pedestrians. The choices that can be identified by means of GPS data are the routes that pedestrians took during the event and whether an individual performed an activity or not (on a binary way) at the area. However, the exact location of the activities is not possible to be identified, since not only the accuracy of the GPS trackers (5metes) is not sufficient enough for this purpose but also it is not possible to capture the stops that a pedestrian does while walking. Stops could indicate the existence of activity. Instead of the exact activity location, the GPS data can indicate the activity area where an activity is performed.

Thus, taking into consideration the potentials and the limitations of the RP data collected during the event, an assessment on their capability to capture the choice behaviour and the factors that influence the choice behaviour of pedestrians during the event will be presented in the next Section (Section 3.3).
3.3 Usability of data on selecting the choice behaviour for this study

In this Section, an assessment on the usability of collected data is performed regarding their capability to capture the choice behaviour and the factors that may influence pedestrian’s choice behaviour during mass events.

Due to long distances between the positions of each pair of the Wi-Fi sensors and cameras sensors, the level of detail for the analysis of pedestrians’ choices is low. This means that the activity locations choice as well as the exact path that pedestrian chooses to follow is difficult to be captured.

As for the GPS devices, they are not suitable for capturing the exact activity locations choice due to horizontally accuracy issues (5 meters) and the weakness of recognizing stops. However, GPS data collection method is able to reconstruct the route choices and provide with travel information. The travel information gives insights into the performance of activity or not by comparing the time it takes to walk along a specific part of the network with the time spent at this part. If the time spent is higher than the walking time of a specific part of the examined network then apparently it can be said that an individual performed an activity\(^2\) at this area (activity area). The processing of the GPS data in order to create routes and extract travel time information will be presented in Section 5.3.

In Section 2.3 the factors that are observed in literature and practice and are relevant to the activity and route choice behaviour of pedestrians during mass events were identified (Table 2.1). However this list is quite extensive. Therefore, these relationships are evaluated based on the possibility of being derived by the collected data. The relationships can be captured by means of the data collected if the influencing factors can be captured. This provides a practical assessment of the data collection methods and their usability. Figure 3.2 illustrates the approach for the selection of the factors that will be used for the quantitative research.

Figure 3.2 Flowchart with selection criteria of factors

\(^2\) In this study, activity is defined whatever action besides walking. In order to perform an activity, it is required first a stop and then the performance of activity. For instance, a pedestrian stops walking in order to buy a coffee or take some photos
The output of this flowchart gives the factors that can be captured by the collected data and will be researched in this study. These factors are shown in Table 3.1. This selection is the first step to identify the factors that can be quantified using the available data. However, the list of this factor can be diminished more based on the availability of data on specific areas of the event.

Table 3.1 Factors that can be captured using the collected data (in grey are the factors that cannot be captured by RP data collection methods or other available tools)

<table>
<thead>
<tr>
<th>Factors</th>
<th>RP data collection methods</th>
<th>Wi-Fi</th>
<th>Counting-cameras</th>
<th>GPS</th>
<th>Survey (with personal details)</th>
<th>Other tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Familiarity</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Emotional State</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Decision style (Herding)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Area presence time (walking time+activity time)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group composition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impulse behaviour</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Time of day or week</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimension of the location (width)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Google Earth</td>
<td></td>
</tr>
<tr>
<td>Visibility/directness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Google Earth</td>
<td>Network configuration</td>
</tr>
<tr>
<td>Amount and type of activities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SAIL program</td>
<td>Direct investigation of the area</td>
</tr>
<tr>
<td>Walking distance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Direct investigation of the area</td>
<td>Google Earth</td>
</tr>
<tr>
<td>Crowdedness</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of attractions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Direct investigation of the area</td>
<td></td>
</tr>
<tr>
<td>Existence of signs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Photos</td>
<td></td>
</tr>
<tr>
<td>Weather protection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Direct investigation of the area</td>
<td></td>
</tr>
<tr>
<td>Program of the activities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Program SAIL</td>
<td></td>
</tr>
<tr>
<td>Weather conditions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Historical weather databases</td>
<td></td>
</tr>
</tbody>
</table>

Area presence time and the time of day or week can be captured by GPS, Wi-Fi and camera data. The crowdedness can be estimated by using the densities over time which can be derived either by using camera data or Wi-Fi sensors.

Familiarity, emotional state, decision style as well as impulse behaviour cannot be captured by the available data. Another data collection approach should be applied for the derivation of these factors.

However, herding behaviour (it belongs to the decision style) which is related to the tendency of pedestrians to follow the crowd, may be related to the crowdedness. This implies that pedestrians during events may not think about the routes and activities thoroughly but they tend to do what the crowd does. This behaviour cannot be captured by the available collected data. Nevertheless, it can be interrelated with the influence of crowdedness on the route and activity choice behaviour.

---

3 In this thesis, activity duration time is the time spent on everything besides walking (watching an event, spending time buying a drink, eating, socializing, resting etc.)
Personal characteristics like age, gender, group composition are derived from the form that the participants of the GPS data collection filled in.

Others factors like areas’ dimension, amount of activities, number of attractions, existence of signs etc. can be captured by either looking at the map or by direct investigation of the area. For the weather conditions the historical weather data can provide useful information.

Considering the aforementioned, it can be said that the GPS data collection method is more a suitable and more accurate way (in comparison with the other available collection methods; Wi-Fi, Cameras) to extract information on the spatiotemporal movements of pedestrians during mass events. In other words, the main advantage of the GPS is that it can generate pedestrian tracks in an easy and not time consuming way. These tracks are useful for the investigation and understanding of pedestrians’ route choice behaviour. Furthermore, GPS data is a useful tool to derive travel time information. However, the data collected by the other methods (Wi-Fi and cameras) in combination with the GPS data could contribute to the improvement of the analysis of choice behaviour by capturing factors like crowdedness.

For capturing the exact activity locations that a pedestrian chose, more detailed data are needed. To be more specific, sensors located nearby the major activities of the event would be needed in order to extract the activity location choices as well as the route choices. This process would introduce additional cost to the operation. Another solution would be the improvement of the GPS devices regarding their accuracy and their ability to recognize stop locations. Lastly the use of survey where people state what they did, chose or experienced could give a better insight into the route and activity choices. However, this method requires the individual’s cooperation and a good perception of what actually happened during the trip.

Given the capabilities and the usability of the available data sources, the choice behaviour that this thesis is focusing on is the choice of performing an activity which implies the activity area choice and the route choice.

3.4 Conclusions
The current Chapter addresses the way that the selection of the under-study choice behaviour of pedestrians was made by using revealed preference data. In literature several revealed data collection methods are used to capture real life information. Direct observation, survey/trip diaries, video cameras, photos and video time lapse movies, Wi-Fi, Bluetooth, GPS, Social media are the findings from literature

The advantages and the limitations of data collection methods (Wi-Fi sensors, counting cameras, GPS) used during the SAIL are summarized in Table 3.2.

<table>
<thead>
<tr>
<th>Data collection methods</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wi-Fi sensors</td>
<td>• Track pedestrians over space&lt;br&gt;• Estimation of crowd density&lt;br&gt;• Travel time information</td>
<td>Not so accurate reconstruction of the routes taken</td>
</tr>
<tr>
<td>Camera</td>
<td>• Visual image of route taken per pedestrian&lt;br&gt;• Estimation of crowd density</td>
<td>Multiple cameras need to be combined and the same person needs to be identified along the entire route → time-consuming, prone to human errors</td>
</tr>
<tr>
<td>GPS</td>
<td>• Generation of pedestrian tracks&lt;br&gt;• More accurate reconstruction of the routes taken than the ones captured by Wi-Fi sensors.&lt;br&gt;• Travel time information</td>
<td>• Limited number of GPS participants at the same location and moment&lt;br&gt;• The level of detail is limited to the activity area and not to the exact activity location.&lt;br&gt;• Missing measurements (due to signal)</td>
</tr>
</tbody>
</table>
The assessment of these methods are done based on their capability to capture activity and route choice behaviour of pedestrians in an accurate way.

It is concluded that the GPS data collection method is more suitable to extract information on route choices by capturing the spatiotemporal movements of pedestrians in comparison with the other available collection methods. In addition, travel time information which is related to the time spent on specific areas (area presence time) can be derived by the process of the GPS data (Section 5.1). However capturing the exact activity locations that a pedestrian has chosen demands more detailed data in comparison with the available data.

Depending on the research objective which is the activity and route choice, the practical suitability of the available revealed preference data to capture the factors that influence these choices is assessed. The quantitative analysis of the relationships between the influencing factors and the activity and route choices will be done in Chapter 6.

In this section, the most common RP data sources are identified and discussed. Such sources are examined in terms of their capability to capture the influential factors of activity and route choice behaviour during large scale events. Before proceeding to the RP data analysis during the SAIL event (Chapter 5) a general introduction to the case study is given in Chapter 4.
4 Introduction to the case study: Sail, Amsterdam 2015

An increasing number of large events are organised in the urban environment. An example is SAIL2015, Europe’s biggest free accessible nautical event, organised every 5 years on Amsterdam which attracts thousands of visitors. This year SAIL event took place during the period of 19th-23rd of August and around 2.3 million national and international visitors watched tall ships in the area around the IJ-port in Amsterdam.

For the municipality of Amsterdam, a pilot has been performed in order to develop a crowd monitoring dashboard. With this dashboard, information is generated and visualised as support tool for the operations during the event and for analysis afterwards. Real-time data have been collected using different kinds of sensors: counting cameras, Wi-Fi sensors and GPS trackers (Section 3.2). This thesis will use SAIL as a case study in order to analyse the activity and route choices of pedestrians by using revealed preference (RP) data collected during the SAIL event.

In Section 4.1 a description of the SAIL area is given, while in Section 4.2 a brief overview of the SAIL event related with the impression on activity and route choice behaviour during the event will be presented. Next in Section 4.3, before the process and analysis of the RP data some conjectures of the outcomes of the RP data analysis regarding pedestrians’ choice behaviour during the event will be presented based on the overview of the SAIL event atmosphere. Lastly, Section 4.4 discusses the conclusions.

4.1 Description of the SAIL area

A great number of sailing ships (tall ships) is in direct proximity to the city centre and more specifically on the IJ Lake behind Amsterdam Central Station. Apart from the tall ships the whole SAIL area is widespread through Amsterdam. The event area is divided into five so-called oceans (Orange, Blue, Green, Red, and White) according to its theme (Figure 4.1). The orange ocean is the heart of SAIL, ships can be visited via the quay and be admired by the water. On the north side of IJ Lake there is a stage constructed on the water which is the SAIL music arena. The Red ocean is the culture centre of SAIL-SAIL organization collaborates with museums and businesses in the centre of Amsterdam. The Green Ocean is located on the north side of the IJ at the shipyard, the White ocean refers to the area around the Eye museum and the Tolhuistuin. Lastly, at the Blue Ocean the Maritime museum is located (SAIL organisation, 2015).

![Figure 4.1 Coloured oceans of SAIL area (SAIL organisation, 2015)](image-url)
For the development of the crowd monitoring dashboard real-time data (Section 3.2) have been collected mainly from some specific parts of the Orange Ocean (Figure 4.1). Hence, the main focus areas of the pilot study of the municipality of Amsterdam are the following: the orange route (Ruijterkade, Veemkade, Piet Heinkade, Javakade, Sumatrakade), and the purple route (Ijplein, Noordwal) (Figure 4.2). At the Orange Route people can visit ships or even admire them by the water. Since real time data were revealed from the aforementioned areas, this thesis will also focus on the choice behaviour of pedestrians at parts of these areas and more specifically at parts of the Orange route.

![Figure 4.2 Part of the Orange Ocean (Orange and Purple route)](image)

4.2 Overview of SAIL event

In this section a general overview and description of the prevailing atmosphere from the perspective of the impression related to activity and route choice is presented.

In the first day of the event (19th of August) the tall ships entered from the quayside of the North Sea canal and the SAIL parade took place. Then the tall ships were moored along the Orange Ocean (Ijhaven). The next days (20th–23rd of August), the tall ships were open to public from 10am to 11 pm. During this 5-day event, several events were held along the orange route (live music concerts, activities for children, etc.). Moreover, every evening was ending up with a breath-taking fireworks display at the Java Island northern area, which was visible even from the Amsterdam train Station.

Regarding the prevailing conditions, having a walk around the SAIL area can give an impression on which areas were of great interest during these days and were mostly chosen to perform an activity. Beginning with Veemkade, it was observed that it was almost always crowded and the pedestrians could not freely walk at most of the length of the route (Figure 4.3). Bottlenecks were created because people were waiting in the queues either to visit the tall ships or have drinks/food at cantinas (Figure 4.4) which were occupying a large part of the road. Another interesting observation was that the flow of pedestrians from west to east direction was much greater in comparison with the one from east to west.
The Java Island was observed to be less crowded than Veemkade. People could walk almost freely at the majority of the parts of both Javakade and Sumatrakade. Sumatrakade was crowded only during the fireworks display late in the evening, since it offered great view. In general, at Javakade, people were waiting to visit tall ships, or were sitting on the grass or next to the sea in order to get rest. Furthermore, it was noticed that the majority of pedestrians at Sumatrakade were walking with direction from west to east, while at Javakade most of them were walking with direction from east to west (direction to the Amsterdam Central Train Station). Moreover, at Sumatrakade the use of pawns along the street enabled the separation of different modalities (Figure 4.5). In other words, not only pedestrians could use the road but also other modes (cars, bikes).

Behind the Amsterdam Train Station at Ruijterkade (Figure 4.6), it was almost always congested; the visitors were creating queues waiting for ferries to have a tour to the Green or White Ocean (Figure 4.1), while at the same time there were pedestrians who were moving straight ahead to the Orange route or were standing still and were watching the ships (taking photos, etc). Thus, the interaction between the passing pedestrians and the pedestrians waiting for the ferry made the free walking along this area impossible.
Following the purple route, pedestrians were able to walk unhindered by crowdedness at most of its parts. However, on Noordwal the route along the dike was crowded because people could watch and admire the ships from this location. Despite the fact that there were other alternative paths that were less crowded pedestrians tend to choose the route via dike, implying their interest in watching or even visiting the tall ships.

Some general remarks regarding the area which could influence or even determine the activity and route choice are presented below:

- Along the event area there were multiple signs (Figure 4.7), as well as an information point just before the divergent point of the Orange and Purple route, where the visitors can be provided with a map (Figure 4.8).
- Due to the crowdedness at some specific parts of the network, there was also the enforcement of crowd management team in order to alleviate the situation. More specifically, on the 22nd of August (early in the afternoon) due to crowdedness at Ruijterkade the pedestrians with direction to the Central Station were navigated via Kattenburg/Dijksgracht (Appendix D) while on the 23rd of August there was police and security enforcement outside the area of ferries in order to ensure that pedestrians are not standing still at this area.
- During the days of the SAIL event the weather was nice, a factor that contributed to the encouragement of the participation to the event.
4.3 Expected outcome of the case study

Based on the literature review (Chapter 2) and the overview of the SAIL event atmosphere (Section 4.2) conjectures are made regarding the expecting outcome of the revealed preference data that were collected during SAIL Amsterdam event 2015. The expectations are related to the activity and route choice of pedestrians during the event.

Expected Preference to perform an activity Since SAIL is a nautical event the major activities are located next to the sea and are related to the tall ships. Apart from these activities, the visitors have the opportunity to enjoy food/drinks or even go to music concerts. The majority of these activities are located at Veemkade and Javakade streets, where congestion is expected. Despite crowdedness, it is expected that pedestrians will have a desire to perform activities due to event state of mind and thus they will visit these activity areas. Furthermore, it is expected that visitors will spend more time visiting tall ships than having a take away food. It was observed during the event that the queues outside the tall ships were much longer than the ones in cantinas.

Expected route choices It is expected that visitors will prefer to follow the signs leading to the Orange route along the sea where most of the activities and attractions are. Considering that the main characteristic of SAIL event is visiting or enjoying the view of the tall ships, it can be expected that visitors will have a tendency to choose routes in order to be close to the attractions of the event. However, it is expected that when they decide to leave the event due to tiredness they choose routes which are not crowded and give them the opportunity to return back as fast and easy as possible. Lastly, on specific parts of the network where the crowdedness is dominant and a pedestrian is not interested in performing any activity he will try to choose a route less crowded and maybe of less interest from attraction point of view (choice of shortcuts for example).

4.4 Conclusions

The conclusions and observations of this chapter are based on the author’s experience at SAIL event. Some basic remarks are explained shortly below.

The main attractions of the SAIL event are the entrances of tall ships where queues were formatting. Furthermore the presence of queues was usual when people were waiting for the ferries to have a boat trip. Regarding the network, the majority of pedestrians has a clear preference on walking close to the attractions which were related to the event, despite the existence of crowdedness.
Another remark regarding the days that the event took place is that the weather conditions remained almost the same (sunny) during all days. Lastly, to the best of author's knowledge no panic situation happened.

It is interesting to see if these expectations will be verified by analysing the GPS data. This chapter provides a general impression on the activity and route choice behaviour during the SAIL event. This overview will lead on the formulating of the hypotheses of the data analysis plan in Chapter 5. The process and the analysis of the GPS data are presented in the next Chapter (Chapter 5).
5 Process and analysis of the GPS data

In Section 3.2 a short description on the way that the Revealed Preference (RP) data were collected during the case study of SAIL event 2015 was given. It is important to be stressed that this thesis elaborates more on the analysis of the RP data retrieved by the GPS trackers that were handing out during the event. The analysis of the Wi-Fi and camera data has already been done and some of the results of the analysis will be used in combination with the GPS data in order to investigate the choice behaviour of pedestrians during the SAIL event.

Section 5.1 presents the approach which is followed in order to process the raw GPS data to a manageable format and create a usable dataset for the investigation of pedestrians’ choice behaviour. Capturing and analysing the factors that may influence the choice behaviour of pedestrians by using RP data is the main challenge of this thesis. After the selection and filtering of the GPS data, a data analysis plan is presented in Section 5.2 where the hypotheses that will be tested are formulated. A quantitative analysis of the GPS data is conducted in Section 5.3. First the sample size of the GPS dataset is defined (Section 5.3.1) and then information about this dataset is given (Section 5.3.2).

Moreover in Section 5.3.3 a quantitative analysis of the GPS data is conducted and some general remarks on pedestrians’ movement patterns and behaviour regarding the understudied area are presented. Thus the following research question is discussed:

5. What information is extracted on the activity and route choice during the SAIL event from the collected data?

In Section 5.4 the GPS dataset is validated. Lastly conclusions are drawn in Section 5.5.

5.1 Processing the GPS data from SAIL

Before elaborating on the approach of processing the GPS data, it is important to be clarified that the main objective of this approach is to extract information regarding the choices of pedestrians during the SAIL event. To be more specific, the choices that can be identified by means of the GPS data are the routes that pedestrians took during the event and their choice to perform an activity or not. The performance of activity can be determined by comparing the time spent in the area derived by the GPS data with the average walking time estimated by the Wi-Fi data. Thus, the approach of processing the GPS data is formulated based on the necessity of extracting the aforementioned objectives. The software tool used for the filtering and analysis of the data is MATLAB.

Figure 5.1 shows a flow diagram of all the steps which are taken in order to process the data. The process that will be explained in the following paragraphs gives some general information about the GPS data, while Appendix B elaborates more on the technical part of the steps that are followed for the processing of the GPS data. Furthermore, Table 5.1 gives an overview of the inputs, the tools and the final output of the processing of the GPS data.

The raw GPS data that were collected during the period of the event (19/08-23/08) as well as the Google forms (Section 3.2) that were filled in by the participants of the GPS data collection were available. These two sources of data are the first inputs that need to be transformed into the required output by multiple steps in the process.
Figure 5.1 Flow diagram of processing the GPS data

Table 5.1 Overview of the processing of the GPS data

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Tools</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw data (GPS Server)</td>
<td>MATLAB, QGIS, Google Earth</td>
<td>Structure/Characteristics of each trip:</td>
</tr>
<tr>
<td>Trip data (Google Forms)</td>
<td></td>
<td>• <strong>Information</strong>: Trip ID, GPS ID,</td>
</tr>
<tr>
<td>Boundary boxes (QGIS)</td>
<td></td>
<td>departure/arrival time, gender, age, group size, area</td>
</tr>
<tr>
<td>Network configuration (Google Earth)-Decision points</td>
<td></td>
<td>boxes, travel time, entry time, route</td>
</tr>
<tr>
<td>Spatial/Boundary boxes (QGIS)</td>
<td></td>
<td>• <strong>Measurement</strong>: Longitude, latitude, time</td>
</tr>
<tr>
<td>Decision points</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network configuration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpolation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The raw data from the server provide a list of all the trips that took place during the period of the event. More specifically, the raw GPS data were retrieved from the server of MyGPStracker, consisting of a large matrix indicating tracks (position over time) of every trip. This data file contains the ID of the GPS device, the latitude, the longitude and the time. Each GPS tracker could have been used more than once by different participants, meaning that each GPS tracker does not necessarily correspond to one trip, but to multiple trips. Hence, the retrieved raw data should be processed further in order to select the subset of the delivered GPS measurements which is related to every participant of the GPS data collection separately. In other words, each participant should be assigned to a unique trip.

As mentioned in Section 3.3, the participants of the GPS data collection had to fill a form with some personal information (gender, age, group size). The answers were reported on a Google form (Excel file). This file also includes information about the departure and arrival time of each pedestrian from the standing point of TU Delft team, as well as the GPS device ID (named by the server). By matching the GPS device IDs given by the raw data file with the ones reported to the Google form file, constrained by the start and end time of the trip of each participant, a link between the raw data and the data from the Google forms is created. This link leads to the assignment of each participant to a unique trip (Step1: Select trip data). After this matching, the visualization of the track of each participant is possible. In Figure 5.2, an example of visualising the tracks of a trip is illustrated.

In the graphical representation of the trajectories of the participants, it was observed that the data were noisy. In order to filter the noise, a simple moving average (SMA) formula was used. Moving average smooths data by replacing each data point with the average of the neighbouring data points defined within the span (Step2: Smoothing data).

Next, a list of boundary boxes with spatial attributes is being derived by the division of the SAIL area (Figure 5.3) according to the network configuration and the activity areas, in order to assign the GPS measurement points to the spatial boundary boxes (Step3: Assign every GPS measurement to a spatial/ boundary box). The size of each box is determined by the amount of details that are needed for the study. The boxes contain intrinsic characteristics depending on the specific features of their location. The discretization into boxes provides information regarding the route choice and the possibility of performing an activity within a box. For the division of the SAIL area into bounding boxes the QGIS software was used.

\footnote{In this thesis, one trip is defined from the time moment one participant takes the GPS device from the stand until the time moment the same participant returns the device to the TU Delft stand.}
After this assignment, a list of boxes that were visited by every participant is created. Thus, there is a box sequence for every trip. Based on the network configuration and the decision points (defined by the route and activity choice) along the SAIL area, a set of route alternatives is created. Each route is characterized by a specific combination of box sequence. In other words, based on the attributes of the boundary boxes, each route has all the attributes of the boxes that belong to the combination of the boxes that formulate the route (Step4: Create routes based on the box sequences).

However, by having a visual inspection of the plots of the pedestrian’s trajectories in combination with area boxes, it is observed that some boxes have few data (missing data—not continuous trajectories). This lack of data points inside the boxes could lead to a wrongly estimation of the time spent in the area boxes. To deal with this issue, an interpolation is applied. The interpolation mainly contributes to the calculation of the entry and exit times to and from the boxes (Step5: Calculate the entry/exit times to/from the boxes for every trip). Then the time spent along different OD pairs specified by various combinations of box sequences is estimated (Step6: Time spent in every box per trip & Step7: Time spent along the routes for every trip).

Finally, the output of this whole process is the creation of a structure where the characteristics of each trip are available. For every trip it is known the gender, the age and the group size of each participant. Apart from these characteristics, the box sequence that every participant visited, the time spent to every box sequence, the entry time to each box belonging to each box sequence as well as the routes formed by the box sequences are known.

5.2 Data analysis plan

For research integrity and quality, a data analysis plan is needed. It guards against data-driven results and allows analyses to be reproduced. This can be done by testing the formulating hypotheses. Four questions are formulating to analyse the RP data collected during SAIL. The data will be able to confirm or refute the hypotheses. The hypotheses and their corresponding questions are presented below:

- In the Orange route, it is expected that pedestrians with direction from Station to Verbindingsdam will choose routes with attractions related to the SAIL event (which are indicated by SAIL organization). Which routes do pedestrians choose along the Orange route?
Along the Orange route with direction from Station to Verbindingsdam, a number of trips are not having as final destination the Verbindingsdam possibly due to the long length of the route (around 2.5 km) and the crowdedness. It is hypothesised that they might choose to stop at intermediate destinations or return back. Which is the distribution of pedestrians over the intermediate final destinations and which are the possible reasons for not arriving at Verbindingsdam?

It is expected that the route taken with direction to the Station is influenced by the previous taken route. More specifically, it is expected that pedestrians who walk with direction to Verbindingsdam first choose the route via Veemkade because the majority of the attractions exist (tall ships, food/drinks, music stages) in this route. On the way back, they most probably prefer taking Piet Heinkade route because it is less crowded than Veemkade. Which is the relationship between the previous taken and the way back chosen routes?

Along the Veemkade route it was seen that more pedestrians choose to perform activities at the start of Veemkade than at the end. More specifically, longer queues where formatting outside the tall ships at the area around the bridge (boxes 9-11-12-14-16, Figure 5.3). Which activity areas do visitors of SAIL event choose to perform an activity along the Orange route? In which areas do they spend most of their time?

The above questions will be answered in Section 5.3.3.

5.3 Analysis of the GPS data

In this Section an extensive analysis of the GPS data is conducted. First, the sample size of the GPS data is determined (Section 5.3.1) and some information by means of descriptive analysis is given about the data (socio-demographic and revealed) collected during the GPS devices distribution (Section 5.3.2). In Section 5.3.3 a quantitative analysis of the GPS data is conducted and some general remarks on pedestrians' movement patterns and behaviour regarding the understudied area are presented.

5.3.1 Examined sample size of the GPS data

The GPS data need to be filtered in order to create a usable dataset dedicated for the acquisition of choice behaviour of pedestrians along the Orange route (where this thesis is focusing on) during the SAIL event.

322 trips were performed by the participants of the GPS data collection. This number represents the total number of GPS devices that was distributed during the SAIL event (19/08-23/08). However, these trips are being assessed based on their capability to create routes along the Orange route. If they are not able, then they have to be removed from the database. The first step in the data selection is to examine if there are trips that are lacking GPS measurements (data points) and thus are not able to create routes. After this filtering 264 trips remain in the database.

The second step is to find out how many trips choose to take a ferry outside Amsterdam Train Station and then return back at the train station without walking at the Orange route, so they are not making any route or activity choice along the understudied area. This number of trips should be removed from the database since they do not contribute to the investigation of the activity-route choice behaviour as it is defined in this thesis. This removal leaves 230 trips in the final dataset. Thus the final dataset (230 trips) is examined on capturing information about the activity and route choice of pedestrians during the SAIL event. The algorithm for the creation of this dataset is shown in Figure 5.4.
5.3.2 Information about the examined dataset

A description of the database which is related to the Orange route area is summarized in Table 5.2. The results of the statistics are visualized in Figure 5.5.

In summary, from the total sample of 230 trips, 60% of them are performed by males while the average age of the participants is fluctuating between 55-64 years old. The most dominant group size consists of 2 persons. In general, it can be said that the average participant of this study is probably middle aged-couples. At this point it should be clear that these data are related to the participant that carried the GPS tracker, so if there is a group of more than 2 persons the only known information is the gender and the age (personal data) of the participant that carried the GPS device. Furthermore, the majority of the trips started in the morning, but this can be explained by the fact that the TU Delft team tended to distribute as much GPS devices during the morning. Besides, there was a limited number of GPS devices (only 110) that could be distributed per day. Therefore, sometimes the TU Delft team had to wait to have some GPS devices back in order to hand them out again on the field.

Regarding the characteristics of the trips performed, it is noticed that the majority of them have a duration of 5-6 hours in average.
Table 5.2 Descriptive Statistics for the examined dataset

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>Max Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>141</td>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td>38</td>
<td>55-64</td>
</tr>
<tr>
<td>45-54</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>65+</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td><strong>Group size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2 persons</td>
<td>124</td>
<td>2 persons</td>
</tr>
<tr>
<td>3 persons</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>4 persons</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>5 persons</td>
<td>16</td>
<td>2 persons</td>
</tr>
<tr>
<td>6 persons</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7 persons</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>8 persons</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><strong>Time of day</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morning</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td>Afternoon</td>
<td>84</td>
<td>Morning</td>
</tr>
<tr>
<td>Evening</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td><strong>Day of the week</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20/08 (Thursday)</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>21/08 (Friday)</td>
<td>77</td>
<td>Friday</td>
</tr>
<tr>
<td>22/08 (Saturday)</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>23/08 (Sunday)</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td><strong>Total trip duration</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2 hours</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2-3 hours</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>3-4 hours</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>4-5 hours</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>5-6 hours</td>
<td>48</td>
<td>5-6 hours</td>
</tr>
<tr>
<td>6-7 hours</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>7-8 hours</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>8-9 hours</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>9-10 hours</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>10+ hours</td>
<td>14</td>
<td></td>
</tr>
</tbody>
</table>
Figure 5.5 Gender, age, time of day, group size, day of the week and trip duration distributions over the 230 trips sample.
5.3.3 General remarks on pedestrians’ movement patterns and behaviour retrieved from GPS data.

In Section 4.2 a general sense of the prevailing atmosphere during the SAIL event as it was experienced by the author was described. However, the investigation of the movement patterns of pedestrians during the SAIL event as are derived from the revealed GPS data gives an interesting input into the better understanding of pedestrians’ behaviour.

As it was mentioned before, the focus area of this thesis is the Orange route. The physical network of the Orange route consists of 2 discretized large areas—the mainland (Veemkade, Piet Heinkade) and Java Island (Sumatrakade, Javakade) - which are connected via Verbindingsdam. However, it was observed that Veemkade area was much more crowded than the Java Island (Figure 5.6, Figure 5.7). Actually, at the Java Island as it was mentioned in Section 4.2, visitors could walk almost freely at the majority of its parts. Another difference between these two areas is that the most important, famous and historical tall ships were found in the Veemkade, implying that the attractiveness of Veemkade was increased in comparison with Javakade. More specifically, during the event, it was seen that more people (longer queues) preferred visiting the tall ships that were moored along Veemkade than the ones on Javakade. Considering the aforementioned, Veemkade area is considered more interesting for the objective of this thesis, since the prevailing conditions at this area (crowdedness) could better capture how pedestrians behave in a constraint space in the sense of choosing routes and deciding whether they want to perform an activity or not.

Thus, in this thesis the investigation of the movement patterns of pedestrians will be focused on the mainland with direction from Amsterdam Central Station to Verbindingsdam and the way back (Figure 5.8). An overview of pedestrians’ movements and choices during the SAIL event for this specific area will be presented in the next paragraphs. This overview of the pedestrian movements will contribute to better understanding of pedestrians’ behaviour during the event.

Along the understudy area, there are multiple decision points where visitors of SAIL event can choose among a number of alternative routes. These decision points can function as the origins of the OD –pairs which are formulating along the understudy area. Since this thesis elaborates on the trips on the mainland with direction from Amsterdam Central Station to Verbindingsdam, the selected OD –pairs are the following: Amsterdam Central Station (origin) - Verbindingsdam (destination) and Verbindingsdam (origin) - Amsterdam Central Station (destination)
Beginning with the direction from the Amsterdam Central Station to Verbindingsdam, from the total dataset of 230 trips, it was found that 184 trips crossed the studied area (with this direction). The rest took the ferry and then returned through the network via the other direction (Verbindingsdam to Amsterdam station) or they chose to walk around the station area (De Ruiterkade) and then walk with direction to the city centre (Figure 5.9).

The aforementioned trips distribution turn out to be some of the dominant pedestrians’ movement patterns with direction from Amsterdam Central Station to Verbindingsdam. Figure 5.10 illustrates these patterns.
Figure 5.10 Visualized common pedestrian patterns (direction from Amsterdam Central Station to Verbindingsdam)-Examples

a. A trip that crosses the understudy area

b. A trip that belongs to the category of taking the ferry and then return through the network on the way back (direction to the station)

c. A trip which belongs to the category of walking along the De Ruijterkade and the city centre
Before investigating the distribution of pedestrians over the route alternatives for the direction from station to Verbindingsdam, it is necessary to define the most common route choices along the under-study area during the event. After having a visual inspection\(^5\) to the overview of all the trips of each day separately, it was found that the most preferred route alternatives are depicted in Figure 5.11.

Route [I] is the main route (Veemkade via Bimhuis) indicated by SAIL, route [II] is the one that takes the shortcut from the Piet Heinkade to the Veemkade via Blauwhoedenveem, route [III] one that takes the Piet Heinkade to the Veemkade via Vriesseveem (just before the bridge) and lastly route[IV] is the one that crosses the Piet Heinkade straight ahead. In Appendix D, there is map which indicates the location of the aforementioned streets.

![Figure 5.11 Chosen route alternatives (direction from the Amsterdam Central Station to Verbindingsdam)](image)

As it was mentioned each route alternative is given as a box sequence. The box sequence of each route alternative is given in Table 5.3:

<table>
<thead>
<tr>
<th>Route</th>
<th>Legend</th>
<th>Box sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route[I]</td>
<td></td>
<td>58-3-62-2-9-11-13-14-16-19-21</td>
</tr>
<tr>
<td>Route[II]</td>
<td></td>
<td>58-61-9-11-13-14-16-19-21</td>
</tr>
<tr>
<td>Route[III]</td>
<td></td>
<td>58-4-11-13-14-16-19-21</td>
</tr>
<tr>
<td>Route[IV]</td>
<td></td>
<td>58-54-10-55-15-17-20-21</td>
</tr>
</tbody>
</table>

A script was written in MATLAB in order to assign the 184 trips to the different route alternatives. For some trips there were missing data on the pedestrians’ tracks making impossible their assignment to any of the available route alternatives. For this reason, 12 trips were removed from the database and finally 172 trips were taken into account on the 1st route choice decision point (outside Amsterdam Central Station). The share of 172 trips over the 4 available routes with destination to Verbindingsdam is presented in Table 5.4. It is clear that Route [I] was the most preferred route, since 75% of the total GPS

---

\(^5\) The main patterns of movements of the participants during the event are visualized by using a useful tool/platform created in Matlab. This tool gives the possibility to have an overview of the tracks of pedestrians during specific time periods (Appendix B)( Figure B.18)
participants chose this route. The way that pedestrians were distributed over the different alternatives is extensively depicted in Figure 5.12.

Table 5.4 Trip share over the 4 routes (direction from station to Verbindingsdam)

<table>
<thead>
<tr>
<th>Route</th>
<th>No of trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route[I]</td>
<td>129 (75%)</td>
</tr>
<tr>
<td>Route[II]</td>
<td>33 (19.2%)</td>
</tr>
<tr>
<td>Route[III]</td>
<td>3 (1.8%)</td>
</tr>
<tr>
<td>Route[IV]</td>
<td>7 (4%)</td>
</tr>
</tbody>
</table>

**Total number of trips** 172

Figure 5.12 Distribution of 172 trips over the 4 alternative routes - Intermediate stops

Figure 5.13 refers to the trips that did not arrive to box 21 (destination), along with the corresponding reason. Figure 5.14 indicates the intermediate destination boxes of those trips along with their distribution. Note that the missing data along the route (20%) were disqualified from Figure 5.14.

It is clear that almost half of the trips walk the full length of the chosen route and arrived at Verbindingsdam (box21) (Figure 5.14). A large share walks until box 16. The main attributes of this area (box16) is that there are some trees and supermarkets giving the sense to the participants that it is a logical point to turn around. Furthermore, the share of trips that had an intermediate destination at box 9 is almost equal to the ones of box 11. Both boxes correspond to areas where pedestrians had the
opportunity to have a boat trip. This could give a possible explanation to the fact these trips were not appeared at the next area boxes.

In summary, it is obvious that not all the trips arrived at Verbindingsdam (box 21). Some of them returned halfway, others chose another route or took a boat to have a boat trip or for some of them their track is not known due to missing GPS data (Figure 5.15).

Figure 5.15 Trip distribution over intermediate stops

a. Trips of Route [I] that stopped at intermediate boxes
b. Trips of Route [II] that stopped at intermediate boxes
c. Trips of Route [III] that stopped at intermediate boxes
d. Trips of Route [IV] that stopped at intermediate boxes

Figure 5.15 presents for each route separately, the trip share over the intermediate boxes. Another interesting remark regarding the trip share over the intermediate boxes is that almost 50% of the trips that chose route [IV] walked until box 55. This may be related to the fact that this area (box 55) connects the Orange area with the city centre. Thus, it seems that some participants decided on their way that they do not want to cross the Orange area but they prefer to take another route leading to the city centre.

However, the share of trips that chose route [III] and route [IV] is low (3 and 7 trips respectively), so the sample size of these routes is deemed inadequate to provide a fairly accurate picture of pedestrians’ routes and activities preferences in the study site.

**Direction from Verbindingsdam to Amsterdam Central Station**

Next to the analysis of the direction from Amsterdam Central Station to Verbindingsdam, the investigations of the trip patterns of the other direction (from Verbindingsdam to Amsterdam Central Station) are presented below.
From the total dataset of 230 trips, it was found that only 138 trips crossed the understudied area with direction from Verbindingsdam to Amsterdam Central Station. The rest took the ferry from the Java Island or public transport from Verbindingsdam and returned back to the central station (so they did not walk). A high percentage of trips do not give clear information of the pedestrians’ movements due to missing data from the GPS devices (Figure 5.16).

An overview of the aforementioned trip patterns are illustrated in Figure 5.18

Regarding the 138 trips that crossed the area with direction from Verbindingsdam to train station, it should be clarified that their origin does not only refer to Verbindingsdam Bridge but also to the other location points between Verbindingsdam and station. As mentioned before, not all the trips with direction from the station to Verbindingsdam arrived at Verbindingsdam but had intermediate destinations. Thus, these destinations are defined as the origins for the way back to the station. From the total dataset of 138 trips that crossed the under-study area, 60% of them is originated from Verbindingsdam (box21), while the rest 40% is divided into lower percentages for trips that originate from points (defined by boxes) that are located between the Verbindingsdam and the station (Figure 5.17).

Figure 5.16 Trip patterns (direction from Verbindingsdam to Amsterdam Central Station)

Figure 5.17 Origins of the trips that crossed the area

a. A trip that crosses the understudy area (Trip30)
b. A trip that belongs to the category of taking the ferry from Java Island and return to the station (Trip260)

c. A trip which belongs to the category of taking Public Transport from Verbindingsdam back to the station (Trip82)

Figure 5.18 Trip patterns (direction from Verbindingsdam to Amsterdam Central Station)

Since in this thesis the selected OD pair is Verbindingsdam (origin) - Amsterdam Central train station (destination), the distribution of the trips that originate from Verbindingsdam (138*60%) over the route alternatives for the direction from Verbindingsdam to the station is selected to be presented. However, before investigating the distribution of the trips, it is necessary to define the most common route choices along the understudy area during the event. After having a visual inspection of all the trips of each day separately, it was found that the most preferred route alternatives can be categorized as follows:

- Via Piet Heinkade (straight); it is the same as the route[IV] but for the opposite direction
- Via Veemkade; it is the same as route[I] but for the opposite direction

---

The main patterns of movements of the participants during the event are visualized by using a useful tool/platform created in Matlab. This tool gives the possibility to have an overview of the tracks of pedestrians during specific time periods (Appendix B, Figure B.18)
• **Via shortcuts** from Veemkade to Piet Heinkade. There is a number of shortcuts which connects the Veemkade with Piet Heinkade. These shortcuts are the following: Blauwhoedenvleem(box61); it is the same as route[II] but for the opposite direction, Vriesseveem(box4); it is the same as route[III] but for the opposite direction, Vemenplein(box12), Withoedenvleem(box5), Klapmutseveem (box7), combination of shortcuts (Appendix D)

The share of 138 trips over the 4 alternatives routes with direction from Verbindingsdam to train station is presented in Table 5.5. Furthermore, the shares of the aforementioned alternatives are illustrated in Figure 5.20. It is clear that the route via Piet Heinkade was mostly preferred from Verbindingsdam to the station. The route via Veemkade was almost equally chosen as the route from Veemkade to Piet Heinkade via shortcuts. In Figure 5.21, it can be seen that the most frequently chosen shortcut is via Withoedenvleem (box5). This can be attributed to the fact that after the point of section between Veemkade and the Withoedenvleem, the part of Veemkade route was too crowded. Thus, it can be assumed that the participants were tending to avoid the crowdedness by taking the shortcut (box5).

### Table 5.5 Trip shares over the route alternatives (direction from Verbindingsdam to train station)

<table>
<thead>
<tr>
<th>Route</th>
<th>Same as (for other direction)</th>
<th>Legend</th>
<th>No of trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Via Veemkade</td>
<td>Route [I]</td>
<td>------</td>
<td>39 (28%)</td>
</tr>
<tr>
<td>Veemkade to Piet</td>
<td>Route [II], Route [III]</td>
<td>------</td>
<td>34 (25%)</td>
</tr>
<tr>
<td>Heinkade via shortcuts</td>
<td></td>
<td>------</td>
<td>57 (41%)</td>
</tr>
<tr>
<td>Via Piet Heinkade</td>
<td>Route [IV]</td>
<td>------</td>
<td>57 (41%)</td>
</tr>
<tr>
<td>(straight)</td>
<td></td>
<td>------</td>
<td>57 (41%)</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td>------</td>
<td>8 (6%)</td>
</tr>
<tr>
<td><strong>Total number of trips</strong></td>
<td></td>
<td>------</td>
<td><strong>138</strong></td>
</tr>
</tbody>
</table>

Figure 5.19 Route alternatives for the direction from Verbindingsdam to the station (Origin: Box 21, Destination: Box58)

---

7 In this thesis a shortcut is defined as a vertical route that connects two parallel roads. In this case the parallel roads are Veemkade Street and Piet Heinkade Street.
Nevertheless, in order to understand the way that the participants choose routes from Verbindingsdam to the station, it is necessary to examine if the previous taken route influences their choice from Verbindingsdam to the station (way back). Hence, it is important to find if there is any relation between the previous taken route and the route that the departing trips from Verbindingsdam with direction to the station choose.

Table 5.6 describes when a certain route is picked on the way from the station to Verbindingsdam which route is chosen on the way back.

A Pearson chi-square test (Table 5.7) is conducted to examine whether there is a relationship between the previous taken route and the chosen route for the direction from Verbindingsdam to the station. The results reveal that there is a significant relationship between the two variables (Chi square value = 52.1, df =32, p <0 .05). In other words, the relationship is considered to be significant with a 95% confidence level. Thus, there is a statistically significant association between previous taken and chosen route (direction from Verbindingsdam to the station).

As mentioned, Veemkade was the main route as suggested by the SAIL organisation and the majority of the attractions (tall ships, food-drinks, sea view) and events were expanded along this route. It is expected that this route would have been crossed at least once by all the dataset of the trips that walked at the understudied area (part of the Orange route). According to Table 5.6, it is clear that when a participant had taken route [I] (Veemkade via Bimhuis) it is more likely to take Piet Heinkade (20.5%) for the way back. This could be explained by the fact that they have already enjoyed the attractions of Veemkade and might prefer a less crowded route for their return. However, it should not be omitted that there is a considerable share of participants (13.3%) that chose to return back from the same route (Veemkade via Bimhuis) as their previous taken route. A possible explanation is that the first time that they crossed Veemkade, they opted for recognition of the area and then on the way back being familiar with this route, they choose it again and decide which activities they want to perform. Another plausible explanation could be relevant to the landscape, as a walk along the ships may be considered more attractive than Piet Heinkade.

Regarding the participants that had chosen route[II] (Veemkade, avoiding the Bimhuis), it is seen that the participants preferred the Piet Heinkade route for their way back, possibly for the same reasons as the ones that had previously chosen route[I].

Another interesting point, is the trend of the participants that took the ferry to visit Java Island and then returned via Verbindingsdam on foot. This group of pedestrians tend to choose Veemkade (14.5%) maybe because they wanted to see the attractions of this route since they did not have this opportunity before.
The number of observations that corresponds to route [III] and route [IV] (Table 5.4, Figure 5.11) is too small and may lead to unreliable conclusions. For instance, it was expected that the participants that took route [IV] (Piet Heinkade -direction from station to Verbindingsdam) would have chosen Veemkade for their way back. However this was not the case, since based on the available observations, it seems that they have taken again the Piet Heinkade. Lastly, only one trip took the route [III] and returned from Piet Heinkade, which means that there was no interest in the Bimhuis area. This is not a concrete conclusion and cannot lead to generalities since the sample is not representative (only one trip).

Table 5.6 Cross table of the trips that walked from the station to Verbindingsdam and the way back

<table>
<thead>
<tr>
<th>Chosen route (direction from Verbindingsdam to the station)</th>
<th>Other</th>
<th>Piet Heinkade (straight) Route[IV]</th>
<th>Veemkade to Piet Heinkade via shortcut 12</th>
<th>Veemkade to Piet Heinkade via shortcut 45</th>
<th>Veemkade to Piet Heinkade via shortcut 5</th>
<th>Veemkade to Piet Heinkade via shortcut 61</th>
<th>Veemkade to Piet Heinkade via shortcut 7</th>
<th>Veemkade (via Bimhuis) Route[IV]</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous taken route (direction from Verbindingsdam to station)</td>
<td>Route[III]</td>
<td>2(2.4%)</td>
<td>17(20.5%)</td>
<td>0(0.0%)</td>
<td>1(1.2%)</td>
<td>6(7.2%)</td>
<td>3(3.6%)</td>
<td>1(1.2%)</td>
<td>11(13.3%)</td>
</tr>
<tr>
<td></td>
<td>Route[II]</td>
<td>1(1.2%)</td>
<td>10(12%)</td>
<td>1(1.2%)</td>
<td>0(0.0%)</td>
<td>0(0.0%)</td>
<td>0(0.0%)</td>
<td>1(1.2%)</td>
<td>0(0.0%)</td>
</tr>
<tr>
<td></td>
<td>Route[III]</td>
<td>0(0.0%)</td>
<td>1(1.2%)</td>
<td>0(0.0%)</td>
<td>0(0.0%)</td>
<td>0(0.0%)</td>
<td>0(0.0%)</td>
<td>0(0.0%)</td>
<td>0(0.0%)</td>
</tr>
<tr>
<td></td>
<td>Route [IV]</td>
<td>0(0.0%)</td>
<td>3(3.6%)</td>
<td>0(0.0%)</td>
<td>0(0.0%)</td>
<td>0(0.0%)</td>
<td>0(0.0%)</td>
<td>1(1.2%)</td>
<td>0(0.0%)</td>
</tr>
<tr>
<td>Take a ferry</td>
<td>2(2.4%)</td>
<td>4(4.8%)</td>
<td>1(1.2%)</td>
<td>2(2.4%)</td>
<td>1(1.2%)</td>
<td>1(1.2%)</td>
<td>1(1.2%)</td>
<td>12(14.5%)</td>
<td>24(28.9%)</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>34</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>23</td>
<td>83</td>
</tr>
<tr>
<td>Total (%)</td>
<td>6.0%</td>
<td>41.0%</td>
<td>2.4%</td>
<td>3.6%</td>
<td>8.4%</td>
<td>6.0%</td>
<td>4.8%</td>
<td>27.7%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 5.7 Chi-Square Tests

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>df</th>
<th>Asymp. Sig. (2-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-Square</td>
<td>52.101a</td>
<td>32</td>
<td>.014*</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>47.600</td>
<td>32</td>
<td>.037</td>
</tr>
<tr>
<td>N of Valid Cases</td>
<td>83</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. 40 cells (88.9%) have expected count less than 5. The minimum expected count is .02.

*Correlation is significant at the 0.05 level (2-tailed).
a Average Time spent given that Route[I] was firstly taken for the way forth

b Average Time spent given that Route[II] was firstly taken for the way forth
In Figure 5.22, it is obvious that with the exception of route[IV] (via Piet Heinkade) as previous taken route (Figure 5.22d), for all the other combinations of routes, the average time spent in the previous taken route (from station to Verbindingsdam) is higher than the average time spent in the way back. This is happening because in all the cases except for d, the previous taken route includes (parts of) Veemkade route. It has been mentioned a lot of times that Veemkade was the route with the most of attractions related to the SAIL event and it was always crowded.
To be more specific, considering route [I] (via Veemkade) as the previous taken route (Figure 5.22a) it can be seen that when pedestrians choose the same route (via Veemkade) to return to the station the average time spent is lower, meaning that pedestrians tend to spent more time when they first cross a route than their way back. Some possible explanations could be that pedestrians may prefer to perform activities when they first meet an attraction or that the second time are more familiar with the route so they do not have the recognition state of mind and they can cross the route more quickly. Another explanation could be the prevailing conditions along the route. Maybe when they first took the route, the area was more crowded than in the way back.

An interesting point regarding route [I] (previous taken route) is that the average time spent for the way forth is similar for all the combinations. However there are some remarkable differences on the average time spent in the routes chosen for the way back. For example, choosing the Piet Heinkade to return back leads to lower total average time spent than returning from Veemkade. Taking into account that the length of all the routes alternatives for the way back are almost equal, it can be assumed that the reason why there are differences in the time spent for every alternatives are related to other factors like the performance of activity, the existence of attractions or crowdedness.

Regarding route [II] (Piet Heinkade via shortcut 61) as the previous taken route (Figure 5.22b) it is obvious that the combination of route [II] with Veemkade to Piet Heinkade via shortcut leads to the lower total time spent to cross the area forth and back. This may happen because the shortcuts give the opportunity to avoid the potential congested parts of the network. However, in case of route [IV] as previous taken route, there is a small difference in the total time spent between the two combinations of the way forth and back. Their difference is derived by the time spent difference of the chosen routes of the way back (since the time spent for the way forth is almost equal for both combinations). It seems that the pedestrians that had chosen Piet Heinkade (the same as route [IV]) to return back spent less time than in the option of Veemkade to Piet Heinkade via shortcut. Two possible explanations can be given for this phenomenon; the one is related to the familiarity with route [IV], since this route is the one that was chosen before and the other explanation is related to the opportunities to perform an activity or to the existence of attractions along the Veemkade before taking the shortcut which may lead to higher time spent. Another interesting observation is that when pedestrians choose to return via Piet Heinkade which is the same route taken for the way forth, they tend to spent more time. This can be attributed to possible different prevailing conditions along this route. As a remainder, route [III] and route [IV] lack of observations and this may lead to unreliable results.

The correlation between the time spent in the previous chosen and the chosen route for the way back is examined. It was found that the time spent in the previous taken route has a positive correlation with the time spent in the way back route, meaning that higher values on time spent in the way back route are associated with higher values on the previous chosen route. If the previous chosen route is route [I] then there is a positive correlation between the time spent in the previous chosen route and the time spent in the way back route, while the time spent in the route [II] as previous chosen route had a negative correlation with the time spent in the route that was chosen to return back. In general, the correlation between the 2 variables (the time spent on the previous taken and way back route) were not statistically significant for all the cases (p> .05).

Lastly, another interesting information that can be derived from Figure 5.22 and Table 5.8 regarding the total time spent in the way forth and back is that all the combinations of route [I] score higher in comparison with the other routes. Having a closer look to the route [I] as previous taken one, it is clear that the time spent in this route is higher comparing with the route [II], route [III], and route [IV]. This may be related to the fact that route [I] not only is a bit longer but also offers more opportunities to perform activities and admire the attractions along its length. Furthermore, the route [I] is the only one
route which includes the Bimhuis area. This area was one of the most crowded places along the Veemkade area. Thus this could be a reason of higher walking time due to the prevailing conditions of this area.

Table 5.8 Average total time spent (in minutes) for the way forth and back

<table>
<thead>
<tr>
<th>Route for the way forth</th>
<th>Route [I]</th>
<th>Veemkade to Piet Heinkade via shortcut</th>
<th>Veemkade (via Bimhuis) Route [I]</th>
<th>Piet Heinkade to the station via city centre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route [I]</td>
<td>154</td>
<td>176.34</td>
<td>198.36</td>
<td>-</td>
</tr>
<tr>
<td>Route [II]</td>
<td>128.91</td>
<td>75.92</td>
<td>-</td>
<td>120.94</td>
</tr>
<tr>
<td>Route [III]</td>
<td>96.44</td>
<td>110.50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Route [IV]</td>
<td>136</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Most preferred activity areas at the understudy area

Along the examined area which is expanded from Amsterdam Central Station to Verbindingsdam, the majority of the activities related to the SAIL event took place along the Veemkade street (part of the Orange route). Some parts of the Veemkade were more preferred as activity areas. It is interesting to examine the share of visitors that choose to perform activities at each activity area. For that purpose, the boundary boxes (Figure 5.3) in which Veemkade part is divided are used. The boxes and the activities that can be performed at each of them are listed in Table 5.9. The differences between the same types of activities (i.e. tall ships) are not presented. To be more specific, some tall ships are more famous and historical than others and it is expected that more pedestrians would visit them. However, this information is not available in this thesis. In case that the differences on the type of activity are considered, it could contribute on the better explanation on the choice of activity areas as well as on the activity time spent.

Table 5.9 Activities that can be performed at the boxes of Veemkade

<table>
<thead>
<tr>
<th>Boxes</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>3+62</td>
<td>Food/drinks (Bimhuis), opportunity for boat trip, buy tickets for the boat trip (Figure 5.23)</td>
</tr>
<tr>
<td>2</td>
<td>Tall ships, canteens (Figure 5.23)</td>
</tr>
<tr>
<td>9</td>
<td>Tall ships, canteens, small music stage (Figure 5.24) (Figure 5.25)</td>
</tr>
<tr>
<td>11+12</td>
<td>Tall ships, enjoy the view from the bridge (Figure 5.25)</td>
</tr>
<tr>
<td>13</td>
<td>Tall ships, canteens (Figure 5.26)</td>
</tr>
<tr>
<td>14</td>
<td>Tall ships, opportunity for boat trip (Figure 5.26)</td>
</tr>
<tr>
<td>16</td>
<td>Boat trip (Figure 5.26)</td>
</tr>
<tr>
<td>19</td>
<td>Small boats</td>
</tr>
<tr>
<td>21</td>
<td>Tall ships (Figure 5.27)</td>
</tr>
</tbody>
</table>
The results of the share of GPS participants that choose to perform an activity at each of the aforementioned box are shown in Figure 5.28. It is clear that more than 50% of the visitors that appeared at each box performed at least one activity. The only exception is box 19, which was the least preferred activity area (only 27% of pedestrians performed an activity). It seems that the activity area which was mostly preferred (80% of visitors performed an activity) is the area which consists of boxes 11 and 12. At this area the pedestrians had the opportunity to visit the tall ships or enjoy the view of the event from the bridge. Another interesting observation is that the share of pedestrians that performed activities decreases after box 14. This could be related to the fact that in large scale events pedestrians do not have full knowledge of the event area thus they may tend to perform an activity when the first meet an activity. Another possible explanation could be that as the walking distance increases the desire to perform an activity decreases.
Average activity time spent in chosen box along Veemkade

The average activity time spent in each of the aforementioned boxes is illustrated in Figure 5.29. The maximum average activity time spent is noticed in boxes 11 and 12 (30 minutes) which are the most preferred boxes to perform an activity. On the other hand, box 19 which was the least preferred to perform an activity presents the lowest activity time in comparison with the rest of the boxes. A possible explanation could be that the two activity areas (box11&12 and box19) offer different type of activities so the activity time spent on them differs. This means that the activity time spent to each box is related to the type of activities that are available and the service that is offered at each box.
Last but not least, the average activity time before the bridge is around 15 minutes while after the bridge is around 20 minutes (except for box 19). This difference is not important but it gives an implication that as pedestrians walk longer distances they tend to spend more activity time on the boxes. However, since it is not possible to derive from the data the type of activity that a pedestrian performed, it is not easy to reach to a conclusion if someone visits a tall ship for 20 minutes or due to his/her tiredness he/she stops walking in order to relax.

More information on the distribution of the activity time in each activity box can be found in Appendix B.

5.4 Validation of the GPS dataset

In Section 5.2 a description was given on how the GPS dataset of the visitors of SAIL area (Orange route) was created. This section addresses the validity of the dataset. The validity of the GPS dataset can be done either by comparing them with the entire population that visited the SAIL area or by comparing the data with other data sources available like the Wi-Fi data.

Unfortunately, there is no way to validate if a representative sample from the entire population that visited the SAIL event is present in the examined dataset because the data about the SAIL visitors was not as extensive. However, it is possible to test the validity of the GPS dataset with the Wi-Fi dataset and determine whether they are significantly different or not. This can be done by assuming that the Wi-Fi data are correct and can be used as ground truth. The data of these technologies will be tested on the number of visitors per day and on the time spent in the area per day. The data technologies cannot be tested per hour since the sample of the GPS participants per hour is really small.

A pair of Wi-Fi sensors was selected in order to define the area where the datasets will be compared. The Wi-Fi sensors were selected based on their capability to estimate the time spent as accurate as possible, since the time spent is the one of the two aspects that will be tested. For this purpose, the selected pair of Wi-Fi sensors (Figure 5.30) covers a relatively small area (400 meters) which is not intersected by shortcuts. Thus the movements of pedestrians and the time spent in this area can be sufficiently defined.

![Figure 5.30 Selected area for the validation of the dataset](image)

Percentage of visitors

It is expected that much more people had enabled the Wi-Fi function in their smartphones than the ones that accepted to participate in the GPS data collection by carrying a GPS device. In Figure 5.31 the percentages of Wi-Fi and GPS users that visited the area per day are being compared. A Kolmogorov test (non-parametric) shows that the distributions of the Wi-Fi and GPS users over the days are not significant
different. The Kolmogorov value is 0.354 (Table 5.10) which is higher than the 0.05 belonging to the 95% confidence interval.

![Graph showing GPS and Wi-Fi users percentage per day](image)

Figure 5.31 Percentages of visitors per day –GPS versus Wi-Fi

<table>
<thead>
<tr>
<th>Most Extreme Differences</th>
<th>Absolute</th>
<th>Kolmogorov-Smirnov Z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>0.250</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>-0.250</td>
</tr>
</tbody>
</table>

Table 5.10 Kolmogorov-Smirnov Test

Asymp. Sig. (2-tailed) 1.000

Time spent in the area per day

The distributions of time spent in the area for both Wi-Fi and GPS users are being compared for each day separately. The results of this comparison are presented below:

![Histogram showing area presence time](image)

DAY 20

Area presence time (20/08); Wi-Fi

Mean = 14.8536 minutes
Std Dev = 11.52 minutes
N = 2506
The Kolomogorov test showed that the distributions over the time spent of the two technologies are similar for each day separately.

5.5 Conclusions

This Chapter addresses the process and analysis of the GPS data. During the GPS data process, the approach of the boundary boxes is introduced. It seems a promising approach for describing simultaneously route and activity attributes of a region. Apart from the spatial information that can be indisputably extracted by the boxes, the travel time information inside the boxes can be derived. By estimating the time spent in each box, gives information about the interesting areas from the perspective of crowdedness or attractions and activities. This approach can be extended to other cases (like other large scale events, shopping malls or train stations) where the combination of route and activity attributes is need.
The collected with GPS devices data were filtered in order to create a usable dataset dedicated for the acquisition of choice behaviour of pedestrians along the Orange route. A total sample of 230 trips were found in the Orange route area. 60% of them are performed by males while the average age of the participants is fluctuating between 55-64 years old. The most dominant group size consists of 2 persons. Furthermore, the majority of the trips started in the morning. Regarding the characteristics of the trips performed, it was noticed that the majority of them have a length duration of 5-6 hours in average.

During the analysis of the GPS data, it was found that the majority of pedestrians choose route [I] when they first start a trip. Route [I] provides a full view of the main attractions of the event and it is indicated by signs by the SAIL organization.

Another interesting finding is the existence of association between the previous taken route and the initial chosen route. A Pearson chi-square test was conducted and it was found that the relationship is considered to be significant with a 95% confidence level. Furthermore, a remark that should not be omitted is related to the percentage of pedestrians that choose to cross the Orange route from station to Verbindingsdam and the way back. It was noticed that a higher percentage of pedestrians cross the area with direction to Verbindingsdam in comparison with the percentage of pedestrians with direction to the station. On the way back (direction to the station) there is a variability; pedestrians choose either to return on foot or take a ferry or use PT. This information is valuable for the next SAIL event and the regulatory bodies should take into consideration the aforementioned options (for example increasing the frequency of PT).

Regarding the activity areas choices of pedestrians, it was found that during the event more than 50% of visitors choose to perform an activity along Veemkade route. However some parts of Veemkade are more preferred (like the area consisting of boxes 11&12) and in some parts visitors tend to spend more time. Thus, identifying the most popular activity areas as well as the time spent in each of them could constitute a useful knowledge for future SAIL organization in terms of distributing the activities in a way of minimizing potential congestion.

Lastly, it was found the GPS and Wi-Fi users have similar distributions over the days and over the time spent for each day separately. However, the representativeness of either Wi-Fi users or GPS user with the total population needs to be validated.

Next Chapter (Chapter 6) addresses to the joint activity route choice model. The estimation of this model is done for a part of the SAIL area. The quantification of the factors that influence this choice aspect will be discussed.
6 Joint activity-route choice model

Part of the objective of this research is to understand and predict which factors influence the activity and route choice behaviour of pedestrians during large scale events. This chapter quantifies the relationships between the activity-route choices with the influential factors by means of discrete choice model.

In the proposed theoretical framework in Section 2.4.2, it was assumed that pedestrians know beforehand if they want to perform an activity or not (predetermined decision) but they do not know the exact location of the activity. They only know the activity area and whether an activity has been performed on a box level. In the case of pre-determined preference for activity, the activity and route choices can be determined simultaneously allowing the conception of a joint–activity route choice model. On the contrary, in the case of no preference for activity a route choice model is estimated.

Considering the main assumption that visitors know beforehand whether they have a preference for performing an activity, it can be said that pedestrians can make a joint decision regarding their preference to perform an activity and route choice (since they want to choose a location of an activity, it inevitably means that there is a preference for an activity).

Section 6.1 discusses the proposed model for the joint activity route choice decision. In section 6.2 the reasons which led to the selection of the area which is used for the estimation of the joint activity route choice model will be addressed. Furthermore, the characteristics of the selected area will be described.

In order to examine which factors influence the joint activity-route choice behaviour an extensive statistical analysis is conducted in Section 6.4 guided by the data analysis plan (Section 6.3). Section 6.5 addresses the results and findings. The relationships between the joint activity–route choice and the influential factors will be quantified by means of a discrete choice model. Finally, Section 6.6 discusses the conclusions of the joint activity and route choice model.

6.1 Practical representation of the joint activity route choice model

A simultaneous decision between activity and route choice is assumed for pedestrians that have a predetermined preference to perform an activity. The proposed joint activity-route choice decision is presented in Figure 6.1.

---

**Figure 6.1 Flow chart of the proposed joint activity-route choice decision**
According to the flow diagram (Figure 6.1) the pedestrian enters to the network and he/she knows beforehand if he/she has a preference to perform an activity or not (main assumption: pre-deterministic decision process). If there is a preference for an activity then a joint activity route utility is calculated. If activity exists but pedestrian chooses to skip the available activity then there is only the route utility. In case of no available activities along the route, the utility will have again the value of the route. However, the existence of activity may have a positive impact on the total route utility during events and should be considered in this research.

In other words, in this thesis a discrete choice model of pedestrians' responses to the existence/availability of activities based on their a-priori knowledge on preference to perform an activity or not is estimated. The model can be used to simultaneously estimate activity and route choice. A more elaborate way to visualize the concept of the proposed joint activity route choice model in the network is illustrated in Figure 6.2.

![Figure 6.2 Representation of the joint activity route choice in the network](image)

The structure that is presented in Figure 6.2 consists of two main sections. The left hand side corresponds to pedestrians with a pre-determined preference for activity. The decision point which is defined as the origin, refers to a joint activity and route choice decision. Thus a pedestrian with a preference for activity enters the decision point and makes a joint activity route choice among a number of \( n \) alternatives. The alternatives are paths which consists of sequence of links. Each link includes the probability of performing an activity or not. The joint approach assumes that the pedestrian chooses a path among all the available alternative paths for a given Origin-Destination (OD) pair, meaning that he/she jointly chooses all the links belonging to that path. Considering the right hand side section, it corresponds to the simple case of route choice. The traveller does not have a preference for activity and therefore the decision point describes route choices among a number of \( m \) alternatives.

In Figure 6.3, a tree approach is presented in order to describe the combination of links from which the number of alternatives of the joint activity route choice are derived. Between the origin and the destination there are nodes which define and delimit the activity areas based on the potential of performing an activity. The nodes are connected with each other via functional links which indicate the simultaneous activity-route choice. The number of alternatives is determined by the point which is set as destination as well as by the number of nodes. It should be stressed that the main assumption of the joint decision is that the pedestrian knows before departing from his/her origin the combination of the links that connect the nodes.
This principle of the tree is sufficiently generic to describe a wide variety of situations and derive a number of alternatives. However, the tree will look different for each case depending on the network configuration/topology. The simplified version of the proposed tree is demonstrated and discussed within Section 6.2.

![Decision tree diagram]

**Figure 6.3 Tree approach: Combination of links in order to derive the number of alternatives**

### 6.2 Selected area for the estimation of joint activity-route choice model

In general, for the estimation of a joint activity-route choice model, the area which will be selected for that purpose should meet some criteria. First of all, the number of available routes and the existence of activity areas along the routes are important factors, because without any alternatives no choices can be made. However, it is important for each route alternative to have enough observations for the estimation of the model. Otherwise, the reliability of the model estimated results may be compromised.

Since the estimation is done for a joint activity-route choice model, the selected area should combine both activity and route choice. Thus, another selection criterion is the diversity (type of activities, different routes and route characteristics, different conditions) of the area. This diversity is easier to be kept in track in small areas than in large ones. The choice of small areas results to get insight into pedestrians’ choice behaviour. Considering the aforementioned, the procedure of selecting the area for the estimation of joint activity-route choice model of pedestrians heading with direction from Amsterdam Central Station to Verbindingsdam during the SAIL event is described below.
As it was mentioned before, the focus area of this thesis is the Orange route. In Section 5.3.3, it is derived that from the total Orange route, the part of Veemkade area (Figure 6.4) is considered more interesting (due to crowdedness and number of attractions) than Javakade for the objective of this thesis.

Figure 6.4 Veemkade area - part of the orange route

The Veemkade area is divided into decision points according to the network configuration. These points are defined based on the parts of the network where pedestrians have the option to change the route that follow (return back, take a shortcut, use PT, take ferry). However, in point B there is no alternative route to change, but it is considered as a point of decision since the area after B is not visible before arriving at this point. In B, pedestrians have the opportunity to decide whether they want to continue or not to the area after B, which was invisible before arriving at this point. Furthermore, point B might act as a point of decision to return or not, which the same applies for all the defined decision points as it was seen in Section 5.3.3.

Figure 6.5 Decision points along Veemkade
Moreover, the most preferred route alternatives (direction from the Amsterdam Central Station to Verbindingsdam) that derived by the analysis of the GPS data in Section 5.3.3 are depicted in Figure 6.6. Only a small percentage of GPS users (3%) decided to change route in the intermediate decision points between A and H. Apart from the small number of trip observations in the intermediate decision points there is the issue of the number of alternatives. In most of these points there is the decision of choosing either Veemkade or Piet Heinkade streets. This decision is highly interrelated with the activity choice (by choosing Veemkade there is the possibility to perform an activity, while by choosing Piet Heinkade there is not the option of activity).

Taking into account the aforementioned, it is implied that the origin of the area that is going to be selected should include the decision point A. This point does not only offer a number of route alternatives but also a sufficient number of trip observations. Lastly considering that the size of the area plays role on assessing the diversity of the area, it is better to choose a small part of the Veemkade. Thus the selected area is extended from point A to point C and is presented in Figure 6.7. This area is the only area where the complexity of choosing routes and the complexity of choosing activity is combined.

Figure 6.6 Most preferred route alternatives (direction Station to Verbindingsdam)

Figure 6.7 Selected area for the estimation of the joint activity route choice. The available activities are indicating for each part of the understudied area.
Along the selected area there are various available activities. Namely, there are 3 discretized activity areas: the AB activity area, BC activity area and area from A to C directly. The AB activity area offers the following activities: enjoying food and drinks at the Bimhuis, buying a ticket for a boat trip and taking some photos, whereas the BC includes the following activities: buying take away food from canteens, watching/visiting tall ships and taking photos (Figure 6.7). Regarding the direct route from A to C, without passing via B, there is no activity relevant to the SAIL event.

In this study, the exact activity locations is difficult to be captured by the available RP data. However, the activity areas can be identified on a box level. Knowing the available activities (related to the event) at activity areas defined by the boxes allows a possible explanation arising from the activity time. Based on the activity time, hypotheses on the type of activities performed can be done. For instance, if the activity time is 5 minutes in the AB area, then the scenario of having food at Bimhuis is excluded since the time is too short for having lunch or dinner in a restaurant. Thus it is more likely performing other activities like taking photos or buying tickets for the boat trip. Still, formulating hypotheses on the type of the performed activities is difficult to be done, since the time share along the activities is not known. For example, if the activity time spent in an area is 20 minutes, it is not known if this time refers to one activity or a combination of activities.

Furthermore, the tree approach discussed in Section 6.1 is implemented to the case study (area AC) in order to derive the joint activity route choice alternatives. In Figure 6.8 the derived tree is presented indicating the number of alternatives. The possible alternatives of joint activity route choices are shown in Figure 6.9. The blue colour indicates the part of the route where there is performance of activity.

![Tree approach for the area of interest (Origin: A, Destination: C)](image-url)
Below an example is given for the better understanding of the tree approach:

**Example**

A pedestrian wants to go from A to C. There are 6 available alternatives to reach destination C.

1\textsuperscript{st} alternative: Via B by skipping the activities both in the first part (AB) and in the second part (BC). This choice is symbolized as 1_2 (not performing activity in both parts).

2\textsuperscript{nd} alternative: Via B by skipping the activity in the 1\textsuperscript{st} part (AB) and performing an activity in the second part (BC). This choice is symbolized as 1_2a.

3\textsuperscript{rd} alternative: Via B by performing an activity in the 1\textsuperscript{st} part (AB) and skipping the activity in the second part (BC). This choice is symbolized as 1a_2.

4\textsuperscript{th} alternative: Via B by performing an activity in both parts. This choice is symbolized as 1a_2a.

5\textsuperscript{th} alternative: From A to C directly by skipping B (not passing via B) without performing an activity.

6\textsuperscript{th} alternative: From A to C directly (skipping B) by performing an activity (if activity exists).

Thus it is implied that despite the fact that there is one physical route from A to C via B, 4 alternatives (1_2, 1_2a, 1a_2, 1a_2a) can be determined based on the performance of an activity or not (skipping the activity. The same applies to route from A to C directly (via shortcut). While there is one physical route, 2 alternatives (5, 5a) can be recognized based on the performance of activity or not. However due to the fact that there is no activity along the shortcut, the alternative 5a does not exist. So the choice set consists of 5 alternatives.
Figure 6.9 Visualization of joint activity route choices in area AC. The choice 06 is illustrated in case that there is an activity in the shortcut.
6.3 Data analysis plan for the descriptive statistics

Before the statistical analysis to be conducted in order to examine which factors influence the joint activity-route choice behaviour, a data analysis plan including the hypotheses to be tested is presented. Four questions are formulating in order to test the hypotheses:

- Since the major activity of the SAIL event is the tall ships, it is expected that the majority of visitors of SAIL event would choose to perform an activity in the BC area where there is opportunity to visit tall ships. Which joint activity route choice the majority of pedestrians prefer?
- It is not expected that the age or gender will have significant influence on the activity-route choice model. However, it is possible that elderly people would prefer shorter routes to perform activities. Furthermore, considering the fact that an individual behaves on a different way comparing when he belongs to a group it is expected that the group size may have an influence on the decision of performing of activity. Another expectation relating to the group size, is that the time spent in the area increases when there are more members on the group. This may happens because the group adjust its speed to the speed of the slower member. Does age, gender or group size have influence on the joint activity route choice? Does the group size has an influence on the performance of activity and the area presence time?
- It is expected that the days do not have influence on the joint activity route choice model since during the days of the event no change occurred along this area (different types of activities during the event), which may lead to different choice patterns along the days. Thus another question is formulating: Does the day play a role in the joint activity-route choice?
- It is expected that the longer the time that was spent before arriving in the understudy part the more participants would choose the route via shortcut due to the shorter length because of their potential tiredness. Hence, it is expected that the time spent before arriving in the decision point influence the activity-route choice later. Is there a relationship between the time that a pedestrian chooses to spend before arriving the understudy part of SAIL area (part AC) and the joint activity-route choice?

The above questions will be answered in Section 6.4.

6.4 Descriptive statistics of the factors included in the estimation of the model

An important part of this research is to examine which factors influence the joint activity-route choice behaviour of pedestrians that visited a large scale event, like SAIL and to which extent. Thus, in this Section the factors that influence the activity and route choice model according to literature (Section 2) and real observations from large scale events will be described in order to be used in this study for the estimation of the joint activity route choice model. Descriptive analysis is conducted to explore the relationships between the joint activity–route choice and the influential factors that were specified in Section 2.3.

As it is mentioned in the previous section the understudied area (Section 6.2) for the estimation of the model is the area from A to C with direction from the station to Verbindingsdam. Regarding the sample size of the GPS data which is related to this specific area, it was found the 162 trips crossed the AC area (trips that arrived at box2, Figure 5.12, Section 5.2.3).

**Information about the examined dataset**

A description of the database is summarized in Table 6.1.

From the total examined sample of 162 trips, 61% of them are performed by males while the average age of the participants is fluctuating between 45-64 years old. The most dominant group size consists of 2 persons. Hence, it can be said that the average participant of this study is probably middle aged-couples.
Furthermore, the majority of the trips starts in the morning and the most preferred days to visit the area during the event were Friday and Saturday.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Frequency</th>
<th>Max Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>99</td>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;18</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td>26</td>
<td>45-54 &amp; 55-64</td>
</tr>
<tr>
<td>45-54</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>65+</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td><strong>Group size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2 persons</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>3 persons</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>4 persons</td>
<td>31</td>
<td>2 persons</td>
</tr>
<tr>
<td>5 persons</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>6 persons</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>7 persons</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>Time of day</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morning</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>Afternoon</td>
<td>55</td>
<td>Morning</td>
</tr>
<tr>
<td>Evening</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td><strong>Day of the week</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20/08 (Thursday)</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>21/08 (Friday)</td>
<td>53</td>
<td>Friday &amp; Saturday</td>
</tr>
<tr>
<td>22/08 (Saturday)</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>23/08 (Sunday)</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td><strong>Joint activity- route Choice</strong></td>
<td></td>
<td>Choice 2</td>
</tr>
<tr>
<td>1,2 (Choice 1)</td>
<td>36 (22.5%)</td>
<td></td>
</tr>
<tr>
<td>1,2a (Choice 2)</td>
<td>57 (35.6%)</td>
<td></td>
</tr>
<tr>
<td>1a,2 (Choice 3)</td>
<td>9 (5.6%)</td>
<td></td>
</tr>
<tr>
<td>1a,2a (Choice 4)</td>
<td>27 (16.9%)</td>
<td></td>
</tr>
<tr>
<td>5 (via Shortcut)</td>
<td>33 (19.4%)</td>
<td></td>
</tr>
<tr>
<td>5a</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>

Regarding the joint activity route choice behaviour of the visitors, it is noticed that their distribution over the alternatives varies. More than one third of them choose to walk via Bimhuis without performing an activity (take a photo, have drink/food, enjoy the view) in area AB, while they prefer to perform an activity in the second part (area BC), where they can enjoy the view of the tall ships and take some photos.
In addition to this, it is noticed that almost equal percentage of visitors choose Choice 1(1_2) as the ones that choose the Route via shortcut. The main difference between these choices is that in the first one there are various attractions related to the event (like tall ships, food/drinks) while in the second one there is no activity relevant to the event. Lastly only a 16.9% of visitors walk via Bimhuis (Area AB and BC) and take advantage of performing activities in both parts.

This section addresses the factors that will be used for the model development. However, a first analysis on the relationship between the factors and joint activity route choice is provided. Before analysing the potential relationships between the factors and the joint activity route choice, the physical attributes of the routes (independently from the activity element) are provided (Table 6.2).

From the total number of 162 trips 80% of them choose to take the route via Bimhuis, while the rest 20% choose the shortcut in order to arrive at destination C. Considering C as the final destination to perform an activity, it is obvious that the majority of pedestrians prefer the route with the longer distance, meaning that there are other factors (having access to the tall ships, being part of the indicated route) than distance that determine their choice. However it is interesting to examine the influence of distance on the performance of activity. In literature it was found that if there is a choice between two equal services then the closest is often selected, unless there are specific preferences for the other services. In this understudy case, and more specifically in the area AC which consists of 2 parts (AB and BC), it is not easy to compare the quality of the services of the existing activities in each part. Not to mention that the available data sources in this specific area (GPS, cameras) are not able to capture the type and location of activity (Section 3.3). However, it is observed that more people preferred to perform an activity along the BC area instead of the AB area (35.6% versus 5.6%). This can be attributed to the fact that in the BC area, a pedestrian has the opportunity to watch or visit the tall ships which is the major and dominant activity of the SAIL event. Thus it is expected that more visitors would prefer the BC part than the AB part.

Table 6.2 Physical attributes of the route alternatives from A to C.

<table>
<thead>
<tr>
<th>Physical attributes</th>
<th>Length</th>
<th>Width</th>
<th>Directness</th>
<th>Existence of signs</th>
<th>Existence of attractions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Route via Bimhuis</td>
<td>600</td>
<td>24.5</td>
<td>Turn</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Route via shortcut</td>
<td>90</td>
<td>16</td>
<td>Direct route</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

The shares of pedestrians that choose not to perform an activity are distributed either on Route 1_2 or Route 5 (via shortcut). It is seen that pedestrians consider Route 1_2 (Choice 1) more attractive than Route 5 (Choice 5).

In Route 5 (which belongs to the physical network from A to C directly) there is not the option of performing an activity since there is no available activity, while in Route 1_2 the pedestrians choose to skip the available activities along the physical network from A to C via B. Apart from the existence of activities or not, another difference between the 2 routes is the walking distance. The length of Route 5 is around 100 meters while the walking distance (taking the same reference point) of Route 1_2 is 6 times longer (600 meters). It seems that distance does not influence the choice between the 2 routes in a way that the shorter route will be chosen. This means that even if the walking distance of the route is shorter, pedestrians prefer taking a longer route for a combination of reasons like the attractiveness of the environment, the feeling that in case they change their mind regarding the performance of activity, they
will have the option to perform an activity, the existence of signs indicating the route via Bimhuis and the herding behaviour. Thus the attractiveness of the view and the existence of activities compensate for the extra distance.

![Figure 6.10 Distribution of pedestrians that do not perform an activity](image)

**Relationship between socio-demographic characteristics of pedestrians (gender, age, group size) and joint activity-route choice**

A Pearson chi-square test was conducted to examine whether there is a relationship between socio-demographic characteristics (gender, age, group size) and joint activity-route choice. In Table 6.3 the results revealed that there is no statistically significant association (at 95% confidence interval) between the sociodemographic characteristics and the joint activity-route choice. In addition, Table 6.4 presents the distribution of each of the sociodemographic factors over the joint activity route choices.

<table>
<thead>
<tr>
<th>Chi-Square Tests</th>
<th>Pearson Chi-Square Value</th>
<th>df</th>
<th>Asymp. Sig. (2-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>8.409</td>
<td>4</td>
<td>.078</td>
</tr>
<tr>
<td>Age</td>
<td>.403</td>
<td>24</td>
<td>.340</td>
</tr>
<tr>
<td>Group size</td>
<td>14.602</td>
<td>24</td>
<td>.932</td>
</tr>
</tbody>
</table>

Regarding the gender, according to literature women tend to choose more complex routes than men. However, Figure 6.11a, shows that the distribution of both males and females over the joint activity-route choice is almost similar. The main difference is that women tend to prefer more the route via shortcut in comparison with the Route1_2, while for men this is not the case. It is important to be noted that the influence of gender in the joint activity-route choice is difficult to be captured in this study, since it is only known the gender of the participant that carried the GPS device while as it can been seen in Table 6.1 the great majority of participants were not alone, but part of a group. The gender of the other members of the group is not known, so it is not safe to reach to a conclusion regarding the relationship between the gender and the joint activity-route choice.
Table 6.4 Distribution of sociodemographic characteristics over the joint activity route choices

<table>
<thead>
<tr>
<th>Gender</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Route 1_2</td>
<td>Route 1_2a</td>
</tr>
<tr>
<td>Gender</td>
<td>19.0%</td>
<td>44.4%</td>
</tr>
<tr>
<td>Male</td>
<td>24.2%</td>
<td>29.3%</td>
</tr>
<tr>
<td>Age</td>
<td>&lt;18</td>
<td>33.3%</td>
</tr>
<tr>
<td></td>
<td>18-24</td>
<td>16.7%</td>
</tr>
<tr>
<td></td>
<td>25-34</td>
<td>20.8%</td>
</tr>
<tr>
<td>Age</td>
<td>35-44</td>
<td>30.8%</td>
</tr>
<tr>
<td></td>
<td>45-54</td>
<td>18.6%</td>
</tr>
<tr>
<td></td>
<td>55-64</td>
<td>23.3%</td>
</tr>
<tr>
<td></td>
<td>65+</td>
<td>18.2%</td>
</tr>
<tr>
<td>Group size</td>
<td>individual</td>
<td>50.0%</td>
</tr>
<tr>
<td></td>
<td>2 persons</td>
<td>25.0%</td>
</tr>
<tr>
<td></td>
<td>3 persons</td>
<td>21.7%</td>
</tr>
<tr>
<td>Group size</td>
<td>4 persons</td>
<td>12.9%</td>
</tr>
<tr>
<td></td>
<td>5 persons</td>
<td>20.0%</td>
</tr>
<tr>
<td></td>
<td>6 persons</td>
<td>20.0%</td>
</tr>
<tr>
<td></td>
<td>7 persons</td>
<td>50.0%</td>
</tr>
</tbody>
</table>
As for the age, in literature it was found that different age groups make their choices based on different reasons that are valued in a different way (either higher or lower). In Figure 6.11b, it is obvious for all age groups that when the route via Bimhuis is chosen, they prefer to perform an activity at the area after the Bimhuis (area BC). Thus, route 1_2a was mostly preferred by all age groups. Another interesting point is that the shortcut was taken regardless of the age of the participant, but there is a huge difference in the percentage within each of the age group that took this route. For example, only 9% of the age of 65+ chose the shortcut, while around 50% of the 18-24 years old took this route. A possible explanation for this is that younger people have a more explorative behaviour and instead of following the signs of the orange route they took the shortcut. On the other hand, it should not be omitted that the route via shortcut is shorter in length so it would be expected that more elderly people would choose shorter routes, due to their physical characteristics. However, according to the findings this is not the case. Therefore, further analysis in the combination of different factors that influence the choice of pedestrians according to their age may give more reliable deductions.
Finally, regarding the group size, nearly no variation is present in the distribution over the joint activity-route choices. This can indicate that the impact of the group size on the joint activity route choice is smaller than it was expected. Since the group factor was considered according to findings in literature as an influential factor on the decision making, a further analysis in the relationship between the group size and both the performance of activity and the area presence time is conducted below.

**Relationship between group size and performance of activity.**

From the total dataset it was found that 58.1% of the pedestrians performed an activity, while the rest 41.9% preferred either to skip the activities (Route 1_2) or not to perform any activities because they did not have the opportunity given the chosen route (Route 5). The majority of the participants that took part in the GPS data collection was not alone but they were part of a group.

A Pearson chi-square test was conducted to examine whether there was a relationship between the group size and the performance of activity. The results revealed that there was not a significant relationship between the two variables (Chi square value = 0.802, df =6, p > .001). In other words, the probability of performing an activity during the event is not influenced by the group size (Figure 6.12).

**Figure 6.12 Relationship between the group-size and the performance of activity or not**

**Relationship between group size and area presence time**

In Figure 6.13 for each group size, it is noticed that there is a peak in the 20 minutes of the area presence time (or less). Thus, it seems that regardless from the group size that pedestrian belongs there is a preference of spending 20 minutes in the understudy area. This means that the area presence time is not influenced by the group-size. It was expected that the size of the group would have a relationship with the area presence time since the walking speed of the group would be influenced by the heterogeneity of age or the physical condition of the members of the group. Moreover, in case of large group, it is possible to take more time to decide whether to perform an activity or not, due to difference in the preferences between the members of the group. However, this is a conjecture and it needs further research on how the size and composition of the group could influence the choices of pedestrians during the event.
The relationship between the joint activity-route choice and the day is investigating. It was found that the distributions of the joint activity-route choices over the days are similar (according to Pearson chi-square test). The significance is 0.08, which is higher than 0.05 belonging to the 95% confidence interval (so there is no effect). This was expected since during the days of the event no changes occurred along this area (different types of activities during the event), which may lead to different choice patterns along the days. A clear difference between the days is that on the 22nd of August (Saturday) a higher percentage of visitors chose to walk via Bimhuis area and perform activities along this route (area AC and area BC) (Figure 6.14). However, it cannot be said that during this specific day occurred something different from the rest days of the event.

**Relationship between day and joint activity-route choice**

![Figure 6.13 Distribution of the area presence time over the group-size](image)

![Figure 6.14 Distribution of joint activity route choices along the days](image)
Relationship between time of day and joint activity-route choice

For this analysis each day (20-08-2015 until 23-08-2015) is divided in 4 parts: morning (09:00-12:00), afternoon (12:00-17:00 pm), evening (17:00-21:00) and night (21:00-24:00). The categorization “night” was excluded due to limited amount of observations (only one observation). A sample of less than 25-30 measurements may lead to bias on the reliability of estimates error (over or underestimate the true error).

The relationship between time of day and joint activity-route choice is examined by conducting a Pearson chi-square test. The results showed that there is a statistically significant association (at 95% confidence interval) between the time of day and the joint activity-route choice (Table 6.6).

Table 6.5 Distribution of joint activity route choices along the days

<table>
<thead>
<tr>
<th>Time of day</th>
<th>Route 1_2</th>
<th>Route 1_2a</th>
<th>Route 1a_2</th>
<th>Route 1a_2a</th>
<th>Route via shortcut</th>
</tr>
</thead>
<tbody>
<tr>
<td>morning (09:00-12:00)</td>
<td>23.5%</td>
<td>34.6%</td>
<td>9.9%</td>
<td>18.5%</td>
<td>13.6%</td>
</tr>
<tr>
<td>afternoon (12:00-17:00)</td>
<td>12.5%</td>
<td>41.1%</td>
<td>1.8%</td>
<td>17.9%</td>
<td>26.8%</td>
</tr>
<tr>
<td>evening (17:00-21:00)</td>
<td>40.0%</td>
<td>24.0%</td>
<td>0.0%</td>
<td>8.0%</td>
<td>28.0%</td>
</tr>
</tbody>
</table>

Figure 6.15 Time of day distribution over the joint activity-route choice
Table 6.6 Relationship between time of day and joint activity-route choice

<table>
<thead>
<tr>
<th>Chi-Square Tests</th>
<th>Pearson Chi-Square Value</th>
<th>df</th>
<th>Asymp. Sig. (2-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-Square</td>
<td>18.006a</td>
<td>8</td>
<td>.021</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>19.577</td>
<td>8</td>
<td>.012</td>
</tr>
<tr>
<td>Linear-by-Linear Association</td>
<td>.227</td>
<td>1</td>
<td>.634</td>
</tr>
<tr>
<td>N of Valid Cases</td>
<td>162</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.15 illustrates that in the morning (09:00-12:00) and in the afternoon (12:00-17:00) the majority of the participants choose to perform an activity at the area BC by taking Route 1_2a. In the evening until late night, it seems that pedestrians tend to choose routes (Route 1_2 and Route 5) where they do not perform an activity. A reason why they skip the activities could be the available time that they had in order to walk around the whole area.

Relationship between signs and joint activity-route choice

In literature it was found that during mass events, some information measures by means of signs for instance, which are related to the main attractions, routes crowdedness etc., can have a large influence on route choices of pedestrians and that it could even steer or spread the crowd over the network (Zomer, 2014). In the SAIl case and more specifically for the understudied area AC it was seen that there were signs indicating the Orange route part of which is the route via Bimhuis. These signs were just indicating the route and not the activities along the route. However, due to the nature of the event which is related with tall ships, since the signs were indicating the route along the sea it is implying that along this route (part of the orange route) there would be the opportunity to watch the tall ships. Moreover, it should not be omitted the fact that in many cases during mass events, pedestrians have a recognitive state of mind and are lacking of knowledge on where the activity locations are in order to perform an activity. The same may happens during the SAIL event. Thus, this lack of knowledge of the area could be a reason for a pedestrian to follow the route indicating by the signs which passes via the activities.

A comparison between the pedestrians that chose the route via Bimhuis and the route via the shortcut is done. It is found that from the total number of 162 trips 80% of them chose to take the route via Bimhuis which is indicating by signs, while the rest 20% chose the shortcut in order to arrive at destination C; the shortcut is not indicated by signs. Therefore it seems that the existence of signs influences the route choice.

Relationship between the time that the GPS device was handed out and the start time of the trip in the understudy part of SAIL area (part AC)

The majority of the participants appeared at the understudy area (AC area) a few minutes later (around 20 minutes later) after the handing out of the GPS device. Regarding the number of trips that appeared at the AC area 1-4 hours later after the distribution of the GPS devices, a percentage of 67% took the route via Bimhuis and the rest 33% the route via the shortcut. It would be expected that the longer the time that was spent before arriving in the understudy part the more participants would choose the route via shortcut due to the shorter length because of their potential tiredness. Hence, it is implied that the time spent before arriving in the decision point does not influence the route choice later.
At this point, it would be interesting to see if the time spent before influence the performance of activity or not. As it was mentioned, from the total number of participants that spent 1-4 hours in another area before having arrived at the AC area, 67% of them took the route via Bimhuis and 2/3 of them performed an activity along this route. It seems, that performing an activity has stronger influence than walking a shorter distance via the route with the shortcut.

The relationship between the time that a pedestrian choose to spent before arrives at the understudy area and the joint activity-route choice is examined by conducting a Pearson chi-square test. The results revealed showed that there is no statistically significant association (at 95% confidence interval) between the two examined variables.

Table 6.7 Distribution of joint activity route choices along the days

<table>
<thead>
<tr>
<th>Chi-Square Tests</th>
<th>Pearson Chi-Square Value</th>
<th>df</th>
<th>Asymp. Sig. (2-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-Square</td>
<td>6.724a</td>
<td>4</td>
<td>.151</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>6.416</td>
<td>4</td>
<td>.170</td>
</tr>
<tr>
<td>Linear-by-Linear Association</td>
<td>5.197</td>
<td>1</td>
<td>.023</td>
</tr>
<tr>
<td>N of Valid Cases</td>
<td>162</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relationship between the joint activity route choice and the area presence time (time spent) in the understudy part of SAIL area (part AC)

As it was mentioned the area presence time consists of the walking time and the activity duration time. Note that the activity duration is the time spent on everything besides walking, so walking time can reflect the prevailing conditions and the activity duration time can be an indicator for the performance of activity or not.
The pedestrian can choose between 5 alternatives of joint activity route choices. A pedestrian has a different time spent towards each of these alternatives. The data collected by means of GPS devices shows the time spent over the chosen alternatives. The total number of observations per alternative differs. In Figure 6.17 the distribution of the time spent over the chosen alternatives as it was derived from the GPS data is illustrated.

However, the area presence time consists of the travel (walking) time and the activity duration time. Thus, these 2 factors are going to be investigated together. Note that the activity duration time is the time spent on everything besides walking (watching an event, visiting a tall ship, spending time buying a drink, eating), so walking time can reflect the prevailing conditions and the activity duration time can be an indicator for the performance of an activity or not.

The total number of observations per joint activity route choice alternative differs. This is the reason why the bars are higher in the Route1_2, Route 1_2a, and Route via the shortcut. An interesting observation from Figure 6.17 is the occurrence of peaks. At around 15 minutes a peak arises in the number of observations for the cases Route1_2 (skipping the activities), Route 1_2a (performing activities in the second part (BC)) and Route 1a_2 (performing activities in the first part (AB)). Thus, it can be said that performing an activity in either of the parts (AB or BC) or not performing an activity along the route via Bimhuis, the peak is the same, meaning that the pedestrians of this sample have a high preference for being in the area for around 15 minutes regardless from performing an activity or not. This implies that the majority of pedestrians tend to have a trade-off between the walking time and the activity duration time so that the total time spent in the area will be around 15 minutes.

![Histograms showing area presence time distribution for different route choices.](image-url)
Taking into account the choice of performing activities in both parts of the area (Route1a_2a) there is a peak in 30 minutes, which is related with the peaks found in the routes that pedestrian perform activities in one of the parts of the route via Bimhuis. Lastly, for the direct route (via shortcut) a peak is visible at around 2 minutes. This happens because the length of the route is short (90 meters) and there is no activity to perform. Thus, for the shortcut the area presence time is equal with the walking time.

The average area presence time (time spent) also differs per alternative. Since the distribution of the area presence time is not normal but skewed, then the median is taken as a better measure to compare the average of the alternatives. Route 1a_2a and the Route 1a_2 have the highest area presence times (36:30 minutes and 22:24 minutes respectively). Regarding the average area presence time in Route 1_2a is 6 minutes less than the time spent in Route 1a_2 implying that pedestrians prefer to spend more time in the option of performing activities in the first part (AB). As for the Route 1_2 and Route via the shortcut, where the pedestrians do not perform any activities, the area presence time is equal to the walking time. Since Route 1_2 is much longer than the one with the shortcut, it is expected that it takes more time to walk along Route 1_2.

The standard deviation also differs per alternative. Route1_2 and Route via shortcut have the lowest standard deviation. This means that the spread in area presence time is limited. In essence, for these 2 alternatives the area presence time is equal to the walking time since no activity is apparently performed.

The standard deviation of Route 1a_2 is 42:19 minutes and for Route 1a_2a is 28:20 minutes. The average area presence time to the Route1a_2 was lower than Route1a_2a, but the spread is larger. This can indicate that this factor indeed influences the joint activity route choice.

**Relationship between the joint activity route choice and the walking time in the understudy part of SAIL area (part AC)**

Figure 6.18 presents the distribution of the walking time over the chosen alternatives. The walking time can reflect the prevailing conditions. It can be noticed that at 15 minutes a peak arises in the number of observations that have chosen the joint activity route choice alternatives (Route1_2, Route1_2a, Route1a_2, Route1a_2a) that belong to the same physical network. The walking time ranges from 8-15 minutes.
Regarding Route 5, most of the pedestrians walked to this area for 2.5 minutes.

The walking time is independent from the performance of activity or not, but apart from personal factors, it is influenced by the prevailing conditions during the day.

Figure 6.18 Walking time distribution over the joint activity route choices
This is the reason why an investigation in the relationship between the walking time and the time of day is presented below. It is obvious that for every alternative the average walking time increases during the day. Hence, it can be said that the time of day influences the walking time. However, the distributions of the average walking time along the day for each alternative do not differ significantly.

Regarding the alternatives that are derived from the physical route via B, the average walking time during the morning ranges from 11 to 13 minutes, during the afternoon between 12:55-13:00 minutes and during the evening between 13:21-13:31. Figure 6.19 shows that with the exception of the route via the shortcut, all the other joint activity-route choices have almost equal average walking time for each time of day (morning, afternoon, evening).

**Figure 6.19 Distribution of average walking time according to the time of day for each alternative**

**Relationship between the joint activity route choice and the activity time spent in the understudy part of SAIL area (part AC)**

The activity time was derived by combining the data from the GPS and the Wi-Fi sensors.
An interesting observation from Figure 6.20 is the occurrence of peaks. Each joint activity-route choice has a different peak regarding the duration of activity. At 2-3 minutes a peak arises in the number of observations for the case Route1_2a, while for the case of Route1a_2 a peak arises at 6 minutes and at more than an hour. This difference in the peaks implies that the pedestrians showed a preference to spend more time performing an activity at the part from A to B in comparison with the activity time spent in the part from B to C. This is related to the type of activities that the different parts of the route offer. For instance, eating or drinking at a restaurant demand more time than waiting in a canteen to buy something to eat. As it was aforementioned the first part (from A to B) offers the possibility to sit at the Bimhuis to enjoy food and drinks or coffee, while the type of activities offered in the second part cannot be so time consuming.

As for Route1a_2a there are 2 peaks; one at 14 minutes and another one at 18 minutes. It is interesting to see how the activity time is distributed over the 2 different parts (from A to B and from B to C) given that activities are performed in both parts. For that purpose, the distribution of the activity time over the different part of the area is illustrated in Figure 6.21. It is obvious that for the pedestrians that choose to perform activities in both part of the network there are different peaks in the activity time for each parts. In AB area there is a peak at 12:00 minutes, while in BC the highest frequency is met at 4 minutes. This may implies that when someone has already performed an activity he tends to spent less time in the next activity. To have a better insight into the distribution of activity time in both parts of the activity locations, a diagram indicating the activity time spent (splitted into the 2 parts of the network)for each pedestrian/trip that made the choice of Route1a_2a is presented (Figure 6.22).
Figure 6.21 Distribution of activity times for each area separately (AB, BC) when Route1a_2a is chosen

Figure 6.22 Individual activity times for every pedestrian that chose to perform activities in both parts of the network.

6.5 Estimating the joint activity route choice model

In Section 6.5.1 the interpretation of the factors that influence the joint activity route choice into operational attributes in order to be included in the model is described. In Section 6.5.2 the joint activity route choice model is formulated and finally it is estimated.

6.5.1 Operationalization process of the factors included in the model

After an extensive analysis both on the attributes that are included in the model as well as their relationship with the joint activity –route choice, this section focuses on the translation of the influential factors into operational attributes before being included into the model estimation. Both dummy and scale variables were used in order to express them in a way to be included in the model. The selected factors that are included in the joint activity route choice model and the way that they were quantified are being elaborated below. An overall description of the way that the variables were included in the model is presented in Table 6.8.
Table 6.8 Variable description

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Dummy variable</td>
</tr>
<tr>
<td>Age</td>
<td>Effect coding</td>
</tr>
<tr>
<td>Group size</td>
<td>Effect coding</td>
</tr>
<tr>
<td>Time of day</td>
<td>Effect coding</td>
</tr>
<tr>
<td>Day</td>
<td>Effect coding</td>
</tr>
<tr>
<td>Existence of signs</td>
<td>Dummy variable</td>
</tr>
<tr>
<td>Existence of attractions</td>
<td>Dummy variable</td>
</tr>
<tr>
<td>Crowdedness</td>
<td>Dummy variable</td>
</tr>
<tr>
<td>Area presence time (walking+activity time)</td>
<td>Scale variable</td>
</tr>
<tr>
<td>Activity time</td>
<td>Scale variable</td>
</tr>
<tr>
<td>Walking time</td>
<td>Scale variable</td>
</tr>
<tr>
<td>Time spent at the area before arriving at the decision point</td>
<td>Scale variable</td>
</tr>
<tr>
<td>Distance from the 1st performed activity</td>
<td>Scale variable</td>
</tr>
</tbody>
</table>

**Gender**

The gender is a dummy variable controlling whether the respondent is female or male. If the variable is a female then is given a 1 otherwise it is 0. Dummy coding of independent variables is quite common. Each category’s dummy variable has a value of 1 for its category and a 0 for all others. Dummy coding is used because the effects of this variable on other factors that influence the joint activity-route choice need to be tested. By using dummy coding the interaction can be modelled in the simplest way.

**Age**

As it was mentioned in the Section 6.4, there are seven categories of age groups. Due to the fact that there is a lot of variability in the joint activity route choices, a larger sample might be needed for some of the age groups. For this reason the seven categories are merged into three age groups: under35, 35-54, 55 plus (young adults, adults and elderly respectively). The coding used is effect coding. Having 3 variables means that 2 indicator variables need to be used. This is shown below:
<table>
<thead>
<tr>
<th>Age groups</th>
<th>Indicator 1</th>
<th>Indicator 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 35 (young adults)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>35-54 (adults)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>55+ (elderly)</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

The combination of these indicator variables result in different parameters for each age group.

**Group size**

In Table 6.4 (Section 6.4) the distribution of each group size along the joint activity route choices are presented. However, it is clear that there is a high variability in the distribution and a larger sample is needed for some group categories. Thus, the initial seven categories of group size are merged into five group size categories. The new categorization is the following: individual, 2 persons, 3 persons, 4 persons, more than 5 persons.

The group size factor is also coded using effect coding. With five categories of group size, four indicator variables are introduced. The effect coding is presented below:

<table>
<thead>
<tr>
<th>Group Size</th>
<th>Indicator 1</th>
<th>Indicator 2</th>
<th>Indicator 3</th>
<th>Indicator 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2 persons</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3 persons</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4 persons</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>More than 5 persons</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

**Time of day**

The time of day is also coded using effect coding. With three categories of time of day, this means that two indicator variables are introduced.

<table>
<thead>
<tr>
<th>Time of Day</th>
<th>Indicator 1</th>
<th>Indicator 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>morning</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>afternoon</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Evening</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

**Day**

The time of day is coded using effect coding. With four categories of day, this means that three indicator variables are introduced.
<table>
<thead>
<tr>
<th>Indicator 1</th>
<th>Indicator 2</th>
<th>Indicator 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thursday</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Friday</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Saturday</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sunday</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

Existence of signs-Existence of attractions- Crowdedness

Dummy variables were used for both the existence of signs and the existence of attractions. Thus, if there are signs in the route then the existence of signs is 1 if not is 0. The same applies for the existence of attractions. Regarding the crowdedness, it was seen in Section 3.3 (Table 3.1) that the crowdedness can be captured by Wi-Fi and counting cameras. However, in the selected area for the estimation of the model, the crowdedness is not expressed as a scale variable but as dummy. This is because four out of five alternatives share the same physical network implying that the crowdedness is the same (no variation) for these alternatives. As for the shortcut (the 5th alternative), the estimation of the crowdedness is not possible due to the absence of Wi-Fi sensors or cameras at this location.

Area presence time- Walking time-Activity Time

A pedestrian has a different area presence time towards each of the available alternatives. The data collected by means of GPS data indicate only the area presence time towards the chosen alternative. This means that the calculation of the non-chosen alternatives is necessary.

The area presence time consists of the walking time and the activity duration time. Four out of five alternatives (Route1_2, Route1_2a, Route1a_2, Route1a_2a) share the same physical network but they perceive it on a totally different way depending on their desire to perform an activity. These four alternatives have the same walking time but the activity time for each alternative will probably be different. Regarding the alternative Route 5 (shortcut), the walking times are different from the ones of the aforementioned alternatives.

Beginning with the calculation of the walking time of Route1_2, Route1_2a, Route1a_2, Route1a_2a for each time moment, the Wi-Fi data are used. The values are dynamic so for each pedestrian that enters to the area at a specific time moment the average walking time of the same time instant is derived from the Wi-Fi data. For example, if the chosen alternative is Route1a_2 and the choice is made at 09:30, then the walking time of this alternative is derived by the average walking times from the Wi-Fi sensors for this specific time. The no chosen alternatives (Route 1_2, Route1_2a, Route1a_2a) have the same walking time as the chosen alternative Route 1a_2 since they are referring to the same physical network. On the contrary, the calculation of the walking time of the shortcut (Route5) in case of not being chosen is done by taking the average of the walking times of the pedestrians that choose to walk in this shortcut, the same day and the same time of day. Otherwise if the observations are not enough to derive an average, the walking time is calculated by using the following formula:

\[ Walking\ time(Route5) = \frac{Length\ of\ Route5(shortcut)}{speed} \]

Where speed is assumed to be 3km/h. This assumption is based on the dominant age group (45-54 & 55-64) that participated in the GPS data collection (van Venrooij, 2011).

Regarding the calculation of the activity times of the no chosen alternatives different ways were followed depending on the chosen alternative.
In particular if the chosen alternative is Route 1_2 or Route 5, where the activity times are 0, then an assumption was made that if the same pedestrian has chosen to perform an activity in the 1st part (AB) then he/she would have spent the activity time as indicated by the Wi-Fi sensors (1299-1295 Wi-Fi sensors) that cover the area AB (Figure 6.23). For the no chosen alternative Route 1_2a, where a pedestrian performs activity in the second part BC, it is impossible to derive an accurate activity time value from the Wi-Fi sensors (1295-40, Figure 6.23) for the BC area since they are located in a long distance. For this reason the median activity time of pedestrians that have chosen to perform an activity at BC area was taken during the same day and the same time of day.

In case that Route 1_2a is the chosen alternative, it is assumed that the activity time of the non-chosen Route1a_2a will consist of the activity time spent in the first part (AB) derived from the Wi-Fi sensors while there is an assumption that the pedestrian would choose to spend the same activity time in the second part as he would do in the chosen alternative Route 1_2a. For instance, if someone chooses to spent 15 minutes visiting a tall ship in the second part (BC) of the area, he would also choose to spent 15 minutes in the same part if he had chosen to perform another activity in the first part (drink coffee). This is based on the assumption that pedestrians during the event do not feel restricted by the time (no time constraints) and have an event state of mind.

![Figure 6.23 Locations of Wi-Fi sensors used for the estimation of the activity and walking time.](image)

6.5.2 Model formulation- Results and findings

When a pedestrian considers making a trip, he/she is assumed to have prior knowledge of a set of known travel alternatives. The pedestrian’s decision is modelled by postulating that he/she chooses from a choice set that contains all known travel alternatives. The pedestrian’s choice from this set is driven by utility maximization: he/she compares all known travel alternatives and chooses the one with the highest utility.

In this section the joint activity-route choice model is estimated using the software BIOGEME (Bierlaire, 2003). The multinomial model (MNL) is the most basic model that takes more than 2 alternatives into account. Therefore the MNL model is the starting point in the model estimation and it is estimated as the base model. An overview of the results from the joint activity-route choice models created with BIOGEME are presented. The modelling process applied for joint activity-route choice is the following:

- Test each factor or process individually on its impact or significance for joint route-activity choice.
- Combine the significant factors in a stepwise way in order to create the best model. More specifically, start with the model that has the highest rho-square and combine the factors stepwise.
The first step is to create a model for each of the factors. In each of these basic models alternative specific constants (ASC) are used that can capture all unobserved attributes of the model and show the base preference of the pedestrians towards the alternatives. Therefore, an initial model is defined that contains only the Alternative Specific Constant (ASC). Consequently, the first predictor is introduced. It is determined whether or not the predictor improves the model based on the p-value. If p-value is greater than 0.05 then it is concluded with 95% confidence that the predictor is significantly different from its true value (Inès, 2010). If the predictor improves the prediction then the variable is included in the model, otherwise it is excluded and the next predictor is added.

The ASC for Route 1_2 was kept constant and fixed at zero value which is a reasonable starting value for ASC in BIOGEME.

The results of the initial model (only ASC) are demonstrated below.

<table>
<thead>
<tr>
<th></th>
<th>ASC</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rho-Square (R²)</td>
<td>0.077</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted Rho-Square (adj. R²)</td>
<td>0.062</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Log-Likelihood</td>
<td>-240.583</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant on a 95% confidence interval

With the exception of alternatives Route 1A_2A and the Route via the shortcut, it is clear that the p-values of the rest of the alternatives are less than 0.05 meaning that the ASCs are significant. The ASC of alternative Route 1_2A has the most positive value in comparison with the values of the other alternatives, so this route is most attractive. While ASC of Route 1A_2 has for all attributes the most negative value, so this route is least attractive.

The major challenge of the estimation of the joint activity route choice model is to capture the behaviour of two different types of pedestrians; one that has a preference to perform activities and the other that does not. This means that it is not possible to describe with one beta value the behaviour of two completely different groups of people.

Thus it is necessary to include some of the personal preferences of the pedestrians in a way that the desire to perform an activity or not is reflected. If pedestrians have a desire to perform an activity then it is more likely to choose an option with an activity. In this case the estimation of the coefficient of the walking time will be different than the one of not performing an activity. Thus the walking time is valued on a different way based on the preference to perform an activity.

This means that if a pedestrian chooses to perform an activity then apparently considers the activity time more important than walking, while if he does not have a desire to perform an activity then he/she weighs
higher the walking time than the activity time. In other words, pedestrians depended on their desire to perform an activity, they consider the activity and walking time differently.

In the case of performing an activity the beta of the activity time is expected to be as high as possible and the utility function will be depended on the activity time and on the perception of walking in order to perform an activity. In order to encounter with this issue, a factor of preference to perform an activity is introduced. This factor can have two values depending on the performance of activity or not. If pedestrians perform an activity then it is implied that they have a preference to perform an activity. Thus the factor of preference for activity takes the value 1 and is multiplied by the activity time and the walking time attributes. On the contrary, if people do not prefer to perform an activity, the factor takes a 0 value and the part of the utility which is related with the activity time becomes 0. On the other hand, a beta of the walking time without the purpose of performing an activity will be estimated and it will be multiplied by the walking preference factor which is equal to (1 - activity preference). This is a way to explain how pedestrians will interpret the same utility function in a very different manner depending on whether they have a preference to perform an activity.

The form of the utility function including the attributes multiplied by preference factors (preference for activity, preference for walk) is presented below:

$$ Ui = ASCi + \beta_{Activity Time} * Activity Time_i * preference_{Activity}$$

$$ + \beta_{Walking Time} * WalkingTime_i$$

$$ + \beta_{Walking Time(performing activity)} * WalkingTime_i(performing activity) * preference_{Activity} $$

Table 6.10 shows an overview of the parameters/attributes estimated for each individual model, with the value coefficient, p-value, log-likelihood and adjusted rho-square value. In each of these basic models alternative specific constants (ASC) are used to capture the unobserved attributes. An extensive overview of the model results can be found in Appendix C.

According to the results in Table 6.10, the factors that do not provide any significant contribution to the utility of the alternatives are the following: the socio-economic characteristics (gender, age, group size), the time of day (morning, afternoon, evening), the day (Thursday, Friday, Saturday, Sunday), the time spent before arriving at the decision point of the joint activity-route choice, the distance from the 1st performed activity, the existence of signs and activities and the crowdedness.

The gender is a dummy variable controlling whether the respondent is female or male. The two parameters need to be estimated in order to examine their influence on the joint activity-route choice of pedestrians that are participating in the SAIL event. The results of the statistic test show that the gender does not show a significant influence on the choice of pedestrians. Thus there is no difference in the choice behaviour of males and females. However, in the remainder of this thesis, the pedestrians that filled in the survey and carried the GPS devices were not alone but part of the group. The gender of the other members of the group is not known, so it is not safe to reach to a conclusion regarding the relationship between the gender and the joint activity-route choice. In addition, it seems that the age does not influence the choices of the participants. Similarly as the gender, the only known age is the age of the person that carried the GPS device, while the choices derived from the GPS data analysis could indicate a group behaviour instead of an individual, since almost all the participants were part of a group (Table 6.1). Regarding the group size, it does not influence the joint activity route choice model, and there is no difference in the choice behaviour of pedestrians that belong to different group sizes.
<table>
<thead>
<tr>
<th>MNL</th>
<th>Parameters</th>
<th>Values</th>
<th>Log likelihood</th>
<th>Adjusted $\rho^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Walking time</td>
<td>p-value Coef. 0.0* -0.4</td>
<td>-211.810</td>
<td>0.172</td>
</tr>
<tr>
<td>2</td>
<td>Walking time (performing activity)</td>
<td>p-value Coef. 0.0* 2.98</td>
<td>-207.100</td>
<td>0.190</td>
</tr>
<tr>
<td>3</td>
<td>Activity time</td>
<td>p-value Coef. 0.0* 0.277</td>
<td>-188.296</td>
<td>0.266</td>
</tr>
<tr>
<td>4</td>
<td>Time spent before the decision point</td>
<td>p-value Coef. 0.49 0.0118</td>
<td>-241.008</td>
<td>0.064</td>
</tr>
<tr>
<td>5</td>
<td>Distance from the 1st performed activity</td>
<td>p-value Coef. 0.0* 0.00116</td>
<td>-245.292</td>
<td>0.059</td>
</tr>
<tr>
<td>6</td>
<td>Signs</td>
<td>p-value Coef. 0.16 0.258</td>
<td>-240.547</td>
<td>0.066</td>
</tr>
<tr>
<td>7</td>
<td>Crowdedness</td>
<td>p-value Coef. 0.16 0.258</td>
<td>-240.547</td>
<td>0.066</td>
</tr>
<tr>
<td>8</td>
<td>Existence of Activity</td>
<td>p-value Coef. 0.16 0.258</td>
<td>-240.547</td>
<td>0.066</td>
</tr>
<tr>
<td>9</td>
<td>Time of day (morning)</td>
<td>p-value Coef. 1 -1.71e-015</td>
<td>-241.251</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>Time of day (afternoon)</td>
<td>p-value Coef. 1 -1.39e-015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Day (Thursday)</td>
<td>p-value Coef. 1 1.13e-015</td>
<td>-240.583</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>Day (Friday)</td>
<td>p-value Coef. 1 -3.18e-016</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Day (Saturday)</td>
<td>p-value Coef. 1 -3.18e-016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Groupsize (individual)</td>
<td>p-value Coef. 1 9.85e-017</td>
<td>-241.251</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>Groupsize (2 persons)</td>
<td>p-value Coef. 1 -1.75e-015</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Groupsize (3 persons)</td>
<td>p-value Coef. 1 -3.09e-016</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Groupsize (4 persons)</td>
<td>p-value Coef. 1 -5.55e-016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Gender (Female)</td>
<td>p-value Coef. 1 2.94e-015</td>
<td>-240.583</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>Gender (Male)</td>
<td>p-value Coef. 1 -4.63e-015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Age (18-34)</td>
<td>p-value Coef. 1 1.81e-015</td>
<td>-241.251</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>Age (35-54)</td>
<td>p-value Coef. 1 1.24e-015</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant on a 95% confidence interval

In addition, the time of day turn out to not provide significant contribution to the utility of the alternatives. However, according to the statistical analysis conducted in Section 6.3, it was found that there is a statistically significant association between the time of day and the joint activity route choice. It seems that during morning and afternoon the pedestrians tend to have a preference for performing an activity, while during the evening they mostly preferred the routes Route 1_2 and Route 5(via shortcut). A possible reason why these findings are not captured by the model may lies on the way that the time of
day was operationalized by using the effect coding. As for the influence of the day in the choice behaviour is not significant since there was slightly no difference on pedestrians’ choices during the days of the event.

Furthermore, the time that pedestrians spent before arriving at the decision point where the joint activity-route choice model is estimated does not influence their decisions afterwards. Hence, spending either 10 minutes or 2 hours for example in the event before arriving at the decision point does not determine the choices of pedestrians. Nevertheless, the statistical analysis in the previous Section showed different results. More specifically, pedestrians that do not want to perform an activity and have spent more than an hour before arriving at the decision point they prefer to take the shortcut instead of Route 1_2. This choice can be related to the feeling of tiredness.

The coefficients of the existence of signs, crowdedness and activities have not been estimated significantly different from zero. These attributes are important part of pedestrians’ choice behaviour and could explain why some choices are more preferable than others. The lack of a significant result in this case is due to lack of variation between the alternative choices regarding the existence of signs, crowdedness and activities, it is found that these attributes are not significant. In essence, four out of five alternative choices address to the same physical network and share the same route characteristics. Among these characteristics are the aforementioned attributes. This means that there is no difference between the alternatives regarding the existence of signs, activities and crowdedness.

Regarding the significant factors, only the walking time, the activity time and the distance from the 1st performed activity were found significant on a 95% confidence interval. Thus performing an activity or walking along a certain route, will provide utility (or will induce cost) to the pedestrian. Two types of walking time are estimated. The one refers to pedestrians that do not want to perform an activity and the other to pedestrians that want to perform an activity. This categorization of the walking time suggests that the two groups of pedestrians perceive the values of walking time in a different way. If there is no preference to perform activity, the walking time has a negative value implying that pedestrians tend to prefer walking less time. On the contrary, if there is a preference to perform an activity the walking time has a positive value. This means that the utility of an alternative is going to increase if the walking time increases. In the event, it was noticed that more pedestrians preferred to perform an activity at the BC area than the ones that chose AB area, suggesting that they have a preference to walk longer in order to perform an activity. In BC area there was the option to visit and watch tall ships which is the main and one of the most relevant activities to SAIL event.

In accordance with the way that pedestrians with preference to perform an activity perceive walking time, the coefficient of the distance from the 1st activity is positive. The positive value reveals a preference to walk longer distances in order to perform activities in the BC area and after, where the activities are more related to the SAIL event (existence of tall ships). Lastly, as expected the coefficient of the activity time has a positive sign meaning the value of the utility function increases with the increase of the activity time.

The combination of the significant attributes into one model and the estimation of this model is the next step. The distance from the 1st performed activity was used for the calculation of the walking time. Therefore, a correlation is present between these two factors. This correlation is significant on a 99% confidence interval (Table 6.11). Since there is a correlation between these two factors, they cannot be combined. Due to the higher adjusted $\rho^2$ and the higher added value to the model the walking time was chosen to be used in the model instead of the distance from the 1st performed activity.
Table 6.11 Correlation table of the walking time versus the distance from the activity performed

<table>
<thead>
<tr>
<th></th>
<th>WALKING_TIME</th>
<th>DISTANCE_ACTIVITY</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WALKING TIME</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>1</td>
<td>.348**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>162</td>
<td>162</td>
</tr>
<tr>
<td><strong>DISTANCE_ACTIVITY</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>.348**</td>
<td>1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>162</td>
<td>162</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

The best model is a combination of the remaining factors: walking time and activity time. The utility function for alternative i is then the following:

\[
U_i = ASC_i + \beta_{Activity Time} \cdot Activity Time_i \cdot preference_{Activity} \\
+ \beta_{Walking Time} \cdot Walking Time_i \\
+ \beta_{Walking Time(performing activity)} \cdot Walking Time_i(performing activity)_i \cdot preference_{Activity}
\]

(2)

Where \(Preference_{Activity} = 0\) or \(1\)

The activity preference/desire can be directly derived from the database, since in the dataset it is known whether a pedestrian performs an activity or not. In other words, if someone performs an activity then apparently there is a desire/ preference for an activity.

The general information of the joint activity-route choice model is presented in Table 6.13. The model can be tested on its performance using the likelihood ratio test. This value is provided in the table for the comparison of this model with the null situation. A comparison of this model with the best individual model is also possible. This was the model of the activity time. The following formula is used for the estimation of the likelihood ratio:

\[-2(LL(2) - LL(1)) > \chi^2\]

The likelihood of the best individual model is -187,967 (5 parameters) compared to the likelihood of -175,577 for this model. This means that the resulting likelihood ratio is 170,304. This is higher than 5.99 on a 95% confidence interval. Therefore the combined model is significantly better than the individual model. The adjusted rho-square of this model is also improved compared to the individual models.

The utility parameters that are estimated for the combined model are provided in Table 6.13.
Table 6.12 General information on the combined MNL model

<table>
<thead>
<tr>
<th>Model:</th>
<th>Multinomial Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of estimated parameters:</td>
<td>5</td>
</tr>
<tr>
<td>Number of observations:</td>
<td>162</td>
</tr>
<tr>
<td>Number of individuals:</td>
<td>162</td>
</tr>
<tr>
<td>Null log-likelihood:</td>
<td>-260.729</td>
</tr>
<tr>
<td>Final log-likelihood:</td>
<td>-172.833</td>
</tr>
<tr>
<td>Likelihood ratio test:</td>
<td>175.791</td>
</tr>
<tr>
<td>Rho-square:</td>
<td>0.337</td>
</tr>
<tr>
<td>Adjusted rho-square:</td>
<td>0.318</td>
</tr>
<tr>
<td>Final gradient norm:</td>
<td>+1.267e-002</td>
</tr>
<tr>
<td>Diagnostic:</td>
<td>Convergence reached...</td>
</tr>
<tr>
<td>Iterations:</td>
<td>238</td>
</tr>
</tbody>
</table>

Table 6.13 Utility parameters for combined MNL model

<table>
<thead>
<tr>
<th>Rho-Square (R2)</th>
<th>0.337</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted Rho-Square (adj. R²)</td>
<td>0.318</td>
</tr>
<tr>
<td>Final Log-Likelihood</td>
<td>-172.833</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistical Testing on Parameters</th>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC Route1_2</td>
<td>0 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC Route1a_2</td>
<td>-3.47</td>
<td>-5.80</td>
<td>0.00*</td>
</tr>
<tr>
<td>ASC 4 Route1a_2a</td>
<td>-3.31</td>
<td>-6.28</td>
<td>0.00*</td>
</tr>
<tr>
<td>Beta_ActivityTime</td>
<td>0.216</td>
<td>7.20</td>
<td>0.00*</td>
</tr>
<tr>
<td>Beta_Walking Time (performing activity)</td>
<td>1.45</td>
<td>30.92</td>
<td>0.00*</td>
</tr>
<tr>
<td>Beta_WalkingTime</td>
<td>-0.0643</td>
<td>-2.62</td>
<td>0.01*</td>
</tr>
</tbody>
</table>

Significant on a 95% confidence interval

The ASC for Route 1_2 is fixed to zero. All attributes are significant on the 95% confidence interval. The coefficients of the walking and activity time are positive for pedestrians that have a preference to perform an activity. Otherwise if there is no interest in performing an activity then only the walking time has impact to the utility. In this case, the walking time provides negative utility, implying that higher walking times decrease the utility of that alternative. Despite the fact that the walking time of alternative 1_2 is...
much higher than alternative 5, the percentages of pedestrians that did not have a desire to perform an activity were almost equally distributed to route 1_2 and route 5 (22.5% and 19.4% respectively). This means that other factors may compensate for the high walking times of route 1_2 and lead to this outcome.

Two examples of pedestrians that belong to two different kinds of groups (one that has a preference to perform an activity and the other that does not) are presented below:

1st example Desire for activity

Pedestrian A chooses to watch the tall ships. This means that he took the Route 1_2a. The activity times and the walking times for every joint activity route choice alternative are the following:

<table>
<thead>
<tr>
<th>Route</th>
<th>Activity Time</th>
<th>Walking Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_2</td>
<td>0.00</td>
<td>10.38</td>
</tr>
<tr>
<td>1_2a</td>
<td>4.43</td>
<td>10.38</td>
</tr>
<tr>
<td>1a_2</td>
<td>7.99</td>
<td>10.38</td>
</tr>
<tr>
<td>1a_2a</td>
<td>12.42</td>
<td>10.38</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Which is the impact of the total utility of pedestrian A?

Since pedestrian A performs an activity (visiting tall ships) then there is apparently a desire for activity, thus the preference for activity will be equal to 1.

\[
U_i = ASC_i + \beta_{ActivityTime} \times ActivityTime_i \times PreferenceActivity + \beta_{WalkingTime} \times WalkingTime_{(performing \ activity)i} \times Preference \ Activity + \beta_{WalkingTime} \times Preference \ Activity \times PreferenceWalk
\]

\[
U_i = ASC_i + \beta_{ActivityTime} \times ActivityTime_i \times 1 + \beta_{WalkingTime_{(performing \ activity)}} \times WalkingTime_i \times 1
\]

\[
U_{Route1_2} = 0 + 0.216*0 + 1.45 * 10.38*1 = 15.05
\]

\[
U_{Route1_2a} = 0.216 * 4.43 + 1.45 * 10.38*1 = 16
\]

\[
U_{Route1a_2} = -3.47 + 0.216 * 7.99 + 1.45 * 10.38*1 = 13.3
\]

\[
U_{Route1a_2a} = -3.31 + 0.216 * 12.42 + 1.45 * 10.38 = 14.4
\]

\[
U_{Route5} = 0 + 0.216*0+ 1.45*2.5 = 3.62
\]

Pedestrian A has a probability of choosing:

\[
P_{Route1_2} = 23.36\%
\]

\[
P_{Route1_2a} = 60.39\%
\]

\[
P_{Route1a_2} = 3.58\%
\]

\[
P_{Route1a_2a} = 12.19\%
\]

\[
P_{Route5} = 0.48\%
\]

Finally, Pedestrian A chooses the Route 1_2a.
2nd example No desire for activity

Pedestrian B chooses to take Route5 (via shortcut), where there is no activity. This means that there is no desire for activity. The activity times and the walking times for every joint activity route choice alternative are the following:

<table>
<thead>
<tr>
<th></th>
<th>Route 1_2</th>
<th>Route1_2a</th>
<th>Route1a_2</th>
<th>Route1a_2a</th>
<th>Route5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity time</td>
<td>0.00</td>
<td>2.50</td>
<td>16.70</td>
<td>19.20</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Which is the impact of the total utility of pedestrian B?

Since pedestrian B takes the shortcut then there is not a desire for activity, thus the preference for activity will be equal to 0.

\[
U_i = ASC_i + \beta_{ActivityTime} \times ActivityTime_i \times PreferenceActivity + \beta_{WalkingTime} \times WalkingTime_i \times PreferenceActivity
\]

\[
U_i = ASC_i + \beta_{ActivityTime} \times ActivityTime_i \times 0 + \beta_{WalkingTime} \times WalkingTime_i \times (1 - 0)
\]

\[
= ASC_i + \beta_{WalkingTime} \times WalkingTime_i \times 1
\]

\[
U_{Route1_2} = -0.0643 \times 13.72 = -0.88
\]

\[
U_{Route1_2a} = -0.0643 \times 13.72 = -0.88
\]

\[
U_{Route1a_2} = -3.47 - 0.0643 \times 13.72 = -4.35
\]

\[
U_{Route1a_2a} = -3.31 - 0.0643 \times 13.72 = -3.98
\]

\[
U_{Route5} = -0.0643 \times 1.65 = -0.11
\]

Pedestrian B has a probability of choosing:

\[
P_{Route1_2} = 23.60\%
\]

\[
P_{Route1_2a} = 23.60\%
\]

\[
P_{Route1a_2} = 0.73\%
\]

\[
P_{Route1a_2a} = 1.06\%
\]

\[
P_{Route5} = 51.01\%
\]

Finally Pedestrian B chooses Route5 (via shortcut)

The two examples shows the impact of the activity time and walking time depending on the desire of an activity or not. The activity time spent by a pedestrian ranges from 1 minute to 120 minutes. Hence, the impact of the activity time is between 0.216 and 25.92. The walking time ranges from 1.5 minutes to 24 minutes, thus the impact of the walking time is between -0.096 and -1.5 for the case of not performing an activity while for the case of performing an activity, the range is fluctuating between 2.18 and 34.8.
The activity time and the walking time are the most dominant variables that influence the joint activity route choice. Due to a significant correlation the distance from the first performed activity was not taken into consideration in the combined model, though it does influence the choice when modelled individually.

**Models without ASC**

Since there is lack of variation in the characteristics of the choice alternatives, it is expected that the ASC values are very dominant in the choice model estimation. Hence, the models are estimated without ASC values in order to find out the influence of the attributes individually (Table 6.14).

<table>
<thead>
<tr>
<th>MNL</th>
<th>Parameters</th>
<th>Values</th>
<th>Log likelihood</th>
<th>Adjusted $\rho^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Walking time</td>
<td>p-value Coef</td>
<td>-0.130</td>
<td>-247.653</td>
</tr>
<tr>
<td>2</td>
<td>Walking time (performing activity)</td>
<td>p-value Coef</td>
<td>1.48</td>
<td>-239.977</td>
</tr>
<tr>
<td>3</td>
<td>Activity time</td>
<td>p-value Coef</td>
<td>0.00485</td>
<td>-247.019</td>
</tr>
<tr>
<td>4</td>
<td>Time spent before the decision point</td>
<td>p-value Coef</td>
<td>0.00476</td>
<td>-250.340</td>
</tr>
<tr>
<td>5</td>
<td>Distance from the 1st performed activity</td>
<td>p-value Coef</td>
<td>0.00112</td>
<td>-254.711</td>
</tr>
<tr>
<td>6</td>
<td>Signs</td>
<td>p-value Coef</td>
<td>0.18</td>
<td>-260.075</td>
</tr>
<tr>
<td>7</td>
<td>Crowdedness</td>
<td>p-value Coef</td>
<td>0.18</td>
<td>-260.075</td>
</tr>
<tr>
<td>8</td>
<td>Existence of Activity</td>
<td>p-value Coef</td>
<td>0.18</td>
<td>-260.075</td>
</tr>
<tr>
<td>9</td>
<td>Time of day (morning)</td>
<td>p-value Coef</td>
<td>1</td>
<td>-260.729</td>
</tr>
<tr>
<td></td>
<td>Time of day (afternoon)</td>
<td>p-value Coef</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Day (Thursday)</td>
<td>p-value Coef</td>
<td>1</td>
<td>-260.729</td>
</tr>
<tr>
<td></td>
<td>Day (Friday)</td>
<td>p-value Coef</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Day (Saturday)</td>
<td>p-value Coef</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Groupsize (individual)</td>
<td>p-value Coef</td>
<td>1</td>
<td>-260.729</td>
</tr>
<tr>
<td></td>
<td>Groupsize (2 persons)</td>
<td>p-value Coef</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Groupsize (3 persons)</td>
<td>p-value Coef</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Groupsize (4 persons)</td>
<td>p-value Coef</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Gender (Female)</td>
<td>p-value Coef</td>
<td>1</td>
<td>-260.729</td>
</tr>
<tr>
<td></td>
<td>Gender (Male)</td>
<td>p-value Coef</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Age (18-34)</td>
<td>p-value Coef</td>
<td>1</td>
<td>-260.729</td>
</tr>
<tr>
<td></td>
<td>Age (35-54)</td>
<td>p-value Coef</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
There is one more extra significant parameter value (time spent before the decision point) than the models estimated with ASC. It is noticed that the coefficient values of the activity time, walking time, existence of activities, crowdedness and signs decreased showing that they become less dominant without the ASC.

6.6 Conclusions
A simultaneous decision between activity and route choice was assumed for pedestrians that had a predetermined preference to perform an activity.

In general for the estimation of a joint activity-route choice model, the area which will be selected for that purpose should meet some criteria: availability of route alternatives, sufficient observations for each alternative, selection of area which combines both activity and route choice, diversity in the area.

Based on the aforementioned criteria the area depicted in the following figure was selected for the estimation of the model.

In order to examine which factors influence joint activity-route choice behaviour an extensive statistical analysis is conducted. The main findings of this analysis are the following:

- No variation in the distribution of sociodemographic factors over the joint activity route choice.
- During morning and afternoon higher likelihood of performing activities than evening.
- When someone has already performed an activity, he tends to spent less time in the next activity.
- Higher preference for areas where the major event activity exists (tall ships).

The relationships between the joint activity-route choice and the influential factors were quantified by means of a discrete choice model. Since there are two types of pedestrians’ behaviour; one that has a preference for performing activities and the other that does not, it is not possible to describe with one beta value different behaviours. It is concluded that the introduction of personal preferences (desire for activity) could reflect the difference between the types of pedestrians.
If pedestrians have a desire to perform an activity then it is more likely to choose an option with an activity. In the case of performing an activity the beta of the activity time is expected to be as high as possible and the utility function will be depended on the activity time. In order to encounter with this issue, a factor of preference to perform an activity is introduced. This factor can have 2 values depending on the performance of activity or not. If pedestrians perform an activity then it is implied that they have a preference to perform an activity, thus the factor of preference for activity takes the value 1 and it is multiplied by the activity time attribute.

The form of the utility function including the attributes multiplied by preference factors is presented below:

\[ U_i = ASC_i + \beta_{Activity\ Time} \cdot Activity\ Time_i \cdot \text{preference}_{Activity} + \beta_{Walking\ Time} \cdot Walking\ Time_i + \beta_{Walking\ Time(performing\ activity)} \cdot Walking\ Time_{(performing\ activity)}i \cdot \text{preference}_{Activity} \]

The only significant factors found are the following: Activity time, Walking time, and distance from the 1st activity.
7 Conclusions, discussion and recommendations

In this Chapter conclusions are derived based on the findings of this research. This is done by providing an answer to the research question (Section 7.1). Next, the discussion part is addressed in Section 7.2, where the findings and issues of this research are discussed. Finally recommendations are given both for science (theoretical purposes) (Section 7.3) and practice (Section 7.4)

7.1 Conclusions

The objective of this research is to explain and predict how pedestrians make an activity and route choice decision during large scale events. The main research question identified in this research is the following:

‘Which factors influence the activity and route choice behaviour of pedestrians during large scale events and to which extent?’

In order for the research question to be answered, the investigation has been divided into several steps. Initially, the literature survey provides useful insight regarding the factors that influence the choices of pedestrians during large scale events. According to observations and experience derived from large scale events, the factors were categorized in the following groups: personal, system, event and external. The creation of a link between influential factors and pedestrians’ choices has been attempted in several theoretical frameworks relevant to pedestrian behaviour studies. The integration of elements contained in these studies addresses the creation of a proposed theoretical framework within this thesis. According to Root and Recker (1981), the behaviour of pedestrians is preliminarily determined during the pre-travel phase, and is modified and adjusted during the trip in response to unforeseen situations. Hoogendoorn et al. (2001) with the normative theory suggested the modelling the joint route and activity approach since these choices are interrelated. Ton (2014) proposed two different types of pedestrians in train stations, dividing them in terms of their familiarity to the location. The distinction attributed predetermined decisions to familiar pedestrians, while unfamiliar have been proposed to behave intuitively.

Regarding the quantification of the relationships between pedestrians’ choices and influencing factors, discrete choice analysis is being deployed as a useful tool for the analysis and prediction of decision making during large scale events. Within this work, the data collection for the quantification is performed by Revealed Preference data method. Among the tested methods (GPS, Wi-Fi sensors, cameras), GPS tracking is perceived as the most suitable for the investigation. The performance of Wi-Fi sensors heavily depends on their density over the event, while cameras demand continuous human intervention and are prone to errors while observing the pedestrian movements. In contrast, GPS trackers may accurately capture the spatiotemporal movements while offering accurate information regarding the travel and area presence time. Ultimately, data fusion may possibly result to a high detailed dataset.

While processing the GPS data for the SAIL event case study, the box approach is introduced. Accordingly, the under-study area is separated in spatial boxes, while the area presence in each box is monitored. The approach appears promising as it simultaneously describes the route and the activity attributes of strictly defined areas. On one hand, the area presence time pinpoints to the “interesting” areas in terms of crowdedness, attractions and activities, while on the other, the box sequence results to an effective description of the route followed in a specific trip. Besides large scale events, it is likely that the approach can be easily adapted to other cases, such as shopping malls or train stations.

The box approach is proven to be an invaluable tool for the statistical analysis in the case study of SAIL event. Their adequate sizing results to the high exploitation of the level of detail given by the available dataset. The data filtering and consequent processing by means of the box approach provides useful observation and conclusions for the SAIL event case study. Apparently, the main attraction of the SAIL event (i.e. the tall ships) is observed to attract most of the attention of the visitors. More specifically, the
majority of pedestrians chose Route [I] during the start of their trip. Route [I] offered the full view of the main attraction and was indicated by the sign guidance of SAIL organization. Moreover, the statistical analysis suggests the association between the initial route and the route for the way back. In most cases the initial trip route is found to be different than the way-back route. For instance, the visitors who followed Route [I] while initially visiting the tall ships, returned to the train station via a shortcut or alternative travel options (public transportation or ferry).

The final part of this work assesses the construction of a joint activity-route model for pedestrians' choices during the SAIL event. The analysis is based on the predetermined preference of activities for pedestrians. In this way, a visitor knows beforehand whether he is willing to perform an activity but is uncertain about the activity location. The quantitative model is performed for a small area of the event which shows a relatively high diversity in routes and activities. According to the model estimation, the activity time and the walking time are shown to be significant factors for the joint activity-route choice. If there is a preference to perform an activity, the activity time and the walking time have a positive influence on the joint activity-route utility. This means that the average pedestrian would prefer spending more time walking in order to perform activities. In case of not performing an activity the walking time has a negative value implying that pedestrians tend to prefer walking less time. However, except for the walking and activity time the descriptive statistical analysis results to additional factors which are not captured by the model. Apparently, there is a high preference for the areas of the main activity of the SAIL event, which simultaneously determines the route and the activity choice. Additionally, the performance of an activity seems to affect the next activity, as the time spent for the latter is relatively lower. Finally, it is concluded that the sociodemographic characteristics of pedestrians do not influence the joint route-activity choices.

7.2 Discussion

In this section a discussion of the research is addressed. The main points that are going to be discussed are the following:

- Suitability of the revealed data collection method for this research - limitations
- Generalization of the boundary box method
- Generalization of the results in other large scale events
- Modelling jointly the activity and route choice of pedestrians during large scale events

Suitability of the revealed data collection method for this research - limitations

The suitability of the RP data is assessed based on their ability to extract information about the activity and route choice behaviour of pedestrians during large scale events.

Real-time data have been collected using different kinds of sensors: counting cameras (equipped with Wi-Fi sensors), Wi-Fi sensors and GPS trackers.

Beginning with the Wi-Fi sensors, they can only track the pedestrians that have their Wi-Fi function enabled. However, the reconstruction of the possible routes taken by the Wi-Fi users is disputable due to long distances between the Wi-Fi sensors in the SAIL area. It can only be ascertained that the location of an individual is within a range of positions. This limits the suitability of the method to reconstruct pedestrians’ route. Using several closely-located Wi-Fi sensors could help triangulate a more precise location. This implies that many sensors would be needed in order to track a person over a long path, which makes the process costly. However, Wi-Fi data could be used for the estimation of crowd density as well as for the acquisition of travel time information.

As for the cameras, they are able to count the number of pedestrians at given location and time. Except for the density which can be derived from camera data, information on the activity location choice of
pedestrians can be deduced only locally. On the other hand, considering the long distances between the camera sensors in some parts of the SAIL network it is difficult to deduce the routes that pedestrians took, by observing their movements. Thus, camera sensors are not the suitable tool for accurate reconstruction of pedestrians’ route choices; however they give insight regarding activity locations and derive the density.

The GPS devices, due to horizontally accuracy issues and the weakness of recognizing stops, are not suitable for capturing the activity locations choice. Furthermore, some GPS devices lost connection during the event so there were missing data. However, GPS data collection method was able to reconstruct the route choices and provide travel information.

Considering the aforementioned, the GPS data collection method could lead to a more accurate way (in comparison with the other available collection methods; Wi-Fi, Cameras) to extract information on the spatiotemporal movements of pedestrians during mass events is the GPS tracking. In other words, the main advantage of the GPS is that it can generate pedestrian tracks in an easy and not time consuming way. These tracks are useful for the investigation and understanding of pedestrians’ route choice behaviour. However, the data collected by the other methods (Wi-Fi and cameras) in combination with the GPS data could contribute to the improvement of the analysis of choice behaviour by capturing factors like crowdedness.

By using RP data only visible behaviour can be investigated, leaving the individuals’ intentions and motives as well as other personal characteristics and socio-demographic data unknown. If these characteristics were known then the assumption of predetermined and impulsive decision could be captured. This means that there would be a better insight into how pedestrians’ make decisions during events and which their event-state of mind is.

**Generalization of the boundary box method**

Throughout the present work, the boundary box method is introduced in order to combine area presence time data with location attributes. The under-study space is divided in strictly defined segments, each of which has various characteristics such as available activities. In this way, a box refers to an area and its local attributes, while a box sequence may describe a particular route. The deployment of the method requires the adequate location data (i.e. GPS) over time intervals. The boxes’ design is firmly connected to the network topology and may vary in size depending on the targeted level of detail. Their accurate division can potentially highlight interesting or problematic areas in a network. Furthermore, the boundary box approach allows the comparison of boxes, as well as the merging of boxes with similar characteristics. Even though the concept of boxes is hereby applied for the SAIL event case study, minor adjustments may allow its potential application in other cases such as various large scale events, train stations, shopping malls etc. In general, the boundary box method operates as a way of discretising a given network and at the same time attributing particular characteristics to its segments.

**Generalization of the results in other large scale events**

The relationships found in the joint activity route choice model of pedestrians during large scale event are probably applicable on other mass events. The layout of the event area and the type of visitors may be different. However the approach that it was followed regarding the modelling procedure and the derivation of alternatives (decision tree) can be applied and tested in other cases.

It is unknown if the findings and conclusions of this thesis are applicable for the organization of other mass events (such as King’s day in Amsterdam). New data should be collected and compared with the findings at SAIL to give more insight into the general influences on pedestrians’ choices during large scale events.
Modelling jointly the activity and route choice of pedestrians during large scale events

A joint activity – route choice behaviour is based on the assumption that pedestrians can decide beforehand for the whole route and activities performed along the route. However, modelling jointly the activity and route choice cannot give realistic predictions for longer period of time and for large areas. Real-life human choice behaviour and more specifically the activity behaviour of pedestrians during large scale events depends on how nice the chosen activity is and this can determine whether someone wants to stay or perform another activity. Thus, applying a joint decision needs small part of the network (like the examined case, area AC) and short period of time in order to have control of the assumptions made.

7.3 Recommendations
Theoretical recommendations

The proposed theoretical framework includes many choices that a pedestrian can make during a large scale event. However, the relationships between the choices made and the factors that influence these choices have not yet quantified. More specifically, the decision making of a pedestrian before the start of the trip as well as the quantification of the relationships before the trip, could offer opportunities for further research.

Discrete choice models are based on the concept of utility maximization which assumes that the decision makers select the alternative with the highest utility. This however is not always the case, as optimizing the choice costs a lot of effort. A research by Zhu (2008) suggests that not all alternatives are known by people or considered by them (bounded rationality). Even if they are considered, pedestrians are bounded in their ability to choose the best alternative. Therefore, this methodology seems promising.

The GPS data are revealed preference data. This means that the data shows what has been done by the pedestrians. This means that only data during the trip are collected. However, revealed data do not define if pedestrians’ decisions and choices are predetermined or intuitive. Stated preference data before and after the trip would give more insight into the planning procedure. The combination of stated and revealed data would give better insight into pedestrian’s choice behaviour.

Regarding, the common physical links that are shared the alternatives, the issue of the correlated error components is arose. This violates the multinomial logic (MNL) model assumption of independently distributed errors across alternatives. From a statistical point of view, a MNL route choice model will tend to assign counter-intuitively high probabilities to routes that share common network links. From a behavioural point of view, it can be said that the MNL considers overlapping routes distinct alternatives; whereas, pedestrians may consider such routes jointly as minor variants of a single alternative. There are two options to overcome the overlapping routes problem (Frejinger and Blerlaire, 2007). A correction factor can be applied to partially adjust the utilities for overlap, leading to the Path-Size Logit (PSL) model. Alternatively, more complex model forms may be specified that allow for correlated errors, including the multinomial probit model, mixed logit models, and closed-form members of the generalized extreme value (GEV) class of models.

Practical recommendation

During the SAIL event it was observed that pedestrians tend to go where people goes-herding behaviour. This behaviour should be considered, when large scale events are organized and an area of a city or even the whole city changes according to the need of the mass event. This means that crowd managers should focus on this herding behaviour by introducing the appropriate signs and other information measures. Hence, studies on the herding behaviour during events would give the available knowledge for the introduction of the appropriate crowd management measures.
As for the data collection during the event, it is suggested that stated preference data in combination with revealed data would give a better understanding and prediction on pedestrian behaviour. With GPS data, it is difficult to capture the time spent in an activity while the accuracy of the GPS devices varies which lead to not safe results regarding the exact activity location and type of activity. Additionally, the Wi-Fi sensors were located far away from each other which makes almost impossible extracting information about small areas. For these reasons, it would be interesting during the SAIL, when pedestrians return back their GPS devices (after finishing their trip), to declare which activities they performed, where and in which sequence. This information in combination with revealed data would be a useful database which can be used to extract safe conclusions on the way pedestrian choose activities and routes during a mass event.

Another recommendation regarding the data collection which seems promising is the use of Virtual and Augmented reality. However these innovations are in their infancy so it is needed to be developed for use in modelling pedestrians' behaviour during large scale events.

The estimation of other choice models (Latent Class or Mixed Logic models) could be used to compare the results in order to find the best model fit. Furthermore, regarding the examined dataset should be more extended and it would be interesting to be tested if it is representative for the population visited the SAIL event.

The results of this thesis do not provide definite policy suggestions. However, by stating the limitations of the RP data collection the regulatory bodies will be more critical on choosing the most suitable and appropriate collection method for their scope. Moreover, the findings from the RP data collection could be used in combination with other data collection methods in a complementary way (for example using stated preference data in the form of trip diary).
References


Hoogendoorn, S. P., 2015. Unravelling slow mode travelling and traffic with innovative data to a new transportation and traffic theory for pedestrians and bicycles, s.l.: s.n.


SAIL organisation, 2015. *SAIL*. Available at: https://www.sail.nl/2015


Spek, S. v. d., 2007. Legible City – Walkable City – Liveable City: Observation of Walking Patterns in City Centres Introductory paper, Urbanism On Track – Expert meeting on the application in urban design and planning of GPS-based and other tracking-based research. Delft, s.n.


Thurau, 2013. Datenauswertung - Beobachtungen Laufen, Bern: SBB CFF FFS.


Ton, D., 2014. NAVISTATION: a study into the route and activity location choice behaviour of departing pedestrians in train stations, s.l.: TU Delft, Delft University of Technology.


Appendix A - More information on the theoretical frameworks

Some of the theoretical frameworks found in literature that introduce a relationship between the choices of pedestrians is elaborated below.

Theoretical framework of Root & Recker (1981)

Root & Recker (1981) introduced a theoretical framework of decision taken which consists of two phases: the pre-travel phase and the travel phase (Figure A.1). In the pre-travel phase the activities are chosen and an activity program is planned. The decisions in this phase influence the choices made during the travel phase. The travel phase starts with the execution of the first activity included in the activity program. Depending on the occurrence of random disturbances (i.e. actual activity durations exceeding expected durations, actual travel times exceeding expected travel times, etc.) the actual activity pattern may not reproduce the activity program. If the activity pattern is different from the activity program then the activity program needs adaptations before the next activity carried out. These adaptations depend on various factors such as the size of the remaining travel budget, the amount and the distribution of remaining travel budget and flexibility, the expectations concerning the level of service afforded by future activity site arrivals and the number of activities remaining to be completed. The current levels of these variables are inputs into the simulation to determine the response to the random disturbance(s). After each travel/activity response is executed, a feedback loop that connects the next travel/activity decision and the planned activity program indicates that (a part of or all) the planned activity program may be readjusted by the individual prior to the execution of the next activity. This provides information about the possibilities to execute the activity sequence.

![Figure A.1](image)


Hoogendoorn, et al (2001) introduce a framework that shows different levels of pedestrian behaviour (Figure A.2). The strategic level includes long-term decisions of the pedestrian, such as choice of activities (as well as the activity order) to be performed and choice of destination. At the tactical level, pedestrians make short-term decisions about the order in which the activities will be performed based on the prevailing conditions. Activity scheduling, activity location choice and route choice are then determined on the tactical level, and thus during the trip. The tactical and strategic level influence each other. The operational level pertains to immediate decisions of the pedestrian concerning his behaviour.
Interactions with other pedestrians play an important role at this level. An overview of these decision levels, related processes and interactions between different decisions levels are given in Figure A.2 (Hoogendoorn, et al., 2001).

Figure A.2 Levels in pedestrian behaviour based on Hoogendoorn et al. (2001)

Hoogendoorn & Bovy (2004) assume that pedestrians choose the activity location and route simultaneously. They use the concept of utility in their research. Pedestrians try to maximize their utility by optimising the route and activity location(s). Based on this normative theory, route choice, activity area choice, and activity scheduling are simultaneously optimized using dynamic programming for different traffic conditions and uncertainty levels. With respect to the activity schedule, it is assumed that it is fixed. From all the feasible activity schedules the optimal one is chosen. As mentioned before, no consensus exists in literature about how people determine the activity schedule (Ettema, et al., 1993).

**Theoretical framework of Hoogendoorn & Bovy (2005)**

Hoogendoorn & Bovy (2005) also introduce a relationship between activity scheduling and the route and activity location choice (Figure A.3). A simultaneous choice process between route choice and activity location choice is introduced. The main difference in this framework compared to the framework of Hoogendoorn, et al (2001) is that the strategic level decisions (activity choice) influence the decisions made on a tactical level but not the other way around. In Hoogendoorn, et al (2001) the strategic level decisions were also influenced by decisions on a tactical level (feedback relation). This could mean that the experience gained during that trip is taken into account for the next trip.

Figure A.3 Route choice process for the individual traveller (Hoogendoorn & Bovy, 2005)
Theoretical framework of Ton (2014)

Figure A.4 Framework for familiar travellers (Ton, 2014)

Figure A.5 Framework for unfamiliar travellers (Ton, 2014)
Appendix B - GPS data processing

This appendix elaborates more on the GPS data processing. The process is illustrated in Figure (flow diagram) and includes several steps.

The main inputs for the 1st step of the process are the raw data (from the server) and the trip data (from the Google forms). The raw data were retrieved by the server of My GPStracker and were delivered in mat. format. The file with the raw GPS data includes one matrix with a size of 399496 x 4. Each column of the matrix corresponds to the ID of the GPS device, the latitude, the longitude and the time (epoch time) (Figure B.1).

The trip data includes personal information about the participants of the GPS data collection like the gender, age and group size. This information was reported on a Google form table (excel file) during the event. This table also contains information about the departure and arrival time of each participant from the standing point of TUDelft team, as well as the Trip ID, the GPS ID (labels on the devices), IMEI-number and GPS ID (server).

In order to extract information about the movements of each participant during the event the combination of the raw and trip data is necessary. The file with the raw data includes some measurements that are irrelevant to the data collection for the SAIL event. More specifically, the server was receiving signals from the GPS devices before the event (during their transportation from Delft to Amsterdam). They were also sending signal even if there were not carried by a participant. Furthermore, each GPS device could have been used more than once by different participants, meaning that each GPS tracker does not necessarily correspond to one trip, but to multiple trips. Hence, the retrieved raw data should be processed further in order to select the subset of the delivered GPS measurements which is related to every participant of the GPS data collection separately. In other words, each participant should be assigned to a unique trip. The tool used for the filtering and analysis of the data is MATLAB. Thus, both raw and trip data were loaded to MATLAB and their corresponding times were converted to matlab time in order to have a convenient form for the processing of the data.

By matching the GPS device IDs given by the raw data-file with the ones reported to the Google form-file, constrained by the start and end time (departure and arrival time from/to the standing point) of the trip of each participant, a link between the raw data and the data from the Google forms is created. This link leads to the assignment of each participant to a unique trip (Make trip data).

After this matching, the visualization of the track of each participant is possible Figure B. since the coordinates and the corresponding times of the remaining measurements in the GPS dataset were assigned to the 322 pedestrians that participated in the GPS data collection.

---

8 In this thesis, one trip is defined from the time moment one participant takes the GPS device from the stand until the time moment the same participant returns the device to the stand.
Figure B.1 GPS raw data (1st column: GPS ID, 2nd column: epoch time, 3rd column: longitude, 4th column: latitude)

Figure B.2 Excel file which includes information about the participants that carried a GPS device

Figure B.3 Trajectories of pedestrian with trip ID 254
In the graphical representation of the pedestrian trajectories it was observed that the data were noisy. To cope up with the noise, a simple moving average (SMA) formula was used. Moving average smooths data by replacing each data point with the average of the neighbouring data points defined within the span.

According to MATLAB, this process is equivalent to lowpass filtering with the response of the smoothing given by the difference equation

\[ y_s(i) = \frac{1}{2N+1} (y(i+N) + y(i+N-1) + \cdots + y(i-N)) \]

where \( y_s(i) \) is the smoothed value for the \( i \)th data point, \( N \) is the number of neighboring data points on either side of \( y_s(i) \) and \( 2N+1 \) is the span.

The span is the number of points of which the average is taken. In addition, by evaluating the distance of each point up to the moving average of the route that was measured some outliers were removed. A threshold value determines the points which should be deleted. For the span and threshold value different values are tried out and the one which gave intuitively (best fit on the map of Amsterdam) the best results were chosen (span: 7 and threshold value: \( 0.1 \times 10^3 \)). Figure B.4 explains the way that the outliers were found and removed.

Figure B.4 The first graph shows the outliers that were above a threshold value (above orange line). The second graph shows in binary values the points that are deleted.

In addition, the frequency of the GPS signals was examined for each participant’s trip. A closer look on the plot of the trajectories indicates that there are missing data for a number of trips. This means that the measurement frequency can sometimes be more than every 5 seconds which is the frequency given by the manufacturer. For instance, from the frequency measurement distribution illustrated in Figure B., it can be seen that only 27% of the measurements were taken every 5 seconds, while in Figure B.5 the GPS device performed very well since more than 90% of the measurements were taken every 5 seconds.
Figure B.5 Frequency distribution of a poorly GPS device (trip ID 59)

Figure B.6 Frequency distribution of a well performing GPS device (trip ID 254)

Figure B.6 shows the distribution of time measurements for all the GPS devices that were distributed during the event in order to check the accuracy of the GPS data regarding the frequency of the signal sent. It is clear that only 60% of the GPS measurements had frequency every 5 seconds. It should be noted that having measurements every 10, 15, 20 seconds etc (multipliers of 5) seems rational since the devices are set to receive measurements every 5 seconds. This can be explained by the fact that the communication between the satellites and the receiver might have been broken making the estimation of the position of the pedestrian impossible.
However, in Figure B.7 it can be seen that there are some measurements taken at time intervals different from multiples of 5 seconds (16, 31 seconds) which is difficult to be explained. A possible reason could be that the settings for the minimum distance between two points is too large and therefore many points were discarded. Nevertheless, this does not explain why certain devices have more measurements every 5 seconds than others, since all settings are the same for all the devices used in the field.

After the aforementioned general remarks on the examined GPS data, the process of creating routes is next. For this purpose, a list of boundary boxes with spatial attributes deriving by the division of the SAIL area (Figure B.8) according to the network configuration and the location of activities is used, in order to assign the GPS measurement points to the spatial boundary boxes. For the division of the SAIL area into bounding boxes the QGIS software was used.

To be more specific, each bounding box has an ID (Figure B.8) and the geometrical attributes of the bounding boxes were exported to a csv file in order to be loaded to MATLAB. Then a MATLAB code was written in order to assign the GPS measurement points to the spatial boundary boxes (Figure B.8). After this assignment, a list of boxes visited by every pedestrian was created. For instance, Figure B.9 presents
a part of the sequence of the boxes visited (2nd column) by pedestrian ID=86. In the 1st column there is the matlab time while in the 3rd and 4th column the longitude and latitude are given respectively. The orange arrows indicate the first time that the participant appears in the box.

Figure B.9 GPS measurement points assigned to the boxes (right: Pedestrian ID=254, Left: Pedestrian ID=86)

Figure B.10 Boxes visited by pedestrian ID=86. The 1st column indicates the time moment that a participant is located in each box (2nd column). The 3rd and the 4th columns give the longitude and the latitude respectively.

After this assignment, a list of boxes visited by every participant is created. Thus, there is a box sequence for every trip. Based on the network configuration and the decision points (defined by the route and activity choice) along the SAIL area a set of route alternatives are created. Each route is characterized by a specific combination of box sequence.

Along the Orange route of SAIL area, there are multiple decision points where visitors of SAIL could choose among a number of alternative routes. Considering the locations of Points Of Interest (POI) along the Sail area (Figure B.11, Figure B.12), the network configuration and the main patterns of the route and activity choices of the participants during the event, the set of route alternatives is defined. The main patterns of movements of the participants during the event are visualized by using a useful tool/platform created in Matlab. This tool gives the possibility to have an overview of the tracks of pedestrians during specific time periods (Figure B.13).
Having a visual inspection of the plots of the pedestrian trajectories in combination with area boxes, it was observed that some boxes had few data (missing data—not continuous trajectories). This lack of data points inside the boxes could lead to a wrongly estimation of the time spent in the area boxes. This is
because the entry point and the exit point from the box did not correspond to the actual entry and exit
time. In order to overcome this issue, an interpolation was applied and the entry and exit times to and
from the boxes are calculated. Then the time spent along different OD pairs specified by various
combinations of box sequences is estimated.

The methodology for the estimation of entry and exit time to and from the boxes is as following:

Assume that box(i) (blue) and box(i+1) (orange) are two consecutive boxes visited by a pedestrian.

1. A line was drawn in MATLAB connecting the last GPS measurement point in box(i) and the first point
   located in box(i+1)
2. The intersection point between the line and box(i+1) was found
3. The distance \( x_1 \) between the last GPS measurement point located in box(i) and the intersection point
   was calculated
4. The distance \( x_1 + x_2 \) and the time difference \( t_1 - t_2 \) between the last GPS measurement point of box(i)
   and the first point of box(i+1) were calculated
5. The time that the pedestrian visits box(i+1) is given by the formula:

\[
\tau_{\text{entry}} = \frac{x_1}{x_1 + x_2} (t_2 - t_1) + t_1
\]

6. Due to the fact that the boxes were drawn as close as possible (their distance is zero) the
   \( \tau_{\text{entry}}(\text{box}_{i+1}) = \tau_{\text{exit}}(\text{box}_i) \).

After the implementation of the aforementioned approach the time spent in every box can be calculated by:

\[
\text{time spent in the box}_{i} = \tau_{\text{exit}}(\text{box}_i) - \tau_{\text{entry}}(\text{box}_i)
\]

Finally, the output of this whole process is the creation of a structure where the characteristics of each
trip are available. For every trip it is known the gender, the age and the group size of each participant.
Apart from these characteristics, the box sequence that every participant visited, the time spent to every
box sequence, the entry time to each box belonging to each box sequence as well as the routes formed by
the box sequences are known.
Distribution of activity time at each box of part of the Orange route
Figure B.14 Activity time distribution for each box of a selected part of Orange route (Veemkade)
Appendix C - Biogeme results- Joint activity-route choice model

In this appendix an overview of the results from the joint activity-route choice models created with Biogeme are presented. The presented models are MNL models. The modelling process applied for the joint activity-route choice is the following:

1. Test each factor individually on its significance on the choice model
2. Start with the models with the highest rho square and combine factors stepwise

**MNL- GENDER**

<table>
<thead>
<tr>
<th>Rho-Square (R²)</th>
<th>0.077</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted Rho-Square (adj. R²)</td>
<td>0.054</td>
</tr>
<tr>
<td>Final Log-Likelihood</td>
<td>-240.583</td>
</tr>
</tbody>
</table>

Statistical Testing on Parameters

<table>
<thead>
<tr>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1 0 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2 0.460</td>
<td>2.16</td>
<td>0.03*</td>
</tr>
<tr>
<td>ASC 3 -1.39</td>
<td>-3.72</td>
<td>0.00*</td>
</tr>
<tr>
<td>ASC 4 -0.288</td>
<td>-1.13</td>
<td>0.26</td>
</tr>
<tr>
<td>ASC 5 -0.0870</td>
<td>-0.36</td>
<td>0.72</td>
</tr>
<tr>
<td>Beta_GENDER_F -2.94e-015</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Beta_GENDER_M -4.63e-015</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Significant on a 95% confidence interval

\[ U_i = ASC_i + \beta_{GENDER,M} * GENDER_M + \beta_{GENDER,F} * GENDER_F \]

**MNL- AGE**

<table>
<thead>
<tr>
<th>Rho-Square (R²)</th>
<th>0.075</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted Rho-Square (adj. R²)</td>
<td>0.059</td>
</tr>
<tr>
<td>Final Log-Likelihood</td>
<td>-241.251</td>
</tr>
</tbody>
</table>

Statistical Testing on Parameters

<table>
<thead>
<tr>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1 0 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2 0.577</td>
<td>3.45</td>
<td>0.00*</td>
</tr>
<tr>
<td>ASC 3 -1.27</td>
<td>-3.64</td>
<td>0.00*</td>
</tr>
<tr>
<td>Beta_AGE_E1 1.81e-015</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Beta_AGE_E2 1.24e-015</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Significant on a 95% confidence interval
\[ U_i = ASC_i + \beta_{AGE,E1} \times AGE_E1 + \beta_{AGE,E2} \times AGE_E2 \]

**MNL- GROUP SIZE**

Rho-Square (R^2) \hspace{1cm} 0.075  
Adjusted Rho-Square (adj. R^2) \hspace{1cm} 0.052  
Final Log-Likelihood \hspace{1cm} -241.251

Statistical Testing on Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1 (fixed)</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2</td>
<td>0.577</td>
<td>3.45</td>
<td>0.00*</td>
</tr>
<tr>
<td>ASC 3</td>
<td>-1.27</td>
<td>-3.64</td>
<td>0.00*</td>
</tr>
<tr>
<td>Beta_GROUP_E0</td>
<td>9.85e-017</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Beta_GROUP_E1</td>
<td>-1.75e-015</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Beta_GROUP_E2</td>
<td>-3.09e-016</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Beta_GROUP_E3</td>
<td>-5.55e-016</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Significant on a 95% confidence interval

\[ U_i = ASC_i + \beta_{GROUP,E0} \times GROUP_{E0} + \beta_{GROUP,E1} \times GROUP_{E1} + \beta_{GROUP,E2} \times GROUP_{E2} + \beta_{GROUP,E3} \times GROUP_{E3} \]

**MNL-ACTIVITY TIME**

Rho-Square (R^2) \hspace{1cm} 0.278  
Adjusted Rho-Square (adj. R^2) \hspace{1cm} 0.266  
Final Log-Likelihood \hspace{1cm} -188.296

Statistical Testing on Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1 (fixed)</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 3</td>
<td>-3.82</td>
<td>-5.63</td>
<td>0.00*</td>
</tr>
<tr>
<td>ASC 4</td>
<td>-3.90</td>
<td>-6.20</td>
<td>0.00*</td>
</tr>
<tr>
<td>Beta_ACTIVITY_TIME</td>
<td>0.277</td>
<td>7.84</td>
<td>0.00*</td>
</tr>
</tbody>
</table>

*Significant on a 95% confidence interval
**MNL-WALKING TIME**

Rho-Square ($R^2$) 0.188

Adjusted Rho-Square (adj. $R^2$) 0.172

Final Log-Likelihood -211.81

Statistical Testing on Parameters

<table>
<thead>
<tr>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2</td>
<td>0.593</td>
<td>3.24</td>
</tr>
<tr>
<td>ASC 3</td>
<td>-1.25</td>
<td>-3.52</td>
</tr>
<tr>
<td>ASC 5</td>
<td>-2.69</td>
<td>-4.85</td>
</tr>
<tr>
<td>Beta_WALKING_TIME</td>
<td>-0.4</td>
<td>-6.96</td>
</tr>
</tbody>
</table>

*Significant on a 95% confidence interval

$$U_i = ASC_i + \beta_{WALKING_TIME} \times WALKING_TIME_i \times preferenceWalk$$

**MNL-WALKING TIME WHILE PERFORMING ACTIVITY**

Rho-Square ($R^2$) 0.206

Adjusted Rho-Square (adj. $R^2$) 0.190

Final Log-Likelihood -207.110

Statistical Testing on Parameters

<table>
<thead>
<tr>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2</td>
<td>0.593</td>
<td>3.24</td>
</tr>
<tr>
<td>ASC 3</td>
<td>-1.25</td>
<td>-3.52</td>
</tr>
<tr>
<td>ASC 5</td>
<td>1.32</td>
<td>4.61</td>
</tr>
<tr>
<td>Beta_WALKING_TIME(performing activity)</td>
<td>2.98</td>
<td>32.55</td>
</tr>
</tbody>
</table>

*Significant on a 95% confidence interval

$$U_i = ASC_i + \beta_{WALKING_TIME(performing activity)} \times WALKING_TIME(performing activity)_i \times preferenceACTIVITY$$
**MNL- TIME SPENT BEFORE ARRIVE AT THE DECISION POINT**

Rho-Square ($R^2$) 0.076

Adjusted Rho-Square (adj. $R^2$) 0.064

Final Log-Likelihood -241.008

Statistical Testing on Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1</td>
<td>0 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2</td>
<td>0.659</td>
<td>3.18</td>
<td>0.00*</td>
</tr>
<tr>
<td>ASC 3</td>
<td>-1.44</td>
<td>-3.37</td>
<td>0.00*</td>
</tr>
<tr>
<td>BETA_TIME_BEFORE</td>
<td>0.0118</td>
<td>0.69</td>
<td>0.49</td>
</tr>
</tbody>
</table>

$U_i = ASC_i + \beta_{TIME\_BEFORE} \times TIME\_BEFORE_i$

**MNL- DISTANCE FROM 1st ACTIVITY**

Rho-Square ($R^2$) 0.059

Adjusted Rho-Square (adj. $R^2$) 0.048

Final Log-Likelihood -245.292

Statistical Testing on Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1</td>
<td>0 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2</td>
<td>0.760</td>
<td>4.27</td>
<td>0.00*</td>
</tr>
<tr>
<td>ASC 3</td>
<td>0.503</td>
<td>2.23</td>
<td>0.03*</td>
</tr>
<tr>
<td>BETA_DISTANCE_ACTIVITY</td>
<td>0.00116</td>
<td>2.94</td>
<td>0.00*</td>
</tr>
</tbody>
</table>

$U_i = ASC_i + \beta_{DISTANCE} \times DISTANCE\_ACTIVITY_i$

**MNL-EXISTENCE OF ACTIVITY**

Rho-Square ($R^2$) 0.077

Adjusted Rho-Square (adj. $R^2$) 0.066

Final Log-Likelihood -240.547

Statistical Testing on Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1</td>
<td>0 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2</td>
<td>0.481</td>
<td>2.60</td>
<td>0.01*</td>
</tr>
</tbody>
</table>
\[ U_i = ASC_i + \beta_{\text{EXISTENCE ACTIVITY}} \times \text{EXISTENCE ACTIVITY} \]

**MNL-SIGNS**

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1</td>
<td>0 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2</td>
<td>0.481</td>
<td>2.60</td>
<td>0.01*</td>
</tr>
<tr>
<td>ASC 3</td>
<td>-1.36</td>
<td>-3.82</td>
<td>0.00*</td>
</tr>
<tr>
<td>BETA_SIGNS</td>
<td>0.258</td>
<td>1.41</td>
<td>0.16</td>
</tr>
</tbody>
</table>

\[ U_i = ASC_i + \beta_{\text{SIGNS}} \times \text{SIGNS} \]

**MNL-CROWDEDNESS**

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1</td>
<td>0 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2</td>
<td>0.481</td>
<td>2.60</td>
<td>0.01*</td>
</tr>
<tr>
<td>ASC 3</td>
<td>-1.36</td>
<td>-3.82</td>
<td>0.00*</td>
</tr>
<tr>
<td>BETA_CROWDEDNESS</td>
<td>0.258</td>
<td>1.41</td>
<td>0.16</td>
</tr>
</tbody>
</table>

\[ U_i = ASC_i + \beta_{\text{CROWDEDNESS}} \times \text{CROWDEDNESS} \]

**MNL-TIME OF DAY**

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1</td>
<td>0 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2</td>
<td>0.481</td>
<td>2.60</td>
<td>0.01*</td>
</tr>
<tr>
<td>ASC 3</td>
<td>-1.36</td>
<td>-3.82</td>
<td>0.00*</td>
</tr>
<tr>
<td>BETA_CROWDEDNESS</td>
<td>0.258</td>
<td>1.41</td>
<td>0.16</td>
</tr>
</tbody>
</table>

\[ U_i = ASC_i + \beta_{\text{CROWDEDNESS}} \times \text{CROWDEDNESS} \]
### Statistical Testing on Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1</td>
<td>0 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2</td>
<td>0.577</td>
<td>3.45</td>
<td>0.00*</td>
</tr>
<tr>
<td>ASC 3</td>
<td>-1.27</td>
<td>-3.64</td>
<td>0.00*</td>
</tr>
<tr>
<td>Beta.TIME_OF_DAY_E1</td>
<td>-1.71e-015</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Beta.TIME_OF_DAY_E2</td>
<td>-1.39e-015</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Significant on a 95% confidence interval

\[ U_i = ASC_i + \beta_{TIME\_OF\_DAY\_E1} \times TIME\_OF\_DAY\_E1 + \beta_{TIME\_OF\_DAY\_E2} \times TIME\_OF\_DAY\_E2 \]

### MNL-DAY

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rho-Square (R²)</td>
<td>0.077</td>
</tr>
<tr>
<td>Adjusted Rho-Square (adj. R²)</td>
<td>0.050</td>
</tr>
</tbody>
</table>

Final Log-Likelihood: -240.583

### Statistical Testing on Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Robust t-test</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC 1</td>
<td>0 (fixed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASC 2</td>
<td>0.460</td>
<td>2.16</td>
<td>0.03*</td>
</tr>
<tr>
<td>ASC 3</td>
<td>-1.39</td>
<td>-3.72</td>
<td>0.00*</td>
</tr>
<tr>
<td>ASC 4</td>
<td>-0.288</td>
<td>-1.13</td>
<td>0.26</td>
</tr>
<tr>
<td>ASC 5</td>
<td>-0.0870</td>
<td>-0.36</td>
<td>0.72</td>
</tr>
<tr>
<td>Beta.DAY_E1</td>
<td>1.13e-015</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Beta.DAY_E2</td>
<td>-3.18e-016</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Beta.DAY_E3</td>
<td>-3.18e-016</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Significant on a 95% confidence interval

\[ U_i = ASC_i + \beta_{TIME\_OF\_DAY\_E1} \times TIME\_OF\_DAY\_E1 + \beta_{TIME\_OF\_DAY\_E2} \times TIME\_OF\_DAY\_E2 \]
Appendix D - (MAPS)