

Student thesis series INES nr 428

# Using Geographically Weighted Regression (GWR) to explore spatial variations in the relationship between public transport accessibility and car use

A case study in Lund and Malmö, Sweden

**Johanna Andersson**

---

2017  
Department of  
Physical Geography and Ecosystem Science  
Lund University  
Sölvegatan 12  
S-223 62 Lund  
Sweden



Johanna Andersson (2017).

***Using Geographically Weighted Regression (GWR) to explore spatial variations in the relationship between public transport accessibility and car use – A case study in Lund and Malmö, Sweden***

***Tillämpning av geografiskt viktad regression (GWR) för att undersöka rumsliga variationer i förhållandet mellan tillgänglighet till kollektivtrafik och bilanvändning - En fallstudie i Lund och Malmö***

Master degree thesis, 30 credits in *Geomatics*

Department of Physical Geography and Ecosystem Science, Lund University

Level: Master of Science (MSc)

Course duration: *January 2017 until June 2017*

Disclaimer

This document describes work undertaken as part of a program of study at the University of Lund. All views and opinions expressed herein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

# Using Geographically Weighted Regression (GWR) to explore spatial variations in the relationship between public transport accessibility and car use

A case study in Lund and Malmö, Sweden

---

Johanna Andersson

Master thesis, 30 credits, in *Geomatics*

**Supervisor:** Finn Hedefalk

Centre for Economic Demography

Lund University, Sweden

**Exam committee:** Micael Runnström,

Department of Physical Geography and Ecosystem Science

Lund University, Sweden



## Abstract

In Sweden the number of cars per person has increased since the mid-20<sup>th</sup> century. With negative impacts on both health and the environment, private ownership of vehicles represents one of the major challenges in urban transport. To move travelers from privately owned vehicles to public transport has shown to be beneficial in reducing carbon emissions. However, in order to create policies that attract people towards public transport, data of factors influencing transit choice is crucial due to its validation of planning and investments. Previous studies have shown that physical proximity to public transport stations is one of the critical factors when considering transport choice. Consequently, the aim of this thesis is to analyze novel GPS data to investigate the relationship between public transport accessibility and car use in Lund and Malmö, Sweden. By modelling this relationship with the spatial regression model of Geographically Weighted Regression (GWR), regional variations are allowed and investigated. The results in Lund imply a negative association between public transport accessibility and car use, thus suggesting that car use decreases with a higher public transport accessibility. Furthermore, results in Lund indicate that the spatial regression model of GWR is a better fit to the data than the non-spatial regression model of Ordinary Least Squares (OLS). In Malmö, on the other hand, results imply that public transport accessibility does not have significant impact on car use, and suggests that the GWR model is not a better fit to the data than the OLS model. Consequently, the results in Lund and Malmö do not coincide. Nevertheless, in Lund, where model performance is the highest, results imply that car use decreases with a higher public transport accessibility. This study is one of the first to use individual GPS data together with spatial analysis to investigate the relationship between public transport accessibility and car use. Consequently, this study contributes to the literature on the effects of public transport accessibility on car use and on the use of local spatial analyses in accessibility studies. Such knowledge can be utilized in transport planning to reduce car usage.

**Keywords:** Physical Geography and Ecosystem analysis, Geographically Weighted Regression (GWR), Public transport accessibility, travel survey, GPS data, Car usage

## Sammanfattning

Sedan 1950-talet har antalet bilar per person i Sverige ökat. Med en negativ påverkan på miljön såväl som på människors hälsa, representerar privat ägande av fordon en av de största utmaningarna i transportplanering. Att flytta resenärer från privata fordon till kollektivtrafik har visat sig fördelaktigt beträffande att minska koldioxidutsläppen, dock kräver denna typ av mobilisering politisk styrning. För att skapa effektiv policy, som lockar resenärer till kollektivtrafik, krävs information om faktorer som påverkar färdmedel, vilket sedan kan användas till att validera detaljerad planering och finansiella investeringar. Tidigare studier har visat att tillgänglighet till kollektivtrafik är en av de faktorer som har störst påverkan på valet av färdmedel. Därför är syftet med denna studie att analysera individuell GPS data för att undersöka förhållandet mellan tillgänglighet till kollektivtrafik och bilanvändning i Lund och Malmö. Genom att modellera detta förhållande med regressionsmodellen geografisk viktad regression (GWR) möjliggörs och analyseras regionala variationer i relationen. Resultaten i Lund visar att tillgänglighet till kollektivtrafik har en negativ association med bilanvändning, vilket tyder på att bilanvändning minskar med en högre tillgänglighet till kollektivtrafik. I Lund visar även resultaten att den rumsliga modellen GWR är bättre på att modellera data än OLS, som representerar en icke rumslig linjär regressionsmodell. I Malmö har tillgänglighet till kollektivtrafik inte en signifikant relation med bilanvändning, och den rumsliga modellen GWR är inte bättre på att modellera data än den icke rumsliga modellen OLS. Följaktligen så skiljer sig resultaten mellan Lund och Malmö. Men, i Lund där modellen presterar som bäst, så tyder resultatet på att bilanvändning minskar med högre tillgänglighet till kollektivtrafik. Denna studie är en av de första som använder individuell GPS data tillsammans med rumslig analys för att analysera förhållandet mellan tillgänglighet i kollektivtrafik och bilanvändning. Följaktligen bidrar denna studie till litteraturen angående hur bilanvändning påverkas av tillgänglighet till kollektivtrafik och angående användningen av lokala rumsliga modeller i tillgänglighetsanalyser. Denna kunskap kan användas vid transportplanering för att minska bilanvändning.

Nyckelord: *Physical Geography and Ecosystem analysis, Geographically Weighted Regression (GWR), Public transport accessibility, travel survey, GPS data, Car usage*

## **Acknowledgements**

I would like to give a special THANKS to Finn Hedefalk and Emeli Adell. Thank you Finn Hedefalk, for your expertise, commitment and valuable feedback during this entire process. And thank you Emeli Adell, for making this thesis possible and encouraging this work from beginning to end. Furthermore, I would like to thank Leif Linse, Anna Clark and all others at Trivector Traffic that have been involved in the process of developing TRavelVU, thank you for allowing me to analyze new and unique travel survey data. Finally, I would like to thank friends and family, for continuous support and guidance.

## *Table of Contents*

|     |  |    |
|-----|--|----|
| 1   | Introduction.....  | 1  |
| 1.1 | Aim.....   | 3  |
| 1.2 | Hypothesis.....  | 3  |
| 1.3 | Study area.....  | 3  |
| 2   | Background.....  | 5  |
| 2.1 | The incentives behind car use.....                         | 5  |
| 2.2 | GPS in travel surveys.....                                 | 7  |
| 2.3 | Spatial regression.....                                    | 8  |
| 2.4 | GWR in accessibility analysis in the transport sector..... | 13 |
| 3   | Data.....  | 14 |
| 3.1 | Travel survey data.....                                    | 14 |
| 3.2 | Data sets.....   | 16 |
| 4   | Methodology.....   | 18 |
| 4.1 | Data structure and visualization.....                      | 18 |
| 4.2 | Definition and calculation of car use.....                 | 19 |
| 4.3 | Explanatory variables.....                                 | 20 |
| 4.4 | Statistical analysis.....                                  | 23 |
| 5   | Results.....   | 24 |
| 5.1 | Descriptive statistics.....                                | 24 |
| 5.2 | OLS results in Malmö.....                                  | 30 |
| 5.3 | GWR results in Lund.....                                   | 32 |
| 5.4 | GWR results in Malmö.....                                  | 34 |
| 6   | Discussion.....  | 35 |
| 7   | Conclusions.....   | 40 |
| 8   | References.....  | 41 |



# 1 Introduction

## 1.1 Climate change and impacts

Since the industrial revolution, emissions of anthropogenic greenhouse gases (GHG) has continued to increase, in which the 10-year period between 2000 and 2010 is responsible for the highest emissions in history. Additionally, the Intergovernmental Panel on Climate Change (IPCC) reports that in 2014, greenhouse gas emissions were higher than ever (IPCC, 2014). The high levels of carbon dioxide, methane and nitrous oxide in the atmosphere has led to an uptake of energy by the climate system, a consequence that is a major environmental threat with extensive impact on both human and natural systems. Considering the increase in emissions between 2000 and 2010, approximately 11% originated from the transport sector, which represent a sector where the energy demand is highly dependent on technical and urban solutions as well as the behavior of people (IPCC, 2014). The observed warming of the atmosphere and oceans since the mid-20<sup>th</sup> century is a consequence of the anthropogenic greenhouse gas emissions. The 30-year period between 1983 and 2012 is likely to be the warmest period in the Northern Hemisphere over the last 1400 years. Other observed consequences are the decreasing extent of the Arctic sea ice and the Northern Hemisphere snow cover as well as the mass loss of the Antarctic and Greenland ice sheets, contributing to a global mean sea level rise (IPCC, 2014). According to the World Health Organization (WHO) 150 000 deaths per year can associated with climate change and greenhouse gas emissions. Future risks include flooding from coastal storms, where the number of people at risk is estimated to 200 million by the year 2080 (Patz et al., 2005). With high confidence, IPCC predicts future risks of food and water insecurity, reduced income and the loss of ecosystems and biodiversity as a consequence of climate change (IPCC, 2014).

## 1.2 GHG emissions from the transport sector

In European cities, approximately 40% of the greenhouse gas emissions come from motorized vehicles, where in 2014, approximately 66% originated from passenger cars (Alam et al., 2017, Rojas-Rueda et al., 2012). Hence, private ownership of vehicles is one of the major challenges in urban transport with negative impacts on both health and the environment (Trafikanalys, 2015, Rojas-Rueda et al., 2012). In Sweden, the number of cars per person has increased since the mid-20<sup>th</sup> century. Additionally the total mileage for all passenger cars has been increasing from 1999 to 2008, where the total mileage reached a steady state. However, an increase was prominent again in 2014 and if no actions are taken to reduce the car use, the Swedish Transport Administration (*Trafikverket*) predicts that the distance travelled by car will

increase by 25% between 2010-2030 (Trafikverket, 2015, Trafikanalys, 2015). To reduce the environmental pollution, the United Nations Environmental Program (UNEP) has suggested changes regarding public policies that encourages public transport in cities (Rojas-Rueda et al., 2012). Encouraging travelers to use public transport as opposed to privately owned vehicles is beneficial in reducing carbon emissions; however, such urban restructuring demands political governance and planning. To create policies that attract people towards using public transport, data regarding factors that influence the transit choice is crucial due to its validation of detailed planning and financial investments (Pye and Daly, 2015, Chakrabarti, 2017, Steg, 2005).

### 1.3 Car use and public transport accessibility

Previous studies have shown that physical proximity to public transport stations is one of the critical factors when considering transit choice. Furthermore, a study made by Boarnet et al. (2013) reported reduction in car usage for people living close to public transport stations with frequent and reliable public transport (Chakrabarti., 2017, Boarnet et al., 2013). Boarnet et al. (2013) used Global Positioning System (GPS) receivers to collect information about travel patterns. The GPS receiver represents an interesting new data collection method, providing detailed, precise and unbiased data compared to traditional travel surveys based on questionnaires (Rasmussen et al., 2015). However, Boarnet et al. (2013) did not examine regional variations in the relationship between car use and the physical proximity to public transport stations, which if ignored could cause biased parameter estimates and misleading results (Li et al, 2017).

In Sweden, the data collection methods for travel surveys has been limited and to the authors knowledge, there have been no studies using individual GPS travel data to analyze the relationship between car use and public transport accessibility (Allström et al., 2015). This study uses new and detailed GPS data to investigate the relationship between public transport accessibility, in terms of walking distance to public transport stations, and car use in Lund and Malmö. Additionally, it uses Geographically Weighted Regression (GWR) to explore regional variations in this relationship. Consequently, this study contributes to the literature on the effects of public transport accessibility on car use and on the use of local spatial analyses in accessibility studies. Such knowledge can be utilized in transport planning to reduce car usage.

## 1.4 Aim

The aim of this study is to analyze GPS travel survey data from people living in Lund and Malmö, and thereby investigate the relationship between their level of car use and their accessibility to public transport. Additionally, further spatial and socio-economic variables with potential impact on car use will be included in the analysis to obtain deeper knowledge of the incentives of car use. The aim consists of the following four objectives:

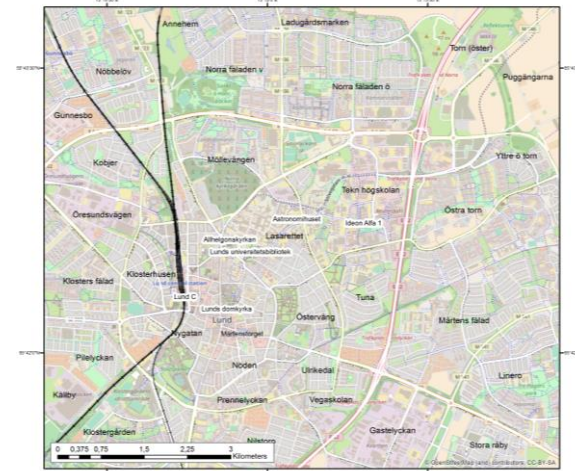
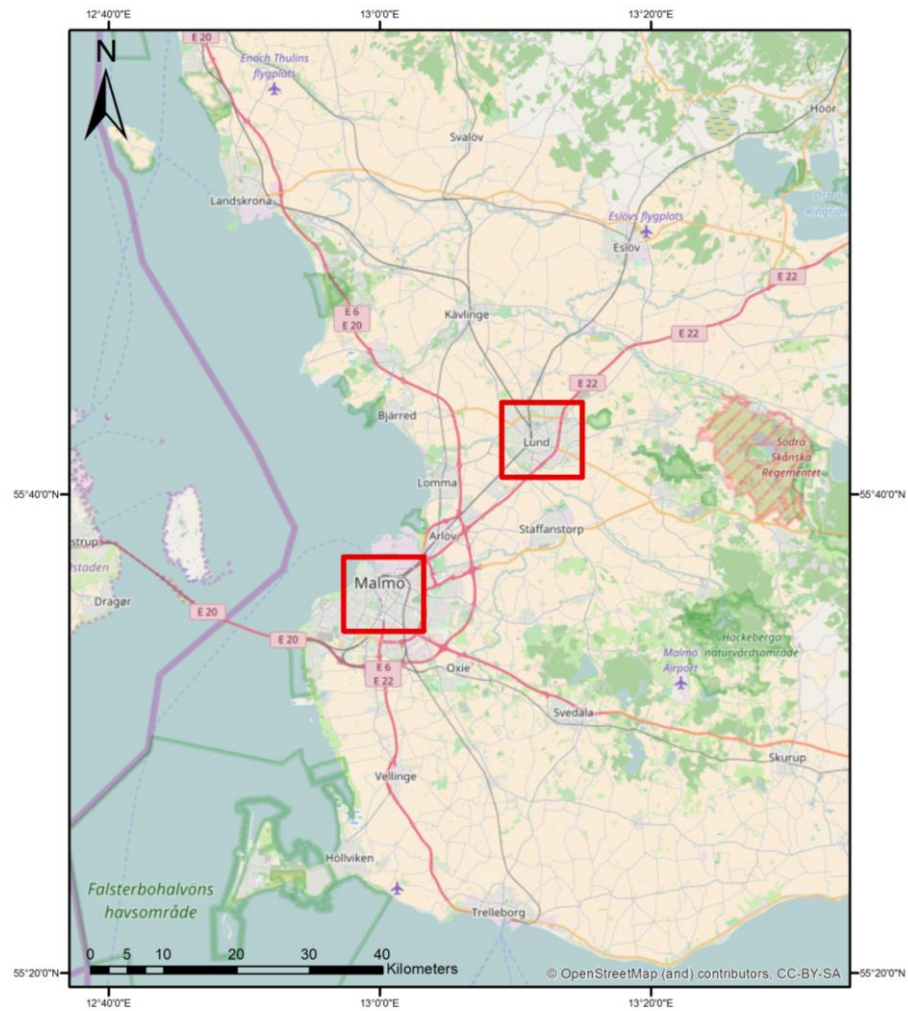
1. Define accessibility to public transport and perform a public transport accessibility analysis in Lund and Malmö.
2. Identify spatial and socio-economic variables with potential impact on car use.
3. Identify the explanatory variables generating the highest model performance when predicting car use in Lund and Malmö, and assess variables with a statistically significant relationship with car use by the use of Ordinary Least Squares (OLS).
4. Apply a Geographically Weighted Regression (GWR) to model the relationship between car use and the explanatory variables. This will generate the highest model performance when predicting car use in order to investigate spatially varying relationships.

## 1.5 Hypothesis

The main hypothesis of this thesis is that the level of car use for the individuals in Lund and Malmö is negatively associated with the accessibility to public transport. It is also expected that this relationship remains also after controlling for other spatial and socio-economic factors. However, because of the diversity within and between Lund and Malmö, it is not expected that the relationships between car use and the explanatory variables is spatially uniform. Thus, a spatial regression model such as the GWR is expected to better predict car use compared to a non-spatial OLS.

## 1.6 Study area

The cities of Lund and Malmö, with a population of 82 476 and 280 407 respectively, were selected as study areas (Figure 1). The cities are located in the south west of Scania and have a population density among the topmost in Sweden. It is common to travel between Lund and Malmö municipalities, and the total number of journeys between Lund and Malmö are the highest in Scania (SCB, 2013, Ullberg, 2013).



Lund



Malmö

Figure 1. The study areas of Lund and Malmö in the south west of Scania, Sweden. ESRI basemap.

## 2 Background

This chapter introduces the background of this thesis and consists of the following four main parts: (1) the incentives behind car use; (2) GPS in travel surveys; (3) Non-spatial and spatial regression; and (4) GWR in accessibility analysis.

### 2.1 The incentives behind car use

One of the critical factors when considering individuals transport choice is their accessibility to public transport, which is recognized as a key criterion when assessing transport policies (Benenson et al., 2011). Additionally, research demonstrates strong correlations between car use, spatial distribution, transport mode cost and socio-economic factors such as income, gender, occupation and age ( Chakrabarti., 2017, Shen et al., 2016, Bastian and Börjesson, 2015).

#### 2.1.1 Public transport accessibility

The emphasis of public transport accessibility is based on the fact easy access to transport is considered to be more important for individuals than the method of transport in question (Benenson et al., 2011). Geographic accessibility measures with regard to public transport is widely considered as the ease of interaction between people and locations and how easy it is to move from one place to another. These measures can be divided into three main types: (1) the physical access to public transport stations; (2) the duration of a journey by public transport; and (3) the access to a destination via public transport, which is a combination of the first two measures (Geurs and van Wee, 2004, Saghapour et al., 2016, Farber and Fu, 2017). The first measure is most commonly used to define public transport accessibility, where both time and distance can be used as a proxy of physical access to public transport stations (Saghapour et al., 2016). The emphasis on the physical access to transit-stops is based on the fact that the time it takes to travel to a public transport stop has considerable impact on the total travel time and, consequently, has substantial impact on what method of transport is chosen (Murray et al., 1998). Accordingly, the greater the proximity to a public transport stations, the greater the chance of them utilizing the services (Hawas et al., 2016). Research shows that walking is the most common way to access public transport stations. Consequently, the walking distance or time to a public transport station has been recognized as the most important variable to define public transport accessibility (Wibowo and Olszewski, 2005). Additionally, it has a clear concept and is easy to comprehend (Lin et al., 2014).

### 2.1.2 Income

Income is recognized as a key variable when predicting car use (Shen et al., 2016). Research regarding travel trends in Great Britain by Stapleton et al. (2017) showed that income explained the travel trends in Great Britain to a great extent, along with a rising fuel cost and urbanization. Additionally, Bastian and Börjesson. (2015) demonstrated that income and fuel prices explained as much as 80% of the aggregated car distances per person in Sweden. Moreover, their results showed that the car distances per person was more susceptible to income in municipalities where income is low and population density as well as the public transport accessibility is high (Bastian and Börjesson, 2015).

### 2.1.3 Gender

Gender is recognized as a variable with impact on car use. Scheiner and Holz-Rau. (2012) demonstrated that there are gender differences in travel habits even in households where there are as many cars as people with a driver's license. This suggests that personal-preferences may play a significant role in travel mode choice. These results may be explained by gender norms. Moreover, research regarding the car use of men and women are contradictory, where some report that women use cars less frequent than men and some report limited gender differences in transport mode choice (Scheiner and Holz-Rau, 2012).

### 2.1.4 Age

Age is another key socio-economic variable that influences car use though the impact of age on car use varies between countries. For example, in Germany retired people tend to walk, bike or utilize public transport, whereas in the USA, retired people tend to increase their car use (Buehler, 2011). Furthermore, Hagenauer and Hekbich. (2017) showed that age is more important than income regarding travel mode choice.

### 2.1.5 Occupation

The transport choice is affected by the complexity of the journey, which is affected by the occupation, job sector and household tasks of an individual (Scheiner and Holz-Rau, 2012). Hence, both residence and workplace environments are important components when considering the choice of transportation of an individual. In respect to the explanatory power of the residence and workplace environments, some argue that the workplace environment is of even greater importance, and thus provide a higher explanatory power in respect to the transport mode choice (Shen et al., 2016). Best and Lanzendorf. (2005) showed that in Cologne, Germany, students used a car in 26% of their trips, whereas full time workers used a car in 51.5% of their trips,

suggesting that employment increases car use. This could be a response of several universities implementing strategies to reduce the use of private cars and increase the use other transport modes, thus altering the student workplace environment (Rotaris and Danielis, 2015).

#### **2.1.6 Proximity to road infrastructure**

The correlation between road infrastructure and the utilization of motor vehicles is widely known, even though the constant growth in both the road network system and the regional travel demand has made it hard to assess the actual impact of road infrastructure on car use (Zhang et al, 2017). However, Duranton and Turner. (2011) show a proportional relationship between extensions of highways and the increase in traffic. Multiple studies highlight density and the adaption to pedestrians as key factors of roads in respect to mode choice, travel distance and trip frequency (Zhang et al, 2017)

#### **2.1.7 Summary**

In summary, public transport accessibility is recognized as one of the most critical factors when considering transport choice, partly based on the fact that transport is a derived need where ultimately accessibility is what matters (Benenson et al., 2011). Additionally, income, preferences, country and the residence and working environment represent other factors with an impact on car use. Consequently, predicting car use is a complicated matter that is much affected by a person's lifestyle, and thus identifying a set of variables that determines transit mode choice is a challenge (Chakrabarti, 2017). Hence, it is relevant to include both public transport accessibility and further socio-economic and spatial variables when predicting car use (Benenson et al., 2011, Shen et al, 2016).

## **2.2 GPS in travel surveys**

Travel surveys are commonly used as a resource in transport planning (Shen and Stopher, 2014). Traditional travel surveys often involve time consuming questionnaires that require detailed information from the respondent. Consequently, the resulting data are only approximations of travel details and are dependent on the respondent's perception of time and distance. Furthermore, the reported trips in traditional travel surveys are usually underestimated (Rasmussen et al., 2015, Xiao et al., 2015). In response, recent travel surveys have incorporated GPS-receivers to collect data, which has resulted in a data collection method with significant advantages compared to traditional travel surveys. One of the key benefits is the reduction of workload for the respondent, enabling data sampling for a longer period

of time (Xiao et al., 2015). Moreover the problem of respondents underestimating time is likely to be reduced, because respondents carry the GPS-receiver relatively continuously. Additionally, data collected with a GPS-receiver will be unbiased, precise and far more detailed compared to traditional travel surveys (Rasmussen et al., 2015).

## 2.3 Non-spatial and spatial regression

The OLS regression represents a global regression technique used to model the linear relationship between a dependent variable and one or more independent variables. Moreover, the OLS regression model assumes the relationship between the dependent and independent variables to be consistent across space, thus it represents a non-spatial regression model. Since spatial data usually possess regional variations and spatial autocorrelation, it is difficult to model spatial data and meet the assumptions of an OLS model. (Gao and Li, 2011). To overcome this limitation and generate higher model performance when modelling spatial data, the spatial regression model of GWR was developed (Fotheringham et al., 2002). GWR accounts for regional variations in the data by allowing local rather than global parameter estimates across a surface. The fundamentals of the method is that spatial autocorrelation is present within the sampled data, thus in GWR, data close to the estimation point  $i$  have more influence of the continuous function at point  $i$  than data further away (Brunsdont et al., 1998). GWR model outputs includes coefficient raster surfaces. One coefficient raster surface is obtained for each explanatory variable included in the model, and represent the change in the dependent variable for every one unit change in the explanatory variable, keeping all other variables constant. These coefficient raster surfaces could be visualized in maps and utilized to inform region wide and local policy (Ali et al., 2007, Fotheringham et al., 2002). Because the OLS regression model is an established and widely used model, the emphasis in this section will be on the spatial regression model of GWR.

### 2.3.1 Geographically Weighted Regression

The complete theory and background of the GWR model is presented in Fotheringham et al. (2002). The GWR is an extension of the global regression model, presented in Equation 1:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (1)$$

where  $y$  represents the dependent variable,  $x$  is the independent variable,  $\beta_0$  and  $\beta_1$  are the intercept and slope coefficients, and  $\varepsilon$  represents the error term. The estimator for



the corresponding variables is presented in Equation 2 (Fotheringham and Oshan, 2016):

$$\beta = (X^t X)^{-1} X^t y \quad (2)$$

where  $\beta$  is a vector of the global parameters to be estimated,  $X$  is a matrix of the independent variables and  $y$  represents a vector of observations that corresponds to the dependent variable. GWR (Equation 3) extends the global regression technique by allowing local instead of global parameters to be estimated, hence making it possible to model regional variations within the data (Fotheringham et al., 2002):

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i) x_i + \varepsilon_i \quad (3)$$

where  $(u_i, v_i)$  represent the coordinates for every  $i^{\text{th}}$  point in space, allowing a continuous surface of parameter values. The fundamentals of GWR is that spatial autocorrelation is present within the sampled data. Therefore, it is assumed that data near to point  $i$  have more influence regarding the estimation of the continuous function at point  $i$  than data further away from  $i$ . In practice, spatial autocorrelation is accounted for by applying a Weighted Least Squares (WLS) to the data. The equation for the GWR estimator is presented in Equation 4:

$$\beta(u_i, v_i) = (X^t W(u_i, v_i) X)^{-1} X^t W(u_i, v_i) y \quad (4)$$

where  $W(u_i, v_i)$  represents a matrix that assigns weights to observations based on their proximity to point  $i$ . Observations close to point  $i$  are assigned higher weights, which correspond to higher influence on the estimation of the continuous function at point  $i$ , and observations further away have lower influence and are assigned lower weights (Brunsdont et al., 1998). The format of the spatial weighting function as well as the number of observations included in the calculation for each spatial location are important aspects of GWR. The Gaussian function is a spatial weighting function commonly applied in GWR, where the weighting of data will decrease with an increasing distance from  $i$  according to a Gaussian curve (Equation 5):

$$w_{ij} = \exp\left(-\frac{d_{ij}}{h^2}\right) \quad (5)$$

where  $w_{ij}$  is the weight of observation  $j$  at point  $i$ ,  $d$  represents the distance from  $i$  and  $h$  represents the size of the kernel or bandwidth. Accordingly, the weight of data will decrease with an increasing distance from  $i$  until excluded from the calculation when

reaching a weight value of zero. There are two types of bandwidths: fixed or adaptive. A fixed bandwidth uses a kernel with a fixed size across the study region, resulting in a weighting function that is applied equally at every calibration point. The adaptive bandwidth uses an adaptive kernel that increases or decreases in size according to the distribution of the data. To select a fixed kernel could potentially cause a problem when modelling data that is not evenly distributed, because few observations will be included in the calculation in regions where data are sparse. Thus, adaptive kernels, with smaller kernels in areas where data are dense and larger kernels in areas where data are sparse, can be applied to data with a non-homogenous distribution (Brunsdont et al., 1998).

### 2.3.2 Criticism of Geographically Weighted Regression

The validity of the GWR method has been discussed, where multicollinearity between the independent variables represent one of the major concerns (Páez et al., 2011). Literature implies that the GWR method is highly susceptible to the effects of multicollinearity between the independent variables. This suggests that the correlation between GWR variables are a consequence of multicollinearity among the independent variables; both considering the local variable estimates for all locations and the specific variable estimates at each location (Fotheringham and Oshan, 2016). Effects of multicollinearity between variables include  $R^2$  overestimates and unreliable parameter estimates. However, research by Fotheringham and Oshan (2016) demonstrated that multicollinearity is not a problem in practice, and that the GWR method is robust to the effects of multicollinearity. Moreover, literature advocates caution when interpreting GWR results from small sample sizes ( $n \approx 160$ ), particularly when assessing the spatial heterogeneity of data (Páez et al., 2011). Nevertheless, to the author's knowledge, only one study has attempted to assess the effect of sample size on GWR model performance. This study, conducted by Devkota et al (2013), showed that regardless of the sample size, 88% of the variables included in the analysis had a non-stationary relationship with the dependent variable (Mitra L. Devkota, 2014).

### 2.3.3 OLS and GWR model output variables

An OLS model assumes the relationship modelled to be consistent across space. Thus, data with regional variations violate the OLS assumption of global stationarity. The Koenkers Breusch-Pagan statistics (BPK) test for heteroscedasticity in the OLS data by computing an OLS regression called LM regression (Equation 6) and then calculating the sample value of BPK (Equation 7):

$$\tilde{v} = Zd + e \quad (6)$$

Where  $\tilde{v}$  is a vector of squared OLS residuals,  $Z$  is a matrix of observed sample values of the independent variables,  $d$  is a column of vector coefficients and  $e$  represent an error term (Honda, 1988).

$$BPK = NR^2 \quad (7)$$

Where  $R^2$  is the coefficient of determination of the OLS regression and  $N$  is the sample size. A statistical significant BPK value reveals non-stationarity and violates the OLS assumption of global stationarity in the relationship modelled (Honda, 1988). Consequently, a statistical significant BPK value indicates that a local regression model such as the GWR, allowing the relationship modelled to vary across space, would be a better fit to the data than an OLS model (Gao and Li, 2011). To assess and compare the performance of the OLS and GWR models, the Adjusted R-squared ( $R^2$ ) and Akaike information Criterion corrected for small sample sizes (AICc) values may be considered. The  $R^2$  represents the coefficient of determination and measures the goodness of fit of the data, with values ranging from 0 to 1. The most general definition of the coefficient of determination is presented in Equation 8:

$$R^2 = 1 - \frac{SSD_{res}}{SSD_{tot}} \quad (8)$$

where SSD is the squares of deviation in a linear regression and the  $SSD_{res}$  and  $SSD_{tot}$  represent the residual square sum and the total square sum, respectively (Tjur, 2009). The Adjusted  $R^2$  (Equation 9) improves the  $R^2$  by normalizing the numerator and denominator according to their degrees of freedom, and hence compensates for the number of variables in the model (Ohtani, 2000):

$$R^2_{adj} = 1 - \left[ \frac{(1 - R^2)(n - 1)}{n - k - 1} \right] \quad (9)$$

with  $n$  representing the number of observations in the data sample and  $k$  representing the number of independent variables included in the analysis. Higher Adjusted  $R^2$  values indicate a better model fit, which corresponds to a higher proportion of the dependent variable variance being accounted for by the independent variables (Gao and Li, 2011). The Akaike information Criterion corrected for small sample sizes

(AICc) is a relative measure of model performance, where a smaller AICc value between two models that includes the same variables represents higher model performance. The definition of AICc is presented in Equation 10:

$$\text{AICc} = -2 \text{ supL} + 2q + 2 \frac{q(q + 1)}{N - q - 1} \quad (10)$$

where supL is the log-likelihood of the model representing the degree of fit and  $q$  is the total number of parameters included in the model (Karen, 2008). Generally, when comparing two models including the same variables, an AICc value differential of 3 or more indicates the model with a lower AICc value to be the better model. Consequently, an AICc value differential of 3 or more between two models indicates the model with the lower AICc value to have a better fit to the observed data (Gao and Li, 2011). Moreover, the variance inflation factor (VIF) can be used to assess multicollinearity between variables in the model and is presented in Equation 11:

$$\text{VIF}_j = -\frac{1}{1 - R^2} \quad (11)$$

where  $R^2$  is the coefficient of determination and the  $\text{VIF}_j$  is calculated for one independent variable  $x_j$ , based on the linear relationship between  $x_j$  and the other independent variables. VIF values much greater than 1 indicates multicollinearity between the variables in the model, (Vu et al., 2015). Moreover, spatial autocorrelation in regression residuals violates OLS assumptions and indicates a biased model (Rogerson, 2001). Moran's I (Equation 12) is one of the most commonly used statistics to test for spatial autocorrelation in datasets.

$$I = \frac{N \sum_i \sum_j W_{ij} Z_i Z_j}{M \sum_{i=1}^n Z_i^2} \quad (12)$$

Where  $N$  represents the number of data points,  $Z_i = x_i - \bar{x}$ ,  $\bar{x}$  is the mean value of  $x$ , and the  $M = \sum_{i=1}^n \sum_{j=1}^n W_{ij}$  together with  $W_{ij}$  are elements of the matrix of spatial proximity,  $M$ , that show the degree of spatial association between the points  $i$  and  $j$ . Further,  $W_{ij}=1$ ,  $j$  represents 1 if the  $k$  nearest neighbours of  $i$  and  $W_{ij}=0$  everywhere else (Kalogirou and Hatzichristos, 2007).

## 2.4 GWR in accessibility analysis in the transport sector

Du and Mulley. (2006) were the first to apply Geographically Weighted Regression (GWR) in the transport sector when the authors explored spatial variations in the relationship between transport accessibility and land values in the United Kingdom. Results indicated that the relationship between transport accessibility and land values was non-stationary, thus motivating the use of the local regression model of GWR that allows regional variations within the data. Du and Mulley. (2006) implied that modelling the relationship between transport accessibility and land values with GWR led to further knowledge about factors resulting in land value uplift. Additionally, Salas-Olmedo et al. (2017) used GWR to explore spatial variations in the relationship between the accessibility to green spaces and socio-economic variables in Chilean cities. Similar to Du and Mulley. (2006) their data possessed regional variations and results indicated that the modelled relationship varied across the city (Salas-Olmedo et al, 2017).

## 3 Data

This chapter introduces the data utilized in this study. Five datasets have been used: (1) GPS travel data; (2) questionnaire data; (3) income area data (4) road network data; and (5) public transport station data.

### 3.1 Travel survey data

Travel survey data represent the basis of the analysis in this study. The travel survey data consists of two types of data: (1) GPS travel data; and (2) questionnaire data. The GPS travel data represents travel data with spatial association from participants, and the questionnaire data represents personal information about participants. Both data types were obtained from TRavelVU, an app developed by Trivector Traffic in order to collect detailed travel survey data in an efficient way. In total, travel survey data from 136 individuals are analyzed in this study, with 67 and 69 individuals living in Lund and Malmö, respectively. The 136 individuals were recruited at public transport stations in Lund and Malmö or by signing up to donate data after getting in contact with Trivector Traffic. Thus, a majority of individuals analyzed in this study are users of the public transport system, which introduces a bias in the data (discussed in section 6).

#### 3.1.1 GPS travel data

By collecting GPS-tracks from users, and classifying the tracks as either trips or activities, the app collects a wide range of information regarding user activity and travel patterns. An event where the user is identified within the same area for a period of time is classified an activity, and a journey between two activities is classified as a trip. If the app identifies a trip, it classifies the trip with the presumable means of transport; such as walking, bike, car etc. The app is able to identify 10 transport modes automatically and the user can manually choose from 18 different means of transport. The first time an activity is identified at a specific location the user categorizes the type of activity, such as work, home, shopping etc. The next time an activity is identified at the same location, the app suggests the same activity, but this can be modified by the user if necessary. To make sure trips are being classified with the correct means of transport, the user approves the data collected for each day and can correct trips with an inaccurate classification. Hence, all data included in this study have been approved by the user. GPS travel data is obtained in the format of GeoJSON, an open standard format that represents geographical features (Butler, et al, 2016).

### 3.1.2 GPS travel data in this study

The GPS tracks analyzed in this study are collected between November the 7<sup>th</sup> 2016 and February 10<sup>th</sup> 2017; however, the number of trips and time period analyzed varies between all users. The data amount and time period analyzed for every individual is dependent on numerous of factors; for example, the time period the app has been downloaded, how often location services were activated, and how active the user has been. Hence, all user datasets will be unique and data amounts will vary considerable between users. The average amount of total trips for a user in Lund and Malmö is 55.24 and 66.61, respectively (Table 1). This could be put in relation to the average number of trips/day/person, which is 2.63 in Lund and 2.57 in Malmö (Ullberg, 2013).

**Table 1.** The average number of trips per user in Lund and Malmö.

| City  | Average trips/person | Standard deviation |
|-------|----------------------|--------------------|
| Lund  | 55.24                | 37.62              |
| Malmö | 66.61                | 70.30              |

### 3.1.3 Questionnaire data

An app questionnaire collecting voluntarily personal information about the participants, including age and occupation, represents the basis of the questionnaire data. The app questionnaire is attached in Appendix 1.

### 3.1.4 Questionnaire data in this study

Information about the user gender, age, main occupation and car accessibility is obtained by the app questionnaire (Appendix 1). All questionnaire datasets where complete except for the age dataset, in which approximately 40% of the data was missing. The management of the missing age data is further discussed in section 4.3.4.

### 3.1.5 Personal Integrity

The integrity of the app users is protected by the Swedish integrity law (Swedish: *personuppgiftslagen*), which was implemented in 1998 to avoid infringement of personal integrity when managing personal information. One important part of *personuppgiftslagen* is the approval of individuals regarding the management of their data (SFS 1998:204. *Personuppgiftslag*). In TRavelVU, this is accounted for by letting the users approve the management of their data before using the app. In respect of their personal integrity, the homes of the individuals analyzed in this study are mapped on an aggregated level.

## 3.2 Data sets

The following datasets have been utilized in this thesis:

- GPS travel data provided by the app TRavelVU in the format of GeoJSON, converted to point and line shapefiles.
- Questionnaire data in text format provided by the app TravelVU
- Income and region data obtained from the Swedish bureau of statistics (SCB), representing the total income from employment in a year for individuals over 20 years old. The income is represented as a median value for sub regions in the Lund and Malmö municipalities. The municipality sub areas are called Small Areas for Market Statistics (SAMS) and is constructed by SCB.
- Car and bike road network data obtained from *Trafikverket NVDB (Nationell vägdatabas)*
- Public transport station data of bus and train obtained from Open Street Map (OSM)

The corresponding data format, reference system and data provider is presented below (Table 2).

**Table 2.** The data formats utilized in this study.

| Datasets                 | Format            | Reference system | Provider          |
|--------------------------|-------------------|------------------|-------------------|
| GPS travel data          | GeoJSON           | SWEREF99         | Trivector traffic |
| Questionnaire data       | Text              |                  | Trivector traffic |
| Income area              | Polygon shapefile | SWEREF 99        | SCB               |
| Road network             | Line shapefile    | SWEREF 99        | Trafikverket      |
| Public transport station | Point shapefile   | WGS 84           | OSM               |

### 3.2.1 Quality of the OSM data

During a systematic quality analysis of the OSM data in England, results proved that OSM data is accurate with an average distance of 6 m between OSM data and data from the high-quality geodatabase OSM (Haklay, 2009). Due to diversity in the quality of the OSM data worldwide, a visual inspection was implemented to assure good quality of the data. Figure 2 and Figure 3 show examples of the visual inspection from Malmö and Lund, where the OSM public transport stations were compared to the stations in an ESRI basemap. No outliers were identified and results indicated that the OSM public transport station data was sufficiently accurate in Malmö and Lund.



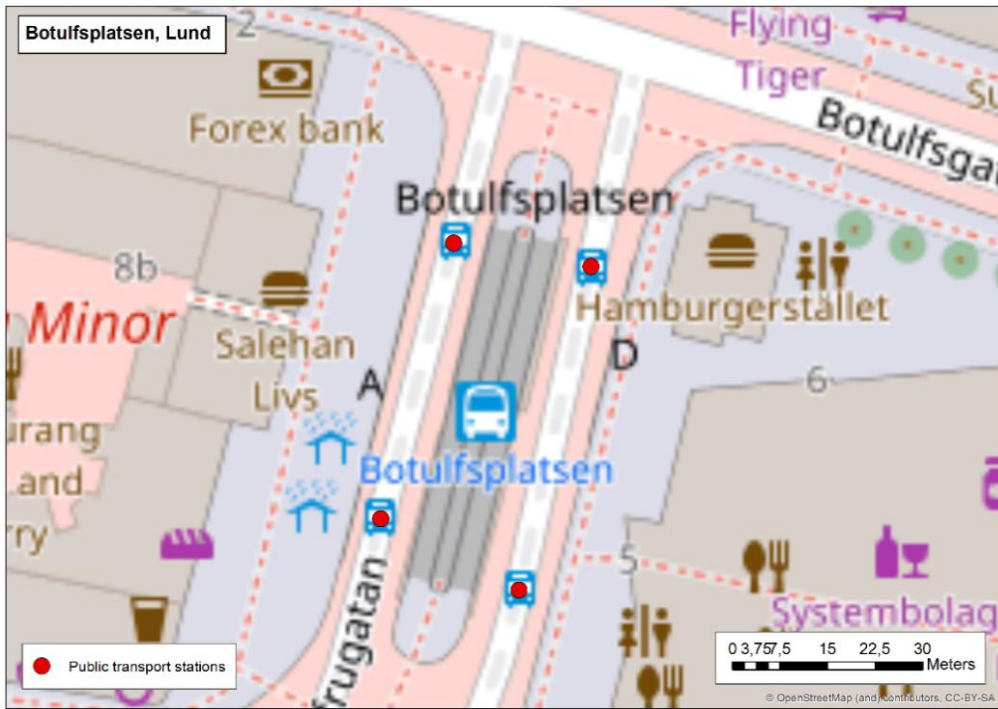


Figure 2. OSM public transport stations in relation to an ESRI basemap at Botulfsplatsen, Lund.

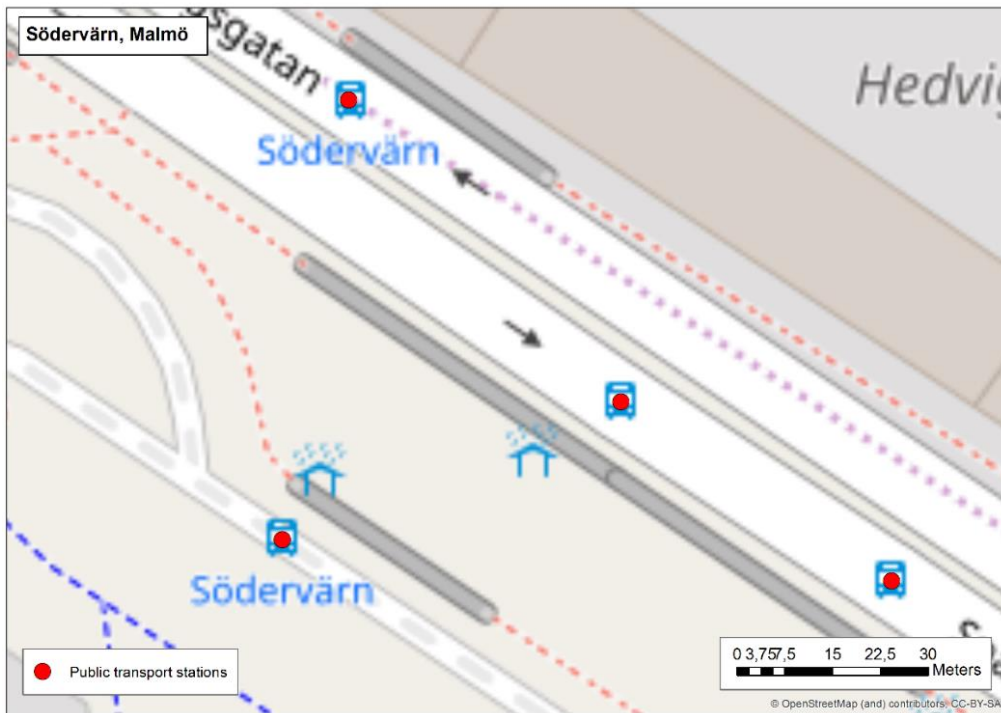


Figure 3. OSM public transport stations in relation to an ESRI basemap at Södervärn, Malmö.

## 4 Methodology

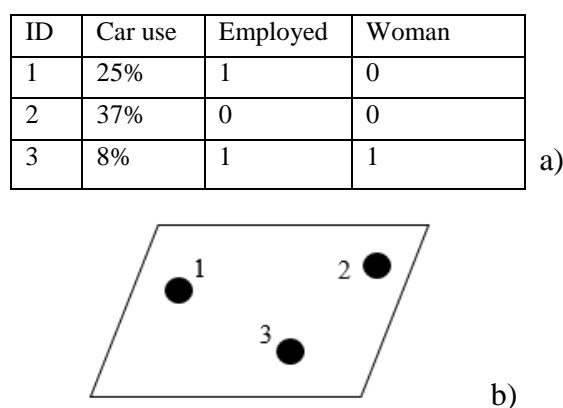
The objective of this study is to investigate how public transport accessibility affects car use. OLS and GWR models will be implemented, specifying car use and public transport accessibility as the dependent and explanatory variables respectively. Furthermore, the relationship between car use and income area, age, gender, car accessibility, main occupation, distance to the city central and distance to the closest highway will be investigated, to obtain further knowledge of incentives of car use. Thus, these variables will also be used as explanatory variables in this study. The methodology of this thesis consists of four main parts: (1) comments of data structure and visualization; (2) computing car use; (3) computing the explanatory variables; and (4) the statistical analyses.

### 4.1 Data structure and visualization

The following sections introduce the data structure and visualization in this study.

#### 4.1.1 Data structure

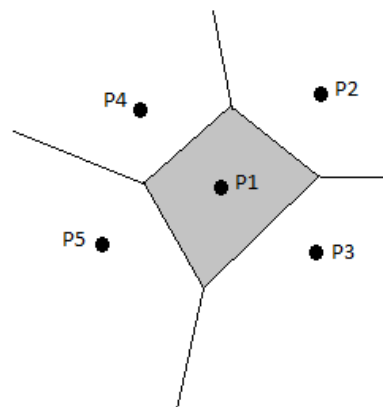
To explore regional variations in the incentives behind car use in Lund and Malmö, the participants needed to be assigned individual information of car use, public transport accessibility, age, gender, main occupation, car accessibility, income, distance to the city central station and distance to the closest highway. This was achieved by creating two point shapefiles, representing the participant residences in Lund and Malmö, and classifying each point with information of car use and information corresponding to the explanatory variables; because each point represented one residence and thus one individual. An example of this is demonstrated in Figure 4 below, where each point represent the residence of one individual. However, unlike the example below, points in the study was assigned with information relating to all explanatory variables, and not just main occupation and gender.



**Figure 4.** Example of the individual data structure in this study, where a) represents attributes to each residence b).

#### 4.1.2 Data visualization

Due to the personal integrity of the individuals analyzed in this study, individual data is presented on an aggregated level using Thiessen polygons. The Thiessen polygons are derived from the topological relationship between points in the Euclidean plane, where one polygon is created for each point in the dataset. The size of a Thiessen polygon depends on the dataset point distribution, because the area of a Thiessen polygon is associated with the closest point in respect to the Euclidean distance. Figure 5 shows an example of a Thiessen polygon created by a point set. Due to the formula of the Thiessen polygons, small polygons correspond to a dense distribution of points and large polygons correspond to a sparse distribution of points in that area (Mu, 2009). Hence, the residence density in this study is represented by the area of the Thiessen polygons, where a large area represents sparse distribution of participant residences and a small area represents a dense distribution of participant residences. In respect to this analysis, multiple residences with very high proximity would represent data visualization that is considerably less aggregated than when residences are sparse. However, there were no such residences in this study.



**Figure 5.** Thiessen polygon created by points P1, P2, P3, P4, and P5.

#### 4.2 Definition and calculation of car use

In order to investigate the relationship between car use and the explanatory variables, the participants needed to be assigned with individual information corresponding to their level of car use. In practice, this analysis was computed by converting the GPS travel data format from GeoJSON to shapefile, making it possible to import the data into the GIS software ArcGIS10.3.1 and analyze the travel data for all participants. Further, the individual level of car use referred to as *Car use*, was computed by applying Equation 13 on all participants' data:

$$Car\ use_i = \frac{No.\ of\ car\ trips_i}{No.\ of\ total\ trips_i} \quad (13)$$

where No. of car trips<sub>*i*</sub> represents the number of trips that are classified as car trips for person *i*, No. of total trips<sub>*i*</sub> represents the total number of trips, with any mean of transport, for person *i* and *Car use*<sub>*i*</sub> represents the car use value for person *i*. Hence *Car use* for each individual represents a percentage value, in which a higher percentage correspond to a higher level of car use. Considering the example in Table 3, person *i* would be classified with a car use value of 33% would he possess the example data.

**Table 3.** Example of GPS travel data of person *i*.

| Individual | No. of car trips <sub><i>i</i></sub> | No. of total trips <sub><i>i</i></sub> | Car use <sub><i>i</i></sub> |
|------------|--------------------------------------|--|-----------------------------|
| <i>i</i>   | 50                                   | 150                                    | 33%                         |

### 4.3 Explanatory variables

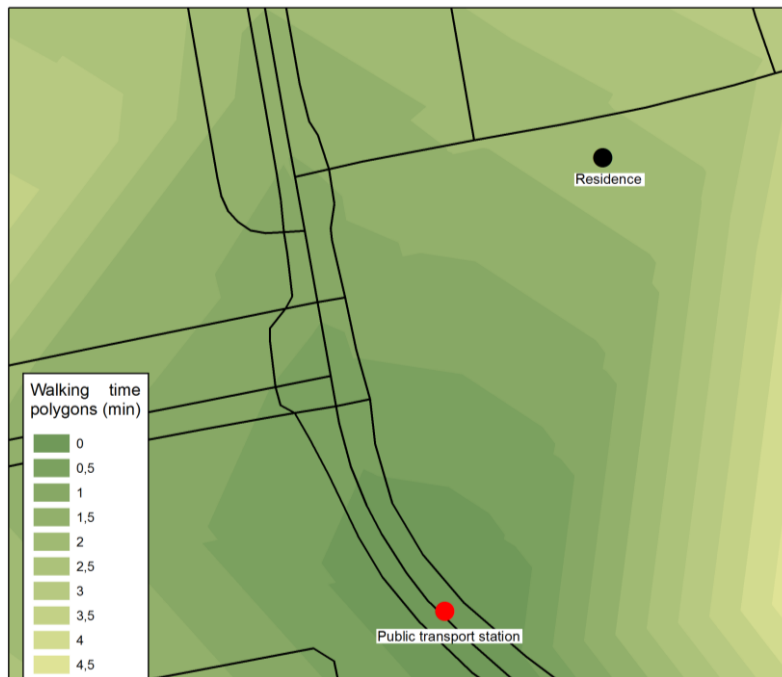
The main explanatory, or independent, variable, of interest in this study is public transport accessibility. Other explanatory variables are gender, main occupation, car accessibility, age, income, distance to the city central station and distance to the closest highway. The following sections describe the variables that were not obtained from the app questionnaire: public transport accessibility, income area, distance to the city central station and distance to the closest highway. Additionally, definition of age is presented below due to missing questionnaire data. Data of main occupation, car accessibility and gender are obtained from the app questionnaire (Appendix 1) and are simply based on the participants answer.

#### 4.3.1 Definition and calculation of public transport accessibility

In this study, the physical relationship between an individual's home and their closest public transport station is used to define public transport accessibility. More specifically, public transport accessibility is defined by the walking time by road from an individual's home to their closest public transport station. This data was obtained by performing a public transport accessibility analysis in ArcGIS10.3, using *Service area* in the *Network analysis* extension. This application computes the distance or travel time to certain facilities by road by utilizing a point layer and road network as input data. Before utilizing *Service area*, each road segment in the road network was classified with the time in minutes it would take to walk from one end of the road segment to the other, using Equation 14:

$$T_i = \frac{\text{Length}_i}{4.68} \cdot 60 \quad (14)$$

where  $\text{Length}_i$  represent the length of road segment  $i$  in km, 4.68 represents the walking velocity in km/h and  $T_i$  represents the minutes it would take to walk from one end of a road segment  $i$  to the other (Pachi, 2005). The walking time to public transport stations was computed using the *service area* tool and allowing the public transport stations to represent facilities. By setting walking time thresholds to 30 seconds, walking time polygons were created for every 30 second increase in the walking time to a public transport station (Figure 6).



**Figure 6.** Walking time polygons from a public transport station

Each participant was thereafter assigned a public transport accessibility value based on the spatial overlap of their residence and the walking time polygons. Because the residence in Figure 6 overlaps with the walking time polygon of 2 minutes, the individual in this residence would be classified with a public transport accessibility of 2 minutes. This indicates that it takes between 2 and 2.5 minutes to walk from their residence to the closest public transport station.

#### 4.3.2 Definition and calculation of income area

In this study, income is defined as the the total income from employment in a year for individuals over 20 years old, represented as median values for sub regions in municipalities. Based on income thresholds obtained from SCB, the income areas were categorized into low (<158'011 SEK), medium-low (158'012-268'312 SEK),

medium-high (268'313-446'641 SEK) and high (>446 641 SEK) (SCB, 2009). In practice, shapefiles representing income of sub regions in the municipalities were imported to ArcGIS10.3.1 and each region was classified to possess a low, medium-low, medium-high or high income. Each individual was assigned an income category based on the income classification of their residence region. Hence, income is presented on an aggregated level and the classification of each individual is based on the income of their region of residence.

#### 4.3.3 Definition and calculation of distance to city central stations and closest highway

The Euclidean distance between the residence of the individuals and the city central stations and closest highway was used as definition of distance in this study. This corresponds to the straight line between the residence of the individuals and the city central station and closest highway. Consequently, the Euclidian distance in km was used to calculate the variables: *Distance to Malmö C*, *Distance to Lund C* and *Distance to highway*, representing the distance to the city central stations and closest highway in Lund and Malmö.

#### 4.3.4 Definition and calculation of age

The proportion of the individuals with missing age data was approximately 40%. Individuals with missing age categories were assigned an age based on the age distribution of the sample population, by utilizing weights. The weights were calculated for each unique age in the population, and represented the proportion of the age population with that unique age. Considering the example in Table 4, 10%, 20%, 50% and 20% of the data obtained have an age of 20, 25, 30 and 35, respectively. The assignment of age was based on the weights; hence, the chance of being assigned an age of 35 would be 20% considering the age population in table 4. Further, age in Lund was categorized in to the groups: <25, 26-30, 31-35, 36-52 and >52, and age in Malmö was categorized in to the groups: <25, 26-30, 31-35, >35. This categorisation was based on the age distribution in Lund and Malmö, which is why the age groups are not exactly the same on both cities.

**Table 4.** Example of an age population.

| Individual | Age | Weights |
|------------|-----|---------|
| $i_1$      | 20  | 10      |
| $i_2$      | 25  | 20      |
| $i_3$      | 30  | 50      |
| $i_4$      | 35  | 20      |

## 4.4 Statistical analyses

In this study, the global regression model Ordinary Least Squares (OLS) and the local regression model Geographically Weighted Regression (GWR) were applied to explore the incentives behind car use. Separate models are estimated for Lund and Malmö, whereas the variables included in both analyses are the same, apart from the age classes 36-52 and >52 years which is applied in Lund exclusively and the age class of >35 that are applied in Malmö exclusively. The input data format were two point shapefiles, holding information about the dependent variable of car use and the explanatory variables included in the analysis: public transport accessibility, age, gender, main occupation, car accessibility, income, distance to the city central and distance to the closest highway.

### 4.4.1 OLS regression models

The Exploratory Regression tool in ArcGIS 10.3.1 runs OLS regression analysis on all possible combinations of the input variables in order to find a properly specified OLS model (Rosenhein et al, 2011). Consequently, the Exploratory Regression tool was used to identify the explanatory variables generating the highest model performance when predicting car use in Lund and Malmö respectively; both in respect to the adjusted  $R^2$ , AICc, BPK and VIF values. The variables generating the highest model performance in both cities were included in OLS models in Lund and Malmö, respectively, to estimate the explanatory variables with a statistically significant ( $p < 0.05$ ) relationship with car use in both cities. By including the same variables in separate models for Lund and Malmö, potential differences regarding the incentives behind car use could be detected between the cities.

### 4.4.2 GWR models

To explore regional variations in the relationships modelled, GWR was applied to the explanatory variables generating the highest model performance in Lund and Malmö, respectively. When specifying the properties for the GWR model, the kernel type was set to adaptive to account for a heterogeneous distribution of data points in Lund and Malmö. Further a Gaussian kernel was used, representing a kernel where the weighting of data will decrease with an increasing distance from  $i$  according to a Gaussian curve. AICc was used to select the optimal number of neighbors, which was set to 30 at both locations.

## 5 Results

This chapter introduces the results in this study and consist of five main parts: (1) descriptive statistics; (2) OLS results in Lund; (3) OLS results in Malmö; (4) GWR results in Lund; and (5) GWR results in Malmö.

### 5.1 Descriptive statistics

The participant mean and standard deviation values regarding the non-categorical, continuous, data in Lund and Malmö are presented in Table 5 and Table 6, respectively.

**Table 5.** Descriptive statistics regarding the non-categorical data in Lund.

| Variable                             | Mean    | Standard deviation |
|--------------------------------------|---------|--------------------|
| Car use (%)                          | 25%     | 25%                |
| Public transport accessibility (sec) | 6.8 min | 4.07 min           |
| Distance to Lund C (km)              | 4.02 km | 1.72 km            |
| Distance to highway (km)             | 2.18 km | 1.42 km            |

**Table 6.** Descriptive statistics regarding the non-categorical data in Malmö.

| Variable                             | Mean    | Standard deviation |
|--------------------------------------|---------|--------------------|
| Car use (C, %)                       | 18%     | 18%                |
| Public transport accessibility (sec) | 3.4 min | 2.15 min           |
| Distance to Malmö C (km)             | 4.01 km | 1.75 km            |
| Distance to highway (km)             | 2.19 km | 1.28 km            |

Lund has a higher average *Car use* of 25% compared to the average *Car use* of 18% in Malmö. Considering *Income area* in both cities, income is slightly higher in Lund where 51% of the users have a medium-high or high income in contrast to Malmö where the corresponding proportion is 23%. Moreover, *Car accessibility* is higher in Lund where 73% of the users have access to a car always or sometimes, compared to 56% in Malmö. In contrast, the mean walking time to a public transport station, *Public transport accessibility*, is twice as high in Lund compared to Malmö, with 3.4 min in Malmö and 6.8 min in Lund. The *Public transport accessibility* standard deviation is 4.07 and 2.15 in Lund and Malmö, respectively. This suggests that *Public transport accessibility* deviates twice as much from the mean in Lund than in Malmö.



The distribution of the participants regarding the categorical, Boolean (0, 1), data in Lund and Malmö is presented in Table 7 and Table 8, respectively.

**Table 7.** Descriptive statistics regarding the categorical, Boolean, data of the 67 participants in Lund.

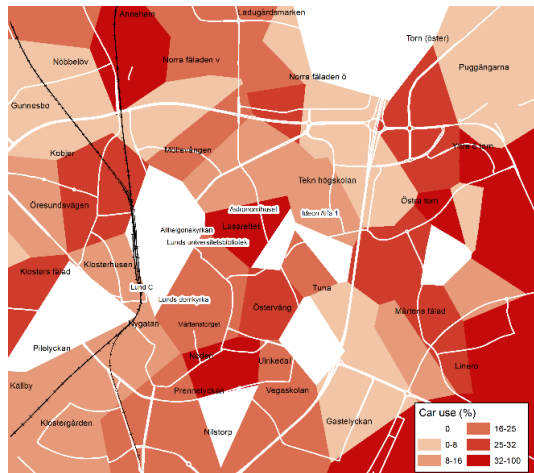
| Variable          | Participant distribution |
|-------------------|--------------------------|
| Income area       |                          |
| -Low              | 27%                      |
| -Medium-low       | 22%                      |
| -Medium-high      | 43%                      |
| -High             | 8%                       |
| Gender            |                          |
| -Female           | 64%                      |
| -Male             | 36%                      |
| Car accessibility |                          |
| -Always           | 49%                      |
| -Sometimes        | 24%                      |
| -Never            | 27%                      |
| Main occupation   |                          |
| -Student          | 40%                      |
| -Employed         | 60%                      |
| Age               |                          |
| <=25 years        | 16%                      |
| 26-30 years       | 25%                      |
| 31-35 years       | 18%                      |
| 36-52 years       | 24%                      |
| >52 years         | 16%                      |

**Table 8.** Descriptive statistics regarding the categorical, Boolean, data of the 69 participants in Malmö.

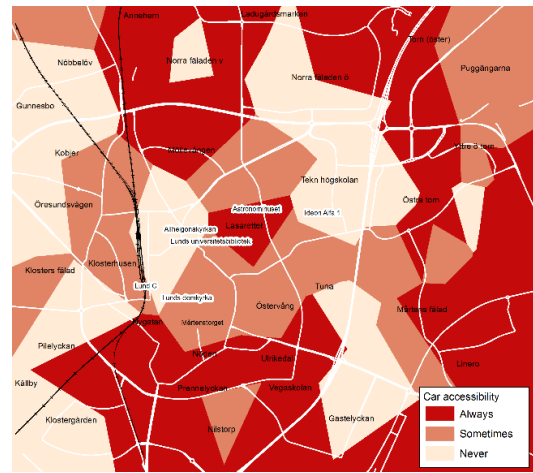
| Variable          | Participant distribution |
|-------------------|--------------------------|
| Income area       |                          |
| -Low              | 15%                      |
| -Medium-low       | 62%                      |
| -Medium-high      | 20%                      |
| -High             | 3%                       |
| Gender            |                          |
| -Female           | 60%                      |
| -Male             | 40%                      |
| Car accessibility |                          |
| -Always           | 19%                      |
| -Sometimes        | 37%                      |
| -Never            | 44%                      |
| Main occupation   |                          |
| -Student          | 46%                      |
| -Employed         | 54%                      |
| Age               |                          |
| <=25 years        | 20%                      |
| 26-30 years       | 31%                      |
| 30-35 years       | 19%                      |
| >35 years         | 30%                      |

Maps displaying the distribution of the variables (Table 5, Table 7) in Lund are presented in Figure 7. It is evident that *Car use* (a) is low or non-existent in areas where *Car accessibility* (b) is defined as never, and high where *Car accessibility* is defined as always. *Public transport accessibility* (c), derived by the public transport accessibility analysis, ranges between 0-9 minutes in most of Lund. However, in the east side of the city, in areas around Östra torn, *Public transport accessibility* ranges between 12-22 minutes. This can be put in relation to a high level of *Car use* (a) in this area.

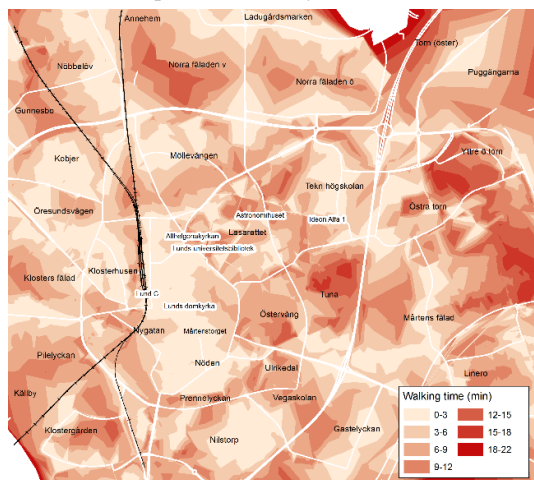
(a) Car use



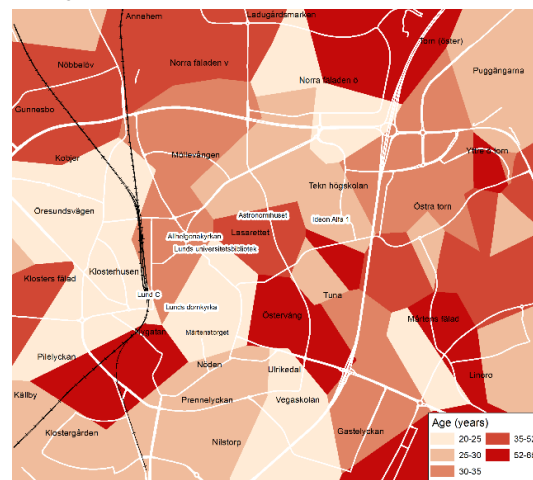
(b) Car accessibility



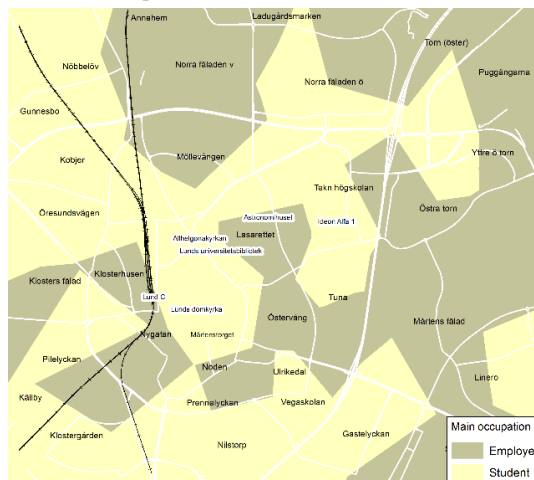
(c) Public transport accessibility



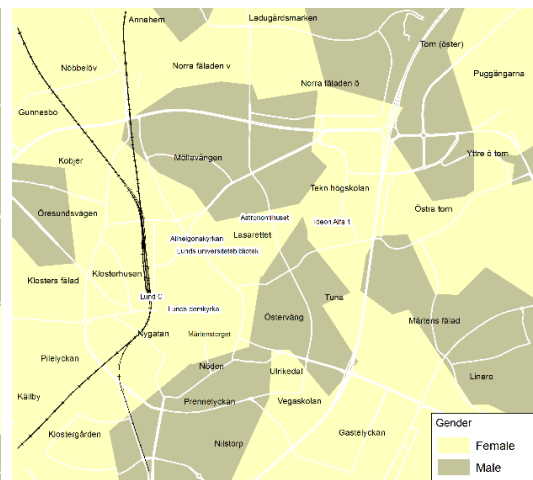
(d) Age



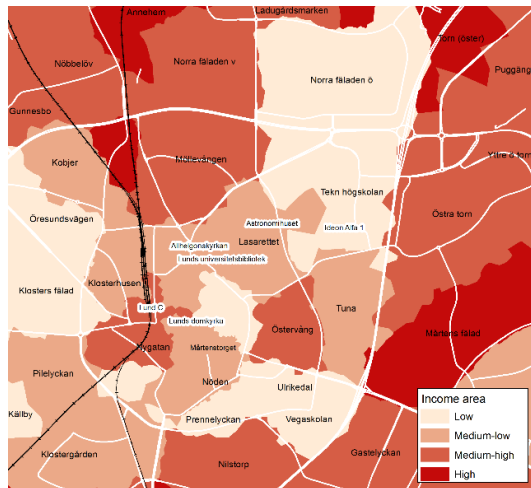
(e) Main occupation



(f) Gender



(g) Income area



(h) Residence density of participants

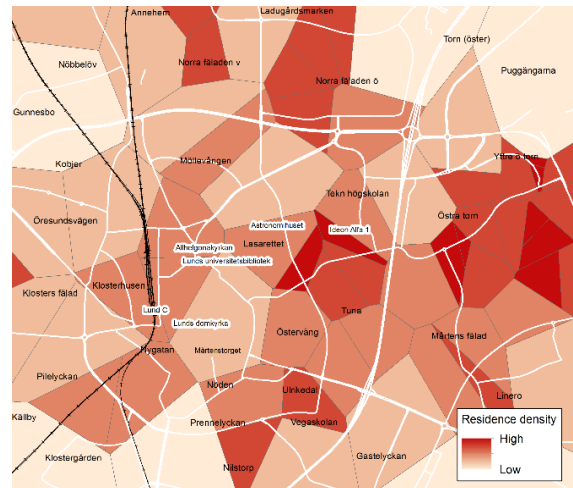
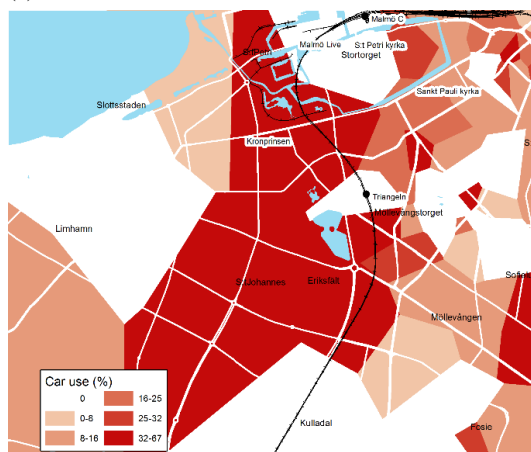


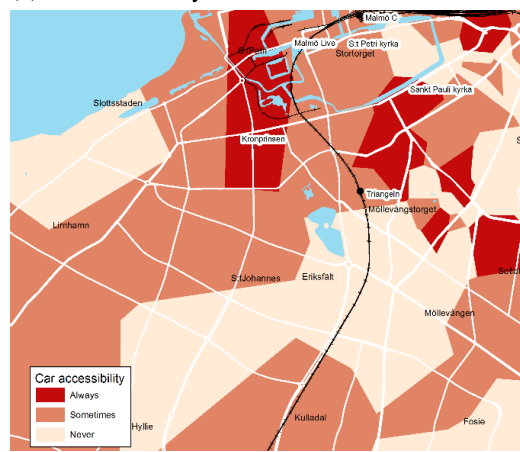
Figure 7. Maps of car use, car accessibility, public transport accessibility, age, main occupation, gender, income area and residence density of participants in Lund.

Maps that show the distribution of the variables in Malmö (Table 6, Table 8) are presented in Figure 8. The association between *Car use* (a) and *Car accessibility* (b) is not as prominent in Malmö as it is in Lund (Figure 7). Considering the *Public transport accessibility* (c), it is relatively uniform across Malmö, ranging between 0-6 minutes. In respect to the *Income area* (g) there is a prominent contrast between the east and west of the city, in which the income in the west side of Malmö are much higher.

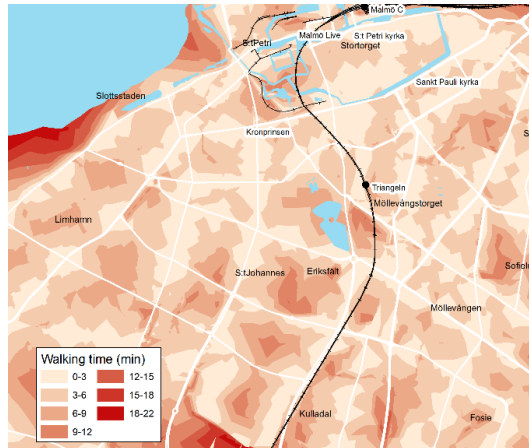
(a) Car use



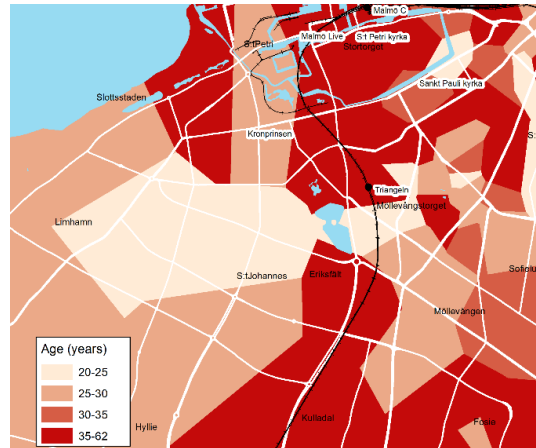
(b) Car accessibility



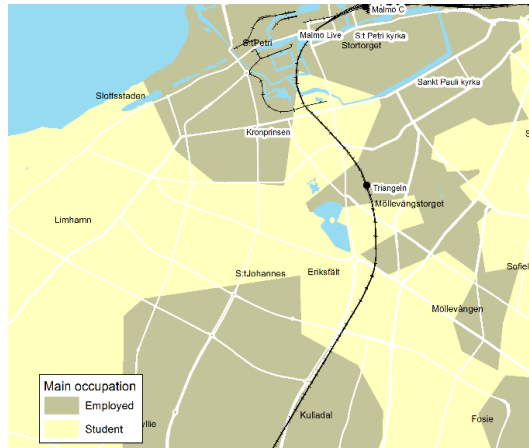
(c) Public transport accessibility



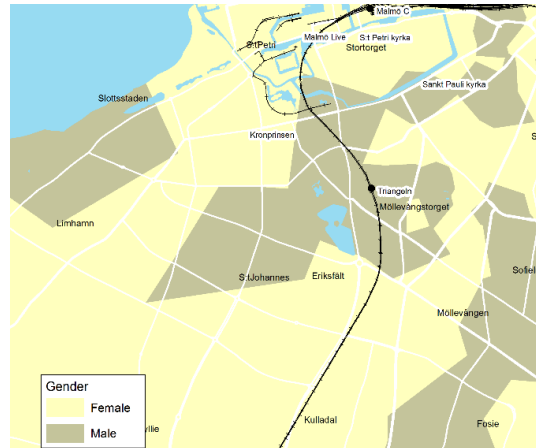
(d) Age



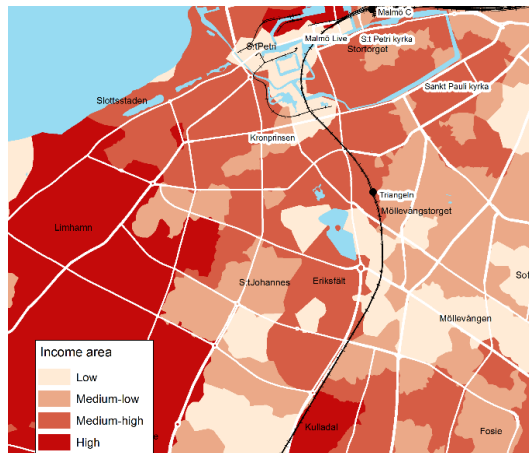
(e) Main occupation



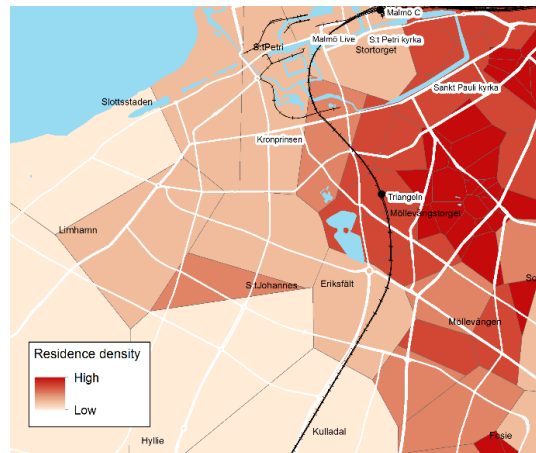
(f) Gender



(g) Income area



(h) Residence density of participants



**Figure 8.** Maps of car use, car accessibility, public transport accessibility, age, main occupation, gender, income area and residence density of participants in Malmö. The light blue areas in the maps represent water.

### 5.1.1 OLS results in Lund

Accessibility to public transport seems to have a relatively strong impact on car use in Lund. Considering the robust probability value below 0.05, *Public transport accessibility* possesses a statistically significant ( $p < 0.05$ ) relationship with *Car use*. (Table 9). The positive coefficient value of *Public transport accessibility* indicates that car use will decrease with a higher accessibility to public transport. More specifically, the coefficient values represent the change in *Car use* for every one unit change in the explanatory variable, holding all other variables constant. Thus, the *Public transport accessibility* coefficient value of 0.02 indicates that with every 30 second increase in walking time to a public transport station, car use increases with 2%.

**Table 9.** Variables generating the highest model performance in Lund. Statistically significant coefficients ( $p < 0.05$ ) are denoted with an asterisk.

| Variable                            | Coefficient | Robust Probability | Robust standard error | VIF    |
|-------------------------------------|-------------|--------------------|-----------------------|--------|
| Public transport accessibility      | 0.0162      | 0.0427*            | 0.0078                | 1.2132 |
| Distance to Lund C                  | 0.0606      | 0.0002*            | 0.0155                | 1.3285 |
| Distance to highway                 | 0.0287      | 0.0694             | 0.0155                | 1.1293 |
| Male                                | -0.0544     | 0.3043             | 0.0525                | 1.1519 |
| Always or sometimes access to a car | 0.1689      | 0.0004*            | 0.0454                | 1.1958 |

Moreover, *Distance to Lund C* and *Always or sometimes access to a car* are also statistically significant. The *Distance to Lund C* coefficient of 0.06 indicates that for each 1 km increase in distance to Lund C, car use increases with 6%. The *Always or sometimes access to a car* coefficient value of 0.17 indicates that if car accessibility changes from never to sometimes or always, car use increases with 17%. The explanatory variables that generate the highest OLS model performance when predicting *Car use* in Lund are *Public transport accessibility*, *Distance to Lund C*, *Distance to highway*, *Male* and *Always or sometimes access to a car*.

The Adjusted  $R^2$  value for the OLS model is 0.41, indicating that 41% of the car use in Lund can be explained by the explanatory variables included in the model (Table 10).

**Table 10.** Lund OLS results

| No. of Observations | R <sup>2</sup> | Adjusted R <sup>2</sup> | Koenker (BP) statistics | AICc   |
|---------------------|----------------|-------------------------|-------------------------|--------|
| 67                  | 0.4585         | 0.4141                  | 0.02874*                | -21,92 |

The AICc value of the OLS model is -21.92 and the VIF values are all close to 1, indicating no multicollinearity between the variables included in the OLS model. The Koenker (BPK) value of 0.03 is statistically significant which implies that the relationship modelled is non-stationary and violates the OLS assumption of global stationarity (Gao and Li, 2011). Since the relationships modelled are non-stationary, the robust probability and standard error values represent the significance and efficiency of each explanatory variable (Table 10).

The results of testing the OLS residuals for spatial autocorrelation measured by Moran's I are presented in Table 11. The z-score of 1.09, compared with the critical z-score of 1.96, reveals the distribution of the OLS residuals not to be statistically significant clustered or dispersed, indicating that the OLS residuals are randomly distributed and that the OLS model is not misspecified and there are no explanatory variables missing.

**Table 11.** Lund Moran's I results

| Critical z-score | z-score | p-value |
|------------------|---------|---------|
| 1.96             | 1.09    | 0.28    |

## 5.2 OLS results in Malmö

Results in Malmö demonstrate no significant effects of accessibility to public transport on car use. Solely *Distance to Malmö C* has a robust probability value below 0.05 and is statistically significant at the 5% level (Table 12). The association between *Distance to Malmö C* and *Car use*, indicates a decrease in car use with an increase in distance to Malmö C. More specifically, the *Distance to Malmö C* coefficient of -0.02 indicates that if distance to Malmö C increases with 1km, car use decreases with 2%. Additional variables possess Robust probability values greater than 0.05, and are not statistically significant at the 5% level.

**Table 12.** Variables generating the highest model performance in Malmö. Statistically significant coefficients ( $p < 0.05$ ) are denoted with an asterisk.

| Variable                       | Coefficient | Robust Probability | Robust standard error | VIF    |
|--------------------------------|-------------|--------------------|-----------------------|--------|
| Public transport accessibility | 0.0112      | 0.4507             | 0.0148                | 1.0855 |
| Distance to Malmö C            | -0.0232     | 0.0359*            | 0.0108                | 1.2987 |
| Employed                       | -0.0842     | 0.0585             | 0.0437                | 1.1280 |
| Low income                     | 0.1125      | 0.1142             | 0.0702                | 1.1266 |

The explanatory variables that generate the highest OLS model performance when predicting *Car use* in Malmö are *Public transport accessibility*, *Distance to Malmö C*, *Low income*, *Employed* and *Low income*.

The adjusted  $R^2$  value for the OLS model is 0.05, indicating that 5% of the car use in Malmö can be explained by the explanatory variables included in the model (Table 13).

**Table 13.** Malmö OLS results.

| No. of Observations | $R^2$  | Adjusted $R^2$ | Koenker(BP) statistics | AICc     |
|---------------------|--------|----------------|------------------------|----------|
| 69                  | 0.1103 | 0.0546         | 0.0096*                | -36.6040 |

The VIF values close to 1 indicate that there is no multicollinearity between the variables included in the OLS model. The Koenker (BPK) value of 0.009 is statistically significant at the 5% level and reveals the relationship modelled to be non-stationary, thus violating the OLS assumption of global stationarity (Gao and Li, 2011). Since the data have a non-stationary structure, the Robust probability and standard error values represent the significance and efficiency of the explanatory variables (Table 13).

The results of testing the Malmö OLS residuals for spatial autocorrelation measured by Moran's I are presented in Table 14. The z-score of 0.16, compared with the critical z-score of 1.96, reveals the distribution of the OLS residuals not to be statistically significant clustered or dispersed, indicating that the OLS residuals are randomly distributed and that there is no explanatory variables missing in the model.

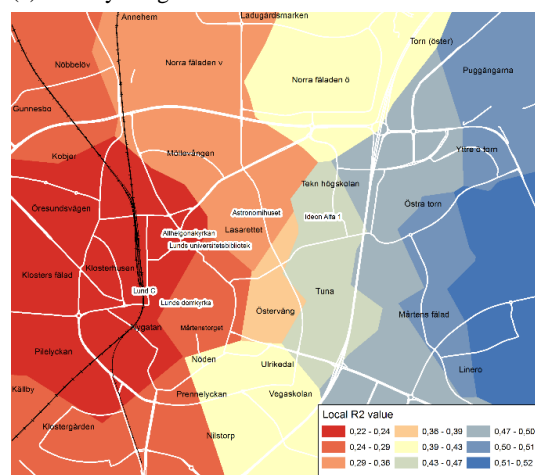
**Table 14.** Malmö Moran's I results.

| Critical z-score | z-score | p-value |
|------------------|---------|---------|
| 1.96             | 0.16    | 0.87    |

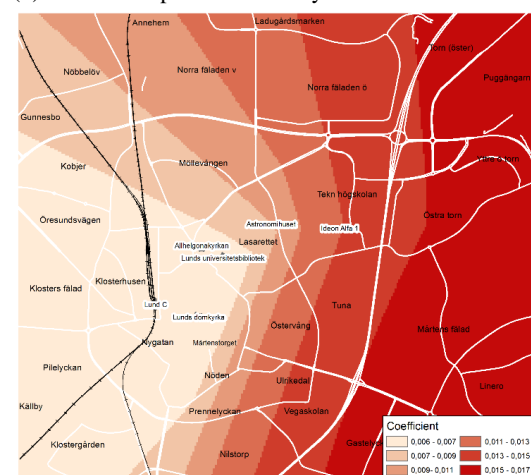
### 5.3 GWR results in Lund

The relationship between *Car use* and the explanatory variables that generated the highest OLS model performance when predicting car use in Lund were modelled using GWR. These variables included *Public transport accessibility*, *Distance to Lund C*, *Distance to highway*, *Male* and *Always or sometimes access to a car*. Regression coefficient surfaces representing the change in *Car use* for every one unit change in the explanatory variables, holding all other variables constant, were obtained using GWR. The following regression coefficient surfaces show statistically significant explanatory variables in Lund; *Public transport accessibility*, *Distance to Lund C* and *Always or sometimes access to a car* (Figure 9). The regression coefficient map of *Public transport accessibility* (b) display the change in *Car use* for every 30 sec increase in walking distance to a public transport station across Lund. It is evident that public transport accessibility has a higher impact on car use in the east than in the west side of Lund, with coefficient values of ranging from 0.02 in the east to 0.01 in the west. In the east side of Lund, in areas such as Östra torn and Mårtens Fälad, *Public transport accessibility* coefficients reach 0.02, indicating a 2% increase in car use for every 30 second increase in walking time to a public transport stop. In areas around Lund C, Allhelgonakyrkan and west of the railway *Public transport accessibility* have lower influence on car use and a 30 second increase in walking time to a public transport stop generates a 1% increase in car use.

(a) Locally weighted  $R^2$  values

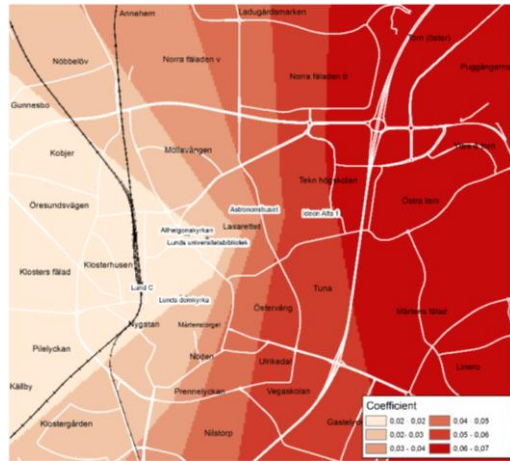


(b) Public transport accessibility

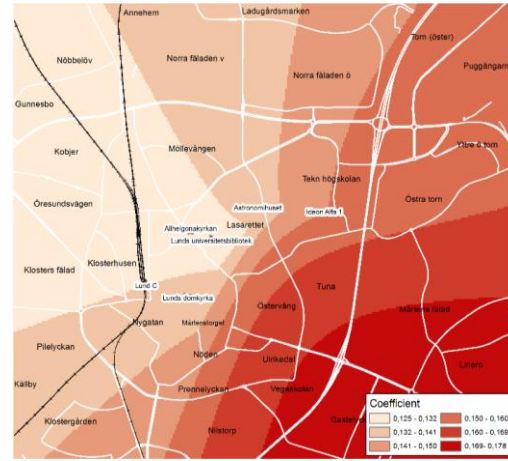




(c) Distance to Lund C



(d) Always or sometimes access car



**Figure 9.** Maps of the locally weighted  $R^2$  values and the regression coefficient surfaces of the statistically significant at the 5% level explanatory variables in Lund obtained by the GWR model: *Public transport accessibility* (sec), *Distance to Lund C* (km) and *Always or sometimes access to car* (Boolean).

The regression coefficient map of *Distance to Lund C* (c) display the change in *Car use* for every 1 km increase in *Distance to Lund C*. It is evident that the highest influence of *Distance to Lund C* is seen in the east side of Lund, in areas such as Östra torn and Mårtens Fälad. In these areas, coefficient values reach 0.07, which indicates that an increase in the distance to Lund C with 1 km would increase car use with 7%. The lowest influence of *Distance to Lund C* is seen in areas around Lund C, Allhelgonakyrkan, Gunnesbo and Klosters Fälad. In these areas, an increase in distance to Lund C with 1 km would increase car use with 2%. The regression coefficient map of *Always or sometimes access to a car* (d) display the change in *Car use* if car accessibility change from never to always or sometimes. *Always or sometimes access to a car* has the highest influence on car use in the south east side of Lund, in areas such as Linero, Gastelyckan and Mårtens fälad. In these areas, coefficient values reach 0.18, indicating that a change of car accessibility from never to sometimes or always would increase car use with 18%. *Always or sometimes access to a car* has lower influence on car use in the north west side of Lund, in areas such as Nöbbelöv, Gunnesbo and Klosters fälad. Coefficient values in these areas ranges between 0.12 and 0.13, indicating that a change in car accessibility from never to always or sometimes would increase car use with 12-13%.

The locally weighted  $R^2$  values obtained by the GWR model (Figure 9 (a) indicates how well the explanatory variables included in the GWR model predicts car use across Lund. With  $R^2$  values ranging from 0.22 to 0.51, it is evident that the GWR model performance varies across Lund, with the most prominent contrast between the east and west side of the city. The highest GWR model performances, with  $R^2$  values of 0.48-0.51, are found in the east side of Lund in areas such as Östra Torn and

Mårtens fälad. The lowest model performances, with  $R^2$  values between 0.22-0.25, are identified around Lund C and on the west side of the railway, in areas such as Klosters fälad and Pilelyckan.. Thus, the explanatory variables included in the GWR model are better at predicting car use in the east than in the west side of Lund

The adjusted  $R^2$  value of the GWR model is 0.44 (Table 15), compared with the adjusted  $R^2$  value of 0.41 for the OLS model (Table 10). This indicates that 44% of the car use in Lund can be explained by the explanatory variables included in the model. The AICc value of the GWR model is -21.38, which is similar to the OLS model AICc value of -21.92 (Table 10). The residual squares of the GWR model is 1.92 and represents the sum of the squared residuals in the model, where residuals are the difference between an observed  $y$  value and the GWR estimated  $y$  value.

**Table 15.** GWR results in Lund.

| No. of Observations | $R^2$  | Adjusted $R^2$ | AICc    | Residual squares |
|---------------------|--------|----------------|---------|------------------|
| 67                  | 0.5337 | 0.4388         | 21.3781 | 1.9210           |

#### 5.4 GWR results in Malmö

The relationship between *Car use* and the explanatory variables that generated the highest OLS model performance when predicting car use in Malmö was modelled using a GWR. These variables include *Public transport accessibility*, *Distance to Malmö C*, *Low income*, *Employed* and *Low income* (Table 13). The regression coefficient surface of the statistically significant explanatory variable *Distance to Malmö C* (b) is obtained by the GWR model and represent the change in *Car use* for every one unit change in the explanatory variable, holding all other variables constant (Figure 10). Thus, the regression coefficient map of *Distance to Malmö C* (b) shows the change in car use if distance to Malmö C increases with 1 km. The continuously negative coefficients values across Malmö imply that an increase in the distance to Malmö C will decrease car use across Malmö. The highest impact of *Distance to Malmö C* is observed in areas around Malmö C and Stortorget, where coefficient values implies a 3% decrease in car use with a 1 km increase in distance to Malmö C. Low impact is seen in areas around Fosie and Slottstaden, where coefficient values implies a 2-3% decrease in car use with a 1 km increase in distance to Malmö C.

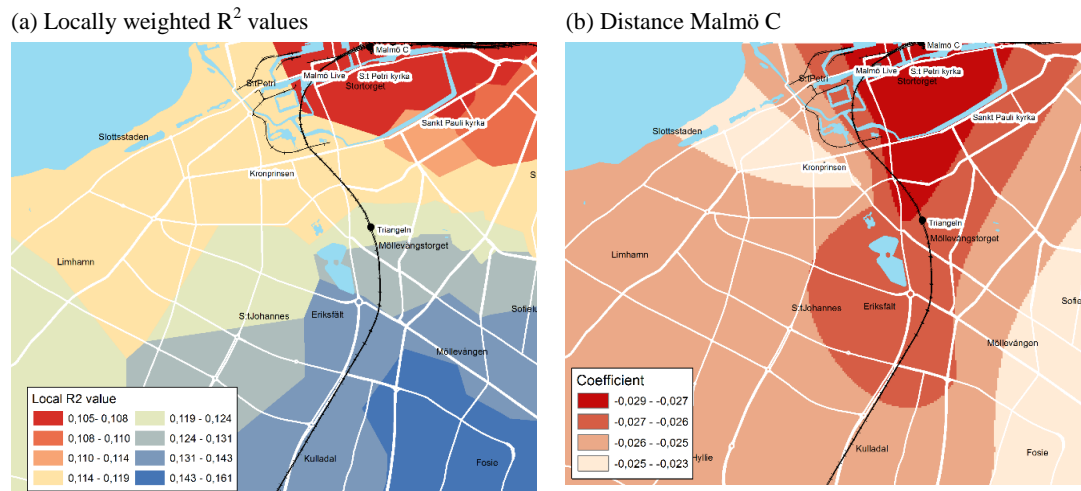


Figure 10. Maps of the locally weighted  $R^2$  values and the regression coefficient surface of *Distance to Malmö C* (km) representing the one statistical significant at the 5% level explanatory variable in Malmö obtained by the GWR model. The light blue areas in the maps represent water.

The locally weighted  $R^2$  values (a) obtained by the GWR model indicates how well the explanatory variables included in the GWR model predict car use across Malmö. With  $R^2$  values ranging from 0.10 to 0.16, the GWR model performance varies across Malmö; with the most prominent contrast between the north east and south east side of the city. The highest GWR model performance, with  $R^2$  values of 0.14-0.16, are found in the south east side of Malmö in areas around Fosie. The lowest model performance, with  $R^2$  values around 0.10, are found in the north east side of the city in areas around Malmö C and Stortorget. Accordingly, the explanatory variables included in the GWR model are better at predicting car use in the south east than in the north east side of Malmö.

The adjusted  $R^2$  value for the GWR model is 0.04, compared with the adjusted  $R^2$  value of 0.05 for the OLS model (Table 16). The GWR adjusted  $R^2$  value of 0.04 indicates that 4 % of the car use in Malmö can be explained by the explanatory variables included in the model. In contrast, the  $R^2$  value of the GWR model is 0.15, compared to the OLS  $R^2$  value of 0.11 (Table 13).

**Table 16.** GWR results in Malmö.

| No. of Observations | $R^2$  | Adjusted $R^2$ | AICc     | Residual squares |
|---------------------|--------|----------------|----------|------------------|
| 69                  | 0.1543 | 0.0427         | -33.9760 | 1.8615           |

The AICc value of the GWR model is -33.97, compared with the OLS model AICc value of -36.60, representing a decrease in the AICc value with 2.63, when moving from the OLS to the GWR model. The residual squares of the GWR model is 1.86 and represents the sum of the squared residuals in the model.

## 6 Discussion

The main hypothesis of this study was that a higher level of public transport accessibility would decrease car use and that this relationship would remain when controlling for further spatial and socio-economic variables. Furthermore, the relationship was not expected to be spatially uniform, thus a spatial regression model such as the GWR was expected to better predict car use compared to a non-spatial OLS. Consequently, the results obtained in Lund indicate some support for the hypothesis; both in respect to the impact of public transport accessibility and in respect to the model performance of the spatial GWR model. The results in Lund demonstrated that car use is negatively associated with public transport accessibility and that the spatial regression model of GWR was a better fit to the data than the non-spatial regression model of OLS. This highlights the importance of using a global regression model to assess statistically significant relationships and the use of a local regression model to examine regional variations within the data; revealing spatial patterns that were not identified by the global model. These results coincide with the suggestions by Fotheringham et al. (2002), implying that the local GWR model would generate higher model performance than a global model when modelling spatial data. On the other hand, results in Malmö do not indicate support for the main hypothesis of this study, considering that accessibility to public transport does not have a significant impact on car use in Malmö. Furthermore, the spatial GWR model was not a better fit to the data in Malmö than the non-spatial model of OLS. Consequently, the results in Lund and Malmö do not coincide. Nevertheless, in Lund, where the model performance is the highest, results imply that car use decreases with higher public transport accessibility.

### 6.1 Comparing results in Lund and Malmö

OLS results in both Lund and Malmö implied the data to possess regional variations, thus motivating the use of the spatial regression model of GWR to allow regional variations within the data (Gao and Li, 2011). Comparing the OLS and GWR models in Lund and Malmö, there are differences both in respect to model structure and performance. Additionally, results vary in respect to modelling the same data with the global regression model of OLS and the local regression model of GWR. The explanatory variables possessing a statistically significant relationship (at the 5% level) with *Car use* in Lund was *Public transport accessibility*, *Distance to Lund C* and *Always or sometimes access to a car*. Both the variables of *Public transport accessibility* and *Distance to Lund C* concerns the physical proximity to public transport stations, thus the association between these variables show support for the theory that transport is a derived need; where ultimately public transport accessibility is what matters (Benenson et al., 2011). However, in Malmö, public transport

accessibility does not have a significant impact on car use and solely the distance to Malmö C has a statistically significant relationship with car use. Furthermore, this relationship implies that car use decreases with an increase in distance to Malmö C and contradicts the results obtained in Lund regarding this relationship. In respect to model performance, it is higher in Lund where both the OLS and GWR models generated adjusted  $R^2$  values greater than 0.4, compared to adjusted  $R^2$  value of 0.04 and 0.05 in Malmö. However, these results are hard to compare due to variations in the model structure between the cities. Moreover, modelling the data in Lund with GWR compared to OLS generated higher model performance in respect to the adjusted  $R^2$  value increasing from 0.41 to 0.44. In Malmö, the adjusted  $R^2$  value decreased from 0.05 to 0.04 when moving from OLS to GWR, however, the  $R^2$  value increased from 0.11 to 0.15. These results are contradictory, thus it is difficult to assess which model that is better at predicting *Car use* in Malmö.

#### 6.1.1 Causes of varying results in Lund and Malmö

A large proportion of the variation in model performance between Lund and Malmö may be a consequence of data limitations. One such limitation was the selection of participants. This selection may be biased because participants were either recruited at public transport stations or decided to donate their data; potentially due to an interest of the transport sector. Another limitation was related to the car accessibility. The initial objective of this study was to only analyze data from car owners and thus analyze the car use of people with larger transport mode choice. However, due to a limited amount of data, all participants' data was included in the analysis, regardless of car accessibility. Consequently, 27% and 44% of the participants in Lund and Malmö, respectively, do not have access to a car. One way to account for this problem was to include the variable of *Car accessibility* in the regression analysis. However, because information about car accessibility is obtained by respondents answering the question “*do you have access to a car?*” there is room for individual interpretation, and individuals might have interpreted it in different ways. Hence, *Car accessibility* might not entirely reflect the real car accessibility of the respondents, something that may affect the model performance. Thus, in Malmö where *Car accessibility* is not statistically significant at the 5% level, the reliability of the variable is doubtful. Considering that Bastian and Börjesson. (2015) demonstrated that income and fuel prices explained 80% of the aggregated car distances per person in Sweden, it is unexpected that variables representing income in this thesis are not statistically significant neither in Lund nor in Malmö. A reason for this result might be that income is presented at an aggregated level and represents the average income for the residence region of an individual. This may level out the effect of income in this study, and perhaps results would have been different if income data represented the

actual income of each individual. Furthermore, correlations between car use and age might not have been detected due to the assignment of age which potentially produced biased results. However, the variations in model performance between Lund and Malmö may also be a consequence of an actual difference regarding the incentives behind car use in both cities, and might indicate that the incentives of car use are context dependent. Moreover, variations in model performance could be due to differences in the accessibility to public transport in both cities. Because the mean walking time to a public transport station is twice as high for the participants in Lund compared to Malmö, perhaps the accessibility to public transport in Malmö is at a level where it does not have a significant impact on car use.

## 6.2 Spatial patterns

Except for generating higher model performance, one of the key benefits of utilizing the GWR model was the computation of coefficient raster surfaces that displayed the relationship between car use and each explanatory variable exclusively (Fotheringham et al., 2002). By analysing the coefficient surfaces of the variables possessing a statistically significant relationship with car use, it is possible to detect trends and inform both local and region wide policy (Ali et al, 2007). In Lund, the highest GWR model performance was identified in the east side of the city, where *Car use* is high, participant density is high and the *Public transport accessibility* coefficients are the highest. In this area, results imply that an increase in *Public transport accessibility* with 30 seconds increase *Car use* with 2%. Thus, to decrease car use in Lund, one could increase the accessibility to public transport stations by constructing new public transport stations, particularly in the east side of the city. Moreover, the GWR model performance in Malmö is highest in the south east side of the city, in areas with both high and low *Car use*. The statistically significant relationship between car use and *Distance to Malmö C* implies that *Car use* decreases with an increase in *Distance to Malmö C*. Since this represents the one statistically significant relationship in Malmö, it is difficult to suggest measures to reduce car use in Malmö.

## 6.3 Further study

To generate higher model performance when predicting car use in Lund and Malmö, individual income data and a variable assessing the time travel time to work with car in relation to public transport can be included in future studies. Because income is recognized as a key variable when predicting car use and explained as much as 80% of the aggregated car distances per person in Sweden, individual income data would presumably generate high performance if included in the analysis (Shen et al., 2016, Bastian and Börjesson, 2015). Additionally, including two variables that represent the

time and cost it would take to travel to work with car in relation to public transport would presumably generate a more realistic model; because journeys to and from work represent the largest proportion of journeys in Scania and transport cost is an important aspect in travel mode choice (Chakrabarti., 2017, Ullberg, 2013). Furthermore, combined with the expansion of the number and data amount of participants, results would be less biased and trustworthy.

## 7 Conclusions

The aim of this thesis was to use Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) to investigate the relationship between car use and the accessibility to public transport in Lund and Malmö. The results in Lund demonstrate that car use is negatively associated with the accessibility to public transport and that the spatial regression model of GWR generated the highest model performance when predicting car use. Furthermore, the highest model performance in Lund was obtained in the east side of the city where car use is high, participant density is high and the influence of accessibility to public transport is the highest. Thus, to decrease car use in Lund, one could increase the accessibility to public transport; particularly in the east side of the city. Results in Malmö show that accessibility to public transport does not have a significant impact on car use and that the spatial regression model of GWR was not a better fit to the data than the OLS. Thus, it is difficult to suggest any measures to reduce car use in Malmö. Consequently, the results in Lund and Malmö do not coincide. However, in Lund, where the model performance is the highest, results imply that car use decreases with a higher public transport accessibility. The reasons for variations between the cities, both in model performance and the explanatory variables that generates the highest model performance, are difficult to assess and emphasizes the need for future studies of the relationship between public transport accessibility and car use in Lund and Malmö. Furthermore, it highlights the need for similar studies in multiple cities, to conclude major variables with an impact on car use.

This study represents one of the first studies to use individual GPS data along with spatial regression analysis to explore how public transport accessibility affects car use, a beneficial method in multiple ways. The study demonstrated the benefits of utilizing detailed individual GPS data, both in respect to model performance and by allowing spatial analysis of the data. Moreover, this study underlines the importance of using global non-spatial analysis to determine statistically significant relationships and the use of local spatial analysis to examine regional variations in the relationship between car use and public transport accessibility; revealing spatial patterns that were not identified by the global model. Additionally, it was beneficial to select two study areas, generating the opportunity to compare results and obtain deeper knowledge of the incentives of car use. Consequently, this study contributes to the literature on the effects of public transport accessibility on car use and on the use of local spatial analyses in accessibility studies. Such knowledge can be utilized in transport planning to reduce car usage.



## 8 References

- ALAM, M. S., HYDE, B., DUFFY, P. & MCNABOLA, A. 2017. Assessment of pathways to reduce CO2 emissions from passenger car fleets: Case study in Ireland. *Applied Energy*, 189, 283-300.
- ALI, K., PARTRIDGE, M. D. & OLFERT, M. R. 2007. Can Geographically Weighted Regressions Improve Regional Analysis and Policy Making? *International Regional Science Review*, 30, 300-329.
- ALLSTRÖM, A., FRANSSON, M., KRISTOFFERSSON, I. & GUNDEGÅRD, D. 2015. MOBILNÄTSDATA SOM INDATA TILL PROGNOSEMODELLER. SWECO, Linköpings universitet.
- BAHBOUH, K., WAGNER, J. R., MORENCY, C. & BERDIER, C. 2017. Travel demand corridors: Modelling approach and relevance in the planning process. *Journal of Transport Geography*, 58, 196-208.
- BASTIAN, A. & BÖRJESSON, M. 2015. Peak car? Drivers of the recent decline in Swedish car use. *Transport Policy*, 42, 94-102.
- BENENSON, I., MARTENS, K., ROFÉ, Y. & KWARTLER, A. 2011. Public transport versus private car GIS-based estimation of accessibility applied to the Tel Aviv metropolitan area. *The Annals of Regional Science*, 47, 499-515.
- BEST, H. & LANZENDORF, M. 2005. Division of labour and gender differences in metropolitan car use: An empirical study in Cologne, Germany. *Journal of Transport Geography*, 13, 109-121.
- BOARNET, M., HONG, A., LEE, J., WANG, X., WANG, W. 2013. The Exposition Light Rail Line Study A Before-and-After Study of the Impact of New Light Rail Transit Service. The Haynes Foundation
- BRUNSDONT, C., FOTHERINGHAM, S. & CHARLTON, M. 1998. Geographically weighted regression - Modelling spatial non-stationarity. *Journal of the Royal Statistical Society Series D: The Statistician*, 47, 431-443.
- BUEHLER, R. 2011. Determinants of transport mode choice: a comparison of Germany and the USA. *Journal of Transport Geography*, 19, 644-657.
- CHAKRABARTI, S. 2017. How can public transit get people out of their cars? An analysis of transit mode choice for commute trips in Los Angeles. *Transport Policy*, 54, 80-89.
- DU, H. & MULLEY, C. 2006. Relationship between Transport Accessibility and Land Value: Local Model Approach with Geographically Weighted Regression. *Journal of the Transportation Research Board*, 197-205.

- DURANTON, G. & TURNER, M. A. 2011. The Fundamental Law of Road Congestion: Evidence from US Cities. *The American Economic Review*, 101, 2616-2652.
- FARBER, S. & FU, L. 2017. Dynamic public transit accessibility using travel time cubes: Comparing the effects of infrastructure (dis)investments over time. *Computers, Environment and Urban Systems*, 62, 30-40.
- FOTHERINGHAM, A. S., BRUNSDON, C. & CHARLTON, M. 2002. *Geographically weighted regression. [Elektronisk resurs] : the analysis of spatially varying relationships*, Chichester : Wiley , cop. 2002.
- FOTHERINGHAM, A. S. & OSHAN, T. M. 2016. Geographically weighted regression and multicollinearity: dispelling the myth. *Journal of Geographical Systems*, 18, 303-329.
- GAO, J. & LI, S. 2011. Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using Geographically Weighted Regression. *Applied Geography*, 31, 292-302.
- GEURS, K. T. & VAN WEE, B. 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, 12, 127-140.
- GODFREY, L. G. 1996. Some results on the Glejser and Koenker tests for heteroskedasticity. *Journal of Econometrics*, 72, 275-299.
- HAGENAUER, J. & HELBICH, M. 2017. A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*, 78, 273-282.
- HAKLAY, M. 2009. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets.
- HAWAS, Y. E., HASSAN, M. N. & ABULIBDEH, A. 2016. A multi-criteria approach of assessing public transport accessibility at a strategic level. *Journal of Transport Geography*, 57, 19-34.
- HONDA, Y. 1988. A size correction to the lagrange multiplier test for heteroskedasticity *Journal of Econometrics* 38, 375-386.
- HONGBO, D. & MULLEY, C. 2006. Relationship between transport accessibility and land value: Local model approach with geographically weighted regression. *Transportation Research Record*.
- IPCC 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. . 151 pp.

- KALOGIROU, S. & HATZICHRISTOS, T. 2007. A Spatial Modelling Framework for Income Estimation. *Spatial Economic Analysis*, 2, 297-316.
- KAREN, K. K. 2008. Encyclopedia of Geographic Information Science.
- LI, C., LI, F., WU, Z., CH, J. 2017. Exploring spatially varying and scale-dependent relationships between soil contamination and landscape patterns using geographically weighted regression. *Applied Geography* (82) 101-114.
- LIN, T., XIA, J., ROBINSON, T. P., GOULIAS, K. G., CHURCH, R. L., OLARU, D., TAPIN, J. & HAN, R. 2014. Spatial analysis of access to and accessibility surrounding train stations: a case study of accessibility for the elderly in Perth, Western Australia. *Journal of Transport Geography*, 39, 111-120.
- MITRA L. DEVKOTA, G. H. A. R. C. 2014. Effect of Sample Size on the Performance of Ordinary Least Squares and Geographically Weighted Regression. *British Journal of Mathematics & Computer Science*, ISSN: 2231-0851, Vol.: 4, Issue.: 1 (01-15 January), 4.
- MU, L. 2009. Thiessen Polygon A2 - Kitchin, Rob. In: THRIFT, N. (ed.) *International Encyclopedia of Human Geography*. Oxford: Elsevier.
- MURRAY, A. T., DAVIS, R., STIMSON, R. J. & FERREIRA, L. 1998. Public Transportation Access. *Transportation Research Part D: Transport and Environment*, 3, 319-328.
- OHTANI, K. 2000. Bootstrapping R2 and adjusted R2 in regression analysis. *Economic Modelling*, 17, 473-483.
- PACHI, A. J., TIANJIAN 2005. Frequency and velocity of people walking. *The Structural Engineer*, 83.
- PÁEZ, A., FARBER, S. & WHEELER, D. 2011. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A*, 43, 2992-3010.
- PATZ, J. A., CAMPBELL-LENDRUM, D., HOLLOWAY, T. & FOLEY, J. A. 2005. Impact of regional climate change on human health. *Nature*, 438, 310-317.
- PYE, S. & DALY, H. 2015. Modelling sustainable urban travel in a whole systems energy model. *Applied Energy*, 159, 97-107.
- RASMUSSEN, T. K., INGVARSDON, J. B., HALLDÓRSDÓTTIR, K. & NIELSEN, O. A. 2015. Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area. *Computers, Environment and Urban Systems*, 54, 301-313.
- ROGERSON, P. A. 2001. *Some Spatial Aspects of Regression Analysis In: Statistical Methods for Geography*

- ROJAS-RUEDA, D., DE NAZELLE, A., TEIXIDÓ, O. & NIEUWENHUIJSEN, M. J. 2012. Replacing car trips by increasing bike and public transport in the greater Barcelona metropolitan area: A health impact assessment study. *Environment International*, 49, 100-109.
- ROSENHEIN, L., SCOTT, L., and PRATT, M. 2011. Exploratory Regression: A Tool for Modeling Complex Phenomena. Retrieved May 5, 2015 from [www.esri.com/news/arcuser/0111/files/exploratory.pdf](http://www.esri.com/news/arcuser/0111/files/exploratory.pdf)
- ROTARIS, L. & DANIELIS, R. 2015. Commuting to college: The effectiveness and social efficiency of transportation demand management policies. *Transport Policy*, 44, 158-168.
- SAGHAPOUR, T., MORIDPOUR, S. & THOMPSON, R. G. 2016. Public transport accessibility in metropolitan areas: A new approach incorporating population density. *Journal of Transport Geography*, 54, 273-285.
- SALAS-OLMEDO, M. H., WANG, Y. & ALONSO, A. 2017. Assessing accessibility with local coefficients for the LUTI model MARS. *Computers, Environment and Urban Systems*, 64, 194-203.
- SCB, 2009. Statistics Sweden's Market Profiles
- SCB 2014. Befolkningens utbildning och sysselsättning 2014. In: SCB (ed.).
- SCB 2015. Tätorter 2015 Befolkning och arealer.
- SCHEINER, J. & HOLZ-RAU, C. 2012. Gendered travel mode choice: a focus on car deficient households. *Journal of Transport Geography*, 24, 250-261.
- SHEN, L. & STOPHER, P. R. 2014. Review of GPS Travel Survey and GPS Data-Processing Methods. *Transport Reviews*, 34, 316-334.
- SHEN, Q., CHEN, P. & PAN, H. 2016. Factors affecting car ownership and mode choice in rail transit-supported suburbs of a large Chinese city. *Transportation Research Part A: Policy and Practice*, 94, 31-44.
- SHENG, J., HAN, X. & ZHOU, H. 2017. Spatially varying patterns of afforestation/reforestation and socio-economic factors in China: a geographically weighted regression approach. *Journal of Cleaner Production*, 153, 362-371.
- STEG, L. 2005. Car use: lust and must. Instrumental, symbolic and affective motives for car use. *Transportation Research Part A: Policy and Practice*, 39, 147-162.
- TJUR, T. 2009. Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination. *The American Statistician*, 63, 366-372.

- TRAFIKANALYS 2015. Peak car i sikte? Statistik och analys över Sveriges personbilsflotta och dess användning. *In: MINISTRY, I. (ed.)*.
- TRAFIKVERKET 2015. Prognos för personresor 2030 Trafikverkets basprognos 2015. *In: MINISTRY, I. (ed.) 1 ed.: Trafikverket*.
- ULLBERG, C. W. O. M. 2013. Resvaneundersökning 2013. 49.
- WIBOWO, S. S. & OLSZEWSKI, P. 2005. MODELING WALKING ACCESSIBILITY TO PUBLIC TRANSPORT TERMINALS: CASE STUDY OF SINGAPORE MASS RAPID TRANSIT. *Journal of the Eastern Asia Society for Transportation Studies*, 6, 147-156.
- VU, D. H., MUTTAQI, K. M. & AGALGAONKAR, A. P. 2015. A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *Applied Energy*, 140, 385-394.
- XIAO, G., JUAN, Z. & ZHANG, C. 2015. Travel mode detection based on GPS track data and Bayesian networks. *Computers, Environment and Urban Systems*, 54, 14-22.
- XU, B. & LIN, B. 2017. Factors affecting CO2 emissions in China's agriculture sector: Evidence from geographically weighted regression model. *Energy Policy*, 104, 404-414.
- ZHANG, Y., ZHENG, S., SUN, C. & WANG, R. 2017. Does subway proximity discourage automobility? Evidence from Beijing. *Transportation Research Part D: Transport and Environment*, 52, Part B, 506-517.

APPENDIX 1.

## App questionnaire

**Are you a man or woman?**

Man      Woman      Other gender identity      Prefer not to answer

**In what year were you born?**

**What is your primary occupation?**

Employed      Student      Other

**Do you have access to a car?**

Always      Sometimes      Never

## **Institutionen för naturgeografi och ekosystemvetenskap, Lunds Universitet.**

Studentexamensarbete (seminarieuppsatser). Uppsatserna finns tillgängliga på institutionens geobibliotek, Sölvegatan 12, 223 62 LUND. Serien startade 1985. Hela listan och själva uppsatserna är även tillgängliga på LUP student papers (<https://lup.lub.lu.se/student-papers/search/>) och via Geobiblioteket ([www.geobib.lu.se](http://www.geobib.lu.se))

The student thesis reports are available at the Geo-Library, Department of Physical Geography and Ecosystem Science, University of Lund, Sölvegatan 12, S-223 62 Lund, Sweden. Report series started 1985. The complete list and electronic versions are also electronic available at the LUP student papers (<https://lup.lub.lu.se/student-papers/search/>) and through the Geo-library ([www.geobib.lu.se](http://www.geobib.lu.se))

- 400 Sofia Sjögren (2016) Effective methods for prediction and visualization of contaminated soil volumes in 3D with GIS
- 401 Jayan Wijesingha (2016) Geometric quality assessment of multi-rotor unmanned aerial vehicle-borne remote sensing products for precision agriculture
- 402 Jenny Ahlstrand (2016) Effects of altered precipitation regimes on bryophyte carbon dynamics in a Peruvian tropical montane cloud forest
- 403 Peter Markus (2016) Design and development of a prototype mobile geographical information system for real-time collection and storage of traffic accident data
- 404 Christos Bountzouklis (2016) Monitoring of Santorini (Greece) volcano during post-unrest period (2014-2016) with interferometric time series of Sentinel-1A
- 405 Gea Hallen (2016) Porous asphalt as a method for reducing urban storm water runoff in Lund, Sweden
- 406 Marcus Rudolf (2016) Spatiotemporal reconstructions of black carbon, organic matter and heavy metals in coastal records of south-west Sweden
- 407 Sophie Rudbäck (2016) The spatial growth pattern and directional properties of *Dryas octopetala* on Spitsbergen, Svalbard
- 408 Julia Schütt (2017) Assessment of forcing mechanisms on net community production and dissolved inorganic carbon dynamics in the Southern Ocean using glider data
- 409 Abdalla Eltayeb A. Mohamed (2016) Mapping tree canopy cover in the semi-arid Sahel using satellite remote sensing and Google Earth imagery
- 410 Ying Zhou (2016) The link between secondary organic aerosol and monoterpenes at a boreal forest site
- 411 Matthew Corney (2016) Preparation and analysis of crowdsourced GPS bicycling data: a study of Skåne, Sweden
- 412 Louise Hannon Bradshaw (2017) Sweden, forests & wind storms: Developing a

model to predict storm damage to forests in Kronoberg county

- 413 Joel D. White (2017) Shifts within the carbon cycle in response to the absence of keystone herbivore *Ovibos moschatus* in a high arctic mire
- 414 Kristofer Karlsson (2017) Greenhouse gas flux at a temperate peatland: a comparison of the eddy covariance method and the flux-gradient method
- 415 Md. Monirul Islam (2017) Tracing mangrove forest dynamics of Bangladesh using historical Landsat data
- 416 Bos Brendan Bos (2017) The effects of tropical cyclones on the carbon cycle
- 417 Martynas Cerniauskas (2017) Estimating wildfire-attributed boreal forest burn in Central and Eastern Siberia during summer of 2016
- 418 Caroline Hall (2017) The mass balance and equilibrium line altitude trends of glaciers in northern Sweden
- 419 Clara Kjällman (2017) Changing landscapes: Wetlands in the Swedish municipality Helsingborg 1820-2016
- 420 Raluca Munteanu (2017) The effects of changing temperature and precipitation rates on free-living soil Nematoda in Norway.
- 421 Neija Maegaard Elvekjær (2017) Assessing Land degradation in global drylands and possible linkages to socio-economic inequality
- 422 Petra Oberhollenzer, (2017) Reforestation of Alpine Grasslands in South Tyrol: Assessing spatial changes based on LANDSAT data 1986-2016
- 423 Femke, Pijcke (2017) Change of water surface area in northern Sweden
- 424 Alexandra Pongracz (2017) Modelling global Gross Primary Production using the correlation between key leaf traits
- 425 Marie Skogseid (2017) Climate Change in Kenya - A review of literature and evaluation of temperature and precipitation data
- 426 Ida Pettersson (2017) Ekologisk kompensation och habitatbanker i kommunalt planarbete
- 427 Denice Adlerklint (2017) Climate Change Adaptation Strategies for Urban Stormwater Management – A comparative study of municipalities in Scania
- 428 Johanna Andersson (2017) Using geographically weighted regression (GWR) to explore spatial variations in the relationship between public transport accessibility and car use : a case study in Lund and Malmö, Sweden
- 429 Elisabeth Farrington (2017) Investigating the spatial patterns and climate dependency of Tick-Borne Encephalitis in Sweden
- 430 David Mårtensson (2017) Modeling habitats for vascular plants using climate factors and scenarios - Decreasing presence probability for red listed plants in



## Scania

- 431 Maja Jensen (2017) Hydrology and surface water chemistry in a small forested catchment : which factors influence surface water acidity?
- 432 Iris Behrens (2017) Watershed delineation for runoff estimations to culverts in the Swedish road network : a comparison between two GIS based hydrological modelling methods and a manually delineated watershed
- 433 Jenny Hansson (2017) Identifying large-scale land acquisitions and their agro-ecological consequences : a remote sensing based study in Ghana