Mobility Index Algorithm

A construction of a mathematical based mobility measurement algorithm based on dynamic data with a design science research approach.

Mikal Fuglestein

Master's thesis in Software Engineering at

Department of Computing, Mathematics and Physics,
Bergen University College

Department of Informatics,
University of Bergen

June 2020

Western Norway
University of
Applied Sciences

UNIVERSITAS BERGENSIS

Abstract

To have good mobility regarding public transport and to nearby facilities such as schools and shops is essential for reducing the use of private cars. One can only say how good mobility is for a given point by measure it. This thesis looks into the problem of mobility measurement. It provides a new way of calculating a mobility index, that is normalized with max-min normalization. A procedure of automatic and dynamic data gathering is in focus, which leads to the opportunity of replacing or expansion of various variables. Our finding is a measurement algorithm, based on theory outline in this thesis, and a tool to detect areas with needed mobility improvement for the city planners. This is a collaboration work with MUST (Mobilitetslaboratorium for utvikling av smarte transportløsninger), and the Municipality of Bergen where we are using data from them to calculate the mobility index.

Acknowledgment

I would first like to thank my skilful and professional supervisor, Rogardt Heldal, from the Western Norway University of Applied Sciences. With his energetic and positive personality, Rogardt has guided and motivated me through the process of this project.

I would also like to thank the dedicated enterprise architect from Bergen Municipality, Tom Osnes Svellingen, who has contributed with necessary data access and great ideas and discussion on the relevant problems. The meetings with Tom have always been helpful with feedback and constructive criticism.

I would also acknowledge the great outcome of having the privilege to participate in the MUST community, with presenting the ideas and problems to the field, and receiving useful inputs and insights around the subject of mobility.

Throughout the process of conducting the thesis, a lot of idea sharing and discussion have been done with my friend and class mate, Kenneth Apeland, who also did a thesis on mobility. I would like to thank him for the constructive conversations and motivation.

Finally, I must express my very profound gratitude to my parents and to my girlfriend, for providing me with support and continuous encouragement throughout my years of study. Line Marie Haug Helgøy, thank you for keeping up with me even if our home has been full of papers, books, notes and coffee cups. The accomplishment of both conducting this thesis, and the years of studying, would not have been possible without them. Thank you.

Contents

List of Figures

List of Tables

Acronyms

API Application Programming Interface

CSS Cascading Style Sheets

cURL client Uniform Resource Locator

FOL First Order Logic

GIS Geographic Information System

HTML Hypertext Markup Language

HTTP Hypertext Transfer Protocol

JSON JavaScript Object Notation

MISQ Management Information System Quarterly

PT Public Transport

MUST Mobilitetslaboratorium for utvikling av smarte transportløsninger

NLOD Norwegian Licence of Public Data

OSM Open Street Map

POI Point of Interest

SNAMUTS Spatial Network Analysis for Multimodal Urban Transport Systems

SQL Structured Query Language

WFF Well-formed Formulae

The Greek Alphabet

upper	lower	
A	α	alpha
B	β	beta
Γ	γ	gamma
Δ	δ	delta
E	ϵ	epsilon
Z	ζ	zeta
H	η	eta
Θ	θ	theta
I	ι	iota
K	κ	kappa
Λ	λ	lambda
M	μ	mu
N	ν	nu
Ξ	ξ	xi
O	o	omicron
Π	π	pi
R	ρ	rho
Σ	σ	sigma
T	τ	tau
Y	υ	upsilon
Φ	ϕ	phi
X	χ	chi
Ψ	ψ	psi
Ω	ω	omega

Chapter 1

Introduction

The population in urban areas around the world is increasing at a high rate. It has reached an all-time high, and the share of urban dwellers, currently at 54 percent, is projected to represent two thirds of the global population in 2050. In Norway, 80 percent of the population is residing in urban areas [27]. Simultaneously, the need for low-carbon and zero-emission transport is getting significant attention. As the metropolitan growth curve rises, it is expected higher transport activities, which is currently a growing part of the global emissions of greenhouse gases. The vision of a place where you can walk to do your shopping, where the air is clean, with lots of parks and green space and where businesses can prosper, is a challenge of mobility planning [31].

Obviously, mobility is important to urban areas, and planning for future well functioning mobility is highly prioritized at many levels. To respond to the national level, Bergen municipality has engaged an arena to focus on the challenges and opportunities of mobility solutions. MUST (Mobilitetslaboratorium for utvikling av smarte transportløsninger) has the vision to become an interactive lab for development and deployment of transportation and mobility solutions for the future. The strategy is based on low barriers and open access to transportation and mobility data. To achieve this, MUST encourages cooperation between public and private businesses to reach solutions that lead to managing climate issues, inhabitant densification and new technologies [15]. The commitment to improve the sustainability of Bergen is a commitment between municipality, industry and academia.

The aim of this thesis is therefore, with available data, to help the city of Bergen stepping further into the future of mobility. By developing a platform, which automatically collects mobility data from different sources, that measures the level of mobility at any given location. When combining multiple transport access points, mobility services and routes, a score of mobility is then calculated. The measurement has to be based on mathematical theories, such that an algorithm can be developed from mathematical models. The score will not only satisfy subjective mobility needs, but also play an important part for the city planners in Bergen.

1.1 Research question

The purpose of the research is to contribute with a new way to calculate mobility index, based on mathematical theories, and a software application built on this theory. We will answer the following research questions in this thesis:

How do available transportation and accessibility data give a calculated measurement of the current state of mobility at a given location?

The main research question is further derived into a set of sub questions, which will be answered throughout the conducting of the research. These questions are defined:

1. **What is a good mobility index?**
   The index is the output of the measurement formula defined in chapter 4. Through a case study, in cooperation with the municipality, the important and essential factors of formulating the measurement will be defined. However, understanding some grounded theories regarding accessibility measurement is necessary.

2. **Who are the most important stakeholders, and what do they want?**
   Since the strategy for this research involves an empirical investigation of the contemporary phenomenon of mobility in a real life context, feedback from the stakeholders is important. The need for an improved measurement of mobility is essential to uncover new and better transportation infrastructure planning.

3. **What does the research solve, and why is it important?**
   If an algorithm measures the quality of mobility at a given location, it is easy to understand what is needed in form of new infrastructure. The research provides a grounded tool which tells how satisfying the picture of the current mobility situation is, and a door opener to reach the goal of the

zero-growth objective. This political goal states that traffic growth in urban areas should take place through cycling, walking and public transport [25].

Chapter 2

Background

The following sections in the current chapter will provide the necessary information to understand the implementation of our application. Throughout the developing of our application we have imported and used necessary API's, libraries and datasets. Instead of "reinventing the wheel", we selected tools to improve the managing, performance and robustness of our system.

2.1 Extended dependencies

Folium Python - Library

Folium is a Python library which connects the strength and power of the Python ecosystem with Leaflet, MapBox and Stamen such that the visualization gets the interactive benefits that JavaScript gives [7]. The library makes it possible to manipulate Python code at the back end, which is generated at the front end, abstracting JavaScript and HTML. The visualization of a map is in this case based on a dynamic gathering of tiles, which is frequently updated interactively. Each tile represent a part of a presented map as an image. Folium collects the tiles from the OpenStreetMap library.

OpenStreetMap - Library

OpenStreetMap is one of the largest geographical data sets in the world. It is a collaborative project where data is voluntarily collected and then entered into the OpenStreetMap PostgreSQL database [26]. The library is trusted and used by a variety of popular services. Notable services who include OpenStreetMap geolocation or map-based components are Facebook, Snapchat, Tesla and the previously Apple maps in iOS. Also, OSM is used in the Bergen Bysykkel online visualization of available cycles.

Overpass API
The API serves data from OSM over the web, and has a capacity of total 1.000.000 requests per day. For the project, the API is mainly used to retrieve positions of points of interest in Bergen municipality. Since it is OSM based, the data gathered from the Overpass API is collected by volunteers and open for all. The API is suited for Python programming and Node.js [22].

Flask - Web framework
Flask is one of the most popular Python web application frameworks. Flask provides a virtual environment, both in development and in production. The environment helps managing all dependencies included while developing, such as Werkzeug, Jinja and ItsDangerous. Werkzeug and Jinja provides the communication to the client. Werkzeug implements the standard Python interface between application and server, and Jinja provides the markup language to render pages that the application serves. ItsDangerous is the safety dependency, which securely signs data and protect Flask's session cookie [23]. A trusted user of this web framework worth noting is Netflix.

Geocoder Python - Library
Geocoder is a geocoding library written in Python. In our application, this library covers one common and simple GIS-task; converting addresses to coordinates. The task is done by a geocoding provider, which owns goe-related JSON schemas. By importing the Geocoder library, the provider and its JSON, is easily accessible [5]. The particular provider for our project is called OpenCage.

OpenCage - API
Provides the JSON schemas containing the needed geodata. OpenCage is buildt on open source, and receives its geographical data from volunteers world wide. For data enrichment, OpenCage has notable costumers as Datawrapper, which serves maps for The New York Times, and Schibsted, the parent company of brands like VG, Finn.no and Bergens Tidene [21].

2.2 Data Lake

During 2018, Bergen municipality together with Bouvet, constructed the data lake project, "Lungegårdsvannet". According to Bouvets web page, the project will bring great benefits for the inhabitants and both private and public businesses in Bergen. This project is based on a national initiative which brings

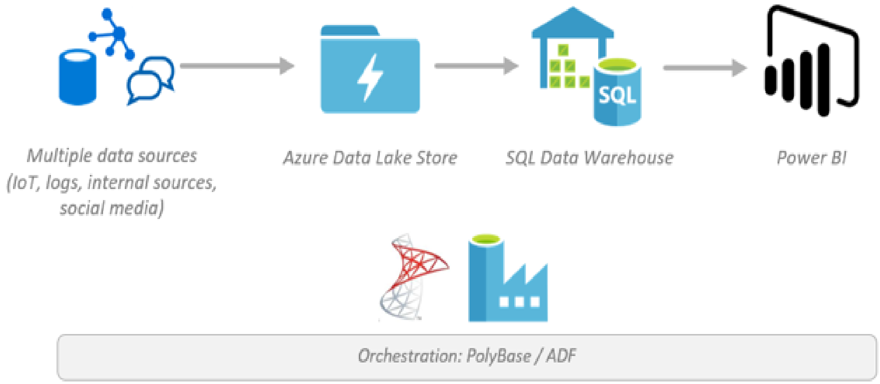

Figure 2.1: Azure Data Architecture: Data warehousing
Microsoft [18]

possibilities to the urban areas regarding big data gathering. One of the goals for the data lake project is the enabling opportunities of cross sharing, comparing and analysis of big data sets. For example, the local planning and building administration has now used data from both the water system and health data to earn better insight of the construction prospect and planning [24]. Another consequence of the data lake is the development of MUST, mentioned earlier in the introduction part.

The containing data in the data lake is gathered from both sensor data and various businesses and services such as Skyss, Bergen Bysykkel, Statens Vegvesen, Bergen Municipality and Motorvognregisteret. Then, the data is loaded in Azure Data Warehouse, and further into data marts in SQL servers [2]. The cloud based solution, provided by Microsoft Azure, is part of the Azure Analytics Platform [18]. As the figure 2.1 shows, the visualization tool Power BI is in that case the end point. In the case of our application, Power BI is replaced by the developed Mobility Index platform, with visualization through rendered HTML.

2.3 Summary

As the figure 2.2 illustrates, the APIs, libraries and frameworks are structured throughout the application. To reach the final production, most of the development has to be implemented in the Python Source Code block, thereby the

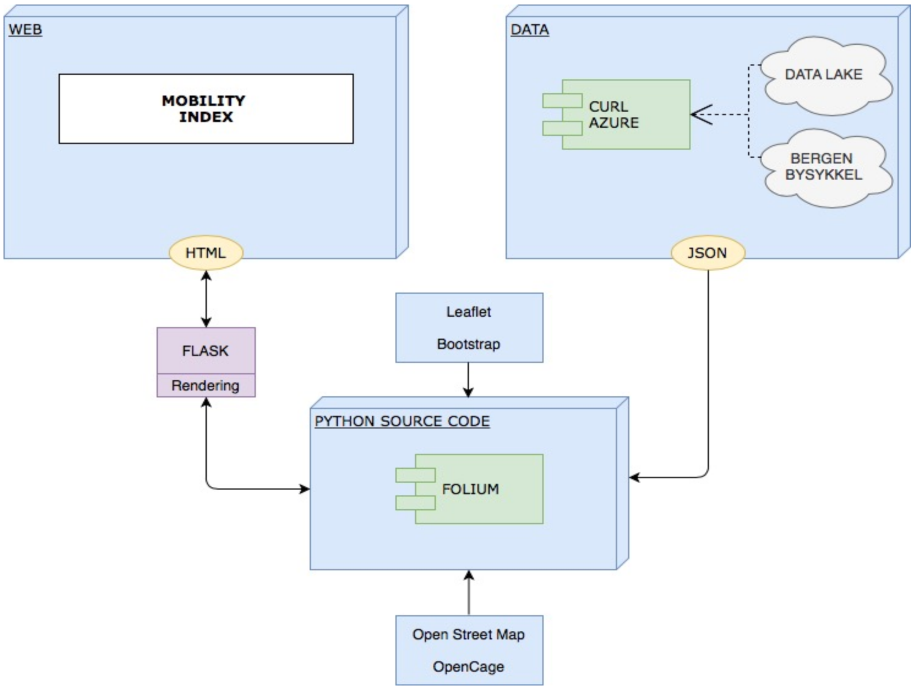

Figure 2.2: Extended Technologies and Dependencies Architecture

programmed formula that calculates the level of mobility. At the visualization part, displayed at the Web block, an interactive map and an input field has to be implemented. With Folium, HTML containing a presentation of a map is generated, and then rendered by Flask. Also, while developing, all dependencies are imported, loaded and updated in a virtual environment.

Chapter 3

Methodology

The scientific research methodology used to solve our problem of the mobility measurement algorithm is **design science research.** The process is based and conducted with design science because of the strategy of the project execution. Namely, based on the development, performance and design of an artifact, which in this case is a technical system. As well as defining stakeholders and end users, brainstorming and evaluating. The method of design science research involves the following two parts to understand, develop and improve the behavior and function of the specific artifact, i.e. our application: [28]

1. The gain of descriptive knowledge.
2. The invention of prescriptive knowledge, based on the descriptive knowledge presented.

The two parts are done in cycles. The American professor in Information Systems and Decision Sciences, and the author of the book Design Research in Information Systems: Theory and Practice, and many papers concerning the matter, Alan Hevner, illustrates the process in figure 3.1.

The left box in figure 3.1 contains the environment. That is in our case the different stakeholders and the MUST community. The design science research box (middle), is the current project, building the artifact, with iterations and evaluations. The last box (right) represents the knowledge base, our understanding of the world, phenomena of the world (observation, classification, measurement and cataloging) and how it makes sense (natural laws, regularities, principles, patterns and theories) [9]. The circles in the figure show the iterative processes, respectively to each case of knowledge.

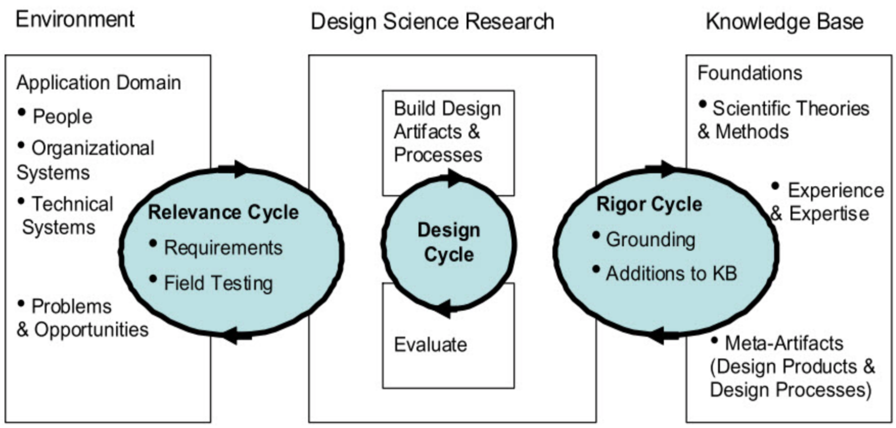

Figure 3.1: Design Science Research: Alan Hevner

A. Hevner [9]

3.1 Seven guidelines of design science research

To clarify how the project is a design science research process, Alan Hevner provides seven guidelines in the 2004 MISQ paper [13]. The guidelines provide a common understanding of the classification and definition of a project as a design science research. Below is a list of the seven guidelines in relation to this research project.

Guideline 1: Design as an artifact

It has to be an artifact, something to be designed and developed as a new contribution to the world. In our case, this artifact and contribution is the mobility index algorithm. A new design and development which specifically measures urban mobility. From descriptive knowledge from mathematics, computer science, urban mobility and accessibility, prescriptive knowledge is gained.

Guideline 2: Problem relevance

The problem, or opportunity, has to be relevant. Which is the objective of design science research, namely to develop technical solutions which are important and relevant to the business. As for the project, solving urban mobility problems regarding transportation are highly relevant and prioritized at every level. It is relevant for improving the inhabitants' perspective on transportation and accessibility, and it is receiving attention and funding from the public sector, both local and national. The relevance and focus can also be drawn to infrastructure planning and environmental issues.

Guideline 3: Design evaluation

It is needed to provide evidence that the artifact is new to the world and making an improvement. To prove that, a continuation of evaluating the design is necessary. The design and ideas were iteratively evaluated together with the municipality as the development went on.

Guideline 4: Research contribution

To keep design science research effective, it must provide clear and verifiable contributions, both to the field and to the knowledge base. Satisfying this guideline is done with conducting this thesis, which aims this research to an academic audience.

Guideline 5: Research rigor

Collecting knowledge base based on fundamental theories, as well as experimental, gives a rigorous design. Throughout the project, the knowledge base is gathered from both academic research and business output from relevant sources in the industry. By the iterative evaluation with stakeholders, the solution comes from an experimental procedure, which makes it a rigorous evaluation.

Guideline 6: Design as a search process

Our problem is characterized in a design search space, with the challenge of finding the optimal design within this space. This means that the design has to be satisfying for the stakeholders. The chosen method of searching the design space, in this case, is the strategy of search heuristics. Heuristic search refers to a search strategy that attempts to optimize a problem by iteratively improving the solution [17]. As the travelling salesman problem, which is established as a standard testbed for new algorithmic ideas [14], possible solutions were generated. These could either be a point in the problem space or a path from the initial state. The process continued with testing, to see if these possible solutions were satisfying by comparing the state reached with the set of goal states provided by the stakeholders. To build the optimal design, it is necessary to prove why the artifact works, and show that it is new to the world. The mobility index algorithm is a new and improved procedure to measure mobility. It is grounded by mathematics, provable in a logic system, and demonstrated used programming and robust technologies.

Guideline 7: Communication of research

The last guideline of defining design science research tells us that the research has to be communicated to both technology-oriented and management-oriented audiences. By attending workshops and meetings in the local mobility environment we discussed the opportunities for innovation involving both public and private mobility related businesses. This process consisted of two main parts; one part containing continuous meetings with the enterprise architect at the municipality, and another part containing an interactive involvement regarding greater discussion around mobility challenges in Bergen. The research topic was communicated through these interactions. By pitching the problem to the relevant audience, an opportunity for evaluation and experimental feedback was presented.

3.2 Identifying the appropriate quadrant

The book ”Positioning Design Science Research for Maximum Impact” written by Alan Hevner and Shirley Gregor, identifies four research project contexts and potential design science research contributions in a 2 x 2 matrix. The x-axis represents the maturity of the problem context from high to low. The y-axis on the other hand represents the current maturity of artifacts that exist as potential starting points for solutions to the research question, also from high to low [10].

First, what is the maturity of our problem? And second, what is the maturity of the existing prescriptive knowledge? The answers give the correct identification of which of the four quadrants in the matrix our problem belongs to.

3.2.1 Invention

The projects in this quadrant involve new solutions to new problems. Inventing a new solution to a new problem is rare and comes with high risk and high reward. This territory is described as an invention breakthrough, the opposite of the accepted ways of thinking and doing. To find a feasible solution, the process depends on creativity, insight, imagination and cognitive skills. Inventions can be categorized as design science research when it provides an artifact which is applied and evaluated in a real-world context, with new contributions of both descriptive and prescriptive knowledge. Design science research in this quadrant has to contain new research of little or unknown application knowledge of understanding the problem, and provide a non-existing type of artifact as the

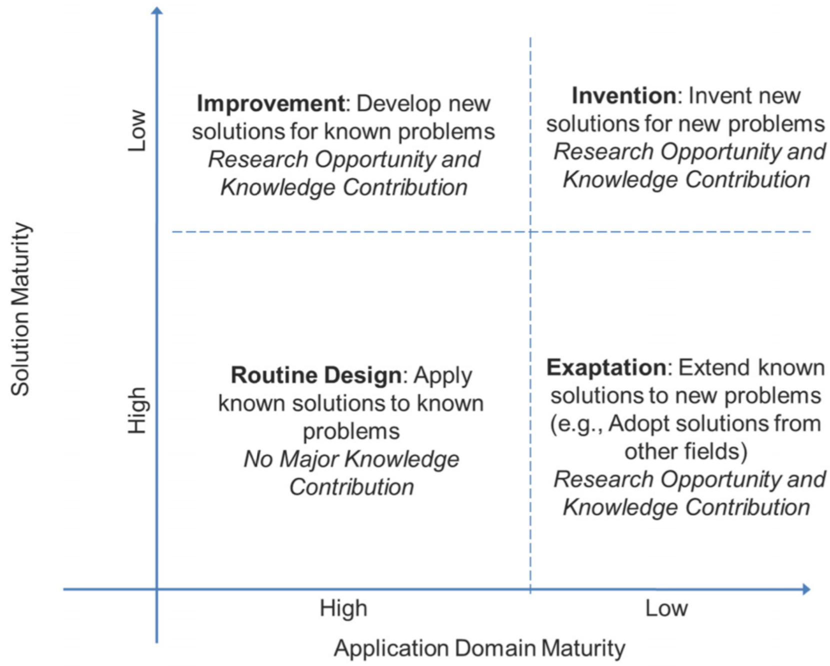

Figure 3.2: Design Science Research Matrix: Alan Hevner
A. Hevner [10]

solution. The non-existing problem makes it even more unnecessary to provide a research question. An example of such invention in this quadrant is the artifact of the first bicycle [10].

3.2.2 Improvements

As figure 3.2 shows, this quadrant is located in the upper left in the matrix, which means that the problem is known. The goal then is to provide a better solution to the known problem, than the former solutions. Here, design science research can provide an improvement to an already existing solution as an artifact to the known problem. An artifact in the form of more efficient and effective products, processes, services, technologies, or ideas. The improvement gives a challenge of proving that the design artifact advances from the descriptive knowledge. Most information technologies and information systems research fall into this quadrant. The level of improvement is therefore more easily evaluated, such as comparing time or efficiency. Examples of such are the improvement of algorithms; search algorithms, encryption algorithms and recommendation algorithms. Design science research in this quadrant allows clear representations of improvements and communicating the new artifact [10].

3.2.3 Exaptation

Extending known solutions to unknown problems may occur for individuals who have experience in multiple disciplines of thought. The term ”exaptation” is originally drawn from biology. It describes an organism which has a trait that was originally used for one purpose, but over time and evolution, that trait is now being used for another purpose. In design science research, we might discover an artifact, which is a solution to a problem, that could also be a solution to a new problem. As for the information technologies and information systems research, this type is common. These projects may exapt existing technologies, for example data gathering by sensor technology or data warehouse, to master a new problem. In exaptation research, it is necessary to demonstrate that the extension of known design knowledge into a new subject is interesting and nontrivial. The new subject must contain challenges that were not contained in the subject in which the techniques have already been applied [10].

3.2.4 Routine design

As for the location of this quadrant (lower left), routine design occurs when known solutions are used to solve known problems. The existing knowledge area and the artifact are well understood. Therefore, situations containing routine design involves non or little research methods. If the design routine leads to new knowledge and artifact discoveries, the procedure will belong in another quadrant.

3.2.5 Where does our research problem belong?

To answer the questions which were asked at the beginning of this section, our research has to be defined and identified by one of the quadrants. Clearly, the field of mobility and its problems and opportunities has a high reach of application domain maturity. In the particular case of the research conducted and presented in this thesis, the solution maturity is seemingly low. The topic of algorithmic measurement gives the proof of research opportunity and knowledge contribution, which categorize this particular artifact in the **improvement** quadrant. What to be improved, is the design artifact of measurement. Design science research is again stated as the fitted research method to our extent. As for the iterative researched descriptive knowledge, the existence of such a measurement formula is to be found in different academic papers and articles. The improvement can be seen as an extension of an accessibility measurement to a mobility measurement, as the artifact. By the great attention of the problem field, the improvement is important not only to our specific case, but to the whole field.

3.3 Validity

The aspect of clarifying if the design research is valid, is defined by the relationship between the research and its findings. Validity measures to what extent the observations reflect the variables that are of our interest. The measurement decides whether our findings are really about what they appear to be about. To avoid logic leaps and false assumptions, the data used to calculate the mobility index value is collected dynamically. From queries on open mobility data, the gathering maintains an objective reference [19].

3.3.1 External validity

By testing the robustness of our findings, they were presented in a group discussion, which measures external validity. The gathered group of relevant individuals in the mobility profession, represents an extension of the stakeholder diversity. As the research question, theory and execution of the implementation were explained, feedback and insight were collected. The common denominator of the understanding of the research findings shows a contribution of new and coveted knowledge to the knowledge base. To conclude with the matter of external validity, it is clear that the development of a mobility index algorithm reaches different interests among the stakeholders. The results of such an algorithm catch an attention to both city planners, innovative decision makers and a public audience. Which states the fact of highly external validity.

3.3.2 Internal validity

Without the knowledge of the researchers, some independent variables can cause a threat to internal validity. As this research does not contain groups for an experiment, the basis of internal validity becomes little. The three categories of internal validity are single group threats, multiple group treats and social threats to internal validity [33]. However, to clarify whether the algorithm is correct, four external readers have verified the basis of the algorithmic execution. None of them found anything directly wrong, some thought the geometric function was a bit overcomplicated for targeting only Bergen. Regardless, the function works generically for larger distances, such that other areas may use the same algorithm concept.

3.3.3 Construct validity

Construct validity relates to the generalization of the result of the experiment to the concept or theory behind the experiment. Some threats relate to the design of the experiment, others to social factors. For the design aspect of construct validity, the threat ”inadequate preoperational explication of constructs” concerns whether the theory behind the constructed research is sufficiently defined [33]. In this case, the mathematical theory which is the basis for the developed algorithm, can be seen as sufficiently defined. The various formulas and calculations are carefully discussed and verified by several mathematicians, and alternative geometric functions are discussed in chapter 4.

Another threat, ”experimenter bias”, defines the effect of the researchers' choices and point of view when constructing the experiments [33]. The theory and the mathematical definition of the algorithm do not receive this threat, due to its abstractness. However, the implementation and declarations of various variables, may fall into the experimenter's bias. The chosen variables of e. g. Γ_2 and Γ_3 (defined in chapter 4), may paint a picture of our view of preferred facilities. For the research conducted, the results reflect more an example than an objective measure, because the algorithm permits other parameters, and is therefore generalized.

Chapter 4

Theory

To understand the complexity and the abstractness of the structure of a class of algorithms, the sections of the algorithm theory needs to be presented. This theory provides a basis for the design of the formal problem, solved according to the rules of syntax. As for structural algorithms concerning geographical subjects, the mathematical terms of linear algebra and vector operations are highly relevant. The great benefit of conducting the solution with programming languages is the capability of trigonometric calculations, though in radians, not degrees. The conversion from degrees to radians is simply because of the requirement of the syntax in programming languages.

4.1 Mobility definition

In this case we see mobility as a combination of how to travel and its costs. Which can be derived to the definition of accessibility. Several early papers have discussed the matter, thereby Hansen, which sees it as “the potential of opportunities for interaction” [11]. A closer and more precise definition, which can be an expansion of Hansen, is Handy and Niemeier [12], noting that this potential is “determined by the spatial distribution of potential destinations, the ease of reaching each destination, and the magnitude, quality, and character of the activities found there”. The view is divided in two aspects, the perspective of individuals and the perspective of buildings and land constructions [20]. The term accessibility expands when the human perspective is included, involving the ability to interact and decide by a non static meaning. The term **mobility** can therefore be defined as a person’s opportunity of interfering in a grid of accessible facilities. We look at mobility from the user’s perspective, including motorists, public transportation, cycling and walking. ”What can I access from my location?" To determinate the level of mobility, we want to locate points of interest close to a given location, and calculate the rate of accessibility. However, to measure mobility we need to look into the various factors of the definition. Factors which can be translated into numbers, and then formulate a mathematical model of measurement.

4.2 Index

The trivial definition of an index, as commonly known, can be defined the same way as a function. As the domain of the project concerns the mobility aspect, a specific function accumulates an index. A comparison to the financial world, an index is a dynamic variable, based on a set of parameters.

4.3 Geographical distance calculations

First, to do mathematical estimations, the earth needs to be defined as a three-dimensional geometrical object. For this calculations, two options are considered: the earth as a sphere or the earth as an ellipsoid.

4.3.1 The earth as a sphere

The basis of a sphere is defined mathematically as a set of points which all contain the same exact distance (r) to one specific point, as a circle, but in three-dimensional space. These are respectively referred to as the radius and the center. A sphere with center (x_0, y_0, z_0), and a radius r, gives the locus of all points (x, y, z), e.i.

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2$$

With a sphere as an assumption of the earth, the next step is to define the appropriate coordinate system, and deriving to the known geographical coordinate system (latitude and longitude). For the spherical coordinate system, an exact position of a point is decided by three values:

1. The radial distance, from the specific point to a fixed origin (Euclidean distance)
2. Its zimuthal angle (polar angle) measured from a fixed zenith direction (opposite direction of the gravitational force at that point)

3. The orthogonal projections azimuthal angle, on a reference plane that passes through the origin and is orthogonal to the zenith, measured from a fixed reference direction on that plane

The three values represent the sphere coordinates: (ρ, θ, φ). To reach the geographical coordinates from the spherical coordinate system, the values need to be converted. First value, ρ, is the altitude value, which for geographical cases, is dropped. Second, θ, refers to the geocentric latitude, conversion:

$$latitude = \theta - 90°$$

Third, φ, refers to the geocentric longitude:

$$longitude = \varphi$$

Calculating distances on the surface of a sphere, by two given sets of position points in geographical coordinates, may now be processed by an algorithmic procedure. A well-known and often used formula for distance calculations for this geometrical object is the Haversine Formula, also used by Google Maps Utils [8].

4.3.2 The Haversine Formula

Versine (versed sine) is a trigonometric function. The function exists in one of the earliest trigonometric tables. Abstractly, the versine of an angle is 1 minus its cosine [3]. The most important related function is half a versine, Haversine formula of navigation. The purpose of the formula may be comparable to the Pythagorean Theorem, namely finding the shortest distance between two location points. The difference, which is highly significant, is the issue of the concept of the Great Circle distance. The distance at the surface of a sphere.

Let the angle from the center to two circle periphery points be Θ

$$\Theta = \frac{d}{r}$$

Where:

- d is the distance between two points on the sphere
- r is the radius of the sphere, in this case the earth radius: $r = 6378.137$ km [32]

Haversine allows us to compute the distance based on two sets of coordinates:

- Latitude point 1: φ_1, and point 2: φ_2.

- Longitude point 1: λ_1, and point 2: λ_2.

The law of Haversine gives:

$$hav(\Theta) = hav(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)hav(\lambda_2 - \lambda_1)$$

Then, as the above Haversine formula with the central angle, latitude and longitude, the $hav(\theta)$ function is:

$$hav(\theta) = \sin^2\left(\frac{\theta}{2}\right) = \frac{1-\cos(\theta)}{2}$$

Which computes half a versine of the angle θ. Further, to compute the distance d, apply the inverse haversine (archaversine) to $h = hav(\Theta)$

$$d = (r)archav(h) = (2r)arcsin\left(\sqrt{h}\right)$$

$$d = (2r)arcsin\left(\sqrt{hav(\Theta) = hav(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)hav(\lambda_2 - \lambda_1)}\right)$$

$$= (2r)arcsin\left(\sqrt{\sin^2\left(\frac{(\varphi_2 - \varphi_1)}{2}\right)\cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{(\lambda_2 - \lambda_1)}{2}\right)}\right)$$

[3]

One of the reasons for the ubiquitous of this formula is its translation to the interpreted, object-oriented, high-level programming languages. To explain the triviality of the concept of this formula, it is represented as pseudo code below:

Algorithm 1: Haversine Formula

```
Result: Calculated distance between two sets of coordinates
initialization;
R = 6378.137 km
 lon1, lat1, lon2, lat2 = radians[lon1, lat1, lon2, lat2]
 dlon = lon2 - lon1
 dlat = lat2 - lat1
 a = sin(dlat/2)^2 + cos(lat1) * cos(lat2) * sin(dlon/2)^2
 c = 2 * asin(sqrt(a))
 d = R * c

 return d
```

Accuracy error is considered to be 1 meter [16].

4.3.3 The earth as an ellipsoid

By deforming a sphere by directional scaling, an ellipsoid (spheroid) is obtained. Instead of a set of points which all contain the same exact distance (axes), the three values are not equivalent, namely an ellipsoid. The standard form of the implicit equation of the spheroid:

$$\frac{x^2 + y^2}{a^2} + \frac{z^2}{c^2} = 1$$

Where:

$$a, c \in R > 0$$

The earth is better approximated by a spheroid than a sphere, simply because the density and the balance of gravitational force and centrifugal force. In other words, because earth is rotating.

4.3.4 The Vincenty's Formula

The Vincenty's formula accounts for the ellipsoid model of the earth. It is far more accurate than the Haversine formula, and even more mathematically complex [16]. As the Haversine formula, two sets of coordinates are given as input to calculate the distance between them. Where ϕ_1, ϕ_2 are the latitude of the points, and L_1, L_2 the longitude of the points. Vincenty's uses the reduced latitudes (latitude on the auxiliary sphere): $U_1 = arctan((1-f)tan(\phi_1))$, $U_2 = arctan((1-f)tan(\phi_2))$, where f is the flattening of the ellipsoid (measure of the compression of a sphere along the diameter). The difference of the two longitudes is presented as $L = L_2 - L_1$. The variable σ is the angular separation between the points, σ_1 is the angular separation between the point and the equator, and σ_m is angular separation between the midpoint of the line and the equator. The distance has the notation s. α_1, α_2 is the azimuths, which is not returned in our algorithm.

$$\sin(\sigma) = \sqrt{(\cos(U_2)\sin(\lambda))^2 + (\cos(U_1)\sin(U_2) - \sin(U_1)\cos(U_2)\cos(\lambda))^2}$$

$$\cos(\sigma) = \sin(U_1)\sin(U_2) + \cos(U_1)\cos(U_2)\cos(\lambda)$$

$$\sigma = arctan2\left(\sin(\sigma)\cos(\sigma)\right)$$

$$\sin(\alpha) = \frac{\cos(U_1)\cos(U_2)\sin(\lambda)}{\sin(\sigma)}$$

$$\cos(2\sigma_m) = \cos(\sigma) - \frac{2\sin(U_1)\sin(U_2)}{\cos^2(\alpha)} = \cos(\sigma) - \frac{2\sin(U_1)\sin(U_2)}{1-\sin^2(\alpha)}$$

$$C = \frac{f}{16}\cos^2(\alpha)[4 + f(4 - 3\cos^2(\alpha))]$$

$$\lambda = L + (1-C)f\sin(\alpha)\sigma + C\sin(\sigma)[\cos(2\sigma_m) + C\cos(\sigma)(-1 + 2\cos^2(2\sigma_m))]$$

$$u^2 = \cos^2(\alpha)\left(\frac{a^2 - b^2}{b^2}\right)$$

$$A = 1 + \frac{u^2}{16384}\left(4096 + u^2[-768 + u^2(320 - 175u^2)]\right)$$

$$B = \frac{u^2}{1024}\left(256 + u^2[-128 + u^2(74 - 47u^2)]\right)$$

$$\Delta\sigma = B\sin(\sigma)\{\cos(2\sigma_m) + \frac{1}{4}B\Bigg(\cos(\alpha)[\cos(-1 + 2\sigma_m)] - \frac{B}{6}\cos[2\sigma_m][-3 + 4\sin^2(\sigma)][-3 + 4\cos(2\sigma_m)]\Bigg)\} \quad (4.1)$$

$$s = bA(\sigma - \Delta\sigma)$$

$$\alpha_1 = arctan2(\cos(U_2)\sin(\lambda), \cos(U_1)\sin(U_2) - \sin(U_1)\cos(U_2)\cos(\lambda))$$

$$\alpha_2 = arctan2(\cos(U_1)\sin(\lambda), -\sin(U_1)\cos(U_2) + \cos(U_1)\sin(U_2)\cos(\lambda))$$

[3]

The value of interest in this case, is the distance s, $s = bA(\sigma - \Delta\sigma)$, where $b = (f-1)a$, and a represents the length of semi-major axis of the ellipsoid (radius at equator).

4.4 Theory of the Mobility Index Algorithm

To reach the goal of computing an index, as defined earlier, the different mobility aspects have to be mathematically translated into numbers. Next, the different factors will be weighted and summarized. As the determination of distance calculation is decided, the available access points are covered. The theory of this computation is defined in three formulas, where the mobility measurement is the sum of those three. The first formula is the following;

$$\Gamma_1 = \log_e \left(\sum_{n=1}^{n_m} F \right) * 5,\ F = \forall(dis(\lambda_1, \varphi_1, [\lambda_2, \varphi_2]))$$

While

$$dis(\lambda_1, \varphi_1, [\lambda_2, \varphi_2]) \leq 1$$

Where Γ_1 is the number of all nearby transportation access points and mobility services. Using the computation for logarithm makes the $n+1$ access point less weighted than n, for all $[n_1, n_2 ... n_m]$. The function $dis(\lambda_1, \varphi_1, [\lambda_2, \varphi_2])$ gives all causes when λ_1 and φ_1 receives a distance of 1 or less kilometer. The set $[\lambda_2, \varphi_2]$ is all coordinates with a transportation access point or mobility service. These are all bus stops, tram stops, ferry access and train stops, as well as all stations for city bikes. The function will either be an implementation of the Haversine formula or the Vincenty's formula. One could argue that the two different formulas may be unnecessary with the short distance calculations, and the Pythagoras distance formula may give a satisfiable result. As the Pythagoras distance is executed in the two-dimensional plane, Haversine and Vincenty's have a circular symmetry around the axis.

The quantifier $\forall$ gives every case given, when the distance function calculates ≤ 1 kilometer as the distance. The formula is weighted by multiplying a factor of 5. The constant is an approximation of the necessity of this mobility aspect, based on a GIS-analysis done by Bergen municipality. The report, including walking accessibility in Bergen and the importance of public transportation availability, gives the results of weighing regarding different facilities. This constant reflects a weigh of 12%, which include all public transport and mobility services reach within walking distance [1].

The next formula considers the accessibility of points of interest. This estimation is given in the cost of travel; time;

$$\Gamma_2 = \frac{\left(\frac{p_1 t_1 + p_2 t_2 + ... + p_n t_n}{p_n} * -1 \right)}{10}$$

Where p_1 is position and $p_1 t_1$ is the time it takes to reach the position. Since lowest cost of travel is preferred, the time average is multiplied with -1. The value 10 is the dividend, and represents the weighting of the formula. The time of travel is calculated regarding to public transportation.

Finally, the last formula represents all facilities which are reachable in walking distance;

$$\Gamma_3 = \left(\sum_{n=1}^{n_m} G\right) + log_e \left(\sum_{n=1}^{n_m} H\right)$$

$$G = \left(\frac{p_1}{p_1} + \frac{p_2}{p_2} + ... + \frac{p_n}{p_n}\right), \ \ p_1 \neq p_2 \neq ... \neq p_n, \ p \neq 0, \ pt \leq 15_w$$

$$H = (p_1 + p_2 + ... + p_n), \ \ p_1 \neq p_2 \neq ... \neq p_n, \ \ pt \leq 15_w$$

Where G and H represent all different accessible services (shops, pharmacies, schools, kindergartens etc). When we have different multiple services, the numbers are added and the sum is calculated logarithmic (for H). The notation for function G may be defined in different ways, the purpose for G is, however, to give the outcome of all different amenities available. $pt \leq 15_w$ defines a walking distance of 15 minutes or less. To maintain the benefit of choice, $(\sum_{n=1}^{n_m} G)$ and $log_e\,(\sum_{n=1}^{n_m} H)$ are added, and gives Γ_3. The denominators in G can be 0, but the function factors in G which has 0 as denominator, will of course be excluded.

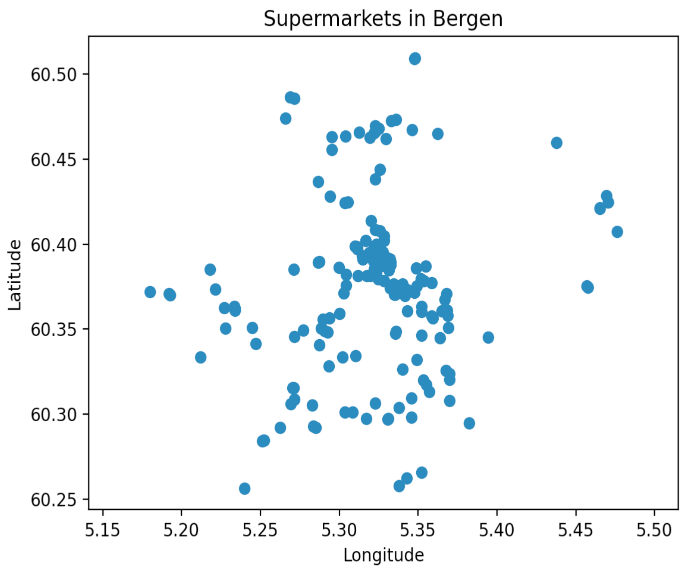

Figure 4.1: Plotting the amenity: Supermarket

The figure 4.1 shows all collected locations of the amenity ”Supermarket” across Bergen municipality. The x-axis points the longitudes, and the y-axis points the latitudes. If a point in the graph is located within 15 minutes of walking (Γ_3), it is caught by the formula which defines the algorithm of Γ_3.

4.5 Feature scaling

In order to make sense of the values returned by the mobility index algorithm, a range has to be normalized. The range of all features should be normalized so that each feature contributes approximately proportionately to the final level of measurement. The method chosen to normalize the range is the rescaling (min-max normalization) method. The range retrieved with this method has the interval of [0, 1]. To reach more readable numbers, the value is multiplied with a factor of 10.

$$x' = \frac{x - min(x)}{max(x) - min(x)} * 10$$

To find the $max(x)$ and $min(x)$ values, we need to run all addresses in Bergen municipality on the mobility index algorithm, and sort out the highest and lowest number. The reason for scaling the value is not only the benefit of a more understanding value, but also the aspect of comparing. Since the algorithm is based on dynamic data, it could work in other cities than in the study. Then we got a mathematical grounded score of comparing between cities or areas in other parts of the country.

4.6 Is the formulae provable in a logic system?

According to the professor and author of the book ”Introduction to Mathematical Logic”, Michal Walicki, the fundamental goals of all scientific inquiry is to achieve precision and clarity of a body of knowledge [30]. The book gives three postulates on this matter:

- all assumptions of a given theory are stated explicitly;
- the language is designed carefully by choosing some basic, primitive notions and defining others in terms of these ones;
- the theory contains some basic principles - all other claims of the theory follow from its basic principles by applications of definitions and some explicit laws

In other words, it boils down to formulating an axiomatization in a formal system. A well known example of such an axiomatization, is the Euclidean geometry. Euclid explained four postulate axioms, such that he could show how many geometrical statements which could be logically derived from a set of principles [30]. These postulates were supposed to be intuitively obvious:

- A1: Given two points, there is an interval that joins them.
- A2: An interval can be prolonged indefinitely.
- A3: A circle can be constructed when its center, and a point on it, are given
- A4: All right angles are equal

The current section is given to show soundness of the constructed formulae. By soundness of a system, the system (i.e. our formulae), is trusted, and thereby correct. A more specific definition of such a system, is when e.g. Γ_1 calculates the correct answer, and a set of coordinates is given, it has soundness only if the coordinates given really are the only coordinates which are correct by the premises of the formula of Γ_1.

The following sub sections will provide provability of the formulas previously constructed. In first order logic (FOL), the formulas are given in a set Σ, where the user can make detailed choices, but in a uniform way. In FOL, the variables range only over individual elements from the interpretation domain, and its atomic formulae consist of applications of relation symbols. If $A \in WFF_{FOL}$ then $\exists x A \in WFF_{FOL}$, only when x is an individual variable $x \in V$ [30].

4.6.1 Γ_1

The first formula gives the number of the nearby transportation access points and mobility services, and the benefit of differential reach. But is it the case that:

$$\models \forall x(L \rightarrow S)?$$

Which states that at every location given, we can find a nearby transportation access point or mobility service. To show provability let M, v be arbitrary such that

$$M \models \forall x(L \rightarrow S)$$

$$\forall x(L \rightarrow S) \Leftrightarrow \exists x L(x) \rightarrow S$$

$$\llbracket \exists x L(x) \rightarrow S \rrbracket_v^M = 1$$
$$\Leftrightarrow \llbracket \exists x L(x) \rrbracket_v^M = 0 \vee \llbracket S \rrbracket_v^M = 1$$
$$\Leftrightarrow \llbracket \neg \exists x L(x) \rrbracket_v^M = 1 \vee \llbracket S \rrbracket^{M_v} = 1$$
$$\Leftrightarrow \llbracket \forall x \neg L(x) \rrbracket^{M_v} = 1 \vee \llbracket S \rrbracket_v^M = 1$$
$$\Leftrightarrow (For\ all\ a\ \in M, \llbracket \neg L \rrbracket_{v[x \mapsto a]}^M = 1) \vee \llbracket S \rrbracket_v^M = 1$$
$$\Leftrightarrow For\ all\ a\ \in M, (\llbracket \neg L \rrbracket_{v[x \mapsto a]}^M = 1 \vee \llbracket S \rrbracket_v^M = 1)$$
$$\Leftrightarrow For\ all\ a\ \in M, (\llbracket \neg L \rrbracket_{v[x \mapsto a]}^M = 1 \vee \llbracket S \rrbracket_{v[x \mapsto a]}^M = 1)$$
$$\Leftrightarrow For\ all\ a\ \in M, (\llbracket L \rrbracket_{v[x \mapsto a]}^M = 0 \vee \llbracket S \rrbracket_{v[x \mapsto a]}^M = 1)$$
$$\Leftrightarrow For\ all\ a\ \in M, \llbracket L \rightarrow S \rrbracket_{v[x \mapsto a]}^M = 1$$
$$\Leftrightarrow For\ all\ a\ \in M, \llbracket \forall x (L \rightarrow S) \rrbracket_{v[x \mapsto a]}^M = 1$$
$$M \models \forall x (L \rightarrow S)$$

After the result of showing that the formula is provable, we see easily that the formula returns 0 if there is not any nearby transportation access points or mobility services. For Γ_3, we could express the existence of amenities the same way as the transportation access points and mobility services are expressed in this formula.

4.6.2 Γ_2

The second formula gives the representation of the average time of reaching points of interests in Bergen municipality. For this formula, when the input address is given, there exists an average time to reach a particular point of interest. The abbreviation of Γ_2 can be written as:

$$A \rightarrow \exists x T$$

Now, by showing provability, let M, v be arbitrary:

$$A \rightarrow \exists x T \Leftrightarrow \exists x (A \rightarrow T)$$
$$M \models_v A \rightarrow \exists x T \Leftrightarrow \llbracket A \rightarrow \exists x T \rrbracket_v^M = 1$$
$$\Leftrightarrow \llbracket A \rrbracket_v^M = 0 \vee \llbracket \exists x T \rrbracket_v^M = 1$$
$$\Leftrightarrow \llbracket A \rrbracket_v^M = 0 \vee there\ is\ an\ a\ \in M\ s.t.\ \llbracket T \rrbracket_{v[x \mapsto a]}^M = 1$$
$$\Leftrightarrow there\ is\ an\ a \in M\ s.t.\ (\llbracket A \rrbracket_v^M = 0 \vee \llbracket T \rrbracket_{v[x \mapsto a]}^M = 1)$$
$$\Leftrightarrow there\ is\ an\ a \in M\ s.t.\ (\llbracket A \rrbracket_{v[x \mapsto a]}^M = 0 \vee \llbracket T \rrbracket_{v[x \mapsto a]}^M = 1)$$
$$\Leftrightarrow there\ is\ an\ a \in M\ s.t.\ (\llbracket A \rightarrow T \rrbracket_{v[x \mapsto a]}^M = 1)$$
$$\Leftrightarrow \llbracket \exists x (A \rightarrow T) \rrbracket_v^M = 1$$
$$M \models_v \exists x (A \rightarrow T)$$

Chapter 5

Implementation

This chapter explains the implementation of the mobility index algorithm, which has its mathematical base of the formulas constructed and explained in chapter 4. The implementation builds on this mathematical ground, and accumulates in a software engineered system, which gives the promised solution of the technological demonstration. The chapter provides examples of dynamic data gathering and the implementation of each formula. Our application is implemented in Python, with libraries containing JavaScript, HTML and CSS, kept in a virtual environment.

5.1 Challenges regarding code-translation

Translating mathematical formulas to machine readable code bears different challenges. As the theoretical notation of the human readable mathematical language can be expressed with abstract and unknown variables and constants, the execution has to be done by a machine readable language. The abstractness of the mathematical language brings challenges to the process of defining Python code with real arguments. Variables such as λ and φ are the expression of coordinates, p_n is a point consisting of coordinates as attributes, and $p_n t_n$ is the value of time it takes to reach the point. All of these variables need to be in the space of natural numbers to calculate the formulas, and therefore has to be translated into machine readable language with correct syntax.

The challenge of retrieving these natural numbers lies in the gathering of data, which should represent the different variables. First, λ and φ are gathered in two different ways, for two different purposes, respectively. The first set is easily collected from the user input, but the second set, which is a set of sets of coordinates, has to be found every time the formula of Γ_1 is called. Also, $[\lambda_2, \varphi_2]$ is the set that contains different sources of gathering sets of coordinates. Bearing in mind the importance of a dynamic approach, the challenge is solved by requesting data directly on the sources with generic queries. In this way, the soundness of the system is held at our end of the execution.

Another challenge is how and where we can collect the updated data regarding all amenities, especially with change or extension in mind as a possible feature. The formula (Γ_3) states the amenities abstractly as $[p_1, p_2, \ . \ . \ . \ p_n]$, but finding and translating the points into latitudes and longitudes, becomes the next challenge. So, "how does one get all amenities imaginable and its coordinates?" Well, luckily the existence of voluntarily open source contributions is regarding this topic (as well). The benefit and the main reason for choosing such an approach of selecting data about amenities is the magnitude of open source possibilities. The greatest open source related to this matter is the OSM. The huge open database is an enormous advantage, as the different data is continuously updated by the very inhabitants who interest and relate to the information. Every contributor to this open database system keeps track of the reality which regards the containing data. Conveniently, the OSM Overpass API gives the opportunity of querying the chosen data, i.e. amenities.

Defining theoretical, mathematical formulas brings challenges, due to the power of high theoretic abstractness. Nevertheless, the Python Programming Language is highly capable of executing complex procedures, and the challenges mentioned above only concerns the aspect of variable findings. As the challenges are far from impossible, the implemented functions have one main feature in common, all arguments are a set of coordinates.

5.2 Dynamic data gathering

Most of our data handling is asynchronously changed over time, which is necessary to reach the full mobility aspect. Therefore, the data is collected in json format directly from the source, with the use of cURL requests. The data may be modified and updated by its provider. These modifications will automatically be the new basis of the algorithms procedures. By the "Norwegian license of public data" (NLOD), the different mobility providers give a direct access by an HTTP request.

Easy example of sending data between its provider through a web browser and our system, via cURL request:

```
headers = {
    'Client-Identifier': 'IDENTIFIER',
}

cycle_stations = requests.get('https://gbfs.urbansharing.com/
    bergenbysykkel.no/station_information.json', headers=headers)

cycle_stations = cycle_stations.json()
```

By the request we receive the response as a json object.


```
{
 "last_updated": 1540309984,
 "data": {
    "stations": [
       {
          "address": "Kaigaten",
          "capacity": 19,
          "lon": 5.327895003562276,
          "name": "Raadhuset",
          "station_id": "223",
          "lat": 60.39168071231371
       },
```


We can now easily pick our preferred attributes for calculating a distance from the input location to the nearby city cycles in Bergen. As for the algorithm which finds the nearby mobility points, longitude and latitude is set as arguments. By using this dynamic data gathering approach, our mobility index algorithm finds all cycles updated at the source, at any given time.

Another dynamic data collection approach, different from json format gathering, is the connection to the data lake. Microsoft Azure provides managing data objects through the Python package Azure Blob storage, which is a client library.

Also, as reaching local amenities by walking is an important part of the mobility picture. The amenities are collected from the Open Street Map database, by the Overpass API. Queries on the data give the selected amenities with its attributes, such as latitude and longitude.

```
pharmacy_query = """
[out:json];
area["name"="Bergen"][admin_level=7];
(node["amenity"="pharmacy"](area);
 );
out center;
"""

def get_coordinates_by_overpass_query(query):
    overpass_url = "http://overpass-api.de/api/interpreter"
    response = requests.get(overpass_url,
                            params={'data': query})
    data = response.json()

    coords = []
    for element in data['elements']:
        if element['type'] == 'node':
            lon = element['lon']
            lat = element['lat']
            coords.append([lon, lat])
        elif 'center' in element:
            lon = element['center']['lon']
            lat = element['center']['lat']
            coords.append([lon, lat])

    return coords
```

5.3 The Mobility Index Algorithm

An abstract mathematical model for the algorithm is constructed and defined, and covers the theory behind the implementation. The mobility index algorithm builds on the formulas, which gives the starting points and a base of the development. The control flow begins with only one input from the user; an address. The address is then converted into computer-readable data, i.e. coordinates, by the built in Python Geocoder and based on online, dynamic geographical data by OpenCage. It is essential that the procedures are dynamical, such that the necessary future changes will not rely on changes in the source code. Next, this given location is one of two arguments in the distance function. The distance function converts decimal degrees into radians, before running the Haversine formula on two sets of coordinates, as explained in chapter 4. For this implementation, the Haversine formula is chosen over the Vincenty's formula. The reason for implementing Haversine over Vincenty's is its better running time and its satisfying accuracy of distance calculation.

5.3.1 Implementation of Γ_1

The first equation of three in total, Γ_1, finds nearby transport points and mobility services. The conversion from mathematics to Python code is, in this case, rather straight forward, after the sets of coordinates are collected. The formula Γ_1 from the Theory Chapter is implemented by dynamic programming.

$$\Gamma_1 = \log\left(\sum_{n=1}^{n_m} F\right) * 5,\ F = \forall(dis(\lambda_1, \varphi_1, [\lambda_2, \varphi_2]))$$

While

$$(dis(\lambda_1, \varphi_1, [\lambda_2, \varphi_2])) \leq 1$$

From the formula above, several Python functions are defined. Functions that calculate the distance, collecting nearby stops and mobility services, all with latitude and longitude as arguments. As the formula states, the sum of all nearby stops and mobility services within one kilometer is multiplied with the factor 5.

```
def gamma1_equation(lon, lat):
    v = len(get_nearby_stops_places(lon, lat))
    w = get_nearby_cycle_stations(lon, lat)
    x = v+w
    if x > 0:
        return math.log(v+w)*5
    else:
        return 0
```

From this point, with the location as the argument, the overall mobility picture within a radius of one kilometer from the given input, is collected.

5.3.2 Implementation of Γ_2

The second formula defined in chapter 4 calculates the average time to reach all points of interest. The POIs are defined in a static list of json objects, and the travel time is decided by the Journey Planner API of Entur. The travel time is reached by calling a function which sends a query to the Journey Planner API, by two sets of longitude and latitude as arguments.

```
def trip(from_lon, from_lat, to_lon, to_lat):
        query = QUERY_JOURNEY.format(**{
            'from_lat': from_lat, 'from_lon': from_lon,
```

```
            'to_lat': to_lat, 'to_lon': to_lon
        })
        r = requests.post(API_URL,
            json={'query': query},
            headers=h
        )
        d = json.loads(r.text)['data']['trip']['tripPatterns'][0]

        duration = d['duration']

        return duration
```

The query sent to the API is modified and manipulated with the preferred coordinates, which gives the wanted duration. However, the syntax is decided by Entur, and gives an output in json format. The travel time is defined in seconds, which makes the specific function easier in the mathematical continuation of the equation.

```
QUERY_JOURNEY = """{{
  trip(
    from: {{
       name: ""
           coordinates: {{
            latitude: {from_lat},
            longitude: {from_lon}
            }}
    }}
    to: {{
       name: ""
           coordinates: {{
            latitude: {to_lat},
            longitude: {to_lon}
            }}
    }}
  )
  {{
    tripPatterns {{
      duration
      }}
    }}
}}"""
```

```
def gamma2_equation(lon, lat):
    duration_list = get_duration_list(lon, lat)
    Sum = sum(duration_list)
    gamma2 = -1*((Sum/60)/len(duration_list))/10

    return gamma2
```

Finally, the formula of Γ_2 is converted to Python code, bearing in mind the factor of -1, explained in the previous chapter.

5.3.3 Implementation of Γ_3

The formula of Γ_3 calculates the accessibility to nearby amenities. As the formula given in chapter 4, it is defined by adding the sum of the different amenities with the logarithm of the total sum of all amenities. The procedure of calculating what is nearby, is the same as in Γ_1, the Haversine formula of distance between two sets of coordinates. The amenities included in the data are supermarkets, pharmacies, schools, kindergartens, restaurants, bars and the government regulated liquor stores. These are collected in json format by sending quieries on the Open Street Map Overpass API.

$$\Gamma_3 = \left(\sum_{n=1}^{n_m} G\right) + log_e\left(\sum_{n=1}^{n_m} H\right)$$

$$G = \left(\frac{p_1}{p_1} + \frac{p_2}{p_2} + ... + \frac{p_n}{p_n}\right), \;\; p_1 \neq p_2 \neq ... \neq p_n, \; p \neq 0, \; pt \leq 15_w$$

$$H = (p_1 + p_2 + ... + p_n), \;\; p_1 \neq p_2 \neq ... \neq p_n, \;\; pt \leq 15_w$$

The code snippet below defines the calculation of the formula above in Python code.

```
def gamma3_equation(lon, lat):
    G = 0
    H = get_bars(lon, lat) +
        get_kindergartens(lon, lat) +
        get_pharmacy(lon, lat) +
        get_restaurants(lon, lat) +
        get_schools(lon, lat) +
        get_supermarkets(lon, lat) +
        get_vinmonopol(lon, lat)

    if get_bars(lon, lat) > 0:
        G = G+1
    if get_kindergartens(lon, lat) > 0:
        G = G+1
    if get_pharmacy(lon, lat) > 0:
        G = G+1
    if get_restaurants(lon, lat) > 0:
        G = G+1
    if get_schools(lon, lat) > 0:
        G = G+1
    if get_supermarkets(lon, lat) > 0:
        G = G+1
    if get_vinmonopol(lon, lat) > 0:
```

```
        G = G+1
    if H > 0:
        return G+math.log(H)
    else:
        return G
```

Reaching a measured mobility level depends on the accessible mobility opportunities. The highest level of accessibility is therefore defined at the greatest transportation hub, located at the city center.

5.3.4 Summarize Γ_1, Γ_2 and Γ_3

Now as we have the three functions representing the three formulas, the return values need to be summarized. Common for the three functions are the arguments, which are the longitude and latitude from the converted input address. Then, by defining a final function, which summarize the three numbers from the three formulas respectively, the index is calculated.

```
def get_mobility_index(lon, lat):
    index = gamma1_equation(lon, lat) +
             gamma2_equation(lon, lat) +
             gamma3_equation(lon, lat)

    return index
```

The function above is called when the Folium map is generated, rendered by Flask, and presented as a web application, simultaneously as the input address is given.

Chapter 6

Interpretation and results

To explain the findings and the results of the development, an example with an arbitrary input, and step by step calculations will be presented. First, the formulas have to be provable. If not, the value is of course 0 for our formulae.

6.1 Step by step calculation

We calculate the formula with arbitrary real numbers:

6.1.1 Γ_1

$$\Gamma_1 = \log_e \left(\sum_{n=1}^{n_m} F \right) * 5,\ F = \forall(dis(\lambda_1, \varphi_1, [\lambda_2, \varphi_2]))$$

While

$$(dis(\lambda_1, \varphi_1, [\lambda_2, \varphi_2])) \leq 1$$

In our calculation, the address "Borgermester Platous gate 9" is the input. From the Python Geocoder and OpenCage API, we have:

$$\lambda_1 = 5.330642,\ \varphi_1 = 60.385468$$

The set of $[\lambda_2, \varphi_2]$ is the longitude and latitude of all transportation points and mobility services, gathered from the Data Lake and dynamically from Bergen Bysykkel, given in json. The Haversine formula selects the set of coordinates within 1 kilometer. The list of coordinates has a length of 34 transportation points and 20 city cycle stations, which gives:

$$F = \forall(dis(\lambda_1, \varphi_1, [\lambda_2, \varphi_2])) = 54$$

$$\left(\sum_{n=1}^{n_m} F\right) * 5 = 270$$

$$\log_e(270) = 5,598421958998$$

$$\Gamma_1 = 5,598421958998$$

6.1.2 Γ_2

We follow the same procedure as the previous section by continuing with the same arbitrary numbers. The calculation of the formula gives:

$$\Gamma_2 = \frac{\left(\frac{p_1t_1+p_2t_2+...+p_nt_n}{p_n} * -1\right)}{10}$$

With ”Borgermester Platous gate 9” as input, we have:

$$p_1t_1 + p_2t_2 + ... + p_nt_n = 125,41\overline{6}$$

$$p_n = 7$$

$$\frac{\left(\frac{125,41\overline{6}}{7} * -1\right)}{10} = -1,791\overline{6}$$

$$\Gamma_2 = -1,791\overline{6}$$

6.1.3 Γ_3

As Γ_3 finds nearby amenities, the formula is provable by the structure M as in Γ_1. The local amenities at the given input address are gathered by queries on the OSM Overpass API. With the address ”Borgermester Platous gate 9”, we have:

$$\Gamma_3 = \left(\sum_{n=1}^{n_m} G\right) + log_e\left(\sum_{n=1}^{n_m} H\right)$$

$$G = \left(\frac{p_1}{p_1} + \frac{p_2}{p_2} + ... + \frac{p_n}{p_n}\right), \;\; p_1 \neq p_2 \neq ... \neq p_n, \; p \neq 0, \; pt \leq 15_w$$

$$H = (p_1 + p_2 + ... + p_n), \;\; p_1 \neq p_2 \neq ... \neq p_n, \;\; pt \leq 15_w$$

Amenities	p
Supermarkets	23
Schools	14
Kindergartens	6
Restaurants	76
Liquor stores	1
Bars	20
Pharmacies	6

Table 6.1: Amenities: Γ_3

$$G = (\frac{23}{23} + \frac{14}{14} + \frac{6}{6} + \frac{76}{76} + \frac{1}{1} + \frac{20}{20} + \frac{6}{6}), \;\; p_1 \neq p_2 \neq ... \neq p_n, \;\; pt \leq 15_w$$

$$G = 7$$

$$H = (23 + 14 + 6 + 76 + 1 + 20 + 6), \;\; p_1 \neq p_2 \neq ... \neq p_n, \;\; pt \leq 15_w$$

$$H = 146$$

$$\left(\sum_{n=1}^{n_m} G\right) + log_e \left(\sum_{n=1}^{n_m} H\right) = 7 + \log_e(146)$$

$$= 7 + 4,9836066217$$

$$\Gamma_3 = 11,9836066217$$

6.1.4 The mobility index calculation

After the above step by step calculations, we now have all three values from the three formulas, respectively. The **sum** of Γ_1, Γ_2 and Γ_3 will give us the total result of the mobility index, at the given address. If we continue with the example address in the previous calculation, we have:

$$\Gamma_1 + \Gamma_2 + \Gamma 3$$

$$= 5,598421958998 - 1,791\overline{6} + 11,9836066217$$

$$= 15,790361914031333 \approx 15,8$$

As the application renders HTML to present the result of the calculations done in the implementation, the users perspective is taken care of. Figure 6.1 shows the output of the mobility index algorithm, with the address "Borgermester Platous gate 9" as input, with a mobility index value equal to the calculations done above.

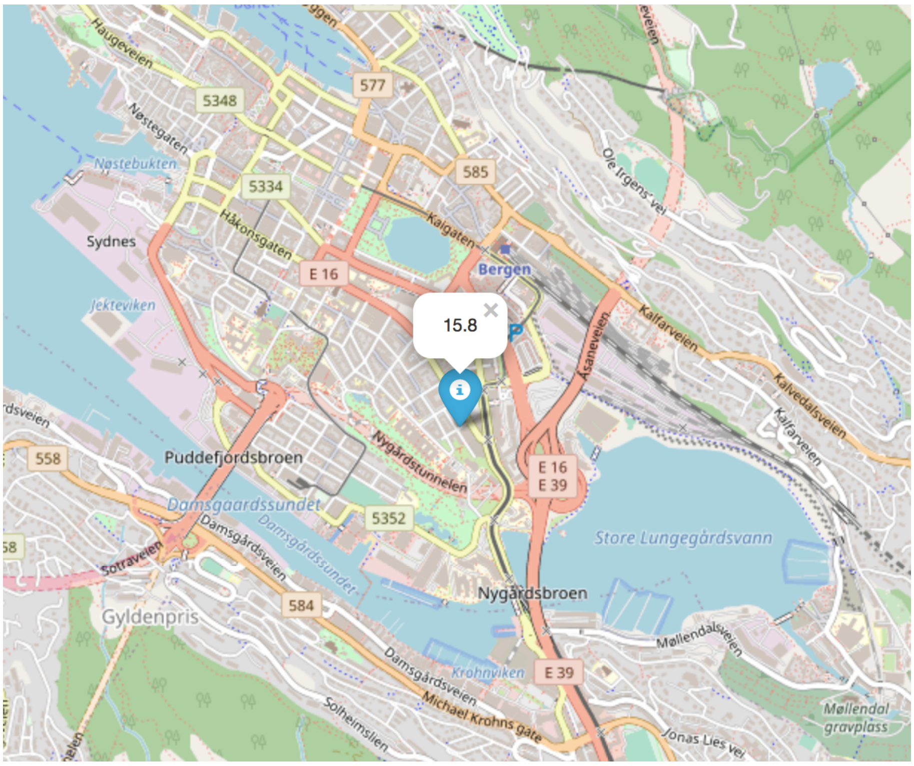

Figure 6.1: Result of the mobility index algorithm at "Borgermester Platous gate 9", without max-min normalization

With a measured level of mobility at 15.8, how do we know if it is good or bad? To answer the question, we have to consider scaling the values.

6.2 Scaling

In chapter 4 we presented the max-min normalization formula of defining a proportionally scale of measurement. To explain the result of the formula, we will execute the formula with the findings of real numbers, and show the new score of mobility.

6.2.1 Max-min normalization

After running addresses on the algorithm containing an assumption of reaching low numbers, the output became a negative value. This is because Γ_1 and Γ_3 were zero, and Γ_2 was a significant low negative.

Min **of** Γ_1

$$\lambda_1 = 5.246196, \varphi_1 = 60.243800$$

$$F = \forall(dis(\lambda_1, \varphi_1, [\lambda_2, \varphi_2])) = 0$$

$$\left(\sum_{n=1}^{n_m} F\right) * 5 = 0$$

$$\log_e(270) = False$$

$$\Gamma_1 = 0$$

Min **of** Γ_2

$$\Gamma_2 = \frac{\left(\frac{p_1t_1+p_2t_2+...+p_nt_n}{p_n} * -1\right)}{10}$$

$$p_1t_1 + p_2t_2 + ... + p_nt_n = 609$$

$$p_n = 7$$

$$\frac{\left(\frac{609}{7} * -1\right)}{10} = -8,7$$

$$\Gamma_2 = -8,7$$

Min **of** Γ_3

$$\Gamma_3 = \left(\sum_{n=1}^{n_m} G\right) + log_e\left(\sum_{n=1}^{n_m} H\right)$$

$$G = \left(\frac{p_1}{p_1} + \frac{p_2}{p_2} + ... + \frac{p_n}{p_n}\right),\ \ p_1 \neq p_2 \neq ... \neq p_n,\ p \neq 0,\ pt \leq 15_w$$

$$H = (p_1 + p_2 + ... + p_n),\ \ p_1 \neq p_2 \neq ... \neq p_n,\ \ pt \leq 15_w$$

$$H = (0+0+0+0+0+0+0),\ \ p_1 \neq p_2 \neq ... \neq p_n,\ \ pt \leq 15_w$$

$$H = 0$$

$$Since\ \ H = 0,\ \ then\ \ G = 0$$

$$\left(\sum_{n=1}^{n_m} G\right) + log_e\left(\sum_{n=1}^{n_m} H\right) = 0,\ \ since\ \ log_e(0) = False$$

$$\Gamma_3 = 0$$

By the calculations above, the $min(x) = -8,7$. The same procedure is done by finding the $max(x)$, which is 16.5. By the max-min normalization formula, we receive the scaled score by using the three values of $(\Gamma_1 + \Gamma_2 + \Gamma_3)$, $max(x)$

and $min(x)$.

To define the new score of mobility (with normalization), the same address is used as the first example, in 5.1.4 The mobility index calculation.

$$x' = \frac{x - min(x)}{max(x) - min(x)} * 10$$

Where x' is our normalization value.

$$x' = \frac{15,790361914031333 - (-8,7)}{16,5 - (-8,7)} * 10$$

$$x' = 9,7$$

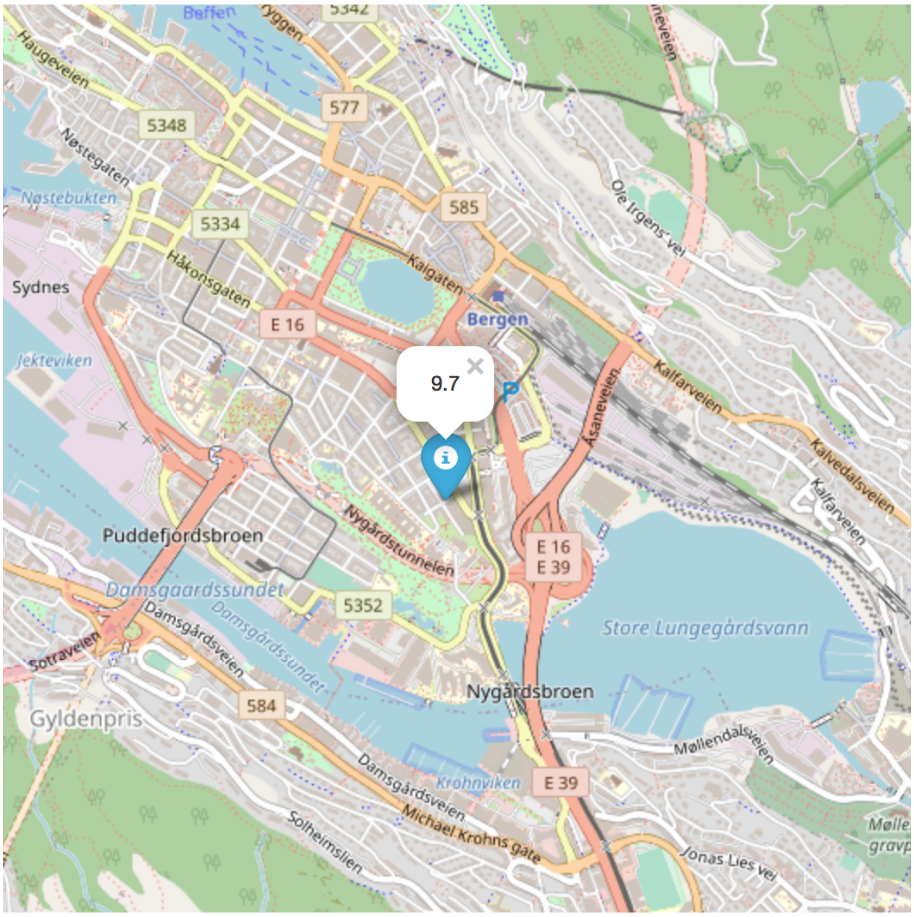

Figure 6.2: Result of the mobility index algorithm at "Borgermester Platous gate 9", with max-min normalization

As the result shows in the figure 6.2, "Borgermester Platous gate 9" as input gives a score of 9.7 out of 10.

The next figure (figure 6.3) shows the result of the mobility index algorithm looping through 18 arbitrary addresses in Bergen. The addresses are converted into coordinates, and plotted on the generated map.

Figure 6.3: Result of the mobility index algorithm running 18 arbitrary addresses

The different scores presented define the mobility, concerning all factors mentioned in chapter 4. As expected, the greatest variety of mobility access points and transportation services, and different facilities, is located in the city center. The values are only a correct score if the quality of accessing all different types of facilities is equally necessary for each individual. Therefore, the score can be seen as a preference of the potential destinations, and the magnitude, quality, and character of the activities found there. However, the score 1,9 on the map, makes the interest of a further investigation of the possibilities in the area. Especially compared to the higher scores plotted around this point.

Chapter 7

Feedback from MUST

In this chapter, we will present the feedback given from the MUST community on our contribution regarding presentations held in the MUST workshops. Throughout the research, close and constructive cooperation is conducted together with Bergen Municipality. As mentioned in chapter 1, MUST is an open arena for idea sharing, concerning innovation in the field of mobility.

7.1 Presentation

At the beginning of the design science research process, we contacted the responsible for running the MUST gatherings, such that our project could be presented to the local mobility environment. To reach high effectiveness, two presentations of the concept were held. One presentation at the early stage of the research, and one towards the end project.

7.1.1 First presentation

The presentation aimed to clarify if there was any interest in our project, among the very stakeholders related to the field. Also, with the opportunity of pitching the concept of the mobility index algorithm, useful feedback, and insight could be collected. As for the chosen methodology for this research, such a presentation fitted perfectly concerning the iterative process of knowledge gathering from the environment.

First, the concept of measuring mobility at any given location in Bergen was pitched. Along with the explanation of our idea, three questions towards the MUST participants were asked:

1. What is the necessity of a measured mobility index?
2. Do you know any similar systems/tools like this?
3. What data and factors are important to build a good measurement?

7.1.2 Second presentation

The second presentation was held at the near end of the research process, to give our interpretation and findings conducted since the last workshop. An abstract description of how the algorithm measures mobility was given, by presenting the underlying formulas of the application. Also, two examples of results of the mobility index algorithm were shown, at two different addresses, respectively. Finally, three questions were asked to the MUST contributors:

1. Do you see the results as expected and valid for the different locations?
2. Are there any missing data or factors to improve the measurement?
3. Are there any extension aspects?

7.2 Feedback

The feedback from the two workshops was different. The concept of the first workshop was to pitch an idea, then everybody in the room takes part in the discussion of it using design thinking principles. In this meeting, there were more than 30 people involved from Bergen municipality, Skyss that awards contracts for the operation of bus services, Bergen Light Rail, express boat services and ferries, plus many other companies involved in building mobility solution. The second workshop was to get feedback on the solution that we produced to some of the main stakeholders (5 people) within MUST.

7.2.1 Feedback on the first presentation

As far as they knew, there is by now no existing tool compared to our application. The mobility measurement can be used to detect potential areas for new infrastructure, regarding both residents and public transportation. Others believe that the mobility index could be used as an extension tool in housing advertising. The index can nevertheless be used as a part of reaching the zero growth objective, as it presents the large differences within the municipality. Bergen can therefore use it to plan better transportation infrastructures in more remote areas, as well as making the residents choose public transportation options over private vehicles.

To build a good measurement, most of the members wanted factors such as city bikes, bus, tram, kindergartens, as well as ferries. It was however important for them that most of the factors could be reached within walking distance.

7.2.2 Feedback on the second presentation

All of the members within the MUST community, that participated in the second presentation, believed that the results were reliable. All of them agreed that it would most likely be a higher mobility score in the city area, and a lower score in districts further away from the city center. Also, as the mobility index algorithm gives different measurements, depending on what time of day the execution runs, we have a door opener for new and interesting analysis. There were nevertheless raised a question regarding the walking distance for functionally disabled residents, which has not been taken into consideration in the index.

Chapter 8

Answers to Research Questions

A research question with three sub questions were presented in chapter 1. In the current chapter, the questions will be answered.

What is a good mobility index?

In order to answer the question, we need to know what ”good” qualifies to, and what is defined by ”mobility” in our case. From chapter 3, the base and concept of the equations behind the algorithm are measured with high validity and done by rigorous design. The generalisability which gives external validity, can be concluded with a qualification of ”good” mobility. In other words, a good mobility is the common agreement of high reach of accessibility, concerning the three aspects of Γ_1, Γ_2 and Γ_3. I.e. (on a score from 0-10) an aggregation of:

1. (Γ_1): the number of public transport and carbike services within 1 km
2. (Γ_2): the average time to reach POIs with public transport
3. (Γ_3): the number of amenities that can be reached within 15 minutes walking

However, the output of the algorithm, the mobility index, can be understood as good when the value makes sense for human reading. Therefore, we introduced the max-min normalization in chapter 4.

The mobility index is based on dynamic data gathering, the Entur Journey Planner API and mathematical provable formulas. The level of mobility is then measured and normalized, such that the output gives **a good mobility index.** How good the index is, is decided by the factors behind accessibility and the definition of mobility.

Who is the most important stakeholders, and what do they want?

The question points out which use-cases are convenient to our result of research. As explained in chapter 3, the contribution of new knowledge to the world is done by iterations together with the municipality. The variety of important stakeholders is real considering this project is the first to combine academic resources with the given possibilities of the involvement with Bergen municipality regarding MUST. A measurement of mobility is an important contribution to the vision of MUST itself, and can therefore find its stakeholders within this particular forum. The mobility index algorithm is acknowledged by participants in the local transportation innovation environment, and reached an interest by the municipality. Another important group of stakeholders can be found in the real estate business. By presenting the current mobility level of a particular sale object, a more precise and grounded information is added.

What does the research solve, and why is it important?

The mobility index algorithm behind the developed application is the outcome of this research. As a tool, the algorithm can clarify certain locations which do not hold the expected level of mobility opportunities regarding public transportation. Another important feature of the algorithm is the possibility to modify the data. Since it is dynamically gathered, the data can be manipulated such that the index received gives a response to a fictional mobility picture (i.e. scenario testing). With either including new data or removing existing data, the index will give its representation. In a city planner perspective, the possibility of executing prognosis of the effect on planned infrastructure regarding public transportation or city cycles is important.

How do available transportation and accessibility data give a calculated measurement of the current state of mobility at a given location?

To answer the main research question asked in chapter 1, we have to look at the phase of methodology, development, calculations and the effect of the results. Throughout this thesis, our aim were to define a tool such that the city of Bergen could step further into the future of mobility. The developed system, which defines the algorithm containing the measurement formulas, gives now a mathematical based level of mobility from a human perspective. By gathered data, defining the opportunities of accessibility, and with complex distance calculations combined with relevant APIs, the result satisfies the goal. Through the design science research, the interest of the different stakeholders was clear, and the tool is considered necessary for both city planners and innovative decision makers. The possibilities of clarifying areas of opportunities is highly relevant for both reaching local and national mobility goals. The mobility index algorithm is able to detect the mobility status at chosen time intervals, such that a clearer situation picture is presented about the needs in the mobility improvements. Also, the level of mobility can be calculated with fictional data sets, such that strategic city planners will make their decisions based on provable mathematical formulas.

Chapter 9

Related work

The current chapter represents related academic work, focused on measurement formulas and the impact of the findings. The final section of the chapter looks closer into the comparison and interpretation of our research and findings.

9.1 Spatial Network Analysis for Multimodal Urban Transport Systems

According to Curtis et. al., the developed accessibility instrument will fill a significant gap in planning for accessibility by providing a planning support tool that can be used to inform strategic land use and transport planning [4]. The GIS-based tool, SNAMUTS, estimates the relationship between public transport network configuration, performance and service standards. Also, it evaluates the geographical distribution or clustering of land use activities in an urban area. By strategic planning documents, SNAMUTS combines the use of transportation, location and the public transport route into nodes and segments. The tool makes then some particular definitions:

- Minimum service standard
- Activity nodes
- Travel impediment

Minimum service standard
Defines the transportation service frequency by time, weekdays and weekends respectively. The purpose is to set a standard of design for future accessibility by public transport.

Activity nodes
Points to the major transport activity centers, which appear in the structural city transport planning documents. This defines the main routes with high-frequency tram or bus lines. The activity nodes are obviously close to relevant infrastructure, such that resident and job are attached.

Travel impediment
Every route segment contains an impediment value which consists of the average travel time divided by the number of services per hour, separately for each direction. This number is then multiplied by a factor of 8, to arrive at more readable numbers. The travel impediment (proxy distance) between any two activity nodes on the network is thus made up of the sum of the impediment values on each route segment passed along the path. Another indicator adds to this by considering the transfer penalty on public transport.

There are also operational aspects such as closeness, centrality degree, nodal connectivity and network coverage, which all give a network structure across a metropolitan area. The authors focus on these three problems:

- Which activity centers could best be intensified?
- Which centers should perform a regional role and which ones a local role?
- Where should public transport investment (infrastructure, service improvement) go?

Result
The development is now an important and well suited tool used to inform decisions regarding both public transport network configuration and land use intensification concerning accessibility. Also, the tool provides information on detail oriented metropolitan strategic planning, with identifying gaps on public transportation. This leads to important knowledge regarding prioritising future expansion of public transport infrastructure in the area.

During the research the authors experienced the development or use of traditional implementation on transport models, to be little concerning analytic planning techniques and decision making. Therefore, their aim was to construct a system which could be used without any previous knowledge or expertise of complex mathematical models and how to manage them. The figure 7.1 shows how accessibility changes with the implementation on a new rail corridor and a bus network reconfigured to act as a feeder service.

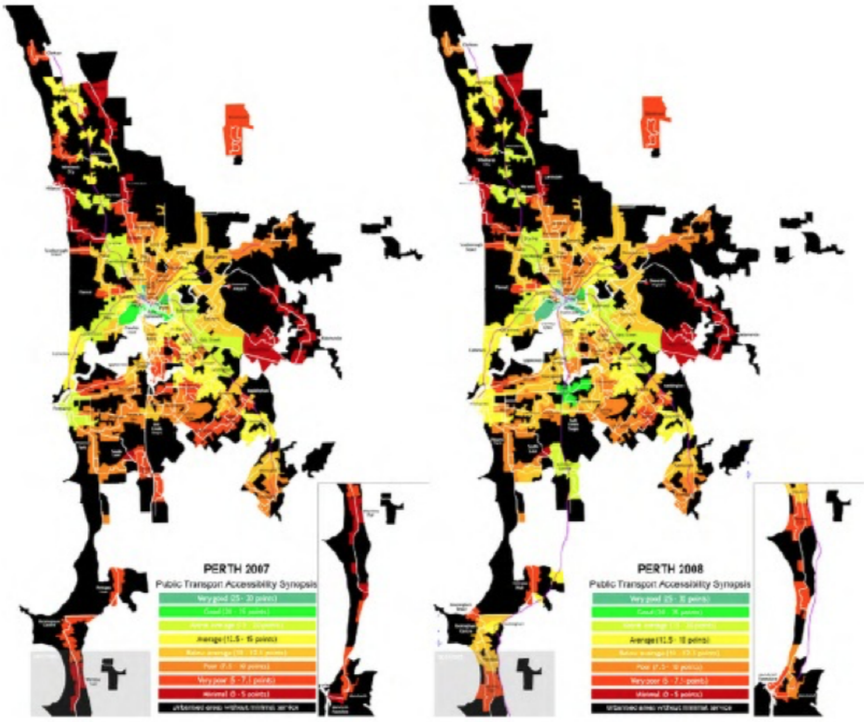

Figure 9.1: SNAMUTS result of Perth, Australia
S. Curtis et. al. [4]

9.2 Measure of urban accessibility provided by transport services in Turin: a traveller perspective through a mobility survey

The highly relevant paper, written by Riccardo Ceccatoa, Francesco Deflorioa, Marco Diana and Miriam Pirraa, investigates and reflects around theory of mobility and calculates a measurement of accessibility, based on transport data. The calculations of both active and passive accessibility are based on travel speed, for both car and public transport (PT) as travel modes in Turin (Italy).[6]

Accessibility measurement

The researchers' method of calculating the accessibility measurement is based on the ratio between the minimum distance between the two centroids from the road network $(dist_{i,j})$ and the minimum travel time between the two zones $(time_{i,j})$:

$$Acc_{j,i} = \frac{dist_{i,j}}{time_{i,j}},\ i, j = 1, ..., N;\ i \neq j$$

$Acc_{j,i}$ represents the average time per distance between locations, which leads to the second equation:

$$Acc_{act,s} = \frac{\sum_{i=1}^{N} Acc_{s,j}}{N},\ s \neq j$$

$Acc_{act,s}$ calculates the average accessibility based on the number of zones (N). The numerator in the formula is the sum of the measured accessibility at any given zone, which gives an overall measure of the transport services in Turin.

Results

As the researchers calculated the measure on both car and public transport, some interesting comparisons were made.

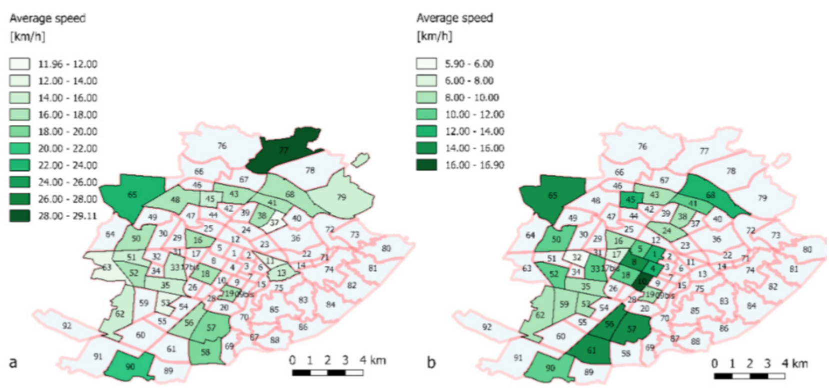

Figure 9.2: (a) accessibility on car and (b) accessibility on public transportation, Turin, Italy

R. Ceccatoa et. al. [6]

The figure 7.2 represents (a) accessibility on car and (b) accessibility on public transportation. Regarding public transportation, the tram in Turin plays an important role, considering the quotients and the dividend in the equations. A comparing aspect to (a), is the road-network, which for zone 77 contains high free flow speed infrastructure, with high speed road. This allows reaching the city with high speed and short travel time.

9.3 Effective accessibility: Using effective speed to measure accessibility by cost

According to Vale, his paper describes accessibility measurement by cost [29]. The measurement is based on speed, which is distance divided by time. Through the case study described in the paper, done in Lisbon, the author included the social aspect of income in the calculations of measurement.

Accessibility measurement
The following formula is defined for effective speed. It does contain a social concept, as the total time needed to create the conditions to be able to travel at a certain speed (for each travel mode) depends on income, which is a clear indication of social group. Effective speed is given by the expression:

$$ESpeed_{m,inc} = \frac{atd}{\frac{ttc}{h_{wage}} + h_{other} + h_v}$$

where $ESpeed$ is the effective speed for a mode m, considering an individual with a total annual income inc. The annual total distance travelled atd is the numerator, ttc is the total transport cost associated with atd, divided by h_{wage}, which is the hourly wage. The variable h_{other} is the amount of hours devoted to maintaining the vehicle, and h_v is the total time spent travelling in the vehicle. The reason for the social concept added to the formula is simply that individuals with high wages need to work fewer hours to cover the expense of owning and using different travel modes.

The next formula measures effective accessibility, using a cumulative opportunity measure:

$$A_i = \sum_{j=1}^{n} O_j f(C_{ij})$$

$$\begin{cases} f(C_{ij}) = 1 & if \quad C_{ij} \leq \delta \\ f(C_{ij}) = 0 & if \quad C_{ij} > \delta \end{cases}$$

where A_i defines the accessibility of place i. O_j are opportunities found at place j and C_{ij} is the cost of travelling between i and j. δ represents the threshold considered in the accessibility measure. To represent opportunities, Vale used job locations, effective travel time was set to 30 minutes [29].

Result
With various travel modes and job locations as reachable opportunities, GIS-representations were made. The time-based accessibility is done within in the

interval of 08:00 to 09:15, and three calculations with average income, low income and high income. The figure 7.3 shows the outcome of accessibility measurement on public transport.

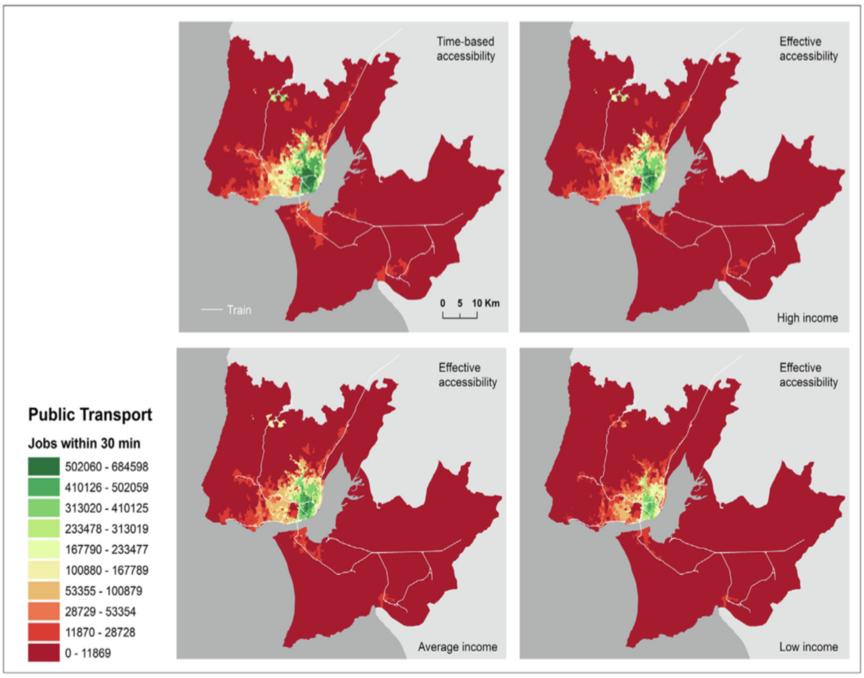

Figure 9.3: Result: Public Transport, Jobs within 30 min, Lisbon, Portugal
D. Vale [29]

As well as public transport, the measurement was also done with private vehicles, cycle and walking, separately. The three figures 7.4 - 7.6 show private vehicles (automobile), cycling and walking.

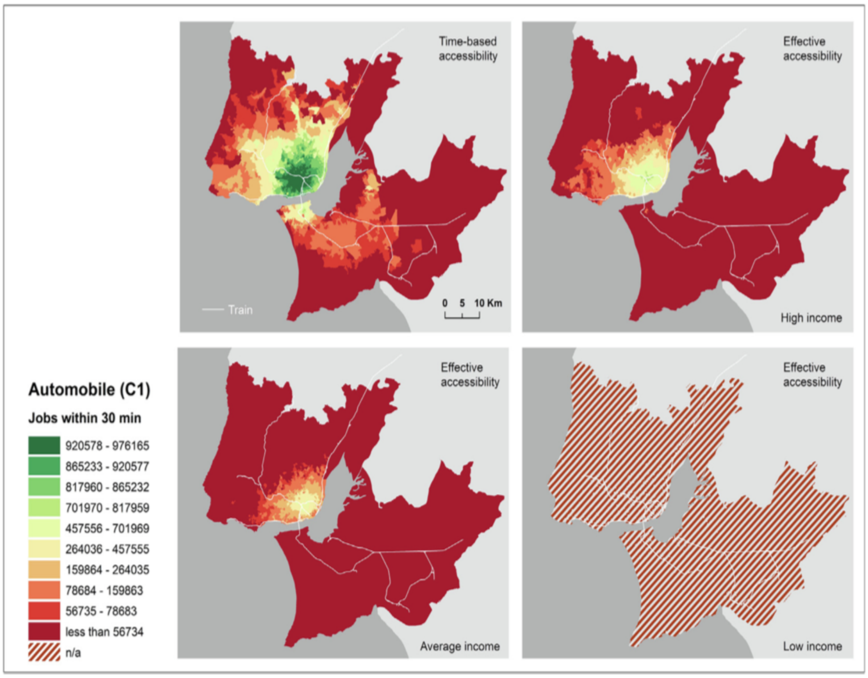

Figure 9.4: Result: Automobile (C1), Jobs within 30 min, Lisbon, Portugal
D. Vale [29]

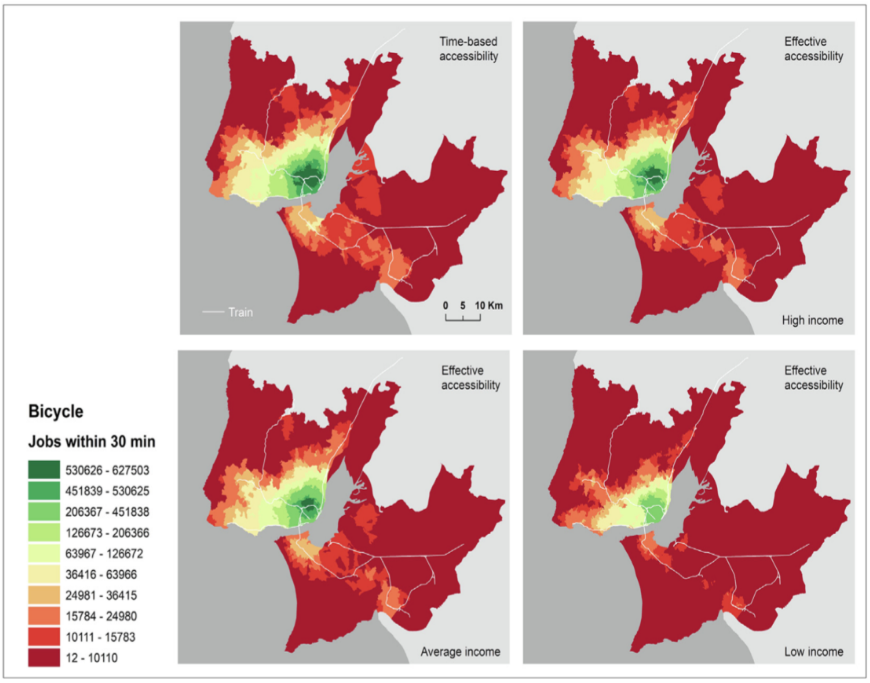

Figure 9.5: Result: Bicycle, Jobs within 30 min, Lisbon, Portugal
D. Vale [29]

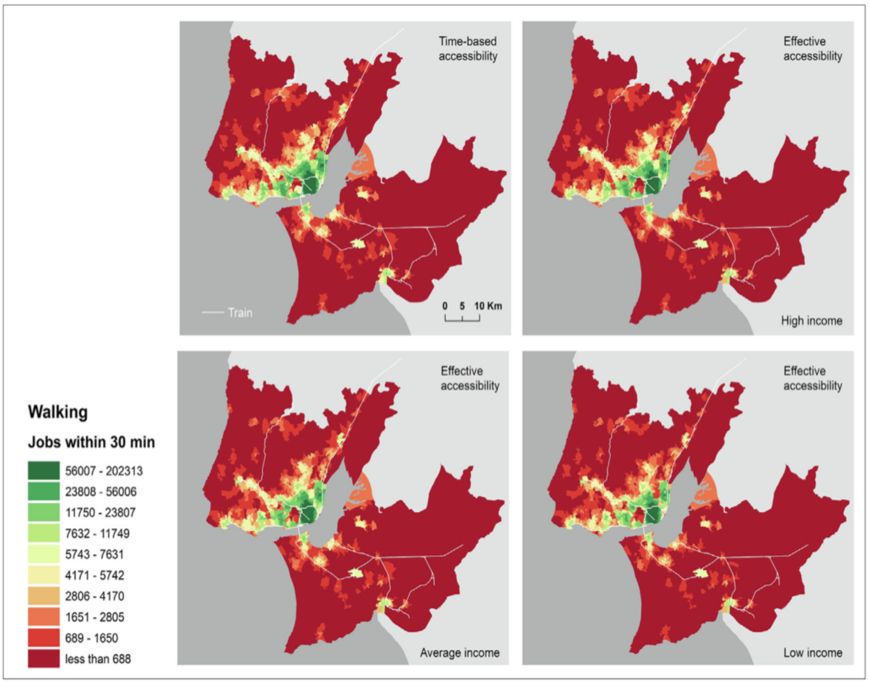

Figure 9.6: Result: Walking, Jobs within 30 min, Lisbon, Portugal
D. Vale [29]

Vale [29] concludes with an existence of high dependence of income rate. Public transport is shown to be slow, compared to the other, which also depends on the calculation of speed. An interesting finding, is the evidence of poor PT accessibility and how important cycling accessibility can be for the lower income areas. Which also motivates to invest in PT in certain areas. However, the effect of cycling gives a high interest of planning investment, considering the overall accessibility measurement.

9.4 Comparison and interpretation

When looking at the related work, the common goal of identifying areas with poor accessibility rate is achieved. As the research contributes with different formulas, accessibility is measured. However, some factors of the formulas are indeed commonly involved, as the values of time and distance. To consider the formulas constructed at the Theory Chapter, three important factors are included; time, distance and the amount of reachable points. We can see a common way of measurement is speed, with both the Turin-study and the Lisbon-study gives a measurement with $\frac{distance}{time}$. The formula Γ_2 can be related to such a measurement. With the factors:

$$p_1t_1 + p_2t_2 + ... + p_nt_n$$

We collect the travel time from the Entur Journey Planner API, and divides by the value of the amount: p_n. The formula then gives the average travel time. If the distance had been a dividend in our matter, the result of the sum of the total dividend given is just an equal relation. The difference between Γ_2:

$$\Gamma_2 = \frac{\left(\frac{p_1t_1+p_2t_2+...+p_nt_n}{p_n} * -1\right)}{10}$$

and the formula given in the Turin-study calculating average accessibility is only this particular dividend of distance, without mentioning the factor of weighting and negative multiplication.

$$Acc_{act,s} = \frac{\sum_{i=1}^{N} Acc_{s,j}}{N},\ s \neq j$$

In our research, we do not consider the distance as a dividend to be necessary in Γ_2. However, the distance is taken care of in Γ_1 and Γ_3, not as a dividend, but as a factor to calculate time.

The travel impediment formula in the SNAMUTS project consists of the average travel time divided by the number of services per hour [4]. The formula can be seen as an abbreviation of the combination of formulas Γ_1 and Γ_2. Instead of dividing the number of services per hour, the different services available at the given time are added, in our case $(\Gamma_1 + \Gamma_2)$. This gives a different value, but the same purpose. Also, the researchers took their liberty of multiplying the formula of travel impediment with a factor of 8, in our case the factor 5 is multiplied in Γ_3, and Γ_2 contains 10 as a denominator. The constants provide weighting for the different formulas.

Our research differs from the research presented in this chapter, in form of combining three formulas. The Lisbon-study [29] and the Turin-study [6] describes one main formula of measurement, and one sub formula included in the main formula. In our case, summarizing the three formulas, Γ_1, Γ_2 and Γ_3, provides a higher inclusion of varieties of factors. Also, the formulas are independent, which gives the opportunity of variable expansion. The expansion is important in order to further improve the accuracy as well as making specific measurement based on selective weighted factors. Our research which resulted in the calculation of mobility measurement, can be seen as an extension and an improvement of the previous research on accessibility measurement.

An important set of included factors makes the mobility measurement stand out from related measurement formulas. The unique feature of the mobility index algorithm compared to the other related findings is the amount of different factors, and the opportunity of weighting, removing and adding.

Chapter 10

Conclusion and Future work

10.1 Conclusion

In this thesis, we have provided an application given an address in Bergen, that gives a mobility score between 0 and 10, where 0 showing low mobility and 10 excellent mobility. We have provided the mathematical foundation that the mobility measurement algorithm is based on, and we have implemented the solution in Python. We have applied the algorithm on several addresses, and the result looks promising. Our results have created a great deal of interest within MUST, and the participation in the mobility community has contributed with useful feedback and insight.

We used the research methodology design science in this research that required iteratively processes with both the environment and the knowledge base. The return to the environment was our mobility index application, and to the research community, the mathematical foundation behind the mobility index that is communicated via this thesis.

10.2 Future work

During the conducted research, the research problem has proved to be an enormous field of opportunities, and therefore a continuous space of knowledge contributions and extensions. As the mathematical ground is settled for the formulas, the number of different factors to be added is only limited to a man's imagination.

10.2.1 Can we define a new formula to extend the index?

Bearing in mind the formulas, Γ_1, Γ_2 and Γ_3, an interesting extension of the mobility index algorithm could be to define a new formula: Γ_4. Namely, a formula that involves private vehicles. To either combine or compare the output of the formula to the previously presented formulas, it has to contain the same basis as Γ_2.

$$\Gamma_4 = \frac{\left(\frac{p_1 tv_1 + p_2 tv_2 + \ldots + p_n tv_n}{p_n} * -1\right)}{10}$$

All four formulas give different solutions of combination. Comparing Γ_2 and Γ_4 gives the opportunity of finding the difference between public transport and private vehicle at a given position. When it comes to calculating the measurement of mobility, all four formulas can be combined.

10.2.2 Categorization

Is it the case that the measurement is true for all individuals? Of course not, the different amenities and POIs may be weighted differently for different categories of individuals. Creating different categories with groups of individuals with common attributes and life situations, a more fitted and personal level of mobility measurement is gained. By weighting POIs and amenities according to the interests of the different categories, such a mobility measurement is achievable.

10.2.3 Extend the lists of POI and amenities

The two lists are statically presented, where the points of interest are presented as a json object, and the amenities are dynamically gathered through the Overpass API. As the script for getting the data for every amenities contains the same syntax, the opportunity of extension is extremely high.

10.2.4 New analysis and GIS-representations

As an alternative future result of the research conducted, several GIS-representations can be done. With GIS-technologies suited for Python, comparisons and analysis can be presented. Also, new, interesting and relevant research problems can be defined such that the mobility index algorithm will contribute as a part of the solutions. An interesting analysis topic is the understanding of the mobility effect in chosen intervals of time. As the algorithm calculates the mobility measurement at the exact time of execution, a loop of calling the mobility index in different time intervals is possible. By running the algorithm a specific number of times in each time interval, and calculate every average in each interval, the quality of mobility is set in a relation of time (e.g. rush hour, holidays).

Bibliography

[1] T. Andresen, “Gis-basert gangtilgjengelighetsanalyse, bergen kommune (internal document) (2019).”

[2] B. avd Stavanger. Viser vei for en mer presis, enklere og samkjørt kollektivtrafikk (accessed: 15.04.2020). [Online]. Available: https://www.bouvet.no/prosjekter/kolumbus-viser-vei

[3] G. V. Brummelen, *Heavenly Mathematics: The Forgotten Art of Spherical Trigonometry*. [Online]. Available: https://books.google.no/books?id=0BCCz8Sx5wkC&pg=PR7&redir_esc=y#v=onepage&q=haversine&f=false

[4] J. S. Carey Curtis and R. Mellor, “Spatial network analysis for multimodal urban transport systems.” [Online]. Available: http://www.accessibilityplanning.eu/wp-content/uploads/2013/01/1-SNAMUTS-R1.pdf

[5] D. Carriere. Geocoder: Simple, consistent (accessed: 14.04.2020). [Online]. Available: https://geocoder.readthedocs.io/index.html

[6] R. Ceccatoa, F. Deflorioa, M. Diana, and M. Pirraa, “Measure of urban accessibility provided bytransport services in turin: a traveller per-spective through a mobility survey.” [Online]. Available: https://doi.org/10.1016/j.trpro.2020.03.020

[7] S. Coast. Python data, leaflet.js maps. [Online]. Available: https://python-visualization.github.io/folium/

[8] G. D. . Contributers, “Maps sdk for android utility library.” [Online]. Available: https://github.com/googlemaps/android-maps-utils

[9] S. Gregor and A. R. Hevner, “Positioning and presenting design science research for maximum impact.” [Online]. Available: https://ailab-ua.github.io/courses/MIS611D/gregor-2013-positioning-presenting-design-science-research.pdf

[10] ——, *Positioning Design Science Research for Maximum Impact*. [Online]. Available: https://ailab-ua.github.io/courses/MIS611D/gregor-2013-positioning-presenting-design-science-research.pdf

[11] W. G. Hansen, "How accessibility shapes land use, (1959, pp. 73-76)." [Online]. Available: https://doi.org/10.1080/01944365908978307

[12] ——, "Measuring accessibility: An exploration of issues and alternatives, (1997, p. 1195)." [Online]. Available: https://doi.org/10.1068/a291175

[13] A. Hevner, S. March, J. Park, and S. Ram, "Design science research in information systems." [Online]. Available: http://content.schweitzer-online.de/static/catalog_manager/live/media_files/representation/zd_std_orig__zd_schw_orig/014/341/402/9781441956521_content_pdf_1.pdf

[14] H. Hoos and T. Stützle, "Travelling salesman problems, (2005, pp. 357-416)." [Online]. Available: https://doi.org/10.1016/B978-155860872-6/50025-1

[15] B. Kommune. Mobilitetslab for smarte transportløsninger (accessed: 06.04.2020). [Online]. Available: https://www.bergen.kommune.no/hvaskjer/tema/must

[16] J. Lawhead, *Learning Geospatial Analysis with Python*. [Online]. Available: https://ebookcentral-proquest-com.pva.uib.no/lib/bergen-ebooks/reader.action?docID=4520784#

[17] J. Lu and M. Zhang, "Encyclopedia of systems biology, (2013, p. 52)." [Online]. Available: https://doi.org/10.1007/978-1-4419-9863-7_875

[18] J. B. P. Manager and A. Storage. Under the hood: Performance, scale, security for cloud analytics with adls gen2. [Online]. Available: https://azure.microsoft.com/en-us/blog/under-the-hood-performance-scale-security-for-cloud-analytics-with-adls-gen2/

[19] P. L. Mark Saunders and A. Thornhill, *Research Methods for Business Students, Pearson Education, Fifth edition (2009, pp. 167-158)*.

[20] H. J. Miller, "Place-based versus people-based accessibility, (2005, pp. 63-89)." [Online]. Available: https://doi.org/10.1108/9780080460550-004

[21] OpenCage. Opencage geocoder (accessed: 14.04.2020). [Online]. Available: https://opencagedata.com/solutions/data-enrichment

[22] OpenStreetMap. Overpass api (accessed: 12.05.2020). [Online]. Available: https://wiki.openstreetmap.org/wiki/Overpass_API

[23] Pallets. Flask web application frameworkp (accessed: 14.04.2020). [Online]. Available: https://flask.palletsprojects.com/en/1.1.x/installation/#python-version

[24] redaksjonen Bouvet. Et datasjøprosjekt til inspirasjon (accessed: 15.04.2020). [Online]. Available: https://www.bouvet.no/bouvet-deler/et-ordentlig-datasjoprosjekt

[25] K. Solvik-Olsen. National transport plan 2018–2029, oslo (2017, pp. 13-17). [Online]. Available: https://www.regjeringen.no/contentassets/7c52fd2938ca42209e4286fe86bb28bd/en-gb/pdfs/stm201620170033000engpdfs.pdf

[26] team Folium. Open street map (accessed: 14.04.2020). [Online]. Available: https://www.openstreetmap.org/about

[27] UN, "World urbanization prospects (2014, pp. 2-4)." [Online]. Available: https://dx.doi.org/10.18356/527e5125-en

[28] V. Vaishnavi, B. Keuchler, and S. Petter, "Design science research in information systems." [Online]. Available: http://desrist.org/design-research-in-information-systems/

[29] D. Vale, "Effective accessibility: Using effective speedto measure accessibility by cost, lisbon (2020)." [Online]. Available: https://doi.org/10.1016/j.trd.2020.102263

[30] M. Walicki, *Introduction to Mathematical Logic, Word Scientific, Extended edition (2017, pp. 119-124)*.

[31] F. Wefering, S. Rupprecht, S. Bührmann, and S. Böhler-Baedeker, "Guidelines developing and implementing a sustainable urban mobility plan (2013, pp. 6-7)." [Online]. Available: https://www.eltis.org/sites/default/files/guidelines-developing-and-implementing-a-sump_final_web_jan2014b.pdf

[32] D. R. Williams. Earth fact sheet (accessed: 17.04.2020). [Online]. Available: https://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html

[33] C. Wohlin, M. Höst, B. Regnell, P. Runeson, M. Ohlsson, and A. Wesslén, *Experimentation in Software Engineering, Springer, (2012, pp. 105-110)*.