

Government vs Market in Sustainable Residential Development?

Microdata analysis of car travel, CO₂ emission and residence location



Xiaoyun Zhao

To my beloved parents,

*Sure, you are no genius like Albert Einstein who said:
“In the middle of difficulty lies opportunity”.*

*You are simply a father and a mother who have shown me:
In the middle of difficulty lies your love, always.*

献给我挚爱的父母，

岁月染白了您们的双鬓， 您们的爱持久浓烈。



HÖGSKOLAN
DALARNA

Government vs Market in Sustainable Residential Development?

Microdata analysis of car travel, CO₂ emission and residence location

Xiaoyun Zhao

Microdata Analysis
School of Technology and Business Studies
Dalarna University
Borlänge, Sweden
April, 2017

“The issue is always the same:
The government or the market.”
—Ludwig Von Mises

©Xiaoyun Zhao, 2017
Dalarna Doctoral Dissertations No.6
ISBN: 978-91-85941-84-1
Electronic version available at: <http://du.diva-portal.org/>
Illustration on the cover page by Ruimin Ma
Printed by Printeliten, Falun
Dalarna University
SE-79188 Falun, Sweden

Abstract

Increasing car usage and travel demands between residential locations and destinations in order to fulfill the various needs of residents is a primary cause of CO₂ emissions. To win the battle against climate change, a better understanding of the question relating to which urban residential form may most effectively mitigate the CO₂ emissions is the key pathway.

This dissertation is concerned with the above problem and it mainly considers three objectives in providing insights on answering the question. The first objective is to comprehensively and microscopically understand intra-urban car travel behavior. The second objective is to estimate the induced CO₂ emissions from daily intra-urban car travel and to ex-ante evaluate residential plans. The third objective is to assess whether the governmental sustainable residential development objective is aligned with the objectives of the estate market actors. To explore the research questions related to the objectives, a microdata analysis process (data collection, data assessment and transformation, data storage, data analysis and decision-making) is applied and is found essential in gaining access to key variables in exploring the answer of a preferable urban form. The dissertation offers many new solutions to various technical aspects through a microdata analysis process.

The primary contribution of this dissertation is that it outlines an operational model that comprehensively integrates the investors' investment strategy, the residents' choice behavior, and the governmental sustainability objective in the interest of making an ex-ante assessment of residential plans. This ex-ante assessment provides decision-support in sustainable residential development at foremost local level.

The first finding from the implementation of the model on the case study is that the market actors' objectives are, in general, aligned with the government's sustainable residential development objective. The second finding indicates that re-shaping the urban form into a compact city is preferable in mitigating CO₂ emissions, in spite of the fact that the case city is of a polycentric urban form. These findings provide support for those advocating the compact city as the ideal for sustainable residential development, and also provide foresight on settling the answer to the preferred re-shaping of urban forms in climate change.

Acknowledgments

“The best journeys answer questions that in the beginning you didn't even think to ask.”-
Jeff Johnson

I started to understand why it is true that “life is what happens”, when I first began this unexpected but meant-to-be journey of pursuing a Ph.D. degree. Some beautiful paths cannot be discovered without getting lost, and I feel so lucky that I found a group of people who inspired, guided, helped and supported me. Yes, it is they who have accompanied me on this long journey.

Coming towards the end of this journey, I am sure my supervisors, Johan Håkansson and Kenneth Carling are watching and cheering me on across the finishing line, just as they have always supported me during my doctoral studies. Johan, I know that I will have your encouraging and warm smile, as always. I have never told you how grateful I was every time I went to you with usual or unusual requests; you always supported me with your understanding and help. Your encouragement and faith in me always rescued me from the hole of doubt, so I could continue going forward. I have tried to count, but it still remains a mystery, how many cups of coffee you drink every day.

Kenneth, I know that I will still receive your constructive and insightful comments on how I could do better. I will always remember the first lesson from you about the two scientific principles of research: replicability and Occam's razor. Your wisdom and the fact that you always had an answer to my “dead end” questions guided me back in the right direction, so I could keep digging deeper. I am not sure if it is common that a professor could not recognize his own hand writing, but I am sure that it has been, and is still a fun challenge for me. You two are a perfect team to help me, not only in cranking out this dissertation, but also in showing me a way of thinking and solving problems. Working under your supervision is a valuable and unique experience that I would never trade for anything.

This journey could never have been fulfilled without all the administrative and examining work. Thank you, Anders Forsman; from the first day of my doctoral studies you have always been considerate and patient. I know I do not need to worry when I have an administrative problem, because I just tell myself: “*go and talk to Anders, he can, and will, fix it*”. Thank you, Renée Flacking. I am so happy that I have had you as the Ph.D. students' director from my third year; I am also so sad that I can only have you for this short space of time. All the Ph.D. meetings you have arranged were so interesting and broadened what I had thought of as a Ph.D. student. The one-to-one talks let me know that, “*yes, this is the right person I can talk to, and she will make a difference*”, and you sure did. Thank you, Siril Yella, for being my examiner. I remember that you told me: “*Nobody gets a Ph.D. because of being good at remembering, but because of being good at understanding*”. I always use this as a self-check to deepen my understanding.

This journey would have been more difficult without working alongside all the bright colleagues with such a high level of intellectual and people skills. Lars Rönnegård, you showed me how gentle, knowledgeable and caring a person can be. Thank you for organizing the summer school in 2014, and all the helpful information you shared with me. Moudud Alam, you showed me

how smart and hardworking a person can be. Thank you for helping me collect data, with your excellent driving skills. Thank you for letting me supervise a master's thesis with you, I learned a lot. Ola Nääs, you showed me how dedicated a person can be in continuously improving one's teaching skills while already being very good at it. Thank you for sharing a course with me, it inspired me a lot. Daniel Wikström, you showed me how sharp and insightful a person can be in research, while being humorous and passionate in life. Thank you for commenting on my first paper, it helped a lot. Mengjie Han, thank you for all your advice and help; by always being calm in dealing with any situation, you showed me that panic never solves a problem but only makes it worse. Pascal Rebreyend, as my co-author, your perspective showed me how different it can be between subjects, and this enriched my view. Irene Gilsenan Nordin, thank you for all your excellent proof reading services. Hasan Fleyeh, thank you for fixing all the WP numbers and treating me to pizza. Albina Pashkevich, thank you for all the compliments, they brightened my day, every time.

This journey would have had many more detours without your help, Xiangli Meng. All your tips on how to work as a Ph.D. student, and how to find the small fun things amidst the sometimes tedious "monkey job" have guided me greatly. I never got to know how you can get ahead and know all the news, while always being fully devoted to work. I know you will be successful in what you are doing now, as life always rewards a hardworking visionary.

This journey would have been tedious without the laughs and cheeriness you brought me, Catia Cialani. Because of you, I laugh a little harder and think a little more. Because of you, I have realized that *we* are more powerful than *me*. Because of you, I have learnt that passion is more important than rules. Thank you for always being cheerful and understanding. Thank you for being not just a great colleague but a better friend. And, by the way, get the lasagna ready for me, I am coming!

This journey would have been lonely without the company of other Ph.D. fellows. Kristin Svensson and Diala Jomaa, you are both doctors now. Thank you for showing me that enthusiasm and persistence are necessary if I want to reach that goal. Yujiao Li, it is really nice to share an office with you, thank you for your warm messages. Asif M Huq, thank you for your encouragement and the delicious dinner. Åsa Grek, Charlie Lindgren, Ilias Thomas, and Somayeh Aghanavesi, thank you all for sharing your innovative ideas during all the chats and coffee breaks.

This journey would have been unbearable without all the support and joy I received from my dear friends. Amy, you are my best friend and a sister to me. You are the one who knows me as I am, understands where I have been, accepts what I have become, and still gently helps me to grow. You not only bring your friendship to me, but also other good friends, making me feel rich in friendship. Xina, you showed me that being caring and selfless makes a person gentle. Lingtao, you showed me that being strong and independent makes a person fearless. Yuwen, you showed me that being romantic and true makes a person beautiful. Yongling, you showed me that being optimistic and active makes a person brave. Liang you showed me that generous giving makes a person happy. Bin, you showed me that sincere belief makes a person noble. Lei, Jun, Juan and Yang, you showed me that true friendship never fades away over distance or time. You all have a big and beautiful heart,

Acknowledgements

and I wish every day that you experience happiness. Special thanks to Ruimin for drawing the cover page, you are a talented girl, and a great future is waiting for you.

This journey would have been impossible without all the invaluable love from my family. Father and mother, you have dedicated all your love and support to your two daughters, I love you. Sister, thank you for taking care of our parents when I am this far away from China. Stefan, you have patiently endured many, many, long hours alone while I worked on my dissertation. Thank you for always being there, with your love, patience, help and unwavering belief in me, even when I was irritable and depressed. Thank you for creating so many happy and beautiful memories with me throughout this journey. Thank you for becoming a better you, and for helping me become a better me. I love you. Many thanks to your great family; they have always made me feel at home.

This journey is so unique and valuable that I feel I still have so many to thank. If your name appears here, once more, thank you! But I also know how it is, if you opened this dissertation, flipped to the acknowledgements, and find that once again, the author has thanked many people but maybe not you. Believe me, this is not on purpose, I truly thank you:

In my childhood,
the journey was a precious one:
full of wonders and surprises.
In my adulthood,
the journey was a generous one:
full of friendships and love.
Every journey has an end,
Every end is a new beginning.
Now, I am ready.....

Xiaoyun Zhao

March, 2017

List of papers

Papers included

- I. Zhao, X., Carling, K., & Håkansson, J. (2014). An Evaluation of the Reliability of GPS-Based Transportation Data. *Submitted*.
- II. Zhao, X., Rebreyend, P., & Håkansson, J. (2015). How does the complexity of a road network affect optimal facility locations? *Submitted*.
- III. Zhao, X. (2015). On processing GPS tracking data of spatio-temporal car movements: a case study. *Journal of Location Based Services*, 9(4), 235-253.
- IV. Zhao, X., Carling, K., & Håkansson, J. (2016). Residential planning, driver mobility and CO2 emission. *Revised for publication*.
- V. Zhao, X., Carling, K., & Håkansson, J. (2017). On assessing governmental sustainable residential planning and its alignment with residents' and estate investors' objectives. Working papers in transport, tourism, information technology and microdata analysis (2017:01), ISSN: 1650-5581, Dalarna University, Sweden.

Papers not included

- I. Eriksson, P. E., Swenberg, T., Eriksson, Y., Zhao, X. (2016). What do screen-gazing ratios indicate about student engagement in the visual instructional learning situation? *Submitted*.
- II. Eriksson, P. E., Swenberg, T., Eriksson, Y., Zhao, X. (2017). When learners' gazes go off-screen: Exploring learning and the visual attention spectrum using three different but visual information equivalent assembly instructions. *Submitted*.

Contents

Part I	11
1. Introduction	13
1.1 Climate change mitigation and the importance of residential development.....	13
1.2 The need for ex-ante impact assessment methods for sustainable residential development.....	14
2. Research aim and objectives	16
3. The microdata analysis process that governed the development of this dissertation	17
4. Summary of the papers	19
5. Concluding discussion	24
References	25
Part II	29
Paper I: An Evaluation of the Reliability of GPS-Based Transportation Data	33
Paper II: How does the complexity of a road network affect optimal facility locations?	55
Paper III: On processing GPS tracking data of spatio-temporal car movements: a case study	75
Paper IV: Residential planning, driver mobility and CO₂ emission	97
Paper V: On assessing governmental sustainable residential planning and its alignment with residents' and estate investors' objectives	125

Part I

1 Introduction

1.1 Climate change mitigation and the importance of residential development

Sustainable development is the pathway to winning the battle against climate change. Sustainable residential development is at the forefront of this battle because housing and transportation mobility are the two main sources for the generation of greenhouse gas (GHG), especially carbon dioxide (CO₂) emissions (Dhakal, 2009). Particularly, sustainable residential development in urban areas is crucial, as the world's cities make up 2% of the world's total land mass but generate 70% of GHG emissions (Habitat, 2011). The International Energy Agency (2009) projected that transport induced CO₂ emissions globally will grow by nearly 50% to 2030, and by more than 80% by 2050. The primary cause of this projected growth is an increasing travel demand and car usage, particularly so for intra-urban car travelling between residential locations and destinations with facilities that may fulfill various needs of the residents.

Urban form, travel behavior, and vehicle technology are three factors that affect CO₂ emissions induced by car travels (Wright and Fulton, 2005; Hankey and Marshall, 2010). Advances in vehicle and fuel technology, such as electric vehicles and improved fuel efficiency, may mitigate CO₂ emissions, but these mitigations would be offset by the increase in car usage and traffic congestions (Chapman, 2007). There is interdependency between urban form and travel behavior, and thus travel-induced CO₂ emissions (Grazi and Van den Bergh, 2008; Brownstone and Golob, 2009). Therefore, the potential for re-shaping urban form in the interest of mitigating CO₂ emissions has attracted the attention of both researchers and practitioners.

The re-shaping of urban forms may be categorized as follows: *compact city*, *urban sprawl*, and *urban polycenter*, and it is still debated which of these forms is preferable for mitigating CO₂ emissions due to intra-urban car travel (Grazi et al., 2008; Qin and Han, 2013). In the compact city paradigm, car travelling is assumed to be reduced by the proximity of residents and facilities, and the external costs of car travelling, such as CO₂ emissions, are therefore assumed to be modest. In the urban polycenter and urban sprawl paradigms, the monocenter proposition is regarded as unrealistic in the light of today's prevailing urban forms, and it is believed that a compact form would generate substantial congestions. Furthermore, it is purported that residents' heterogeneous and dynamic preferences in choosing residential location and means of transportation must be considered (Mokhtarian and Cao, 2008; Chatman, 2009).

Arguably, the assessment of the sustainability of urban forms requires comprehensive consideration of economic, social and environmental factors, and due to this complication, no consensus exists on which is the preferable urban form (Handy, 2005). Although supply and demand on the (real) estate market, and thus some economic and social factors, are considered, the environmental factor is in focus in this dissertation, as it is evident that a *valid* assessment with regard to all factors simultaneously provides an insurmountable challenge.

The usefulness of determining the relationship between car travel and its induced costs in an urban form under re-shaping is that it can provide guidelines for urban sustainable development. There are three strategies commonly considered for mitigating (or internalizing) the external cost of CO₂ emissions generated by intra-urban car travel:

- 1) Pricing instruments, such as carbon tax or variable user charge (Hensher, 2008).
- 2) Land use and transportation use regulations (Glaeser and Kahn, 2010).
- 3) Passive strategy by foresighting land-use change and its impacts (Verburg et al., 2006).

The first two strategies are difficult to implement as the choice of price instrument and regulations may lead to adverse effects while being off the target, and there are also administrative costs of implementation. The third strategy contributes to more informed policy decisions, as the success of tomorrow's situation depends on decisions made today (Verburg et al., 2006). This is less costly and more flexible, which is important, as the urban form is hard to reshape once it has been developed, and it has lock-in effects on human activities and long-term environmental outcomes (Lefèvre, 2009). In this dissertation, the focus is on the third strategy.

1.2 The need for ex-ante impact assessment methods for sustainable residential development

In most urban planning studies, it is commonly assumed that the residential locations are already fixed, while the consequences of locating various types of facilities have been studied (Dieleman et al., 2002; Krizek, 2003; Norman et al., 2006; Perkins et al., 2009; Lindsey et al., 2011; Carling et al., 2013; Hong and Goodchild 2014). These studies focus on current urban form and its influence on car induced CO₂ emissions. Some of the studies are ex-post evaluations providing information on car travel and CO₂ emissions, but considering the huge lock-in effect it is hard to benefit from this information afterwards. Bourguignon and Ferreira (2003) pointed out that while the ex-post analysis examines the current situation after urban plans have been implemented, it is important to evaluate the potential impact of counterfactual urban plans. It would be helpful for urban planners to have

some estimates of how much each alternative plan would cost, and which aspect would be affected, and by how much, under each of the alternatives.

The European Union Sustainable Development Strategy explicitly reinforced the importance of using a high-quality, ex-ante impact assessment as a tool for improving policy making (Council of the European Union 2006). The strategy states that all major policy decisions should be based on proposals that have undergone an impact assessment, whereby equal consideration should be given to the social, environmental, and economic dimensions of sustainable development. Helming et al. (2011) argued, however, that this requirement for the consideration of all relevant aspects hinders the ex-ante impact assessment due to the complexity, lack of sensitivity to the policy context, and the insufficient integration of the indicator assessment into a sustainability valuation approach. Therefore, to make the ex-ante impact assessment practical it is necessary to have a focus on some aspects. For example, a focus on an ex-ante evaluation of car travel and urban residential planning to enable the quantification of the external cost of CO₂ emissions, which would provide a sub-vision on the residential development and sustainable development achieved more efficiently, and at less cost.

Land-use transport interaction (LUTI) models are attractive for ex-ante evaluation as they may predict the impact of new infrastructure on the transport system (Acheampong and Silva, 2015). However, Johansen et al. (2015) argued that a LUTI model, on the one hand, would be expensive and labor-intensive to develop, and, on the other hand, either too complex to understand, or too simplistic to capture all the urban processes. In line with Helming's et al. (2011) proposition, Kii et al. (2016) pointed out that more flexible and simplified modeling on a local scale that would extract the essence of urban activities is needed to meet the goal of mitigating emissions in urban planning. At this point, it is necessary to recall what Godschalk (2004) has pointed out: "Land use planning faces both an opportunity and a threat. On the one hand, it is widely counted on and expected to deliver both sustainable development and livable communities. On the other hand, it must cope with serious conflicts in the values related to these two beguiling visions."

There are three key actors in residential development, namely, the estate investor and the resident on the micro-level of the market, and (local) government on the macro-level, each with their own objective. In LUTI models, factors at micro-level are mainly in focus, while the macro-level is not considered. As a remedy for this shortcoming, Parker et al. (2003) suggested that MAS (Multi-agent system) models of land-use are well suited for representing complex spatial interactions under heterogeneous and dynamic conditions, recalling that residents and investors often overlook the macroscopic policy constraints exerted by the government in the land-use decision-making process (Bone et al., 2011). However, most studies in the spirit of MAS regard government at the macro-

level, with exclusive power to set the general land-use development, and therefore these studies do not, explicitly, consider the microscopic choice behavior of residents and investors and its influence on the governmental policy (Huang et al., 2014; Zhang et al., 2015).

In the spirit of LUTI and MAS models, this dissertation outlines a model that considers the investors' investment strategy and the residents' choice behavior integrated with the governmental sustainability objectives to be used for ex-ante impact assessment in sustainable, residential development. In addition to outlining the model, its practical implementation is given and demonstrated in a case city.

2 Research aim and objectives

The aim of this dissertation is twofold. The first aim relates to the debate on the preferred urban form (see Section 1.1) for environmental sustainability. The second aim is to enhance the ex-ante impact assessment methodology for sustainable residential development. The dissertation provides insights into the spatial variability of the CO₂ emissions at micro level. It adds new knowledge to existing transport emission and urban planning research. It contributes to the innovations of ex-ante evaluation of urban form, and it generates empirical evidence to help the macro-level policy gain foresight to achieve sustainable residential development. There are three specific objectives to be considered in this dissertation:

Objective One – To comprehensively and microscopically understand intra-urban car travel behavior. The main research questions to be explored for this first objective are: *How can car movements be measured?*, and *How do residents move by car in a city?* To address this first objective, the tasks are: 1) Investigate the reliability of the tools applied in collecting data of car movement. 2) Investigate the complexity of the road network and identify the destinations (i.e. hotspots) of the car travels. 3) Clean and process the data to extract the pattern that depicts residents' car movements.

Objective Two – To estimate the induced CO₂ emissions from daily intra-urban car travel and to ex-ante evaluate residential plans. The research questions to be explored for the second objective are: *How can the processed data of car movements from some residents be applied to estimate a network of induced CO₂ emissions for the city?*, *What could be the change, compared to the current situation in CO₂ emissions in the city, if the residential plans were adopted?*, and *Which subset of residential plans would reduce the CO₂ emissions the most if adopted?* To address this objective the tasks are: 1) Make use of the processed data to identify the hotspots and derive the network of induced CO₂ emissions (NoIEs). 2) Using the NoIEs, compare the CO₂ emissions in the current

situation with those of the residential plans, and estimate the expected change in CO₂ emissions due to the adoption of some, or all, of the residential plans.

Objective Three – To assess whether the governmental sustainable residential development objective is aligned with the objectives of the estate market actors. The research questions to be explored for this final objective are as follows: *How to formulize the objectives of the government, the resident and the investor*, and *How to integrate their objectives into a comprehensive model*. To address this final objective the task is to: 1) Formulize each actor's unique objective in terms of utility, and integrate their objectives into a comprehensive model, 2) Work out an approach for operationalizing the comprehensive model for use at local decision levels, and 3) Implement the model at a local level to make an impact assessment of competing residential development plans.

3 The microdata analysis process that governed the development of this dissertation

To address the aims and the objectives, access to some key variables is required: (1) the intra-urban car movements, (2) the road network, (3) the induced CO₂ emissions, (4) (planned) residential locations and hotspots, and (5) utility estimates of government, resident and estate investor. The process from accessing microscopic data to decision-making contains the following components, data collection, data assessment and transformations, data storage, analysis and decision making, which make up the microdata analysis process. Figure 1 depicts this process which has governed the development of the five papers included in the dissertation, where each paper emphasizes one or two components. Below, I clarify which component(s) has been emphasized in each of the papers.

Paper 1 assesses the reliability of GPS tracking data of car travel by conducting a field experiment with substantial data collection. The conclusions from this first paper serve as a foundation for the data transformation process to derive some key variables of intra-urban car mobility. Paper 2 makes use of a disorganized digital representation of the (gigantic) Swedish road network and this digital representation is transformed into a road network that may be used to address the aims of this dissertation. The work and the findings of this paper provide a guideline for finding the optimal location of hotspots that also serve as another foundation for the identification of hotspots. In Paper 3, an approach for processing GPS tracking data is outlined and applied. The focus of the paper is on data assessment and transformation, and the outcome of this paper is that key variables of intra-urban car movements are obtained. However, this paper also applies visualization techniques which relate to analysis and reporting. In combination, the first three papers achieve the

first objective of this dissertation and provide a platform for the remaining two papers that emphasize decision-making in residential development.

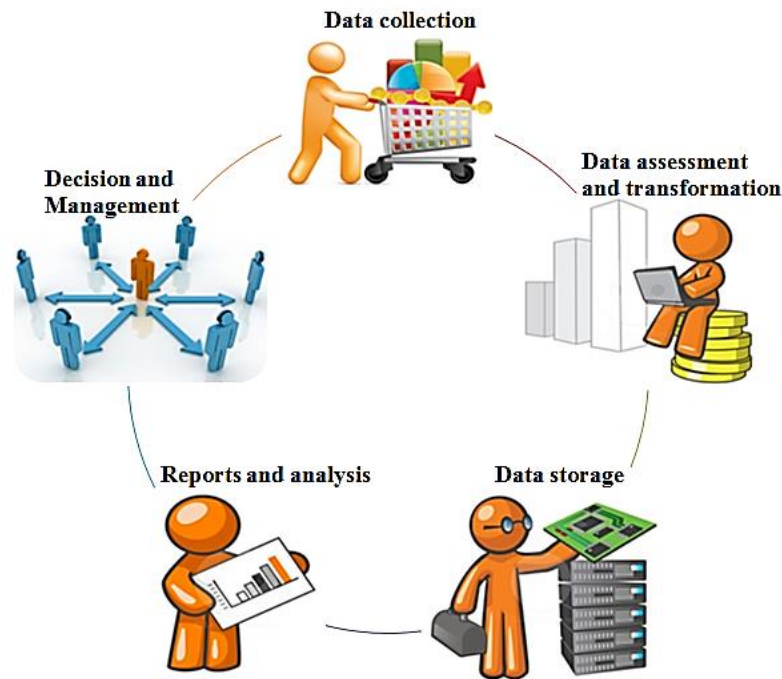


Figure 1: The microdata analysis process

For decision-making in developing residential locations in Borlänge, Paper 4 outlines and implements an ex-ante evaluation of existing residential plans on driver mobility and induced CO₂ emissions. Furthermore, the third and the fourth key variables are derived in this paper, which lays the ground for the local government's management of residential development. This paper realizes the second objective. The virtue of the fourth paper is, however, that it presumes the government to be insensitive to market forces, and may therefore misguide decision-makers by overlooking the possibly conflicting interests of other market actors.

Paper 5 therefore provides an integration of the macro-actor (government) and the micro-actors (residents and investors in the estate market), and provides estimates of each actor's utility (the fifth key variable). The work of the first four papers enables the implementation of the methodology, making the fifth paper a contribution to the decision and management by foresighting sustainable, livable and profitable residential development. This paper fulfills the third objective. The section below discusses the individual papers and how they relate to the components of the microdata analysis process in more detail.

4 Summary of the papers

Paper I: An Evaluation of the Reliability of GPS-Based Transportation Data

Global Positioning System (GPS) tracking technologies have been extensively applied to transportation studies. Based on a review of GPS-based travel studies going back to the late 1990s, Shen and Stopher (2014) stated that compared to self-reported travel diaries, GPS tracking data is more reliable and cheaper than self-reported diaries, though it still has uncertainty issues leading to the general question that Leduc (2008) asks: “How good is the quality of the traffic data?”

Nowadays, the internal system of a portable, inexpensive GPS tracking receiver is designed in a complex way to solve uncertainty issues and to increase accuracy. The U.S. National Coordination Office stated in the performance standard that “well-designed GPS receivers have been achieving horizontal accuracy of 3 meters or better and vertical accuracy of 5 meters or better 95% of the time”. However, it is difficult to choose a so-called “well-designed” GPS receiver given that all manufactures advertise their products as being the best. It is therefore questionable whether a GPS receiver for normal civil use can meet the standards of the U.S. National Coordination Office. The reliability of GPS data needs to be examined before further processing for new areas of applications.

This paper intends to complement the existing literature by contributing to a specific issue that has been little studied, namely, examining the reliability of GPS data collected in a real road network setting. This is done, especially in the validation phase, after data has been transferred into the computer, but before information extraction, such as trip identification, mode detection, or purpose imputation. The questions that are in focus are: *What is the dynamic data reliability with varying transportation modes, road network, environmental conditions and collection settings?*, *How well do the concurrent GPS receivers perform in tracking vehicle mobility?*, and *To what extent can the accuracy provided by the manufactures be trusted?*

This paper outlines a method of assessing the reliability of GPS tracking data by examining how well the data matches the travel information of position and speed. In particular, we applied the method in a real field experiment. In the experiment, we vary the transportation mode, speed, elevation, sampling frequency, filtering level, as well as the receivers. A bicycle, a car, and a bus travelling on pre-set routes with pre-set speeds are tracked by the GPS receiver, GlobalSat BT-338X, with different collection settings. Two other receivers, namely, Magellan SporTrak Pro, and smart phones (Samsung Galaxy S5 Mini), are used to track the bicycle under the same settings for comparison of performances, as well as avoiding results being receiver-specific.

The results from the implementation of the method demonstrate that the GPS tracking data identified the actual positions of the vehicles fairly successfully. The tracked instantaneous speeds are quite accurate, with a tendency towards underestimation. Elevation measurement is highly inaccurate, and we therefore suggest disregarding this parameter in practical use until further investigations. The experimental method can be reproduced, reorganized and reformed into different combinations according to the applications.

Paper II: How does the complexity of a road network affect optimal facility locations?

The road network is a necessary component, not only for facilitating spatial movements of people and goods but also influencing the optimal location of facilities that usually serve as the destinations of the movements. A facility optimally located on the road network is a desired goal in many decision-making processes (Carling et al., 2012, 2015). Peeters and Thomas (1995) claimed that the more complex the road network is, the higher probability of identifying a better solution to the optimal location. Following which, many studies have focused on simplifying the problem through aggregating the candidate nodes or improving the algorithms. However, the road network is often organized hierarchically and asymmetrically with various road levels and spatial structures in order to fulfill transportation needs and to adapt to facility development. The influence of the complexity of road network in finding optimal locations is not well-studied.

This paper aims to investigate how the changes in road network complexity influence the optimal facility locations by applying the widely used p -median model. In addition, to provide further insight into computation complexity and location problems from intra-urban to inter-urban, a detailed sensitivity analysis of four algorithms and various facility numbers is conducted.

The four algorithms are: greedy search (Kuehn and Hamburger, 1963); CPLEX; simulated annealing (Al-Khedhairi, 2008); and imp-Genetic algorithm (Rebreyend et al., 2015). The facility numbers vary from 5 to 50 (with a common difference of 5) to simulate intra-urban and inter-urban location distributions in the case of Dalarna province, Sweden. The size of Dalarna, as well as its structure, is similar to the regions of Vermont and New Hampshire in the US.

The entire road network database provided by the official National Road Database in Sweden consists of 2.6 million roads and 30 million nodes, where 0.15 million roads and 1.8 million nodes are distributed in the Dalarna region. The road network in the Dalarna region has a hierarchical structure composed of ten levels of road. The European highways are at the biggest level, representing the simplest road networks, where there is a limited number of candidate nodes for locating facilities. The local and private streets are at the smallest level and represent the most

complex road network, with a huge number of candidate nodes for locating facilities. Considering the asymmetrical distribution of the road network and the population, network distance is the distance measure in the analysis.

The main result indicates that there is a limit to how complex the road network needs to be to find the best solutions. When the complexity of the road network continues to increase after a certain level the solutions barely improve, but even deteriorate. We found that the choice of algorithms and facility number p has little to no impact on the results, although due to differences in the mechanics of the algorithm, there are some variations among the algorithms. The most complex level of road network (including local streets) is not preferable, especially when the number of the facility is less than 20. The spatial distributions of the optimal locations show that optimal locations found for smaller p remain as part of the optimal locations when p increases.

Paper III: On processing GPS tracking data of spatio-temporal car movements: a case study

GPS devices can track car movements fairly reliably with regard to longitude, latitude, time, speed, and altitude at regular time intervals. This spatial and temporal information can be used for investigating the activities of people and their induced effects. It is possible to use the unprocessed GPS tracking data. However, its usage is limited to rather few aspects, namely those only requiring recorded speed, coordinates, and time.

Little study has attempted to discuss issues related to processing GPS tracking data in detail, let alone provide a procedure or detailed open-source code. The aim of this paper is to address several of the issues arising from processing GPS tracking data and thereby outline a general procedure for the data processing. The study is carried out using real-world GPS tracking data of some 300 cars, originally collected for the purpose of studying CO₂ emissions induced by retailing.

The processing of the GPS tracking data includes a clear definition of movement, a detailed understanding of the capability of the GPS device and the output of the GPS logger, data transformation for summarizing descriptive statistics, data transformation for digital visualization, as well as methods for matching the GPS data to the road network. The data process and transformation provide access to key variables of car movements. The technical documentation of the data processing in this paper is detailed in the interest of readily being replicable on the same, or similar, type of data.

Paper IV: Residential planning, driver mobility and CO₂ emission

Mobility is necessary for people to meet multiple needs, whereby intra-urban mobility between residential locations and hotspots is a primary part. Meanwhile, mobility induces negative externalities like CO₂ emissions. In most urban planning studies, it is commonly assumed that the residential locations are already fixed and studies mainly focus on the analysis of ex-post evaluation of the current urban form and its influence on the environment, and the empirical analyses are mainly done on metropolitan cities. However, while the ex-post analysis examines the current situation after urban plans have been implemented, it is important to evaluate the potential plans and to provide planners with some foresight on the potential influence of the plans.

This paper proposes an ex-ante evaluation method in the spirit of LUTI models, but mainly evaluates the induced CO₂ emissions of car mobility from counterfactual urban residential plans to the primary destinations (i.e. hotspots) on a local scale, where the difference in induced CO₂ emissions of car mobility between the plans is quantified. The method is illustrated by examining a specific urban residential planning case.

Essentially, ex-ante evaluation is a “what-if” analysis (Bourguignon and Ferreira, 2003). In this paper, two main questions are in focus: *What could be the change, compared to the current situation in CO₂ emissions in the city, if the residential plans were adopted?*, and *Which subset of residential plans would reduce the CO₂ emissions the most if adopted?*

To implement this ex-ante evaluation, first, the origins and destinations in the current and counterfactual situations are identified. Second, a network of induced emissions (NoIEs) is constructed by quantifying the emissions induced by moving from the origins to the destinations. The merits of having NoIEs are that the emissions between any origins to any destination could readily be computed, the two main questions above can then be answered. In constructing the NoIEs, Oguchi’s et al. (2002) model is applied since it is particularly suitable for GPS tracking data in estimating the emissions of car travel. A good estimation of speed profile is found crucial for deriving the NoIEs in an urban context. We propose using a systematic triangle function to estimate the speed profile as the benefit of this triangle approximation is that the speed profile can be estimated (fairly accurately) by only one observation of the speed along the road segment, while any kind of elaborated, nonlinear specification requires multiple observations of the speed and would still run the risk of being highly imprecise.

The empirical analysis was done for a case city (Borlänge), and considered 1,291 current residential locations, 50 counterfactual residential locations and 51 destinations in this midsize city.

The quantitative results indicate that urban residential plans can significantly influence CO₂ emissions induced by the car mobility of residents. If the plans are in line with reshaping the urban form towards a compact city, the CO₂ emissions can be mitigated effectively, in spite of the fact that the case city is of a polycentric urban form.

Paper V: On assessing governmental sustainable residential planning and its alignment with residents' and estate investors' objectives

This paper outlines a conceptual model for the interaction of the three key actors in residential development – government, (residential) estate investors, and residents – each actor with their unique objective. Furthermore, the paper develops a methodology to *ex-ante* assess if, and where, the government's sustainability objective may be in conflict with the objectives of the other market actors. The method is applied and demonstrated in a representative Swedish case study, the mid-sized city of Borlänge, which, similar to many other Swedish cities, is currently planning for massive residential development in response to population growth due to, *inter alia*, the influx of immigrants.

In outlining the conceptual model, we define each actor's objective (in utility terms). We regard the government as the macro-actor in residential development, but purport that the government exerts the resident's living preferences and the estate investor's willingness to invest in its utility function. The objectives and behavior of the estate investor and the resident are formulated on a micro-level, meaning that our approach belongs to the tradition of microscopic models frequently applied in the transportation science.

The implementation of the methodology requires access to some key variables or proxy-variables, thereof: (1) the spatial distribution of market value of residencies and the built-up cost, (2) road network and its traffic flow and the corresponding emissions (or other outputs giving rise to negative externalities), (3) availability of data on fractions of green land and water in the surrounding area, and (4) residents' consumption. In the case study, we demonstrate how these variables can be obtained, measured, or derived. The implementation of the methodology also requires a number of parameters to be set. To ensure that the results for the case study are robust for these settings, we conduct a sensitivity analysis where the most reasonable setting (default) is replaced by an aberrant setting.

The application of the methodology in Borlänge shows that market actors' behavior is, in general, aligned with the governmental objective of sustainable, residential development. However, on considering the present residential plans for Borlänge, it seems that the local government is aware

of the market forces and compromises between its sustainability objective and the market actors' preference, in areas where the actors are not in alignment with local government objectives.

5 Concluding discussion

This dissertation is concerned with the question of which urban form may most effectively mitigate CO₂ emissions. This question is presently debated in the literature, and its answer has important implications for sustainable residential planning and development. Although the dissertation offers many new solutions to various technical aspects, its primary contribution is the outline of an operational model that comprehensively integrates investors' investment strategy, residents' choice behavior, and governmental sustainability objective in the interest of making ex-ante assessments of residential plans. This ex-ante assessment provides decision-support in sustainable residential development, foremost, at local level.

To implement the model and obtain valid assessments for the purpose of decision-making, all the microdata analysis components (data collection, data assessment and transformation, data storage and data analysis) need to be highlighted in order to generate five key variables: (1) intra-urban car movements, (2) the road network, (3) induced CO₂ emissions, (4) (planned) residential locations and hotspots, and (5) utility estimates of government, residents, and estate investors. It should be said that the microdata analysis process presents an appropriate framework for providing an answer to the question stated at the beginning of this section. Presumably, the most intriguing findings of this dissertation will probably be its conclusion regarding the concordance between the government aiming for sustainability, and the market actors, in spite of the fact that the latter act in self-interest. However, most time and work has been devoted to the assessment of data quality, data processing and transformation. Without using the microdata analysis framework, I would probably not have fully appreciated the importance and the work required on these components.

The finding from the implementation of the model in the case study is that the market actors' objectives are, in general, aligned with the government's sustainable residential development objective. The results further predict that the residential plans in the center area of the city, upon adoption, will generate less CO₂ emissions than they do today. More broadly, the findings indicate that re-shaping the urban form into a compact city is preferable for mitigating CO₂ emission, in spite of the fact that the case city is of a polycentric urban form.

Hence, this dissertation provides support for those who advocate the compact city as the ideal for sustainable residential development. Of course, analyses on various other cases would provide a more solid ground before settling the ongoing debate on the preferred re-shaping of urban forms.

Hence, for future research, I invite other researchers to apply the model I have developed to other cities. In addition to environmental sustainability that has been the focus of this dissertation, finding an operational way of providing valid ex-ante assessments with regard to social and economic factors would complement this dissertation and offer a comprehensive approach for foresighting sustainable development.

References

- Acheampong, R. A., & Silva, E. (2015). Land use–transport interaction modeling: A review of the literature and future research directions. *Journal of Transport and Land Use*, 8(3), 11-38.
- Bone, C., Dragicevic, S., & White, R. (2011). Modeling-in-the-middle: bridging the gap between agent-based modeling and multi-objective decision-making for land use change. *International Journal of Geographical Information Science*, 25(5), 717-737.
- Bourguignon, F., & Ferreira, F. (2003). Ex ante evaluation of policy reforms using behavioral models. *The impact of economic policies on poverty and income distribution*, Oxford Univ. Press, New York. pp.123-142.
- Brownstone, D., & Golob, T. F. (2009). The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, 65(1), 91-98.
- Carling, K., Han, M., & Håkansson, J. (2012). Does euclidean distance work well when the p-median model is applied in rural areas? *Annals of Operations Research*, 201(1), 83-97.
- Carling, K., Håkansson, J. and Jia, T., (2013), Out-of-town shopping and its induced CO2-emissions, *Journal of Retailing and Consumer Services*, 20:4, 382-388.
- Carling, K., Han, M., Håkansson, J., & Rebreyend, P. (2015). Distance measure and the p-median model in rural areas. *Annals of Operations Research*, 226(1), 89-99.
- Cao, X., Mokhtarian, P. L., & Handy, S. L. (2009). Examining the impacts of residential self-selection on travel behaviour: a focus on empirical findings. *Transport Reviews*, 29(3), 359-395.
- Chapman, L. (2007). Transport and climate change: a review. *Journal of Transport Geography*, 15(5), 354-367.
- Chatman, D. G. (2009). Residential choice, the built environment, and nonwork travel: evidence using new data and methods. *Environment and Planning A*, 41(5), 1072-1089.
- Council of the European Union (2006). *Renewed EU sustainable development strategy*, Brussels, 26 June, 2006.

References

- Dhakal, S. (2009). Urban energy use and carbon emissions from cities in China and policy implications. *Energy Policy*, 37(11), 4208-4219.
- Dieleman, F., & Wegener, M. (2004). Compact city and urban sprawl. *Built Environment*, 30(4), 308-323.
- Glaeser, E. L., & Kahn, M. E. (2010). The greenness of cities: carbon dioxide emissions and urban development. *Journal of Urban Economics*, 67(3), 404-418.
- Godschalk, D. R. (2004). Land use planning challenges: Coping with conflicts in visions of sustainable development and livable communities. *Journal of the American Planning Association*, 70(1), 5-13.
- Grazi, F., & van den Bergh, J. C. (2008). Spatial organization, transport, and climate change: Comparing instruments of spatial planning and policy. *Ecological Economics*, 67(4), 630-639.
- Grazi, F., van den Bergh, J. C., & van Ommeren, J. N. (2008). An empirical analysis of urban form, transport, and global warming. *The Energy Journal*, 97-122.
- Habitat U (2011). *Cities and Climate Change: Global Report on Human Settlements 2011*. London: Earthscan.
- Handy, S. (2005). Smart growth and the transportation-land use connection: What does the research tell us? *International Regional Science Review*, 28(2), 146-167.
- Hankey, S., & Marshall, J. D. (2010). Impacts of urban form on future US passenger-vehicle greenhouse gas emissions. *Energy Policy*, 38(9), 4880-4887.
- Helming, K., Diehl, K., Kuhlman, T., Jansson, T., Verburg, P., Bakker, M., ... & Morris, J. (2011). Ex ante impact assessment of policies affecting land use, part B: application of the analytical framework. *Ecology and Society*, 16(1).
- Hensher, D. A. (2008). Climate change, enhanced greenhouse gas emissions and passenger transport—What can we do to make a difference? *Transportation Research Part D: Transport and Environment*, 13(2), 95-111.
- Hong, J., & Goodchild, A. (2014). Land use policies and transport emissions: Modeling the impact of trip speed, vehicle characteristics and residential location. *Transportation Research Part D: Transport and Environment*, 26, 47-51.
- Huang, Q., Parker, D. C., Filatova, T., & Sun, S. (2014). A review of urban residential choice models using agent-based modeling. *Environment and Planning B: Planning and Design*, 41(4), 661-689.
- International Energy Agency. (2009). *Transport Energy and CO₂: Moving Towards Sustainability*. OECD Publishing.

- Johansen, B. G., Hansen, W., & Tennøy, A. (2015). Evaluation of models and methods for analyzing the interaction between land-use, infrastructure and traffic demand in urban areas. TØI Report, (1415/2015).
- Kii, M., Nakanishi, H., Nakamura, K., & Doi, K. (2016). Transportation and spatial development: An overview and a future direction. *Transport Policy*, 49, 148-158.
- Krizek, K. J. (2003). Residential relocation and changes in urban travel: does neighborhood-scale urban form matter? *Journal of the American Planning Association*, 69(3), 265-281.
- Leduc, G. (2008). *Road traffic data: Collection methods and applications*. Working Papers on Energy, Transport and Climate Change, 1, 55.
- Lefèvre, B. (2009). Long-term energy consumptions of urban transportation: A prospective simulation of “transport–land uses” policies in Bangalore. *Energy Policy*, 37(3), 940-953.
- Lindsey, M., Schofer, J. L., Durango-Cohen, P., & Gray, K. A. (2011). The effect of residential location on vehicle miles of travel, energy consumption and greenhouse gas emissions: Chicago case study. *Transportation Research Part D: Transport and Environment*, 16(1), 1-9.
- Mokhtarian, P. L., & Cao, X. (2008). Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B: Methodological*, 42(3), 204-228.
- Norman, J., MacLean, H. L., & Kennedy, C. A. (2006). Comparing high and low residential density: life-cycle analysis of energy use and greenhouse gas emissions. *Journal of Urban Planning and Development*, 132(1), 10-21.
- Oguchi, T., KATAKURA, M., & TANIGUCHI, M. (2002). Carbondioxide emission model in actual urban road vehicular traffic conditions. *Doboku Gakkai Ronbunshu*, 2002(695), 125-136.
- Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: a review. *Annals of the Association of American Geographers*, 93(2), 314-337.
- Peeters, D., & Thomas, I. (1995). The effect of spatial structure on p-median results. *Transportation Science*, 29(4), 366-373.
- Perkins, A., Hamnett, S., Pullen, S., Zito, R., & Trebilcock, D. (2009). Transport, housing and urban form: the life cycle energy consumption and emissions of city centre apartments compared with suburban dwellings. *Urban Policy and Research*, 27(4), 377-396.
- Qin, B., & Han, S. S. (2013). Planning parameters and household carbon emission: Evidence from high-and low-carbon neighborhoods in Beijing. *Habitat International*, 37, 52-60.

References

- Rebreyend, P., Lemarchand, L., & Euler, R. (2015). A Computational Comparison of Different Algorithms for Very Large p-median models. *Evolutionary Computation in Combinatorial Optimization*, 13-24.
- Shen, L., & Stopher, P. R. (2014). Review of GPS travel survey and GPS data-processing methods. *Transport Reviews*, 34(3), 316-334.
- Verburg, P. H., Rounsevell, M. D., & Veldkamp, A. (2006). Scenario-based studies of future land use in Europe. *Agriculture, Ecosystems & Environment*, 114(1), 1-6.
- Zhang, H., Jin, X., Wang, L., Zhou, Y., & Shu, B. (2015). Multi-agent based modeling of spatiotemporal dynamical urban growth in developing countries: simulating future scenarios of Lianyungang city, China. *Stochastic Environmental Research and Risk Assessment*, 29(1), 63-78.

Part II

Paper I

An Evaluation of the Reliability of GPS-Based Transportation Data

Xiaoyun Zhao[♦], Kenneth Carling, Johan Håkansson

GPS-based data are becoming a cornerstone for real-time transportation applications. Tracking data of vehicles from GPS receivers are however susceptible to measurement errors. The assessment of the reliability of data from GPS receiver is a neglected issue, especially in a real road network setting and in the phase after data transfer but before information identification. An evaluation method is outlined and carried out by conducting a randomized experiment. We assess the reliability of GPS-based transportation data on geographical position, speed, and elevation from three varied receivers GlobalSat BT-338X, Magellan SporTrak Pro and smart phone for three transportation modes: bicycle, car, and bus. The positional error ranging from 0 ± 158 meters, and 74% to 100% with an error within 5 meters depending on the transportation mode and route, there is also a non-negligible risk for aberrant positioning. Speed is slightly underestimated or overestimated with errors around ± 5 km/h except for SporTrak Pro which had an error of -10 km/h. Elevation measurements are unreliable with errors bigger than ± 100 meters.

Keywords: Transportation, GPS tracking, Reliability, Road network

1. Introduction

Global Positioning System (GPS) has emerged for civilian use in the 1990s as the space geodetic technique became accurate and affordable (Zumberge et al., 1995). GPS tracking technologies have extensively been applied in transportation studies, in particular for studying the routes of motorized vehicles (Zito et al., 1995; Quiroga and Bullock, 1998; Murakami and Wagner, 1999). Schönfelder and Antille (2002) presented an approach to collect GPS longitudinal travel behaviour data on humans and described the complexity of their daily life with the interaction between periodicity and variability. Stopher et al. (2007) demonstrated that GPS can be used successfully to supplement travel diary surveys. Kamboj and Dahiya (2011) found standard handheld GPS receiver may be used to measure sag in overhead conductor of power transmission lines along with error estimation technique LSPE. Lindsey et al. (2013) confirmed the feasibility of using GPS for route tracking to identify the specific locations where cyclists ride on a street. In environment control, for instance,

[♦] Xiaoyun Zhao is a PhD student in Microdata Analysis, Kenneth Carling is a professor in Statistics, and Johan Håkansson is a professor in Human Geography. All are at the School of Technology and Business Studies, Dalarna University, SE-791 88 Falun, Sweden. *Corresponding author:* Xiaoyun Zhao, e-mail: xzh@du.se, phone: +46-23-778509.

Carling et al. (2013) and Jia et al. (2013) studied the induced pollutant emissions of CO₂ from car movements by using a GPS tracking data of car movements.

Shen and Stopher (2014) conducted a review of GPS-based travel studies going back to late 1990s, which range from application of GPS travel surveys to methods of processing GPS data. In their review they listed representative studies using dedicated GPS receivers from 14 different countries, as well as four studies regarding smartphones. They stated that data collection based on GPS surveys is more reliable, and cheaper, than self-reported diaries, though GPS data still has some issues that require data processing methods to enhance the reliability of the data.

Gathering information of spatial-temporal mobility by GPS is subject to critical reflections. Leduc (2008) examined recent developments in road transportation data collection and discussed the potentials in providing real-time information for routing and estimating traffic flow and volume. The author also pointed out the bottlenecks of the uncertainty in the GPS technologies still leads to the question: “How good the quality of the traffic data is?”. Moreover, Van der Spek et al. (2009) concluded that GPS offers a widely useable instrument to collect invaluable spatial-temporal data on different scales and in different settings adding new layers of knowledge to urban studies, but the use of GPS-technology and deployment of GPS-receivers still offers significant challenges for future research. Besides, the enormous use of GPS tracking technologies hinges critically on the functioning of the receiver.

Nowadays, the internal system of a portable, inexpensive GPS tracking receiver is designed in a complex way due to the desire for accuracy. The U.S. National Coordination Office stated in the performance standard that “well designed GPS receivers have been achieving horizontal accuracy of 3 meters or better and vertical accuracy of 5 meters or better 95% of the time”. Configuration of a GPS receiver when conducting field tracking is becoming more complicated. However, the receiver can function as an effective and reliable tool for data collection only if it does not affect the nature, quality or authenticity of the data collected (Shoval, 2008; Huang, 2013). Studies on gathering information of trips, travel modes and trip purpose have shown that accuracy varies depending on the methods, attributes and ground truth (c.f. Table2 and Table 3 in Shen and Stopher 2014). The methods reviewed in the tables either performed poorly due to the ambiguity of similar modes or were highly dependent on the “ground truth data” (Zheng et al. 2008; Chang et al. 2015). Moreover, the methods are primarily designed for information identification directly from the data rather than for data evaluation. No doubt that the application of GPS survey has opened a new era for travel data collection, the information identified from the data can be broadly applied. However, it is difficult to choose a so called “well designed” GPS receiver given that all manufactures are advertising their

products to be the best. It is therefore questionable whether a GPS receiver for normal civil use can meet the standard of the U.S. National Coordination Office. The reliability of these GPS data needs to be examined before further processing for the new areas of applications.

In this study we intend to complement the existing literature by contributing to a specific issue that has little been studied, being examining the reliability of GPS data collected in a real road network setting. Especially in the validation phase after data has been transferred into computer, but before information extraction such as trip identification, mode detection, purpose imputation. This means that we will focus on questions like: What is the dynamic data reliability with varying transportation modes, road network, environmental conditions and collection settings? How well do the concurrent GPS receivers perform in tracking vehicle mobility? To what extent can the accuracy provided by the manufactures be trusted?

Following this, the assessment of the reliability of GPS tracking data needs to be scrutinized. This paper outlines a method to examine how well GPS tracking data matches the travel information of position and speed. Specially, we applied the method in a real field experiment. In the experiment, we vary the transportation mode, speed, elevation, sampling frequency, filtering level as well as the receivers. A bicycle, a car, and a bus travelling on pre-set routes with pre-set speeds are tracked by GPS receiver GlobalSat BT-338X with different collection settings. Two other receivers being Magellan SporTrak Pro and smart phones (Samsung Galaxy S5 Mini) are used to track the bicycle under same settings for comparison of performances as well as avoiding the results being receiver specific. The acquired experimental data are freely available¹ for the interest of replicability.

Section 2 provides a review of researches related to examining the reliability of GPS tracking. Section 3 presents the experimental design and the data collection process. Section 4 gives the experimental results. Section 5 ends the paper with a concluding discussion of the findings.

2. Literature review

A thorough search for literatures relevant to the use of GPS-based transportation data was conducted in a former related work (Zhao et al., 2014). There is a vast body of studies reporting on applications of GPS with a brief discussion about the reliability of the data. The discussions contained in these studies do not add any new knowledge to the data reliability and we therefore turn to studies with reliability as the primary concern.

¹ <http://users.du.se/~xzh/>

Obviously the quality of the hardware and the surroundings where the GPS is being used may affect the reliability of the receiver. The starting point is that a GPS receiver requires a clear sight with at least four satellites to determine spatial positions. In urban environments, buildings may partly block satellite signals, forcing the GPS receiver to work with a poor geometric constellation of satellites, thereby reducing the accuracy of the positional estimates. Multipath propagation of the radio signal due to reflection in the surroundings may further lead to decreased positional accuracy without notification by the GPS receiver, thereby reducing the integrity of the navigation solution. The accuracy may be enhanced by advanced hardware chipsets, dual-frequency receivers, carrier-phase measurements supported by augmentation systems (e.g. SBASs, WAAS, EGNOS and MSAS), even combination of the global navigation satellite system (GPS, GLONASS, BeiDou and Galileo, Li et al, 2015). It is possible to have a real-time positional accuracy within decimetres under required conditions; however, those kinds of receivers are too expensive for normal use like in car tracking systems. Moreover, the required conditions do not only call for sophisticated GPS receivers, sensors, vehicles, and map information, but also put requirements on trajectory dynamics and surrounding environment (Skog and Handel, 2009, Li et al. 2015).

Dead Reckoning (DR) system and map matching algorithms integrated with differential GPS (DGPS) are examples of commonly used hybrid systems for enhancing the positioning of vehicles on land (Zhao et al., 2014). The DR system can smooth the error of the GPS and provide continuous positioning even in times when the GPS is unavailable (Meng, et al., 2004). Map-matching has been predominantly applied in post-processing GPS data (e.g., Marchal et al., 2005; Schüssler and Axhausen, 2009a, 2009b). In essence, map matching is to use a digital map of the road network to impose constraints on the GPS navigation and tracking recordings (Skog and Handel, 2009). It has become a popular solution to remedy the inherent error of the GPS when an underlying network is available. Quddus et al. (2007) reviewed the currently existing map-matching algorithms and their limitations. Stopher et al. (2013) proposed adding map editing to manually fix certain data errors besides fixing the cold/warm start issue (Chen et al. 2010).

Stopher and Speisser (2011) conducted tests for five transportation modes under various circumstances and found that the tested GPS devices are accurate enough to be useful as a substitute for self-report surveys. However, they only checked one type of GPS device (BTT08) and neglected the examination of speed. The reviewed studies in Zhao et al., 2014 most relevant to our study are those attempting to assess the reliability of GPS data by comparing them to known conditions. These studies aim to evaluate the reliability of GPS receiver, but are typically not conducted as experiments. They examine one transportation mode, one environment, one aspect of tracked

information, or one configuration of the receiver, the importance of road network is neglected as well. Those studies are also examining the static accuracy using small samples without controlling for external conditions. Studies that examined the reliability of GPS receivers on sport events, animal activities are not reviewed in detail considering there is no road network restriction and the standard for reliability is different from transportation research. A notable exception is the recent work of Schipperijn et al. (2014). They tested the dynamic accuracy of Qstarz Q1000XT portable GPS receiver for the use in public health applications under varying real-world environmental conditions, for four modes of transportation at three levels of sampling frequency. They found that not even a half of the positional recordings were within 2.5 meters of the actual positions with the proportions varying by travel mode and area. Montini et al. (2015) compared the travel diaries generated from smartphones and dedicated GPS devices but with a focus on the performances of sampling frequency, route and activities detection. As claimed by Schipperijn et al. (2014), mobile objects in free-living studies are likely to move dynamically. It is therefore vital to know the dynamic accuracy for various travel modes in changing surroundings. However, Schipperijn et al. (2014) only studied the influence on positional accuracy by changing the sampling frequency of the GPS receiver but neglecting other factors possibly affecting the accuracy.

To conclude, despite the increasing popularity of GPS in active living research, missing data and errors are still the main challenges for GPS studies (Shen and Stopher, 2014). The reliability of current GPS receivers employing different configurations and how they are affected by a variety of conditions for tracking various types of vehicles on real road networks is limited.

3. Experimental design and data collection

Vehicles are restricted by an underlying road network when travelling, various transportation modes are therefore necessary in representing common users travelling on different levels of a road network (c.f. Schipperijn et al. 2014). To examine how well GPS tracking data of vehicles matches an actual route travelled, we therefore consider bicycle, car, and bus being the dominating modes in private transportations; the mode of pedestrian is omitted because bicycles travel on the same level of a road network. In the experiment, the vehicles travel on pre-set routes of known geographical position and elevation with speeds decided in advance. Their mobility is being tracked simultaneously by the GPS receivers when they are travelling.

For the experiments, a standard and integrated GPS receiver that could be broadly used in different vehicles under various circumstances is preferable. Other important features in selecting the receiver are that the receiver is user friendly, easy to operate and has a durable battery. GlobalSat BT-338X, Magellan SporTrak Pro and smart phone (Samsung Galaxy S5 Mini) were chosen after a

survey in the product market. According to the manufacturer, with WAAS enabled, the GlobalSat BT-338X should provide a geographical positioning within an error of 5 meters and a measurement error of speed less than 0.4 km/h while the Magellan SporTrak Pro should have positioning error within 3 meters and a measurement error of speed less than 0.2 km/h. There is no specific claim of GPS accuracy for the smart phone, we take 10 meters for positioning error and 0.5 km/h for speed tracking error (Djuknic and Richton, 2001; Herrera et al., 2010). The manufacturers make no claims about the accuracy in the measurement of elevation for all three types of receivers.

We set intensive sampling intervals of every 1 second, 5 seconds and 30 seconds. Note that the 30 seconds implies that some of the vehicles will easily travel more than 500 meters between recordings. Such setting implies a coarse assessment of the vehicle's mobility pattern. Hence, the levels of sampling frequency represent both dense and sparse data to track position, time, date, speed, and elevation. For BT338-X we also consider both enable and disable data logging when distance is less than the selected radius 20 meters, while the SporTrak Pro and the phone do not equip with this setting.

Table 1: Experimental design (*first phase*) of collecting GPS tracking data

Phase-I (BT338-X)																
Sampling interval		1s				5s				30s						
Receiver No.		3	29	37	36	42	4	14	39	40	77	9	32	74	24	72
Distance radius restriction		0m		20m		0m		20m		0m		20m				
Bicycle & Car	15 km/h															
	20 km/h															
	30 km/h															
	40 km/h															
	45 km/h															
Car	50 km/h															
	60 km/h															
Bus	70 km/h															
	80 - 100 km/h															

We are in possession of 15 BT338-X, 6 SporTrak Pro and 6 Samsung Galaxy S5 Mini; each of them is assigned with a unique identifying number. The experiment is combined with two phases. The first-phase mainly focuses on the influence of the change of transportation modes and travel environment on the data reliability. The second phase focuses on the influence of different types of receivers on the data reliability, which is described in Section 4 below. Table 1 summarizes the main factors and corresponding levels of first phase in the experimental design. In the data transferring phase, the recordings for bicycle, car and bus were transferred separately, and a file was added to

fill in the information of error. The BT338-Xs are randomly assigned to one of three groups of equal size and the sampling interval is set to 1, 5, and 30 seconds respectively to track bicycle, car and bus in the first phase. In each group two randomly selected receivers have the data logging disabled if distance is less than the radius of 20 meters. The data collection of the bicycle and the car is undertaken in Borlänge in Sweden. The data collection of the bus is undertaken along the bus line 151 between Borlänge and its neighbouring city Falun.

It was difficult to fix the speed of the bus in advance as would be preferable. The speed varied along the scheduled route due to the traffic and the behaviour of the drivers. For this reason, only a segment of the route, where the speed varied smoothly between 80 km/h and 100 km/h was used for GPS tracking. Meanwhile the bus trip was filmed. The bicycle followed a strict setting of speeds ranging from 15-50 km/h in six levels. For the car, 15-70 km/h were considered. Travel diaries were used to note unexpected changes in route, speed, and emergent situation. The cyclist and the driver of the car was the same throughout the experiment.

Data for the bicycle was collected at noon in order to reduce the risk of deviation from the protocol caused by other people on the route. Likewise, data collection for the car was undertaken between 3 and 4 in the afternoon to avoid peaks in the traffic. The data collection for the bus was conducted after 6 in the afternoon thereby minimizing the variation in speed due to people waiting at bus stops.

An accurate speedometer of the vehicles is essential for the experiment. We calibrated the car speedometer by riding the bicycle and driving the car side by side and recording the speeds simultaneously. The relationship between the recordings from the bicycle speedometer and the car speedometer by means of linear regression: $Car = 1.0385 * Bike$ with a strong correlation of 0.998. The speedometer of the car was adjusted accordingly in the experiment.

The routes for the experiment were chosen having the need for maintaining a constant speed. As for the car, we also needed to consider the speed limits of the roads while a bicycle may be ridden at any speed on a bicycle path. Figure 1(a) depicts the route for the bicycle with arrows indicating the riding direction. The route is about 2 kilometres and it is a paved bicycle path. The route was used consecutively for each speed at a time. For instance, at the speed of 20 km/h there could be 360, 72, and 12 recordings per GPS receiver for the three levels of sampling frequency. The variation in elevation of the route is only a few meters.

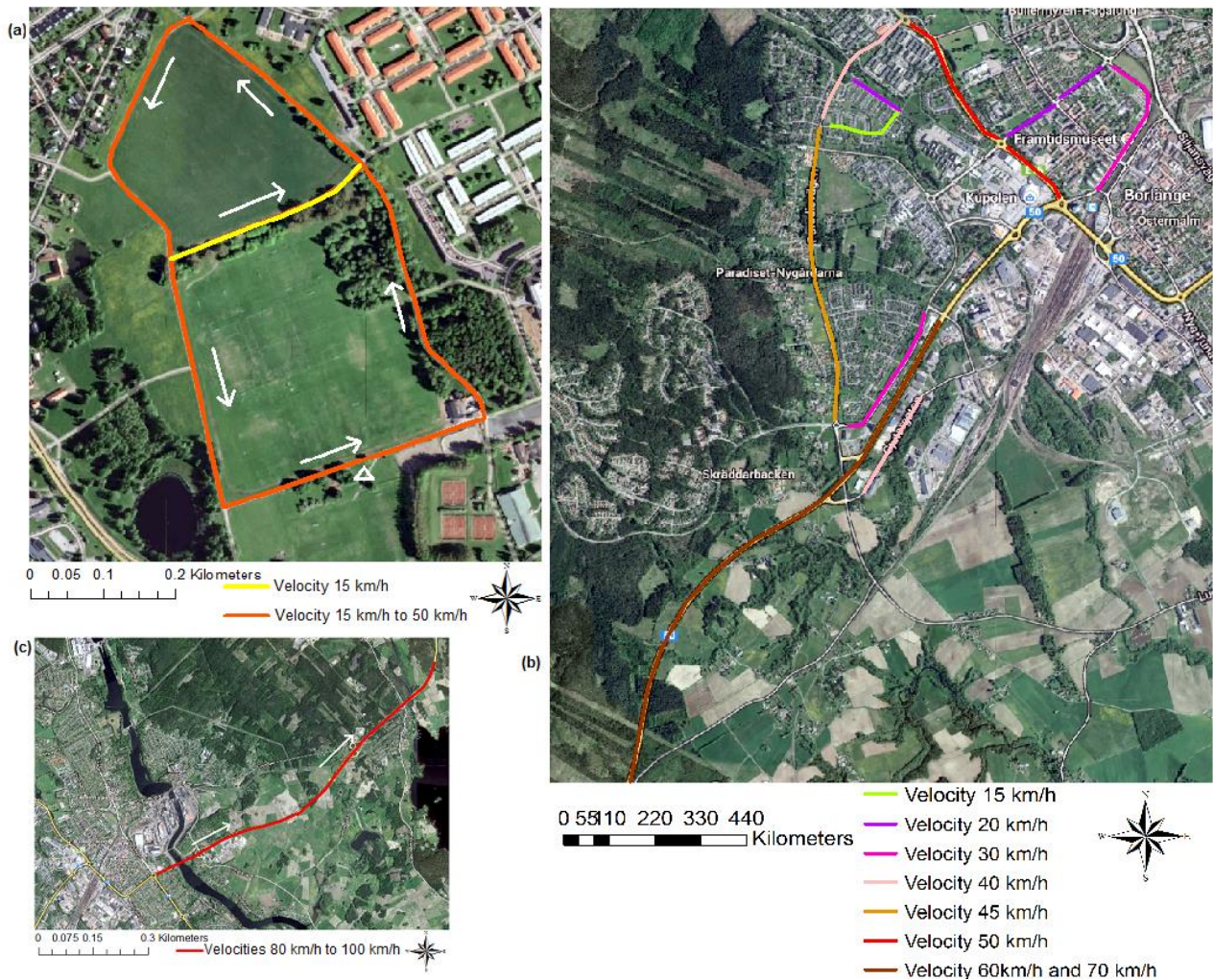


Figure 1: (a) The bicycle route; (b) The car route; (c) The bus route

Figure 1(b) depicts the route for the car. The route is segmented by colour representing the attained speed. The route was travelled 3 to 4 times on both directions to ensure sufficient recordings per cell. The range in elevation is 40 meters, maintaining a constant speed with a car in an ordinary traffic situation is of course difficult. The roundabouts and intersections in Figure 1(b) are identified in advance to highlight it is usually impossible to maintain the speed due to traffic rule and real conditions. The recordings pertaining to segments where the intended speed was not met according to the travel diary were removed. Figure 1(c) depicts the bus route. This route has a variation in elevation with a range of 37 meters.

The original GPS tracking data were kept into DataLogger files. The data were retrieved to a computer by using the software Global Sat Data Logger PC Utility directly after the experiment was completed. The receiver number 4 malfunctioned and did not record any data. The other 14 receivers worked well and we obtained in total 25,901 recordings of the car, 9,224 recordings of the bicycle, and 8,688 recordings of the bus.

4. Experimental results

4.1 Geographical positioning

The position and the trajectory of a car are restricted by the road network (Skog and Handel, 2009). The geographical positions of the mobile object are necessary to identify the objects' trajectory. In the experiment the trajectory of the vehicles is known by the road network and its digital representation. The concordance of the recordings and the road network is measured to be a statistic to assess the reliability of the geographical positioning obtained from the GPS receiver. Ideally the positional recordings should be on the underlying road network² given that the width of the driving road is 14-20 meters and 3.5 meters for the bicycle road.

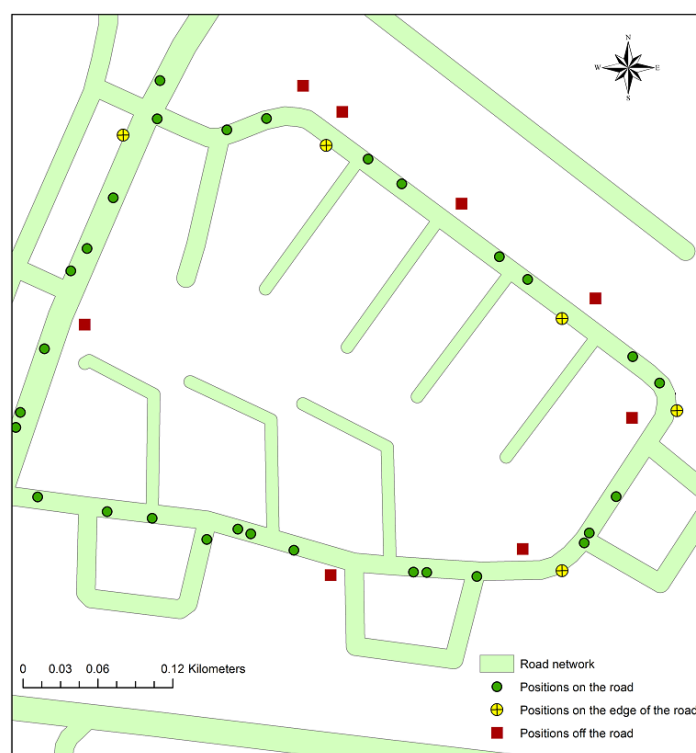


Figure 2: Example of positional recordings and the road network

Figure 2 shows an example of how the positions actually recorded on the car route. The green circles indicate the recordings that match the road network. The yellow circles indicate recordings on the edge of the road network are regarded as matching the road network well enough. The red squares indicate inaccurate recordings off the road network. In this example, 8 of the 42 recordings failed in giving an accurate position of the car, which we suspect how well the positional recordings

² The road network is provided by the National Road Data Base (NVDB) and is operated by the Swedish Transport Agency. The positional error of the road segments used in this study is within 0.2 meter.

match the road network. Figure 3 illustrates the empirical cumulative distribution of the positional error.

Given the width of the roads, almost all the recordings are expected to match the road network. However, people do not usually drive or ride right in the middle of the road, especially on two-lane roads. What’s more, given the real travel circumstances with trees, buildings and other interference for GPS signals, the positional recording are not precisely on the roads. Figure 3 shows that more than 95% of positional recordings for the bus are accurate to be 0-meter error and the biggest error is 28.2 meters. The positioning of the car was accurate that around 80% to be 0-meter error while the biggest error is 158.4 meters. As for the bicycle, the biggest error is 54.8 meters; the recordings from 5-second and 5-second with 20-meter restriction frequently fail to identify its travel positions on the network while for the other settings, 90% are within 5-meter positional error whereas only 30% are with the 0-meter error.

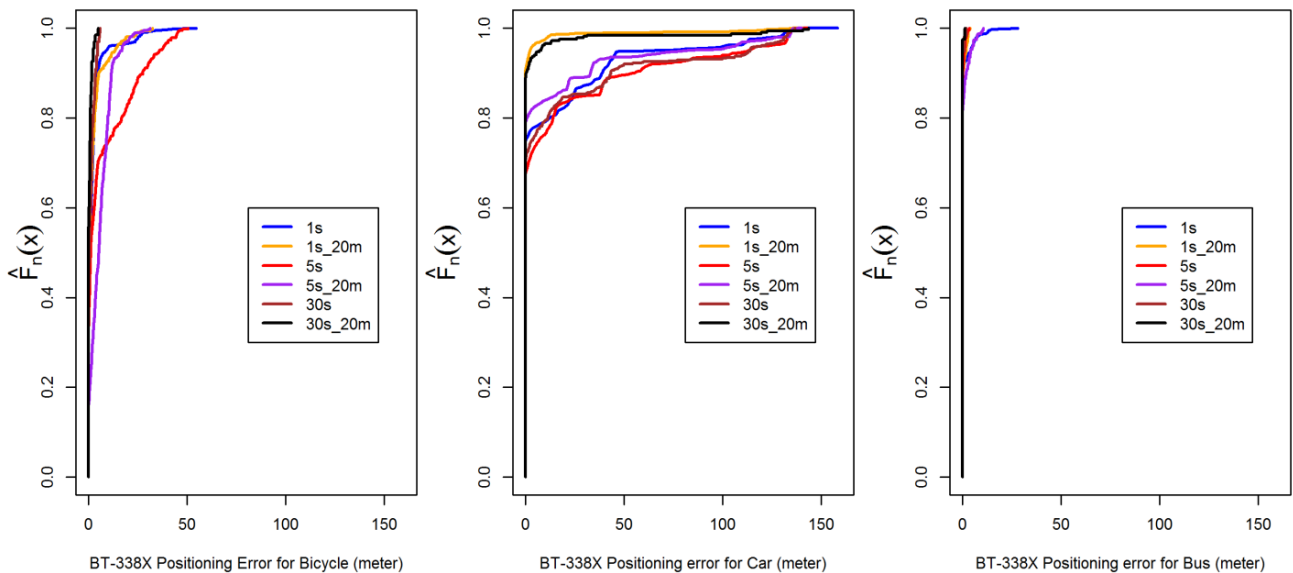


Figure 3: Empirical cumulative distribution of the positioning errors of BT-338X for bicycle, car and bus

As an overall finding drawing on Figure 3, there is no clear pattern emerging from the factors considered in the experiment. Possibly the longest sampling interval tends to lead to better positioning; the setting of the distance of restriction does not have obvious influence in positioning; the receivers generally give higher accuracy in positioning for the bus and the car but tends to have large variation on bicycle. However, we have noted a serial correlation of the recordings implying that an inaccurate recording is likely to be followed by another if the time interval is short. Especially, numerous inaccurate recordings from the first phase experiment are found in the three areas marked with the white circle and the two triangles depicted in Figure 4.



Figure 4: Bicycle and car routes in the secondary experiment

Table 2: Experimental design (*second phase*) of collecting GPS tracking data

Phase-II			
Sampling Interval		5s	30s
Receiver No.	BT338-X	1-6	1-6
	SporTrak Pro	7-12	7-12
	Samsung Galaxy S5 mini	13-18	13-18
Bicycle & Car Speed	15km/h		
	30km/h		

The circled area is nearby power lines located to the north and 200 meters to the east. The areas indicated by triangles have trees with a height of 8-10 meters. We speculated that the positional recordings of the bicycle were interfered by the surrounding environment. We also suspected that the single choice of one specific receiver is partial for the experiment; therefore in the secondary phase of the experiment, we randomly chose 6 BT338-X and added the 6 SporTrak Pro and 6 Samsung Galaxy S5 mini. The speeds 15km/h and 30 km/h and frequency 5s and 30s are chosen for the experiment on the bicycle route and the car route. Table 2 illustrates the factors and corresponding levels the second phase in the experimental design. Figure 4 depicts the two routes travelled by the bicycle; the red route coincides with the route used in the original experiment while the yellow route is a part of the car's route.

Figure 5 gives the empirical cumulative distribution of the positional error for the three types of receivers on the bicycle and the car routes. Although the proportion of accurate recordings on the original bicycle route is higher and with smaller variation, it is still rather low comparing to the car route which is substantially accurate. The biggest error on car route is 27 meters while it is 78 meters on the bicycle route. The errors (≤ 5 meters) of three types of receivers differ but not big than 5% under the same route with same speed and settings. Most inaccurate recordings are identified that happened again at the three areas that are previously identified as problematic. This illustrates that the GPS receiver may generate (infrequent) errors due to the interferences with the surroundings such as trees and built-ups in a non-obvious way (Modsching et al., 2006).

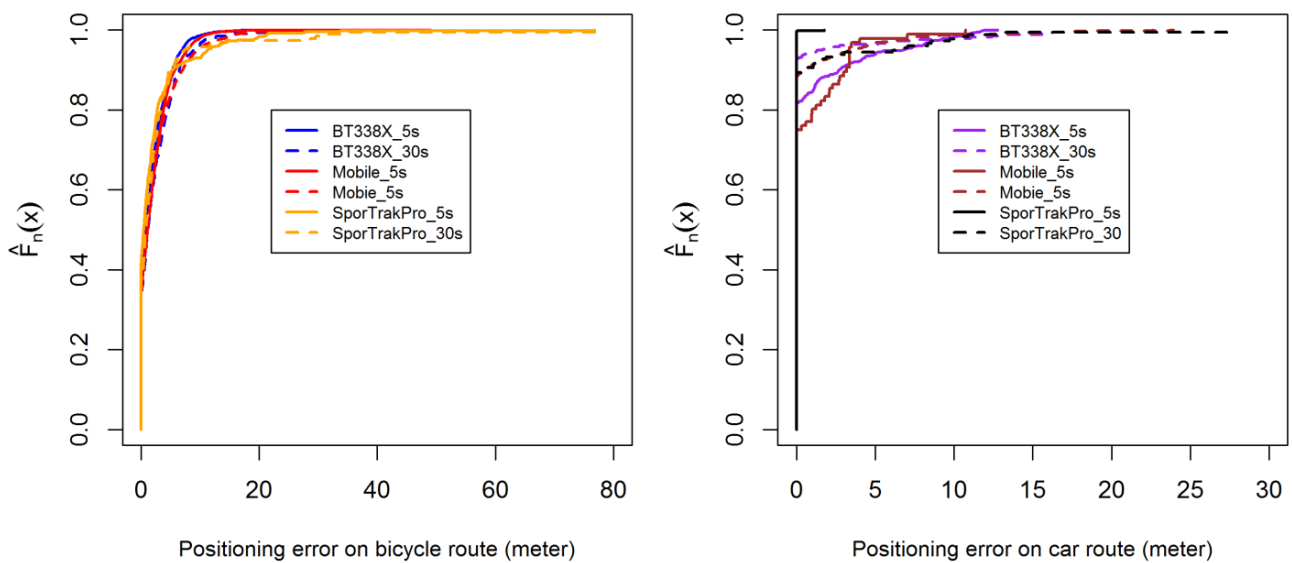


Figure 5: Empirical cumulative distribution of the positioning errors for the three types of receivers on bicycle and car route

4.2 Estimating the speed

It goes without saying that it is more difficult to estimate a changing speed than a constant speed. Drivers (and cyclists) need to adjust their speed in line with the traffic but also at intersections, roundabouts, tortuous locations (Jia et al., 2012) and traffic lights. This is also true in conducting an experiment of this kind. We used the travel diary of the car and the bicycle to delete recordings where the intended constant speeds were not possible to maintain. As for the bus, the films were used for deleting recordings where the speeds were not constant.

Figure 6 illustrates how the recorded speed of the car varies around the pre-set constant speed of a sample from tracking interval 30 seconds of all data of phase-I and phase-II. There is a tendency that the recorded speed is lower in general than the actual speed for all three types of receivers. For BT-338X, the errors are within 5km/h while the manufacturer claimed that the error is within 0.4

km/h. As for the smart phone, the errors are smaller to be within 3km/h; SporTrak Pro had the worst performance with the error to be -10km/h under the test speed 30km/h while the manufacturer claimed that the error is within 0.2 km/h. The speed recordings from SporTrak Pro also exhibited the biggest variation comparing to the other two types of receiver. The analysis of variance (ANOVA) was conducted to formally test for the factors that affect the performances of speed recordings. The response variable is the error between the recorded and the set speed in the experiment. The error increased with the speed.

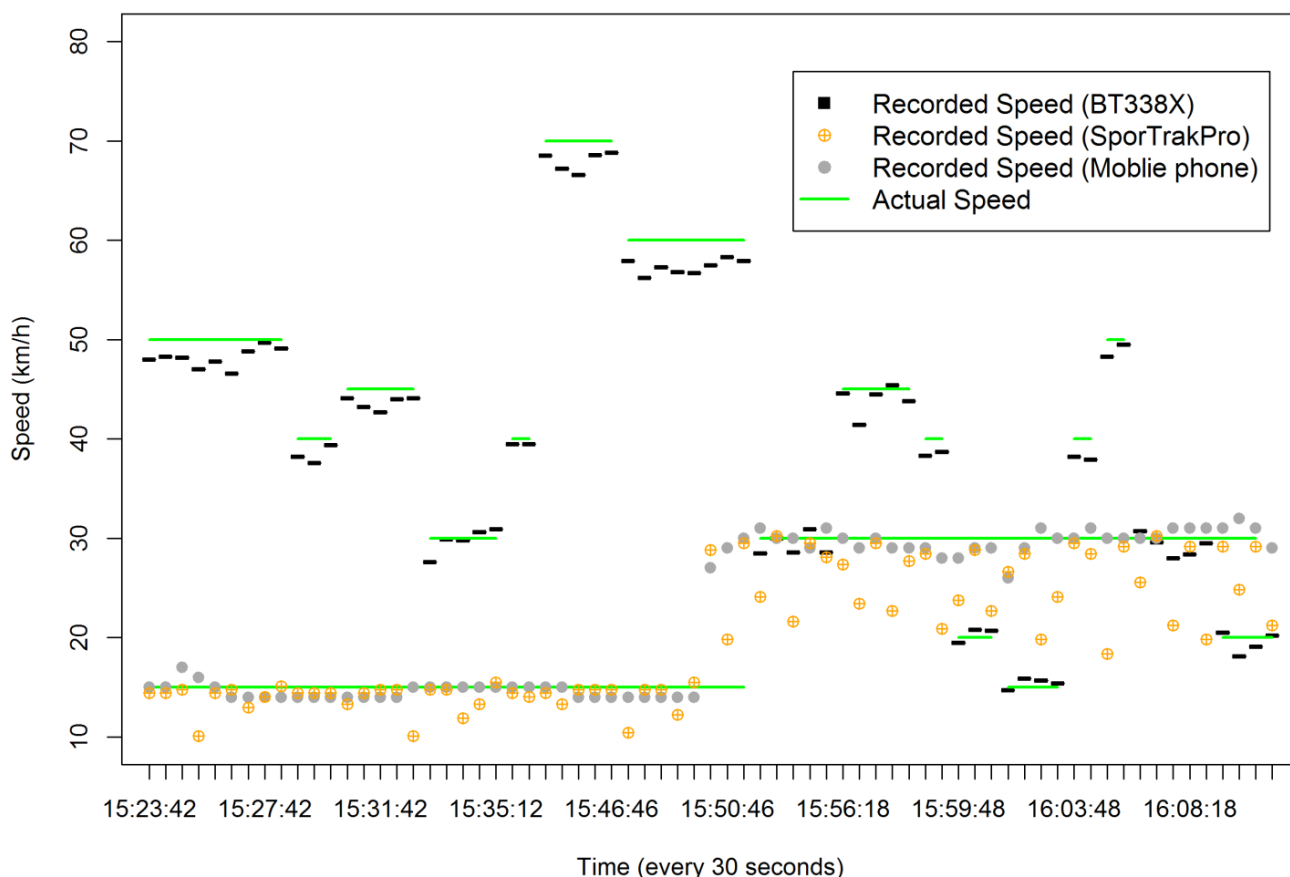


Figure 6: Recorded speed versus actual speed as measured by three types of GPS receivers

Table 3 shows the specific results. There was no significant difference for whether the distance restriction was on or off. The sampling frequency interval was related to the error but less significant. It's found that longer sampling interval was associated with a (marginal) increase in the error. The type of the travel route corresponding to the vehicle type significantly influences the accuracy of the speed recordings. There is significant different performance among the types of receivers. It was strongly significant suggesting that if the positional recording was inaccurate, there will be a greater underestimation of the speed.

Table 3: Analysis of variance table (a) Results from recordings of experiment phase-I; (b) Results from recordings of experiment phase-II

Analysis of Variance Table (a)					
Response: Error					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Set_Speed	1	719.1	719.1	465.5412	< 2e-16 ***
Distance_Radius	1	2.4	2.4	1.5403	0.21459
Time_Frequency	2	7.1	3.6	2.3098	0.09932 .
On_Off_Road	1	119.7	119.7	77.5258	< 2e-16 ***
Vehicle_Type	2	6533.1	3266.5	2114.8093	< 2e-16 ***
Residuals	14346	22158.8	1.5		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Analysis of Variance Table (b)					
Response: Error					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Set_Speed	1	714.3	714.35	225.8096	< 2.2e-16 ***
Receiver_Type	2	321.0	160.49	50.7323	< 2.2e-16 ***
Time_Frequency	1	16.3	16.26	5.1403	0.02341 *
On_Off_Road	1	157.9	157.88	49.9061	1.783e-12 ***
Route_Type	1	61.3	61.32	19.3838	1.086e-05 ***
Residuals	6421	20312.7	3.16		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

4.3 Elevation

Elevation is useful in providing information when two dimensional positioning is not sufficient to identify overlap points on roads with different heights. Moreover, elevation is also influential in travelling cost considering the time and energy use, it is necessary in route scheduling, and environment control. The accuracy of elevation is commonly expected to be poorer than the geographical position due to the requirement of satellites availability for signal strength in estimating elevation.

In order to check the accuracy in the recorded elevation, the geo-information of elevation in Borlänge from the national elevation database (NNH)³ is referred for validation. Each position of the vehicle where a recorded elevation occurred is related to the nearest point in the actual elevation layer. The maximum distance between the recorded position and the actual elevation layer is 21 meters. This is an inconsequential approximation as the road network covered in the experiment does not contain any steep up- and down-hills.

³The elevation data is provided by Sweden's Mapping, Cadastral and Land Registration Authority (www.lantmateriet.se). The elevation model is made by laser scanning and has an average elevation error of 0.1 meter and 0.4 meter in the plane.

The error in recorded elevation with respect to the actual elevation is large for all three types of receiver. No receiver showed substantially good measurement comparing to others. Most of the errors were within the range of -100 meters and 100 meters, but frequently the error exceeded 150 meters. Considering for instance that the bicycle path travelled in the experiment was essentially flat, and even for the car road the elevation change is within 40 meters, such a magnitude in error is enormous and peculiar given error in GPS elevation readings is generally twice as high as horizontal error (Noronha & Goodchild, 2000) and even better (Zandbergen, P. 2009).

5. Concluding Discussion

Current studies have rarely focused on the problem of examining the reliability of dynamic GPS data in the validation phase after data transfer but before further analysis, especially have neglected influences from the variation of road network. This paper focuses on evaluating the reliability of GPS-based transportation data from three different types of portable GPS receivers (including the commonly-used smart phone). The evaluation focuses on data of geographical position, speed, and elevation by tracking vehicles in a complex road network with varying transportation modes, environmental conditions and collection settings in real settings. The experimental method can be reproduced, reorganized and reformed into different combinations according to the applications.

The GPS tracking data identified the actual positions of the vehicles fairly successfully. The three types of receivers performed with not big than 5% differences of accuracy on the same route with same speed and settings. The surroundings of the experiment had no obviously interfering attributes like high built-ups, forests, magnetic fields, and so on; the partially poor identification of the bicycle's positions by trees and in the vicinity of magnetic fields shows however that the positional error of the GPS is highly vulnerable to the surroundings. Overall, the positioning accuracy meets the requirement of applications like routing, mobility pattern recognition, destination imputation and other location based services.

The tracked instantaneous speeds are quite accurate with a tendency of underestimation. The error is monotonically increasing with the speed and the inaccurate position recording. It should however be noted that we did not study the accuracy regarding acceleration and deceleration which are common phenomena in ordinary traffic. More overlapping speeds for different transportation modes should be tested as well. Concerning the recorded elevations in the tracking data, we found it to be highly inaccurate and we suggest disregarding this parameter in practical use until further investigations.

The reliability seems to be unrelated to the sampling frequency. Of course, intensive positional recordings provide more details regarding the mobility pattern. However, it comes at the expense of more aggressive data rendering communication, storage, data processing, data mining, and data analysis. Balancing between these aspects is necessarily specific to the domain of application.

There is drawback of GPS receivers due to a short effective lifespan of the battery (Ryan et al., 2004; Stopher and Speisser, 2011). The data collection part of the experiment in this paper lasted at the most for three hours; the duration of the receivers was not a concern here as the operational time for the receiver is about 11 hours after being fully charged and in continuous mode. However, the lifespan may be a costly drawback in full-scale applications especially for the smart phone (Bierlaire et al., 2013).

Finally, this study examined three specific standard GPS receivers. It would be interesting in the future to conduct further analyses including other types of GPS receiver on larger sample sizes, longer recording periods and more possible environment settings by using the experimental method outlined in this study. What's more, the ground-truth data used for evaluation are inaccurate in a limited tolerance, the choice of the "true value" and the confidence assigned to them has become an general issue to be considered in evaluating the reliability of GPS-tracking data.

References

- Bierlaire, M., Chen, J., & Newman, J. (2013). A probabilistic map matching method for smartphone GPS data. *Transportation Research Part C: Emerging Technologies*, 26, 78-98.
- Chen, C., Gong, H., Lawson, C., & Bialostozky, E. (2010). Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study. *Transportation Research Part A: Policy and Practice*, 44(10), 830-840.
- Chang, Y. J., Paruthi, G., & Newman, M. W. (2015, September). Challenges in using wearable cameras and phone logs to generate ground truth of transportation activities. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (pp. 1361-1366). ACM.
- Djuknic, G. M., & Richton, R. E. (2001). Geolocation and assisted GPS. *Computer*, 34(2), 123-125.
- Herrera, J. C., Work, D. B., Herring, R., Ban, X. J., Jacobson, Q., & Bayen, A. M. (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4), 568-583.

- Huang, D. (2013). *Evidential problems with GPS accuracy: device testing* (Doctoral dissertation, Auckland University of Technology).
- Jia, T., Carling, K. & Håkansson, J. (2013). Trips and their CO₂ emissions to and from a shopping centre. *Journal of Transport Geography*, 33, 135-145.
- Kamboj, S., & Dahiya, R. (2011). Application of GPS for Sag Measurement of Overhead Power Transmission Line. *Intern. J. on Electrical Engineering and Informatics*, 3(3), 268-277.
- Leduc, G. (2008). *Road traffic data: Collection methods and applications*. Working Papers on Energy, Transport and Climate Change, 1, 55.
- Li, X., Ge, M., Dai, X., Ren, X., Fritsche, M., Wickert, J., & Schuh, H. (2015). Accuracy and reliability of multi-GNSS real-time precise positioning: GPS, GLONASS, BeiDou, and Galileo. *Journal of Geodesy*, 89(6), 607-635.
- Lindsey, G. et al. (2013). *Feasibility of Using GPS to Track Bicycle Lane Positioning*. Technical report CTS 13-16. Intelligent Transportation Systems Institute, Center for Transportation Studies, University of Minnesota.
- Marchal, F., Hackney, J., & Axhausen, K. W. (2005). Efficient map matching of large global positioning system data sets: Tests on speed-monitoring experiment in Zürich. *Transportation Research Record: Journal of the Transportation Research Board*, 1935(1), 93-100.
- Meng, Y., Li, Z., Chen, W., & Chen, Y. (2004, December). Reliability and Integrity Issues for Vehicle positioning Systems. In *the 2004 International Symposium on GNSS/GPS, Sydney, Australia*.
- Modsching M., Kramer R., & ten Hagen K. (2006, March). Field trial on GPS Accuracy in a medium size city: The influence of built-up. In *3rd Workshop on Positioning, Navigation and Communication*, 209-218.
- Montini, L., Prost, S., Schrammel, J., Rieser-Schüssler, N., & Axhausen, K. W. (2015). Comparison of travel diaries generated from smartphone data and dedicated GPS devices. *Transportation Research Procedia*, 11, 227-241.
- Murakami, E., & Wagner, D. P. (1999). Can using global positioning system (GPS) improve trip reporting? *Transportation research part c: emerging technologies*, 7(2), 149-165.
- Noronha, V., & Goodchild, M. F. (2000). Map accuracy and location expression in transportation—reality and prospects. *Transportation Research Part C: Emerging Technologies*, 8(1), 53-69

- Quiroga, C. A., & Bullock, D. (1999). Travel time information using global positioning system and dynamic segmentation techniques. *Transportation Research Record: Journal of the Transportation Research Board*, 1660(1), 48-57.
- Quddus, M. A., Ochieng, W. Y., & Noland, R. B. (2007). Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation Research Part C: Emerging Technologies*, 15(5), 312-328.
- Ryan, P. G., Petersen, S. L., Peters, G., & Grémillet, D. (2004). GPS tracking a marine predator: the effects of precision, resolution and sampling rate on foraging tracks of African Penguins. *Marine Biology*, 145(2), 215-223.
- Shen, L., & Stopher, P. R. (2014). Review of GPS travel survey and GPS data-processing methods. *Transport Reviews*, 34(3), 316-334.
- Schipperijn, J., Kerr, J., Duncan, S., Madsen, T., Klinker, C. D., & Troelsen, J. (2014). Dynamic accuracy of GPS receivers for use in health research: a novel method to assess GPS accuracy in real-world settings. *Frontiers in public health*, 2.
- Schönfelder, S. & Antille, N. (2002). *Exploring the Potentials of Automatically Collected GPS Data for Travel Behaviour Analysis: A Swedish Data Source*. ETH, Eidgenössische Technische Hochschule Zürich, Institut für Verkehrsplanung, Transporttechnik, Strassen-und Eisenbahnbau IVT.
- Schüssler, N. and Axhausen, K.W. (2009a). *Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT)*. Working Paper, 568, IVT, ETH Zurich.
- Schüssler, N. and Axhausen, K.W. (2009b). Processing GPS raw data without additional information. *Transportation Research Record*, 2105, pp. 28-36.
- Skog, I., & Handel, P. (2009). In-car positioning and navigation technologies—A survey. *Intelligent Transportation Systems, IEEE Transactions*, 10(1), 4-21.
- Shoval, N. (2008). Tracking technologies and urban analysis. *Cities*, 25(1), 21-28.
- Sun, Z., & Ban, X. J. (2013). Vehicle classification using GPS data. *Transportation Research Part C: Emerging Technologies*, 37, 102-117.
- Stopher, P., FitzGerald, C., & Xu, M. (2007). Assessing the accuracy of the Sydney Household Travel Survey with GPS. *Transportation*, 34(6), 723-741.

- Stopher, P. R., & Speisser, N. (2011). Evaluation of gps device properties for a possible use in future household travel surveys. In *Transportation Research Board 90th Annual Meeting* (No. 11-4271).
- Stopher, P., Greaves, S., & Shen, L. (2013, September). Comparing two processing routines for GPS traces—lessons learnt. In *Proceedings of the 36th Australasian Transport Research Forum ATRF*.
- Theiss, A., Yen, D. C., & Ku, C. Y. (2005). Global Positioning Systems: an analysis of applications, current development and future implementations. *Computer Standards & Interfaces*, 27(2), 89-100.
- Van der Spek S., Van Schaick J., De Bois P. & De Haan R. (2009). Sensing human activity: GPS tracking. *Sensors*, 9(4), 3033-3055.
- Zandbergen, P. A. (2009). Accuracy of iPhone locations: A comparison of assisted GPS, WiFi and cellular positioning. *Transactions in GIS*, 13(s1), 5-25
- Zheng, Y., Li, Q., Chen, Y., Xie, X., & Ma, W. Y. (2008, September). Understanding mobility based on GPS data. In *Proceedings of the 10th international conference on Ubiquitous computing* (pp. 312-321). ACM.
- Zito, R., d'Este, G., & Taylor, M. A. (1995). Global positioning systems in the time domain: How useful a tool for intelligent vehicle-highway systems? *Transportation Research Part C: Emerging Technologies*, 3(4), 193-209.
- Zumberge J. F., Heflin M. B., Jefferson D. C., Watkins M. M. & Webb F. H. (1997). Precise point positioning for the efficient and robust analysis of GPS data from large networks. *Journal of Geophysical Research: Solid Earth* (1978–2012), 102(B3), 5005-5017.
- Zhao, X., Carling, K., & Håkansson, J. (2014). *Reliability of GPS based traffic data: an experimental evaluation*. Working papers in transport, tourism, information technology and microdata analysis (2014:17), ISSN: 1650-5581, Dalarna University, Sweden.

Paper II

How does the complexity of a road network affect optimal facility locations?

Xiaoyun Zhao[♦], Pascal Rebreyend, Johan Håkansson

The road network is a necessary component in transportation. It facilitates spatial movements of people and goods, and it also influences the optimal locations of facilities that usually serve as destinations of the movements. To fulfill the transportation needs and to adapt to the facility development, the road network is often organized hierarchically and asymmetrically with various road levels and spatial structures. The complexity of the road network increases along with the increase of road levels and spatial structures. However, location models locate facilities on a given road network, usually the most complex one, and the influence from the complexity of road network in finding optimal locations is not well-studied. This paper aims to investigate how the complexity of a road network affects the optimal facility locations by applying the widely-applied p -median model. The main result indicates that an increase in road network complexity, up to a certain level, can obviously improve the solution, and the complexity beyond that level does not always lead to better solutions. Furthermore, the result is not sensitive to the choice of algorithms. In a specific case study, a detailed sensitivity analysis of algorithm and facility number further provides insight into computation complexity and location problems from intra-urban to inter-urban.

Key Words: Transportation system; Spatial optimization; Location models; Heuristics

1. Introduction

Movements of people and goods take place between origins and destinations on the road network; therefore, the road network is a necessary component in facilitating these spatial movements. A facility usually serves as a destination of the movements, thus, a facility optimally located on the road network is a desired goal in many decision-making processes since it minimizes travel distance (e.g. Carling et al., 2012). However, the road network is often organized hierarchically and asymmetrically with various road levels and spatial structures in order to fulfill transportation needs and to adapt to facility development. The complexity of the road network increases along with the increase of road levels and spatial structures, thus finding the optimal facility locations could become computationally complicated and time-consuming.

In locating facilities, the road network is usually given (Melkote and Daskin, 2001). Peeters and

[♦] Xiaoyun Zhao is a PhD-student in Micro-data analysis and corresponding author: xzh@du.se, phone: +46 23-778509. Pascal Rebreyend is an associate professor in Computer Science, and Johan Håkansson is a professor in Geography. The authors are all at the School of Technology and Business Studies at Dalarna University, Sweden

Thomas (1995) claimed that the more complex a road network, the higher probability of identifying a more optimal solution. Following which, many studies always work with a most complex road network and focus on aggregating the demand points and using a simple distance measure, such as the Euclidean distance, to simplify the problem and make it computationally treatable (e.g. Schilling et al., 2000; Francis et al., 2009). In addition, the given road network in these studies is simulated, and the number of facility locations is limited to represent various scenarios. In using a real road network for finding optimal locations in rural areas, Carling et al. (2012, 2015) demonstrate that the level of aggregation needs to be more sophisticated and the distance measure needs to be as accurate as possible in order to avoid suboptimal locations in areas where demands and road network are asymmetrically distributed. Few studies use a real road network with changing complexity, along with the road levels and the spatial structures, in finding optimal facility locations in both intra-urban and inter-urban scenarios.

In a pre-test conducted by Rebreyend et al. (2013), the complexity of the road network is found influential in finding the optimal location in rural areas. However, the trade-off among more optimal solutions, computational complexity and spatial distributions of facilities in intra-urban and inter-urban scenarios has not yet been studied.

This paper aims to investigate how the changes in road network complexity influence the optimal facility locations by applying the widely used p -median model. In addition, to provide further insight into computation complexity and location problems from intra-urban to inter-urban, a detailed sensitivity analysis of four algorithms and various facility numbers is conducted.

The four algorithms are: greedy search (Kuehn and Hamburger, 1963); CPLEX; simulated annealing (Al-Khedhairi, 2008); and imp-Genetic algorithm (Rebreyend et al., 2015). Each algorithm works in a different way in solving the p -median model. The facility numbers vary from 5 to 50 (with a common difference of 5) to simulate intra-urban and inter-urban location distributions in the case of Dalarna province, Sweden. The size of Dalarna, as well as its structure, is similar to the regions of Vermont and New Hampshire in the US.

The entire road network database provided by the official National Road Database (NVDB)¹ in Sweden consists of 2.6 million roads and 30 million nodes, where 0.15 million roads and 1.8 million nodes are distributed in the Dalarna region. The road network in the Dalarna region has a hierarchical structure composed of ten levels of road. The European highways are at the biggest level, representing the simplest road networks, where there is a limited number of candidate nodes for

¹ www.nvdb.se

locating facilities. The local and private streets are at the smallest level and represent the most complex road network, with a huge number of candidate nodes for locating facilities.

The remaining part of the paper is organized as follows. In Section 2, we present the p -median model and some related previous studies that guide us in the model choice. Section 3 describes the data applied in the case study and settings for the sensitivity check. Section 4 illustrates the main results, and Section 5 concludes the paper.

2. The p -median model

The p -median model was first introduced by Hakimi (1964), and it has since then been widely used in transportation and location-allocation studies (Farahani, et al., 2012). Given that demand nodes are fixed in the network, the model finds the optimal facility locations that minimize the total travel distance for all demand points to the closest facility. Specifically, the objective function is to minimize the sum of weighted distances between demand points and their respective nearest facilities.

$$\text{Minimize } f = \sum_{i=1}^Q \sum_{j=1}^N w_i * d_{ij} * X_{ij} \quad (1)$$

Subject to:

$$\sum_{j=1}^N X_{ij} = 1, \forall i \in (1,2, \dots \dots Q) \quad (2)$$

$$\sum_{j=1}^N Y_j = p, p \in N \quad (3)$$

Decision variables:

$$X_{ij} = \begin{cases} 1 & \text{if demand at node } i \text{ is allocated to facility } j \\ 0 & \text{if not} \end{cases} \quad (4)$$

$$Y_j = \begin{cases} 1 & \text{if a facility is located at candidate site } j \\ 0 & \text{if not} \end{cases} \quad (5)$$

$$0 \leq X_{ij} \leq Y_j, \forall i \in (1,2, \dots \dots Q), j \in (1,2, \dots \dots N) \quad (6)$$

$$X_{ij} \in \{0,1\}, \forall i \in (1,2, \dots \dots Q), j \in (1,2, \dots \dots N) \quad (7)$$

$$Y_j \in \{0,1\}, \forall j \in (1,2, \dots \dots N) \quad (8)$$

Q = Total number of demand points in the space of interest

N = Total number of candidates for locating facilities

p = Total number of potential facilities

w_i = The weight associate to each demand node i

d_{ij} = The distance between demand node i and potential facility j

The constraint (2) requires that each demand point is assigned to exactly one facility. The constraint (3) ensures that exactly p facility locations are to be chosen among the N candidates. The constraint (6) links the location variables and the allocation variables. Finally, constraints (7), and (8), insure that the location variables(X) and allocation variables(Y) are binary.

To find the optimal location for p facilities to fulfill the demand using the p -median model is NP-hard (Kariv and Hakimi, 1979), and optimal solutions to large problems are difficult to obtain (Al-Khedhairi, 2008). Francis et al. (2009) conducted a literature review of the p -median model, in which about half of the 40 reviewed articles are studies based on real data. The largest number of candidate nodes in the reviewed studies was some 70, 000 candidate nodes. However, p was relatively small (<20) in those studies. The review also showed that almost all distance measures are Euclidean distance and rectilinear distance.

Many recent studies have been developed to illustrate the greater efficiency of network distance versus Euclidean distance, in the analysis of network-constrained objects or phenomena (Yu et al., 2015). Schilling et al. (2000) examined the Euclidean distance, network distance and a randomly generated network distance. Both Euclidean distance and network distance were found to have high computational efficiency and yield better solution quality. The problem scale in their study was, however, small and they did not study a network with different levels of candidate nodes. Although Euclidean distance is most widely used, the network distance in most cases is more accurate in measuring the travel distance between two points, since the Euclidean distance leads to suboptimal solutions in an unpredictable way. Xie and Yan (2008), and Shiode (2011), found that measuring distance by connecting the straight line between locations could possibly overestimate the clustering tendency of the network. The network distance is more appropriate for spatial phenomena or activities constrained by transportation networks, especially in the field of microscopic analysis (Ai et al., 2015). Peeters and Thomas (1995) examined the performance of the p -median model in different network topologies by changing the nature of the links. They found that there is a difference in optimal solutions when the links are changed, but they did not check the computational efforts in finding the optimal solutions. Following them, the topologies of the network have been studied, taking the computational effort into account (Peeters, D., et al., 1998; Melkote and Daskin, 2001; Bigotte et al., 2010; Rahmaniani and Ghaderi, 2013). However, on the one hand, these studies are mainly based on fully connected Euclidean and rectilinear networks or networks with varying numbers of radial and rectilinear arcs. On the other hand, the number of candidate nodes and links are limited, which means that the simulation cannot represent the real road network well.

The complexity of a real road network varies along with the road levels and the spatial structures, which cannot be simply represented by nodes and links in topology. Apart from Rebreyend et al. (2013), very few studies have examined the impact of varying the complexities of road network on the optimal facility locations and their spatial distributions.

3. Data and settings for sensitivity check

3.1 Data

Dalarna is a province located in the middle of Sweden with an area of 28,189 km². The population of Dalarna amounts to 231,934, in December 2015 (Statistic Sweden²). The population is geo-coded and registered on 250 meters by 250 meters squares. The center of each registry square represents one demand point. Each demand point is assigned a weight, corresponding to the population number in that square. There are 15,729 weighted demand points representing the whole population in this region. Figure 1 shows that the population in the studied region is highly asymmetrically distributed, with the majority living in the southeast part.

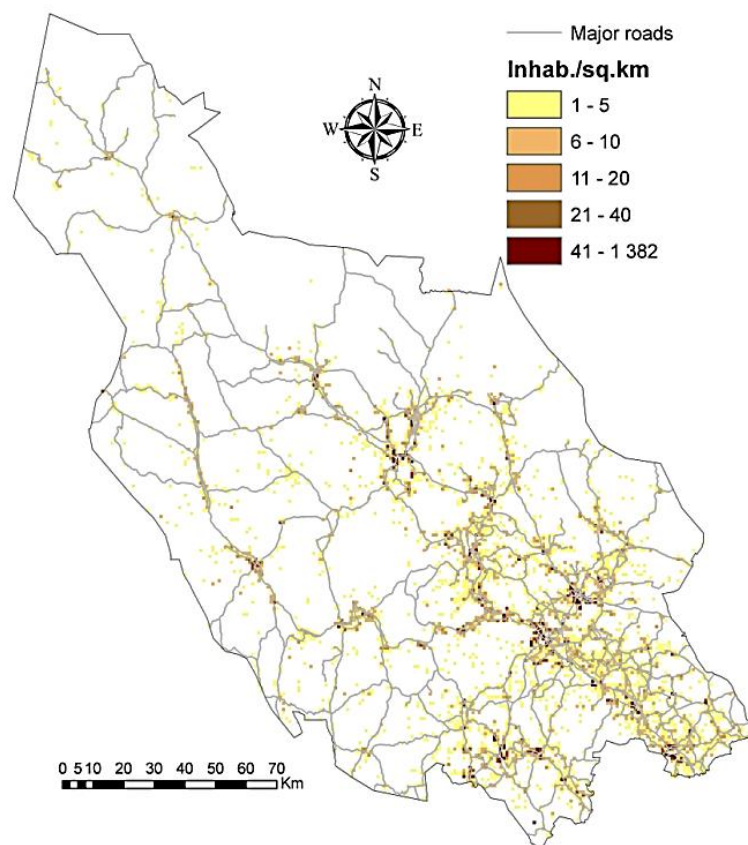


Figure 1. Map of the Dalarna region showing 250-by-250 meter squares of inhabitants

² www.scb.se

Figure 2 shows the complete digitalized representation of the road network in Dalarna, with its hierarchical structure corresponding to 10 road levels. Figure 2(a) shows that the main structure of the road network is composed of levels 0-5, which are European highways, national and regional roads with a total length of 5,479 km. The spatial structure of the road network of these big roads corresponds to the spatial distribution of the population, as illustrated in Figure 1. The European highway forms the simplest road network with a limited number of candidate nodes for locating facilities. While the most complex road network includes all the roads up to the smallest level of the local and private streets, where there is a huge number of candidate nodes for locating facilities. Figure 2(b) shows the small roads at levels 6-9, with a total length of 33,975 km.

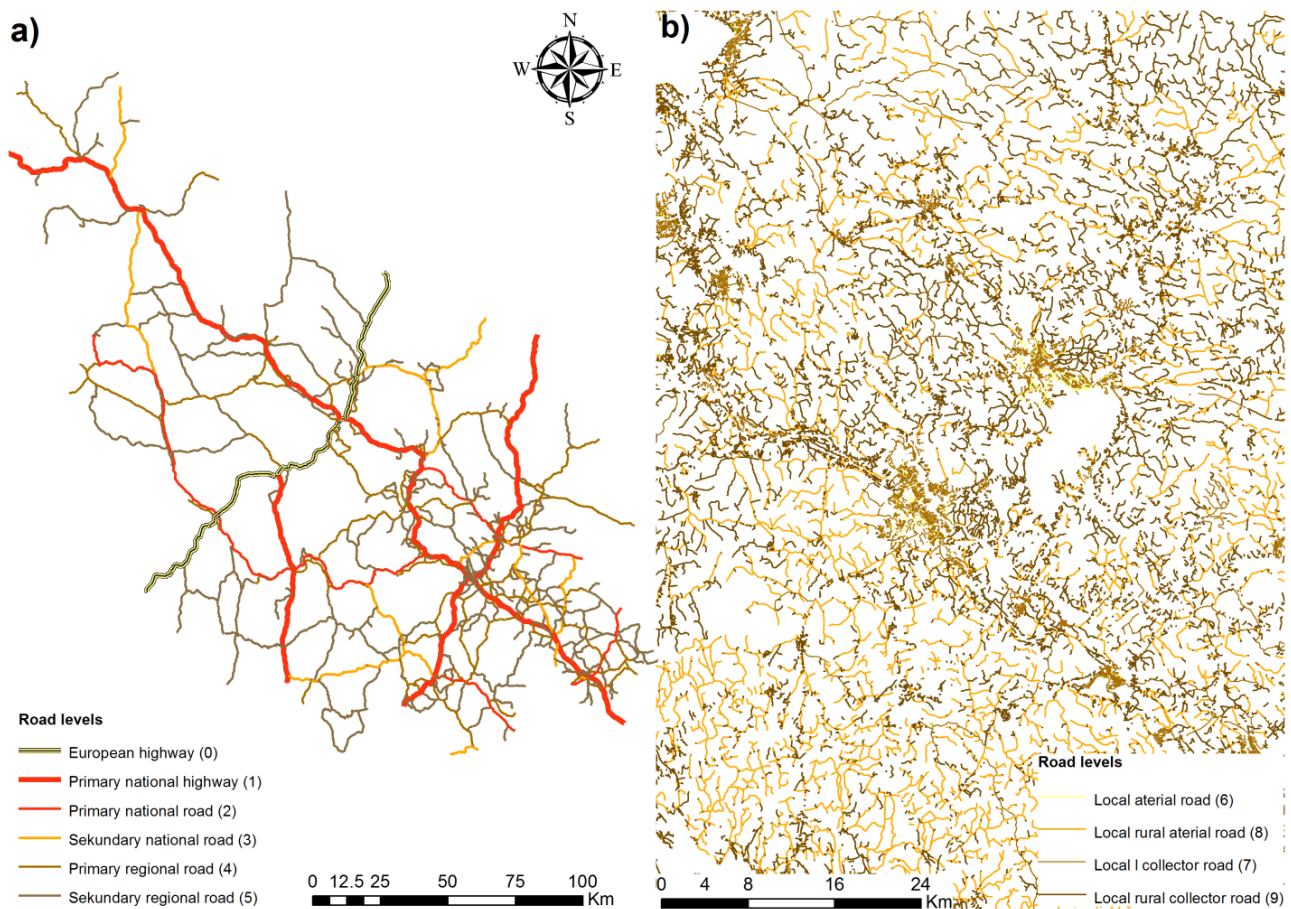


Figure 2. (a) Main structure of the road network corresponding to road levels 0-5 in Dalarna, Sweden
 (b) Local dense road network corresponding to road levels 6-9 in Dalarna, Sweden

Specifically, there are 1,797,939 nodes (the start and the end of the roads), and 1,964,801 road segments in the whole road network in Dalarna. Approximately, one node can be found asymmetrically distributed every 20meters on the road network. In order to have a valid measure to make comparisons between different network densities, the whole network is used to calculate all the network distances between the demand points and the candidate nodes, which can also ensure that

the variation of candidate points from different network densities does not affect the distance measure. However, considering the computation complexity in solving the *NP*-hard *p*-median model, the whole network with all candidate nodes is not practical for use with most locations problems. A 500 by 500 meter grid aggregation on all complexity level roads was conducted for mainly two reasons: one reason is to enable all the complexity levels computationally feasible, and the other reason is to practically stick to the real world situation, since it is very rare to locate a facility every 20 meters.

In order to calculate the network distances, the connectivity of the nodes in the network is first checked. There are 9020 not connected and they are removed. All remaining nodes are tagged to the closest demand origin, or to the closest intersection node. In each grid, at the most, one node is kept as a potential candidate node for optimal location by applying three criteria sequentially. First, we choose the node connects with most road segments. If there are none, we then select the node at the highest complexity level. If there are none there, we at last pick the node which is closest to the center of the grid. Table 1 summarizes the numbers of nodes according to the complexity of the road network; the number of nodes increases when the complexity of the road network increases by including more levels of roads. There are 452 nodes in the simplest road network, and 67, 020 nodes in the most complex road network.

Table 1. Number of nodes in different road network complexities that correspond to the included road levels under the grid aggregation level of 500m by 500m in Dalarna.

Road Densities	Complexity Level	Number of Nodes	Length (km)
0	European highway	452	167
0 – 1	+Primary national highway	1,994	883
0 – 2	+Primary national road	2,909	1,299
0 – 3	+Secondary national road	3,926	1,725
0 – 4	+Primary regional road	6,735	2,923
0 – 5	+Secondary regional road	12,417	5,479
0 – 6	+Local arterial road	12,552	5,631
0 – 7	+Local collector road	20,718	10,964
0 – 8	+Local rural arterial road	45,336	23,086
0 – 9	+Local rural collector road	67,020	39,454

As the coordinates of the residents (demand points) do not perfectly coincide with the nodes in the road network, correspondingly, we use the nearest node in the network to represent the location of residents. The average distance between the residents' node and the nearest network node is 62 meters, which means the approximation does not introduce much difference to the final results.

The travel distance between the demand point and the nearest facility is one key variable that is in the objective function of the *p*-median model, therefore, the distance measure influences the

solutions for optimal facility locations. Carling et al. (2012, 2015) investigated, empirically, the consequences of different distance measures for the optimal location of multiple service centers in rural areas. They stated that the shortest travel time or minimal cost along an existing network intuitively seems to be the most accurate measure for most settings, yet it is infrequently employed. One reason for this is the difficulty and cost associated with collecting data on travel time. Another reason is the complication which arises in modelling the inherent variation in travel time. This paper employs network distance as the distance measure.

3.2 Settings for sensitivity check

Heuristics and approximation algorithms are the predominant techniques used for solving the p -median location problem, as described and explored in the recent literature (Tansil et al., 1983; Rees, 2006; Mladenović et al., 2007; Varnamkhasti, 2012). Four algorithms that work rather differently in solving the p -median model to find the solutions of optimal facility locations are chosen: CPLEX; greedy search (Kuehn and Hamburger, 1963); simulated annealing (Al-Khedhairi, 2008); and imp-Genetic algorithm (Rebreyend et al., 2015).

If the p -median model is formulated as a 0-1 binary programming problem it can then be solved by a Mixed Integer Problem (MIP) solver by using a branch-and-cut approach. The CPLEX from IBM is a commonly used software package for solving optimization problems. Following Rebreyend et al. (2015), some parameters of the solver have been tuned in order to adapt CPLEX to work on large problem instances, specifically, removing default computation time limits, allowing intermediate data storage, and tuning branch-and-cut search tree strategies, according to the manual³.

The standard greedy algorithm for the p -median model was studied by Cornuejols et al. (1977). Resende and Werneck (2004) conducted a constructive greedy algorithm to perform the most profitable move among candidates to get the best local minimum in the path. The Greedy algorithm follows the problem-solving heuristic of finding the locally optimal choice at each stage, with the hope of getting a global optimum. It always chooses the optimal choice at the current stage, rather than considering all other conditions to ensure finding a global optimal. This can be characterized as being ‘short sighted’, but it is easy to implement and can achieve acceptable results within a short time.

Simulated annealing (SA) is one commonly used heuristics for solving the p -median model. Murray and Church (1996) proposed a basic SA algorithm for the p -median model. Levanova and

³ http://www-01.ibm.com/support/knowledgecenter/SSSA5P_12.6.1/ilog.odms.studio.help/pdf/gscplex.pdf

Loresh (2004) studied the SA heuristic that used the 1-interchange neighborhood structure. Carling et al. (2012, 2015), and Han et al. (2013) used tuned SA to solve specific p -median model in a real road network context. The basic idea of SA is not only accepting all the better results in the search process, but also accepting some worse results based on certain probabilities. It is simple to implement and can provide high quality solutions to many problems. The performances of SA are sensitive to the values of control parameters. In this study, we employ the same parameters of SA to the same real world network data as employed by Han et al. (2013). The specific parameter settings and the implementation of SA follow Zhao et al. (2013). This includes their dynamic scheme to update the temperature, which allows the algorithm to have setting for efficiency and accuracy, regardless of the size of the input. After testing with various parameter settings, we found that the scheme used to increase and decrease the temperature works well to avoid the search being trapped in the local optimal for a long time, and finally provides satisfying results.

Genetic algorithms (GAs) are another commonly used heuristics that are designed based on mimicking the evolution process. New solutions are based on previous solutions in ways that are reminiscent of the interaction of genes. Most previous studies used a classical string representation, in which each chromosome is represented as a single string of length p , embedding the index of the selected facilities or nodes. Thorough treatments of GAs can be found in (Davis, 1987, 1991; Reeves, 1993; Dowsland, 1996). Bozkaya et al. (2002) found GAs could produce solutions that are better than exchange algorithms; however, the convergence is very slow. Alp et al. (2003) proposed a GA which is simpler and produces good solutions faster. Following Correa et al. (2001), Rebreyend et al. (2015) proposed an improved genetic algorithm, called imp-GA, to solve large-scale p -median problems. The imp-GA used in this paper follows the description and settings as used by Rebreyend et al. (2015).

The facility numbers vary from 5 to 50 (with a common difference of 5) to simulate the intra-urban to inter-urban location distributions. We run CPLEX and Greedy once, due to the deterministic property of the methods. Since SA is sensitive to its starting point, we randomly select 3 different initial configurations to conduct the SA, and keep the solution with the minimum objective function value (OFV); each run contains 20,000 iterations. As for imp-GA, we set 3 runs with 100 iterations, due to its higher requirement for computation time and computer memory. Max running time is 48 hours for each p and network complexity level. In this paper, the CPLEX version 12.6, Linux 64 bits is used. All the programs are coded with C and compiled using GCC version 4.8.2, and they are launched under a system of Linux (Kernel 3.11-2-amd64). The computer has a memory of 32 G, CPU of Intel Core i7-3770.

4. Results

4.1 Sensitivity to complexity levels

Figure 3 shows how worse the optimal solutions found at each complexity level of the road network compared to the best solution found for a selection of p (results for all tested p can be found in Appendix, Figures A1-A3). Specifically, the improvement of the solution is measured by checking the difference between the OFV in the current complexity level, and the optimal OFV found at the same/other level ($\frac{|\text{current solution} - \text{best solution}|}{\text{current solution}} * 100\%$).

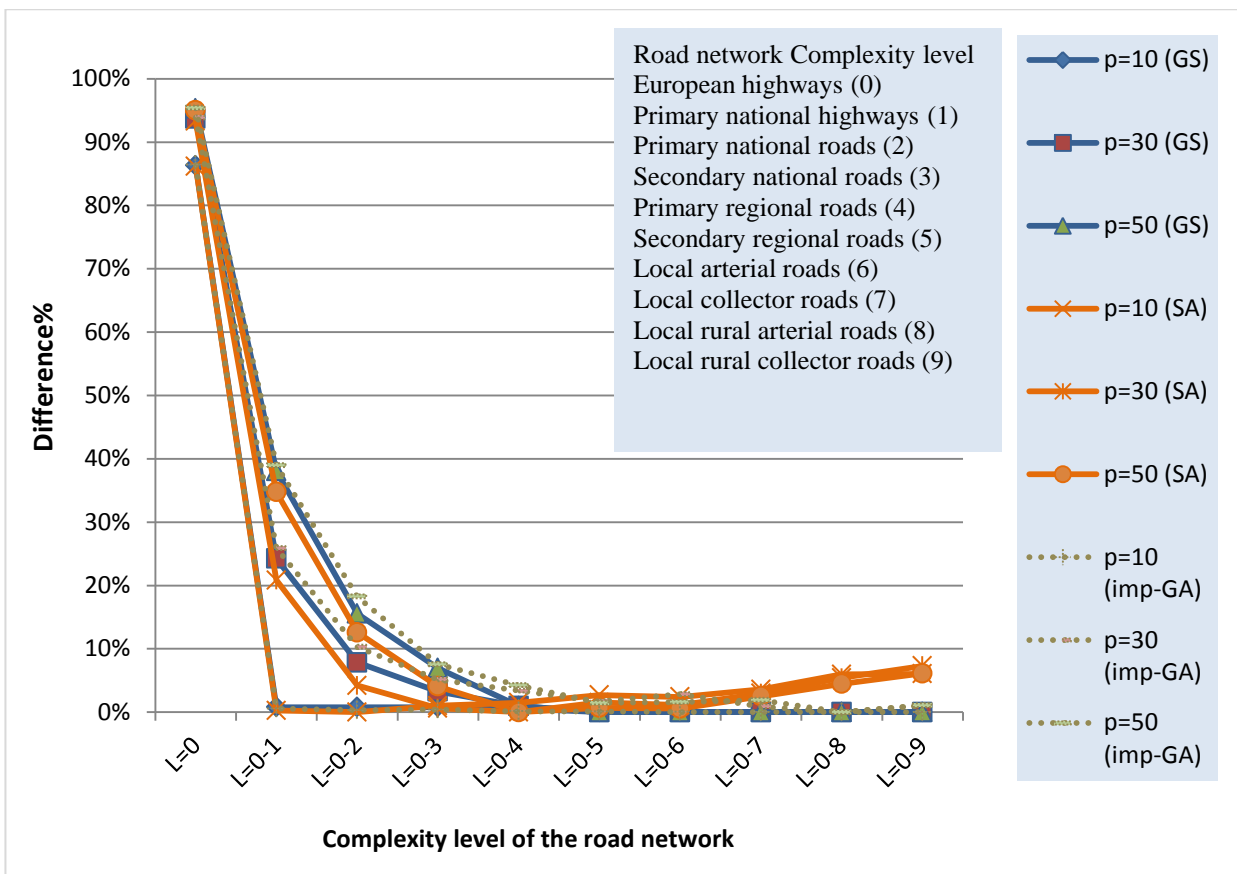


Figure 3. Variations from algorithms Greedy search, SA and imp-GA in excess distances (in percent), compared to the best solutions for an increased complexity level. The x-axis shows the complexity level. The y-axis shows the difference in percentage between the best solution and the current.

The figure reveals that the solutions found from candidate nodes on the simplest road network, namely, the European highways (level 0), are 85 to 95 percent worse than the optimal solutions that can be found. However, the figure also shows that there is a limit to how complex the road network needs to be to find the best solutions. In general, when the complexity of the road network continues to increase after including road level 5, the solutions barely improve, and even deteriorate.

4.2 Sensitivity to algorithms

As is shown in Figure 3, in general, the algorithms behave in a similar pattern of obvious improvement when the complexity increases from the simplest level 0 to levels up to, and including, 0-5 (ca. 12,500 nodes). When the complexity level increases from 0-5, the OFVs from SA deteriorate, the OFVs from the Greedy search neither improve nor deteriorate, and the OFVs from imp-GA show some very small improvements (also see Table 2).

Table2. The optimal objective function value found on different road network densities by tested algorithms for various p .

p	Greedy		SA		imp-GA		CPLEX
	Optimal solution in meters	Included complexity level	Optimal solution in meters	Included complexity level	Optimal solution in meters	Included complexity level	
5	19685.104	0-5	19715.020	0-2	19624.095	0-5	81681 Only provide results at level 0 and fails at upper levels
10	11134.182	0-5	11259.282	0-2	11075.455	0-7	
15	8316.884	0-5	8625.458	0-4	8278.562	0-8	
20	6698.967	0-5	7094.882	0-6	6631.992	0-5	
25	5804.399	0-5	6155.251	0-5	5715.777	0-6	
30	5122.226	0-5	5465.033	0-4	4998.492	0-8	
35	4708.024	0-5	5013.738	0-6	4600.623	0-9	
40	4349.791	0-5	4611.287	0-4	4250.913	0-8	
45	4049.786	0-5	4306.700	0-4	3960.126	0-8	
50	3796.301	0-5	4066.063	0-4	3738.607	0-8	
Results from all 10 complexity levels							

Table 2 further shows how the best OFVs derived from the four algorithms vary among these algorithms. Note that the CPLEX algorithm that gives the exact solutions fails to produce a result higher than complexity level 0. The reason for this is that the location problem addressed in this study is too complex, even when the representation of the road network is at the simplest level (level 0). Imp-GA gives the best solutions. However, for the same facility number, the improvements of the OFVs are rather small compared to Greedy and SA, and the largest improvement is only 2.78%, i.e. approximately 122 meters shorter in network travel distance. Table 2 also shows that the algorithms show a different requirement of network complexity for finding better solutions by including the road levels. However, the differences between the Greedy and imp-GA OFVs are just 1-2 percent. For $p=20$, or larger, the SA gives an OFV that is 8 to 9 percent worse than the OFV from the imp-GA. Whereas, the computation time and computation effort of imp-GA become tremendously large, so that the results cannot be derived for the most complex of the road networks within the time threshold of 48 hours, which is not the case for Greedy and SA.

4.3 Sensitivity to facility numbers

Based on Figure 3 and Table 2, the changes of facility numbers do not introduce obvious changes either in the complexity level or among the algorithms. Figure 4 visualizes the spatial distribution of the optimal locations found by imp-GA.

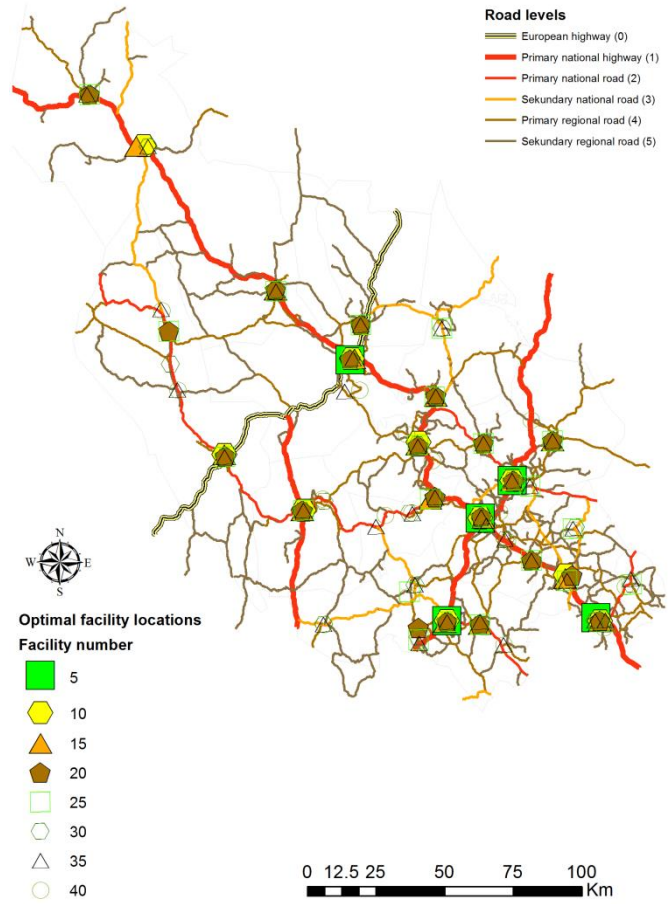


Figure 4. Distribution of optimal loctions (p from 5 to 40) found by imp-GA on the study area Dalarna

In general, the optimal locations found when p is low remain as optimal locations when p increases. The locations are mostly on the major roads in the road network and correspond very well with the complexity structure, as shown in Figure 2(a). This further indicates that the increase in complexity of the road network above level 0-5 does have a limited influence on the spatial distribution of the optimal facility locations. To make the location-allocation practical in transportation and land-use planning, a location that decreases the travel distance a few hundred meters is not appealing, compared to the risk of losing the attraction of customers due to poor accessibility on a denser road network, especially when the road network is asymmetrically distributed.

5. Concluding discussion

This paper aims to investigate how the changes of road network complexity influence the optimal facility locations by applying the widely used p -median model. In addition, to provide further insights on computation complexity and location problems from intra-urban to inter-urban, a detailed sensitivity analysis of four algorithms and various facility numbers is conducted on a specific case study, namely, Dalarna province, Sweden.

The road network has hierarchical road levels and an asymmetrical structure, where the European highway forms the simplest road network with 452 candidate nodes, while the most complex road network includes all the roads up to the smallest level of the local and private streets, with 67, 020 candidate nodes.

The main result indicates that there is a limit to how complex the road network needs to be to find the best solutions. When the complexity of the road network continues to increase after a certain level the solutions barely improve, but even deteriorate. We found that the choice of algorithms and facility number p has little to no impact on the results, although due to differences in the mechanics of the algorithm, there are some variations among the algorithms. The most complex level of road network (including local streets) is not preferable, especially when the number of the facility is less than 20. The spatial distributions of the optimal locations show that optimal locations found for smaller p remain as part of the optimal locations when p increases.

The specific complexity level that is found efficient here might be case-specific, and additional empirical analysis on other cases is needed. In the p -median model, people are assumed to always choose the closest facility and neglect the multiple purpose travel, or the heterogeneous preferences on facilities. Therefore, using the gravity p -median model (Drezner and Drezner, 2007) would provide more insight into the impact of road network complexity on optimal facility locations for aiding the decision-making in location and transportation planning.

References

- Ai, T., Yu, W., & He, Y. (2015). Generation of constrained network Voronoi diagram using linear tessellation and expansion method. *Computers, Environment and Urban Systems*, 51, 83-96.
- Al-Khedhairi. (2008). Simulated annealing metaheuristic for solving p -median model. *International Journal of Contemporary Mathematical Sciences*, 3(28), 1357-1365.
- Alp, O., Erkut, E., & Drezner, Z. (2003). An efficient genetic algorithm for the p -median model. *Annals of Operations Research*, 122(1-4), 21-42.

- Barahona, F., & Anbil, R. (2000). The volume algorithm: producing primal solutions with. *Mathematical Programming*, 87(3), 385-399.
- Bigotte, J. F., Krass, D., Antunes, A. P., & Berman, O. (2010). Integrated modeling of urban hierarchy and transportation network planning. *Transportation Research Part A: Policy and Practice*, 44(7), 506-522.
- Bozkaya, B., Zhang, J., & Erkut, E. (2002). An efficient genetic algorithm for the p-median model. *Facility Location: Applications and Theory*, 179-205.
- Carling, K., Han, M., & Håkansson, J. (2012). Does euclidean distance work well when the p-median model is applied in rural areas? *Annals of Operations Research*, 201(1), 83-97.
- Carling, K., Han, M., Håkansson, J., & Rebreyend, P. (2015). Distance measure and the p-median model in rural areas. *Annals of Operations Research*, 226(1), 89-99.
- Cornuejols, G., Fisher, M. L., & Nemhauser, G. L. (1977). Exceptional paper-location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. *Management Science*, 23(8), 789-810.
- Correa, E. S., Steiner, M. T., Freitas, A. A., & Carnieri, C. (2001). A genetic algorithm for the P-median model. In *Proc. 2001 Genetic and Evolutionary Computation Conf.(GECCO-2001)*, (pp. 1268-1275).
- Davis, L. (1987). *Genetic algorithms and simulated annealing*. Morgan Kaufman Publishers, Inc., Los Altos, CA.
- Davis, L. (1991). *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York.
- Dowland, K. A. (1996). Genetic algorithms-a tool for OR? *Journal of the Operational Research Society*, 550-561.
- Drezner, T., & Drezner, Z. (2007). The gravity p-median model. *European Journal of Operational Research*, 179(3), 1239-1251.
- Farahani, R., Asgari, N., Heidari, N., Hosseini, M., & Goh, M. (2012). Covering problems in facility location: A review. *Computers and Industrial Engineering*, 62(1), 368-407.
- Francis, R., Lowe, T., Rayco, M., & Tamir, A. (2009). Aggregation error for location models: survey and analysis. *Annals of Operations Research*, 167(1), 171-208.
- Hakimi, S. (1964). Optimum locations of switching centers and the absolute centers and medians of graph. *Operations Research*, 12(3), 450-459.
- Kariv, O., & Hakimi, S. (1979). An algorithmic approach to network location problems. II: The p-

- medians. *SIAM Journal on Applied Mathematics*, 37(3), 539-560.
- Kuehn, A. A., & Hamburger, M. J. (1963). A heuristic program for locating warehouses. *Management Science*, 9(4), 643-666.
- Levanova, T. Y., & Loresh, M. A. (2004). Algorithms of ant system and simulated annealing for the p-median model. *Automation and Remote Control*, 65(3), 431-438.
- Melkote, S., & Daskin, M. S. (2001). An integrated model of facility location and transportation network design. *Transportation Research Part A: Policy and Practice*, 35(6), 515-538.
- Mladenović, N., Brimberg, J., Hansen, P., & Moreno-Pérez, J. A. (2007). The p-median model: A survey of metaheuristic approaches. *European Journal of Operational Research*, 179(3), 927-939.
- Peeters, D., & Thomas, I. (1995). The effect of spatial structure on p-median results. *Transportation Science*, 29(4), 366-373.
- Peeters, D., Thisse, J. F., & Thomas, I. (1998). Transportation Networks and the location of Human Activities. *Geographical Analysis*, 30(4), 355-371.
- Rahmaniani, R., & Ghaderi, A. (2013). A combined facility location and network design problem with multi-type of capacitated links. *Applied Mathematical Modelling*, 37(9), 6400-6414.
- Rebreyend, P., Han, M., & Håkansson, J. (2013). How does different algorithm work when applied on the different road networks when optimal location is searched for in rural areas? *Web Information Systems Engineering-WISE 2013 Workshops* (pp. 284-291). Nanjing, China: Springer Berlin Heidelberg.
- Rebreyend, P., Lemarchand, L., & Euler, R. (2015). A Computational Comparison of Different Algorithms for Very Large p-median models. *Evolutionary Computation in Combinatorial Optimization*, 13-24.
- Reese, J. (2006). Solution methods for the p-median model: An annotated bibliography. *Networks*, 48, 125-142. doi:10.1002/net.20128
- Reeves, C. R. (1993). Using Genetic Algorithms with Small Populations. In *Proceedings of 5th International Conference on Genetic Algorithms*, S. Forrest (ed.) Morgan Kaufmann, San Mateo, CA, 92-99.
- Resende, M. G., & Werneck, R. F. (2004). A hybrid heuristic for the p-median model. *Journal of Heuristics*, 10(1), 59-88.
- Schilling, D., Rosing, K., & Reville, C. (2000). Network distance characteristics that affect computational effort in p-median location problems. *European Journal of Operational Research*,

127(3), 525-536.

Shiode, S. (2011). Street-level spatial scan statistic and STAC for analysing street crime concentrations. *Transactions in GIS*, 15(3), 365–383.

Tansel, B. C., Francis, R. L., & Lowe, T. J. (1983). State of the art—location on networks: a survey. Part I: the p-center and p-median models. *Management Science*, 29(4), 482-497.

Varnamkhasti, M. J. (2012). Overview of the algorithms for solving the p-median facility location problems. *Advanced Studies in Biology*, 4(2), 49-55.

Yu, W., Ai, T., & Shao, S. (2015). The analysis and delimitation of Central Business District using network kernel density estimation. *Journal of Transport Geography*, 45, 32–47.

Xie, Z., & Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers. Environment and Urban Systems*, 35(5), 396–406.

Zhao, X., Carling, K., Dan, Z., & Håkansson, J. (2013). *Network density and the p-median solution*. Working papers in transport, tourism, information technology and microdata analysis (2013:25), ISSN: 1650-5581, Dalarna University, Sweden.

Appendix

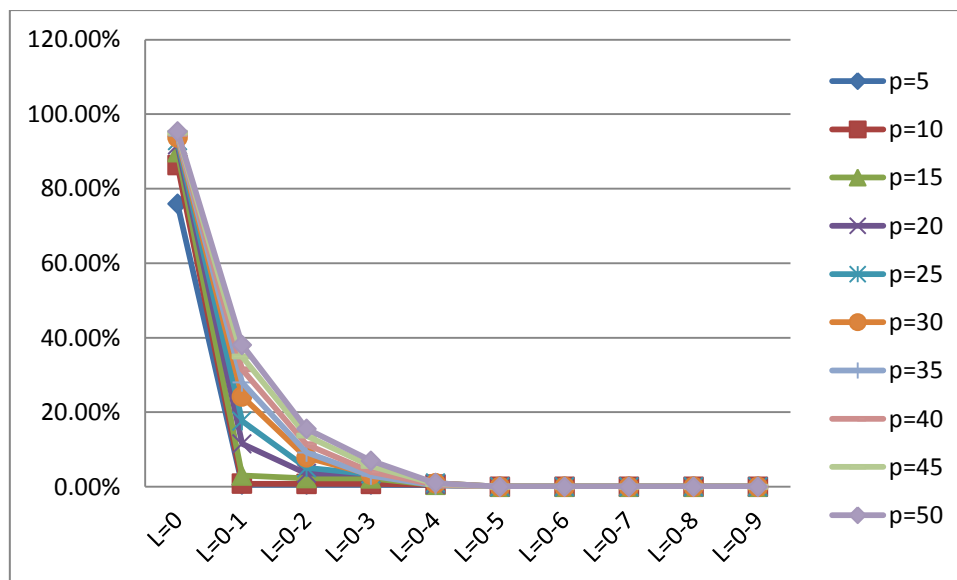


Figure A1. Variations from Greedy in excess distances (in percent) compared to the best solutions for an increased complexity level. The x-axis shows the complexity level. The y-axis shows the difference in percentage between the best solution and the current solution in accordance to $(\frac{[\text{current solution} - \text{best solution}]}{\text{current solution}} * 100\%)$.

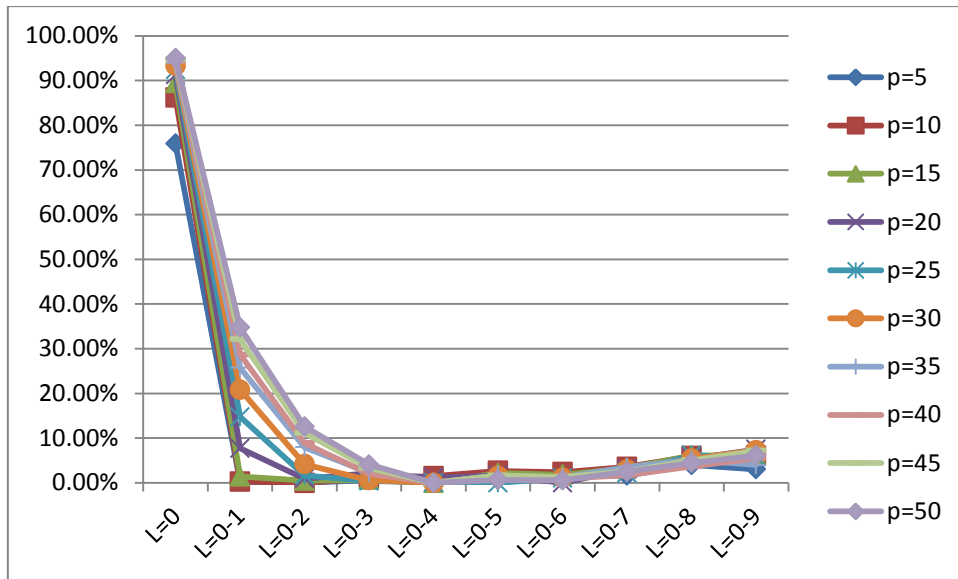


Figure A2. Variations from SA in excess distances (in percent) compared to the best solutions for an increased complexity level. The x-axis shows the complexity level. The y-axis shows the difference in percentage between the best solution and the current solution in accordance to $(\frac{|current\ solution - best\ solution|}{current\ solution} * 100\%)$.

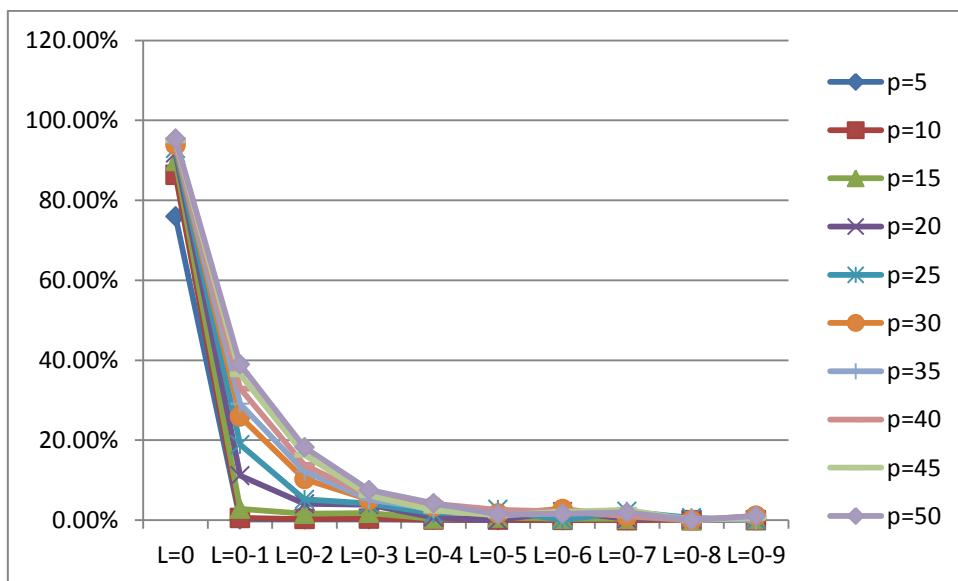


Figure A3. Variations from imp-GA in excess distances (in percent) compared to the best solutions for an increased complexity level. The x-axis shows the complexity level. The y-axis shows the difference in percentage between the best solution and the current solution in accordance to $(\frac{|current\ solution - best\ solution|}{current\ solution} * 100\%)$.

Paper III

ԲՅՈՒՆԵՐԻ
ԲՅՈՒՆԵՐԻ III

On processing GPS tracking data of spatio-temporal car movements: a case study

Xiaoyun Zhao

School of Technology and Business Studies, Dalarna University, Dalarna, Sweden

ABSTRACT

The advancement of GPS technology has made it possible to use GPS devices as orientation and navigation tools, but also as tools to track spatio-temporal information. GPS tracking data can be broadly applied in location-based services, such as spatial distribution of the economy, transportation routing and planning, traffic management and environmental control. Therefore, knowledge of how to process the data from a standard GPS device is crucial for further use. Previous studies have considered various issues of the data processing at the time. This paper, however, aims to outline a general procedure for processing GPS tracking data. The procedure is illustrated step by step by the processing of real-world GPS data of car movements in Borlänge in the centre of Sweden.

ARTICLE HISTORY

Received 14 July 2014
 Accepted 18 September 2015

KEYWORDS

GPS tracking data; visualized map; road network; map-matching

1. Introduction

Global Positioning System (GPS) technology has developed enormously in the last few decades and it continues to improve. The use of a portable device, such as a smartphone or other communication devices with built-in GPS for navigation and orientation, is nowadays common.

GPS devices can also track mobile objects with regard to longitude, latitude, time, speed and altitude at regular time intervals. This spatial and temporal information can be used for investigating the activities of people and their induced effects. It is possible to use the unprocessed GPS tracking data; however, its usage is limited to rather few aspects, namely those only requiring recorded speed, coordinates and time. Several studies have addressed certain issues that arise in processing GPS tracking data. For instance, Kharrat et al. (2008) proposed an algorithm (NETSCAN) for mobile object clustering and applied it in an environment constrained by a network. Giannotti et al. (2011) presented a query and data mining system named M-Atlas, but noted that it is difficult to transform GPS tracking data into mobility knowledge. Etienne, Devogele, and Bouju (2012) provided a method for detecting outliers of spatio-temporal trajectories with primary applicability for travel behaviour analysis.

However, no study has attempted to discuss all issues related to processing GPS tracking data simultaneously, let alone provided a procedure or open source code with detailed

CONTACT Xiaoyun Zhao  xzh@du.se

© 2015 Taylor & Francis

information for doing so. The aim of this paper is to address several of the issues arising in processing GPS tracking data and thereby outline a general procedure for the data processing. Moreover, the program code in the process is public as a reference for use of interest.¹ The study is carried out using real-world GPS tracking data of some 300 cars that were originally collected for the purpose of studying CO₂ emissions induced by retailing.

The processing of the GPS tracking data requires a clear definition of movement, a detailed understanding of the capability of the GPS device and the output of the GPS logger, access to digital data of the road network as well as methods for matching the GPS data and the network. All this is discussed in the paper. Descriptive statistics and visualised maps are used to summarise and illustrate the mobility patterns. The technical documentation of the data processing in this paper is detailed in the interest of readily being replicable on the same or similar type of data. The data in this paper are freely available upon requesting it from the author.

Section 2 of this paper gives an overview of the related literature. Section 3 provides details of the data collection. In Section 4, definition of movement is given and the processing of the data in the plane is described. In Section 5, the movements are further processed to obey the restrictions imposed by the network and the mobility pattern is visualised by maps. Section 6 concludes the paper.

2. Literature review

The application of GPS has increased in location-based services and intelligent transportation system as a consequence of the popularity of portable, low-cost GPS devices. There is a large body of studies that has integrated GPS in the areas of ecology (Steiner et al. 2000; Tuner et al. 2000; Cagnacci et al. 2010), agriculture (Stafford 2000; Auernhammer 2001; Zhang, Wang, and Wang 2002) and sports (Coutts and Duffield 2010; Aughey 2011). The main common feature of these research areas is that there is no underlying road network that confines the mobile objects. However, the road network is a confinement in many mobility studies relying on GPS tracking data (Van Schaick 2010). This paper focuses on GPS tracking data restricted to a road network; there are three broad aspects that have been of concern in this area of research.

Firstly, GPS tracking has been conducted for the purpose of improving the quality and the quantity of travel data. For instance, Wagner (1997), Casas and Arce (1999), Draijer, Kalfs, and Perdok (2000) and Doherty et al. (2001), respectively, have conducted comprehensive data collection with GPS in Lexington, Austin, Quebec City and the Netherlands to test this method vs. ordinary travel diaries. They found that sufficient and valuable travel information could be obtained.

Wolf (2000) checked if GPS data could substitute, rather than supplement, the traditional travel diary. In a later study, Wolf, Guensler, and Bachman (2001) used GPS data to collect travel data in personal vehicles and demonstrated that it is possible to derive trip purpose from the data. Gruteser and Grunwald (2003) studied whether it is technically feasible to reduce the privacy risk in location identification. Leduc (2008) conducted a snapshot of the development of traffic data collection methods and discussed the potentials and challenges related to emerging technologies.

Secondly, using GPS data over a certain period of time to analyse human mobility and travel behaviour on a road network is important. The prime advantage of using GPS is that

it provides real-time spatial and temporal information of the entire trip (Grengs, Wang, and Kostyniuk 2008), upon which it is possible to identify travel time and distance, origin and destination as well as stops. Patterson et al. (2003) applied GPS tracking to classify a user's transportation mode of car, bus or walk as well as to predict the individual's most probable route. Askbrook and Starner (2003), Krumm and Horvitz (2006) and Liao et al. (2007) aimed to understand individuals' outdoor movements using GPS data and to extract individuals' significant places and predict their movements.

Li et al. (2004) inspected the travel time variation in commuting trips, the route choice and the effects on departure time based on GPS data. Zheng et al. (2009, 2010) provided approaches to identify culturally important locations, travel sequences and to differentiate between walking, driving, taking a bus and riding a bike. Huang and Levinson (2012) analysed the influence of movement on a road network and clustered their destinations based on GPS data in the Twin Cities; they found that higher accessibility and diversity of retail services around the destination are more attractive. Schönfelder et al. (2006) concluded that the use of GPS data for travel behaviour analysis could provide unique insight into the structure, size and stability of human activity spaces.

Thirdly, evaluation of GPS data performance is necessary. Positioning technologies based on stand-alone GPS receivers are vulnerable and have to be supported by additional information to obtain the desired accuracy, integrity and availability (Skog and Handel 2009).

It is difficult to obtain accurate GPS data since its performance depends not only on the features of the sensor, the GPS receiver and the vehicle model, but also on the trajectory dynamics and environments. It is even more challenging in urban environments, buildings may block satellite signals, forcing the GPS receiver to work with a poor geometric constellation of satellites, thereby reducing the accuracy of the data (Huang and Tan 2006; Modsching, Kramer, and ten Hagen 2006; Godha and Cannon 2007). Marais, Berbineau, and Heddebaut (2005) found that multipath propagation of the radio signal due to reflection in surrounding objects could lead to decreased position accuracy of the GPS receiver. Schlingelhof et al. (2008) confirmed that the development of intelligent transport system applications and location-based services requires not only higher accuracy GPS, but also better reliability and integrity with auxiliary information.

Map-matching is a commonly used solution to improve the accuracy of GPS data by matching positions and trajectories to a road using a digital map of a road network. Greenfeld (2002), Bruntrup et al. (2005) and Wenk, Salas, and Pfoser (2006) applied an incremental algorithm for matching GPS positions to their most probable locations on a road network. Brakatsoulas et al. (2005) proposed three map-matching algorithms, where the trajectory nature of the data was used to improve accuracy. Mustière and Devogele (2008) provided an approach for matching networks with different levels of detail to determine one-to-many links between networks. Most map-matching studies assumed that the digital map is of high accuracy; however, there are many situations in which this is unlikely to be the case. For instance, White, Bernstein, and Kornhauser (2000) and Ochieng et al. (2009) studied map-matching algorithms to reconcile inaccurate data with a poor digital road network. Quddus, Ochieng, and Noland (2007) conducted a thorough survey of the existing map-matching algorithms and found that enhancement is needed to improve the performance of map-matching in dense urban areas with complex road networks.

To conclude, GPS tracking data have become a reliable source to continuously provide travel data over a certain period. Although high data quality cannot be guaranteed, approaches such

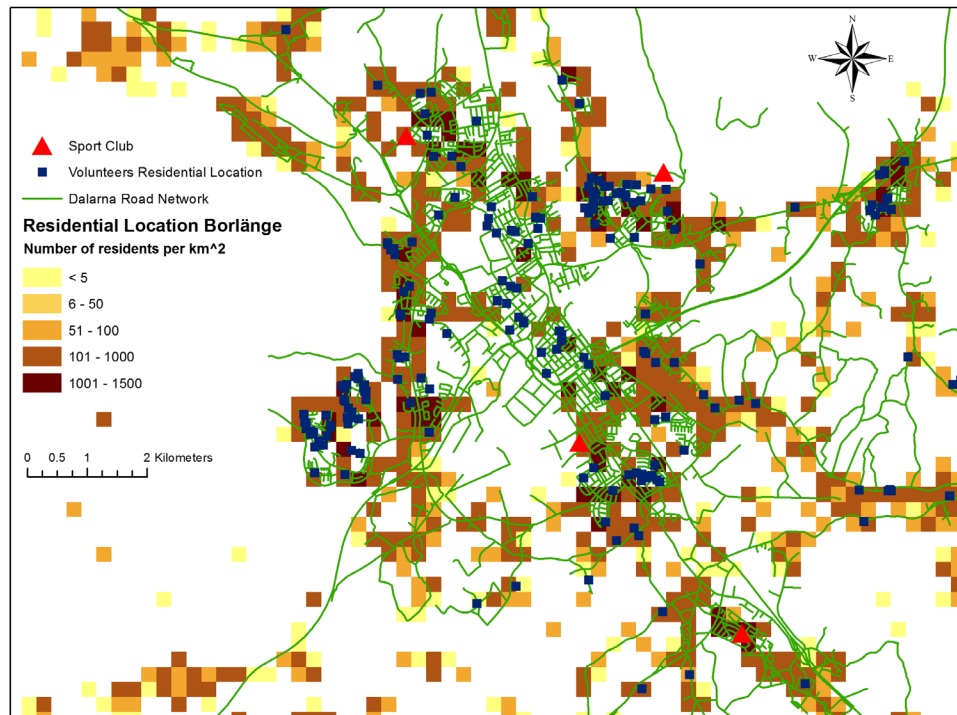


Figure 1. Spatial distribution of the volunteers as well as all the residents in Borlänge.

as map-matching have been widely used in the correction of data inaccuracy. The GPS tracking data have been broadly applied for analysis of travel behaviour and mobility prediction by processing the data; however, studies that have attempted to outline a specific procedure for the data processing and address the related issues are deficient. Therefore, efficiently processing the GPS travel data is crucial to make good use of the current data quantity and quality. Consistently adjusting and modifying the data is necessary to ensure the accuracy and validity for the further data analysis. Precisely illustrating the GPS travel data is helpful to analyse human mobility and predict travel behaviour on a road network.

3. Data collection

The data collection was conducted using a type of standard Bluetooth GPS data logger named BT-338X. Although using GPS devices to replace traditional travel diaries can reduce the collection burden and improve the data quality, there will still be substantial non-response by randomly selecting a sample of the population because it requires consent of the individual to carry the GPS device. We instead successfully negotiated an agreement with four large sports associations (Domnarvets GOIF, Kvarnsveden Hockey, Stora Tuna IK and Torsångs IP) to recruit car-owning volunteers in conducting the data collection. Each association provided approximately 75 anonymous volunteers with their home addresses. A unique ID made up of the association name and a number was assigned to each volunteer.

In total, 89 devices were shared among these volunteers according to a protocol. The device combined a GPS receiver and a data logger with a Bluetooth interface to record their

car movements. Each volunteer's car equipped one device for one or two weeks. The device was always equipped to the same car for the duration of the tracking period. There was no guarantee that the car with the device would only be driven by the registered volunteer because this car could be shared by all the members in the household. This is however not a concern since the car movements were the tracking target.

The volunteers were aware of the atypical situations such as failed to charge or carry the device, device malfunction or car issues. The data collection was undertaken from 29 March to 15 May in 2011 and the successful compliance attained to be 95%. The device activated tracking every 5 or 30 s. The recorded information included date, time, longitude, latitude and speed. There were 309,263 valid positional recordings after removing 5402 invalid ones due to signal loss. The data were stored in 316 log files, one for each volunteer.

Figure 1 illustrates the residential distribution of the volunteers and all the residents in Borlänge. The volunteers are spread out in Borlänge in a pattern similar to all the residents. Due to the requirement that every volunteer must possess a car, the volunteers will appear less concentrated in the centremost area compared to all other residents in general. The four sport associations shown by the red triangles are dispersedly located in the city. Most of the volunteers reside in Borlänge; however, the spatial extension of their movements covered more than half of the entire territory of Sweden (Jia et al. 2012). The focus of this paper is the processing of the predominant movements in Borlänge city.

4. Processing GPS data on the plane

4.1. Data from the GPS logger file

The original GPS tracking data from volunteers were recorded into DataLogger files. Each DataLogger file consists of three main variables, *Date*, *TP* and *positional recording*. The variable *Date* notes the latest date and time when the file was loaded from the device to the computer using the software GlobalSat Data Logger PC Utility. It is in the format of yyyy-mm-dd-tt:mm:ss. The variable *TP* represents the tracks, in which a track is defined as the sequentially linked line based on a number of positional recordings in a specific time period. Each *positional recording* contains the information in the sequence of latitude, longitude, time, date, speed and altitude. The longitude and latitude are referenced by the World Geodetic System 84 (WGS 84) in the degrees:decimal:minutes format and are measured with a precision of 5 m. The time is in the format of tt:mm:ss. The date is in the format of dd-mm-yy. The speed was measured in the unit of km/h. The altitude was not recorded and was assigned the value -1.

Figure 2 shows an example of a DataLogger file from volunteer Domnarvet11. The *Date* shows that the file was loaded at 2011-04-29-13:15:56. The *TP* 1 = 001, 2011-04-05:20:20:27 signifies that the first track was assigned to 001 and it started on date 2011-04-05 and at time 20:20:27. The volunteer Domnarvet11 made 17 tracks in total.

The first track contains 16 positional recordings with numerators from 1 to 16. Specifically, 1 = 60,298,968; 15,282,927; 182,027; 50,411, 6240, -1 indicate that the latitude is 6029.8968, longitude is 1528.2927, the time is 182027 (which is 18:20:27), the date is 50411 (which is 05-04-2011), the speed is 62.40 km/h and the altitude is filled as -1. The listed time is 2 h earlier than the actual local time due to the change of the summer time; therefore, the listed time plus 2 h is the actual local time in recording the positions.

Table 1 shows the number of valid GPS DataLogger files from the volunteers. There are 48 from Domnarvet GOLF, 59 from Kvarnsveden Hockey, 58 from Torsång IP and 71 from

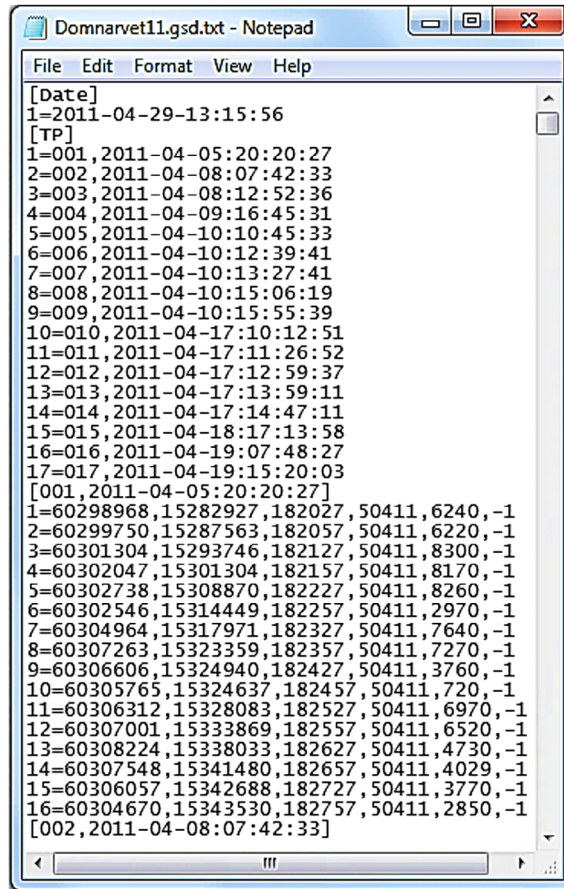


Figure 2. Example of GPS data from volunteer Domnarvet11.

Table 1. Number of valid GPS DataLogger files from each sport association.

	Valid GPS logger files
Domnarvet GOLF	48
Kvarnsveden Hockey	59
Stora Tuna IK A	71
Stora Tuna IK B	80
Torsång IP	58
Total	316

StoraTuna IKA. Additional 80 volunteers from StoraTuna were recruited during the data collection and were assigned as the group of StoraTuna IK B.

Further, we parse the original data into a matrix with eight variables. The *Date* variable is excluded because it does not provide any information regarding the car movements. In this matrix, the variable *TP* is named as *TRACK_ID* and the variable *positional recording* is represented by six variables named as *PR_ID*, *LATITUDE*, *LONGITUDE*, *TIME*, *DATE* and *VELOCITY*. The abbreviation of *PR_ID* means the positional recording ID. The identification for a volunteer is displayed as *USER_ID*. Figure 3 shows this structure and all the variables.

	A	B	C	D	E	F	G	H
1	TRACK_ID	PR_ID	USER_ID	LATITUDE	LONGITUDE	TIME	DATE	VELOCITY
2	1	1	Domnarvet11	60298968	15282927	182027	50411	6240
3	1	2	Domnarvet11	60299750	15287563	182057	50411	6220
4	1	3	Domnarvet11	60301304	15293746	182127	50411	8300
5	1	4	Domnarvet11	60302047	15301304	182157	50411	8170
6	1	5	Domnarvet11	60302738	15308870	182227	50411	8260
7	1	6	Domnarvet11	60302546	15314449	182257	50411	2970
8	1	7	Domnarvet11	60304964	15317971	182327	50411	7640
9	1	8	Domnarvet11	60307263	15323359	182357	50411	7270
10	1	9	Domnarvet11	60306606	15324940	182427	50411	3760
11	1	10	Domnarvet11	60305765	15324637	182457	50411	720
12	1	11	Domnarvet11	60306312	15328083	182527	50411	6970
13	1	12	Domnarvet11	60307001	15333869	182557	50411	6520
14	1	13	Domnarvet11	60308224	15338033	182627	50411	4730
15	1	14	Domnarvet11	60307548	15341480	182657	50411	4029
16	1	15	Domnarvet11	60306057	15342688	182727	50411	3770
17	1	16	Domnarvet11	60304670	15343530	182757	50411	2850

Figure 3. Example of the matrix structure from volunteer Domnarvet11.

Table 2. Descriptive statistics of positional recordings for tracks and volunteers.

	Min	Q1	Median	Q3	Max
TP number per volunteer	1	7	11	17	734
Positional recording number per Track	2	19	79	95	95
per Volunteer	2	278	517	809	66,531
Distance (m) per Track	2	1481	7837	26,921	117,722
per Volunteer	3767	101,712	186,840	349,276	2,471,518
Speed (m/s) per Track	0.01	5.8	11.4	15.3	41.6
per Volunteer	0.03	11.6	13.3	15.9	25.9
Time span (s)	5	5	30	30	342,775

4.2. Descriptive statistics of the processed GPS data

There were 316 volunteers who made 5180 tracks with 309,263 positional recordings according to the reorganised data. Table 2 exhibits that the volunteers made at least 1 and at most 66,531 positional recordings during the tracking period. In total, 73 single positional recordings that cannot compose a track are deleted. The median number of positional recordings in each track is 79, while the minimum is 2 and the maximum is 95. The number of tracks varies from 1 to 734 and 75% of the volunteers have made less than 17 tracks.

The raw time and date were recorded separately in the GPS log file and cannot be used for calculations such as the time span between certain positional recordings or the time differences among tracks. Therefore, the Unix Time Stamp is used to convert the recorded date and time into the number of seconds that have elapsed since 00:00:00 Coordinated Universal Time (UTC), Thursday, 1 January 1970, not counting leap seconds.

The time span between two neighbouring positional recordings was mostly 5 or 30 s if the car did not go to a tortuous location (Jia et al. 2012) or stayed at the same location for a long time. 37.7% of the recordings have a time span of 5 s and 54.3% have a time span of 30 s. The maximum time span was 342,775 s. The reason for the very large time span was that

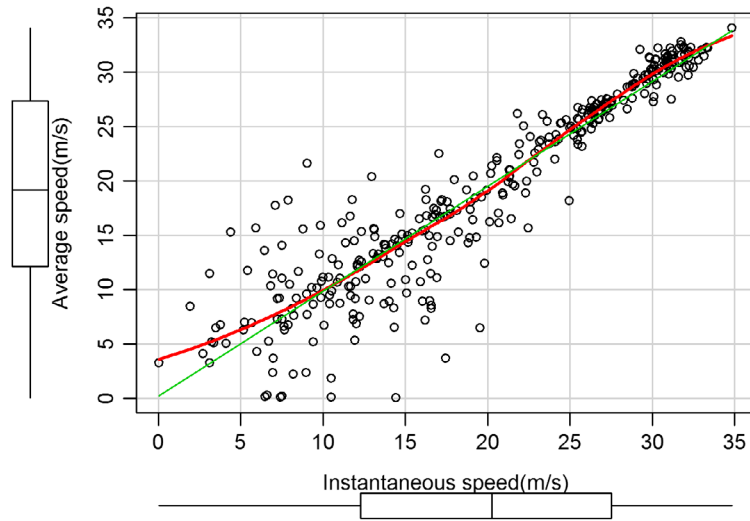


Figure 4. Differences between the instantaneous velocity and the average velocity of a track.

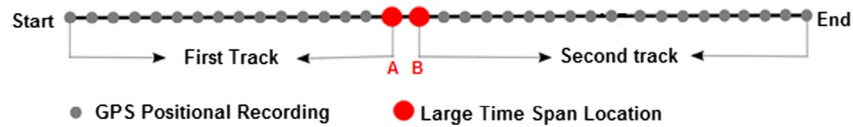


Figure 5. Illustration of redefining a track based on time span between neighbouring positions.

if the car has stopped moving, but the device was kept on, the tracking would pause. If the number of previous recordings in that track was less than 95, the next positional recording would be added when the car continued to move and tracking started again.

The Euclidean distance between two neighbouring positional recordings in one track is calculated and added together. The sum is the distance of this track in the plane. This measurement of the distance underestimates the real network distance that the car has travelled because it is impossible to always directly drive in straight lines from one location to another when travel on the real road network. The underestimation error could become smaller when the positional recordings are more intensive. However, it is difficult to constantly acquire all positions that the car has covered. The distance measure can be varied in this step and should be chosen carefully in this step considering the application focus and the trade-off between the frequency of the positional recordings and the desired accuracy.

In this specific case, the underestimation could be regarded as acceptable, considering that 92% of the positional recordings are tracked with a fairly high frequency of 5 or 30 s. There are huge variations in travel distance as is shown in Table 2. The minimum distance for one track was 2 m, while the maximum was 117,722 m. The total distance that the volunteers had travelled varied from 3767 to 2,471,518 m.

The recorded instantaneous speed is the speed that the car has at the moment of recording. The average speed of the car on a track segment is calculated using the distance and the time length between two neighbouring positional recordings. The average speed of a

volunteer can be derived in the same way. A conversion from km/h to m/s is done in order to be consistent with the measurement of distance (m) and time (s). The median of the average speed for all tracks was 11.4 m/s, while for all volunteers it was 13.3 m/s.

We randomly select 10 tracks from those 5180 tracks, and then generate the scatter plot with the linear regression line between the instantaneous speed and the average speed as shown in Figure 4. Most of the points line up in a fairly straight red line, the slope approximately equals to one compared to the straight green line. The scatter plot indicates that there is a strong positive linear association between the instantaneous speed and the average speed, although the relation is weaker in the low velocities than in the higher ones.

Purposive locations are positions with drastic changes in time, distance or angle along the movement trajectories of the individual volunteers (Jia et al. 2012). It is understandable that a track consists of purposive locations and this leads to the ambiguous issues in defining tracks. It is reasonable to assume a volunteer to stop and finish a certain activity in about or more than 10 min. Therefore, locations where the time interval exceeds a threshold of 550 s are identified and the tracks are thereafter redefined.

If there is no time span over 550 s between two neighbouring positional recordings through the whole track, then keep the information of the start and end points, then assign a *TRACK_ID* to this track. If at least one time span over 550 s is identified and in addition, the distance between the neighbouring positional recordings is less than 2 km, the old track will then be redefined. As is shown in Figure 5, the time span between positions *A* and *B* is larger than 550 s; *A* will be regarded as the end point for the first track, while *B* which happens to be straight after *A* will be regarded as the start point for the second track. This original track will then be segmented into two tracks and each track will be assigned a unique *TRACK_ID*. In total, 6534 time spans are identified and there are 8736 tracks after the redefinition.

5. Processing GPS data on the road network

5.1. Linking positional recordings to tracks

The longitude and latitude of GPS data are referenced by the world geographic coordinate system WGS 84 in the format of degrees:decimal:minutes. We first convert the WGS 84 degrees:decimal:minutes into the WGS 84 decimal degree. Considering that the Swedish projected coordinate system SWEREF99_TM is used in the digital map of Dalarna road network from the National Road Database (NVDB) in Sweden, the transformation from the WGS 84 decimal degree to the SWEREF99_TM is then conducted. Figure 6(a) illustrates the distribution of 309,190 positional recordings from the volunteers; they are intensive and highly overlapped in the centre area. The small enlarged map in Figure 6(a) illustrates how the positional recordings are arranged. Figure 6(b) illustrates the tracks by linking the positional recordings sequentially based on the time of occurrence.

5.2. Matching positional recordings to the road network

As the device did not continuously track the position every second, but rather with 5 or 30 s intervals, it is hard to examine how the car has moved during this time span. Moreover, a standard GPS device is usually sensitive to the surroundings. It cannot continuously provide

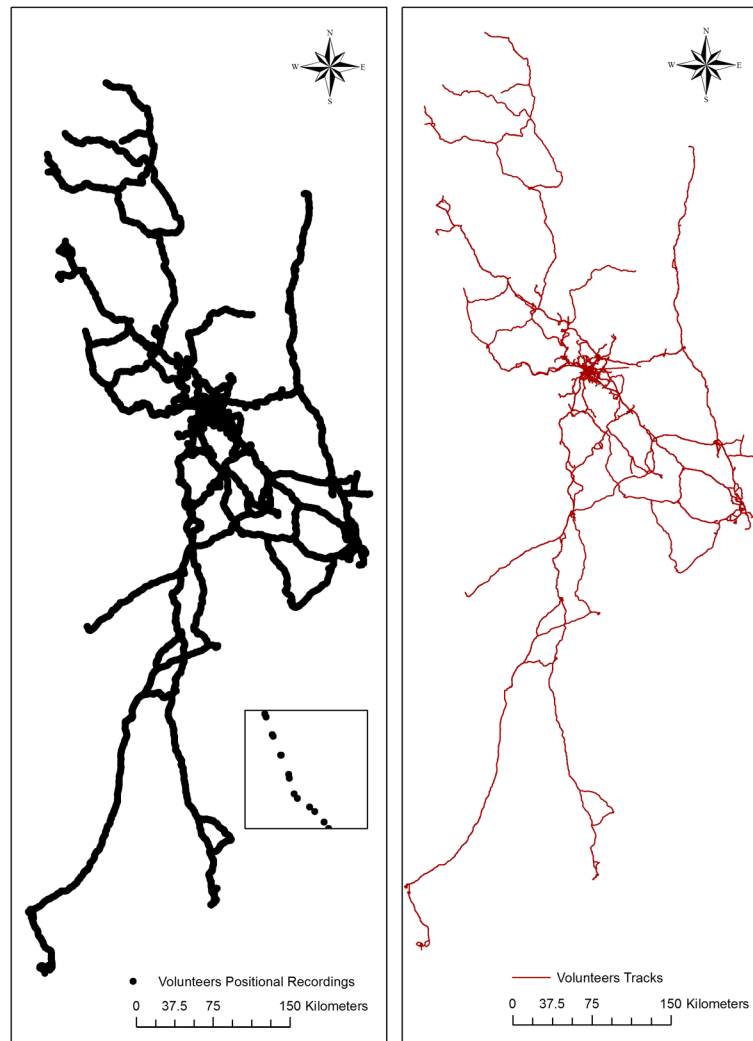


Figure 6. (a) Positional recordings from all the volunteers. (b) Tracks from all the volunteers.

accurate data, but with an error rate of 5 m according to the manual. Now, we define a trip as the link of all the positional recordings over which the car has travelled on the road network. The previously defined tracks on the plane as shown in Figure 6(b) are therefore not identical to the trips of real car movements on the road network.

It is possible to increase the recording frequency and equip more devices on one car to increase the reliability of data. However, that would be problematic due to the increase of control factors. Additional information and post-processing techniques provide the ability to improve the current data performance without inducing any data collection uncertainty. As for the individuals' travel data, the underlying road network provides reliable auxiliary information to verify the data accuracy and improve the usability. The goal is to match the GPS tracking data of the car movements to the real road network using a map-matching algorithm and a spatial join tool.



Figure 7. Positional recordings on the road network.

Before the matching, we verify that not all positional recordings are on the road. As is shown in Figure 7, there are positional recordings such as *a*, *b*, *c* and *d* that are off the road with a certain distance. Tracks from linking such positional recordings would then cause a deviation from the real trips.

Figure 8 illustrates the situation after zooming in on the area that has the highest density of the tracks. In Figure 8(a), it is difficult to see any potential relationship between the data and the road network due to the messy visualisation.

One cause for the messy visualisation is that some movements of the cars are far off the road network due to errors of the positional recordings. The errors vary among different devices. If all the positions that occurred at the same location were recorded correctly and were consistent with the road nodes, the tracks would have been highly overlapping. The distance between each track on the same road would have been less than 14 m, considering the width of the present national two-lane road.

A single track is a track that has only been travelled on once by a single volunteer. Single tracks from a volunteer may occur due to the error in positional recordings. They can also be formed by taking unknown shortcuts or illegal paths since the route choice varies among individuals. Usually, drivers would prefer a shorter distance and an easier path due to fuel consumption, travel time and other costs. A driver may take a shortcut only known to him; therefore, he can avoid taking the detour and the tortuous locations. A driver can also be incorrectly guided if he is not familiar with the roads; he could drive into dead-end roads and then have to turn around. Reasons behind this are complex and difficult to identify. We therefore exclude all single tracks which were only conducted by one volunteer and deviated more than 5 m from the roads, as is illustrated in Figure 8(b). There are 39 single tracks identified from the whole 5107 tracks; thereafter, 5068 tracks remain.

5.2.1. Map-matching

Map-matching is a commonly used approach for correcting off-road positions. Brakatsoulas et al. (2005) concluded that global map-matching algorithms produce better matching results than incremental algorithms, while an incremental method runs fast and performs

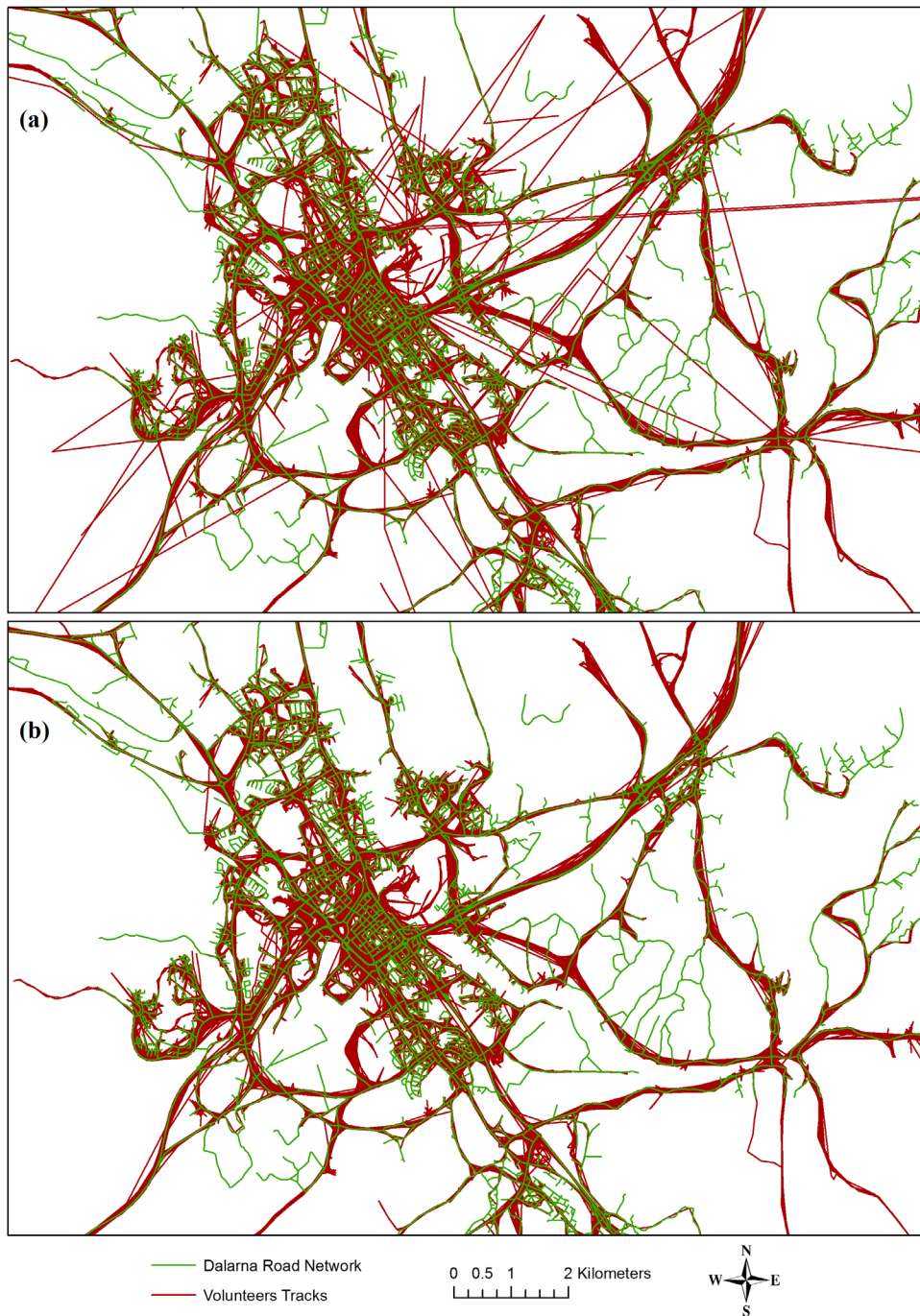


Figure 8. (a) The highest density of tracks with the underlying road network before removing all single tracks. (b) The highest density of tracks with the underlying road network after removing all single tracks.

well when the sampling frequency is within 5 s (Lou et al., 2009). The running time for incremental and global methods is $O(n)$ and $O(mn \log^2 mn)$, where n is the number of positional recordings in a track and m is the total number of edges and vertices in the road network.



Figure 9. An example of map-matching an off-road track: (a) the off-road track, (b) the map-matching correction and (c) the trip after map-match.

Although map-matching will be time-consuming with a large GPS data-set in a complex road network, improvement for decreasing time complexity and increasing robustness is possible. This is a recommended procedure for processing GPS data since it improves the data performance with showing the spatial geometric and topological structures of movements along the road network.

In this paper, 92% of the data have a high sampling frequency of 5 or 30 s; running map-matching on all the positional recordings in the whole road network will have high

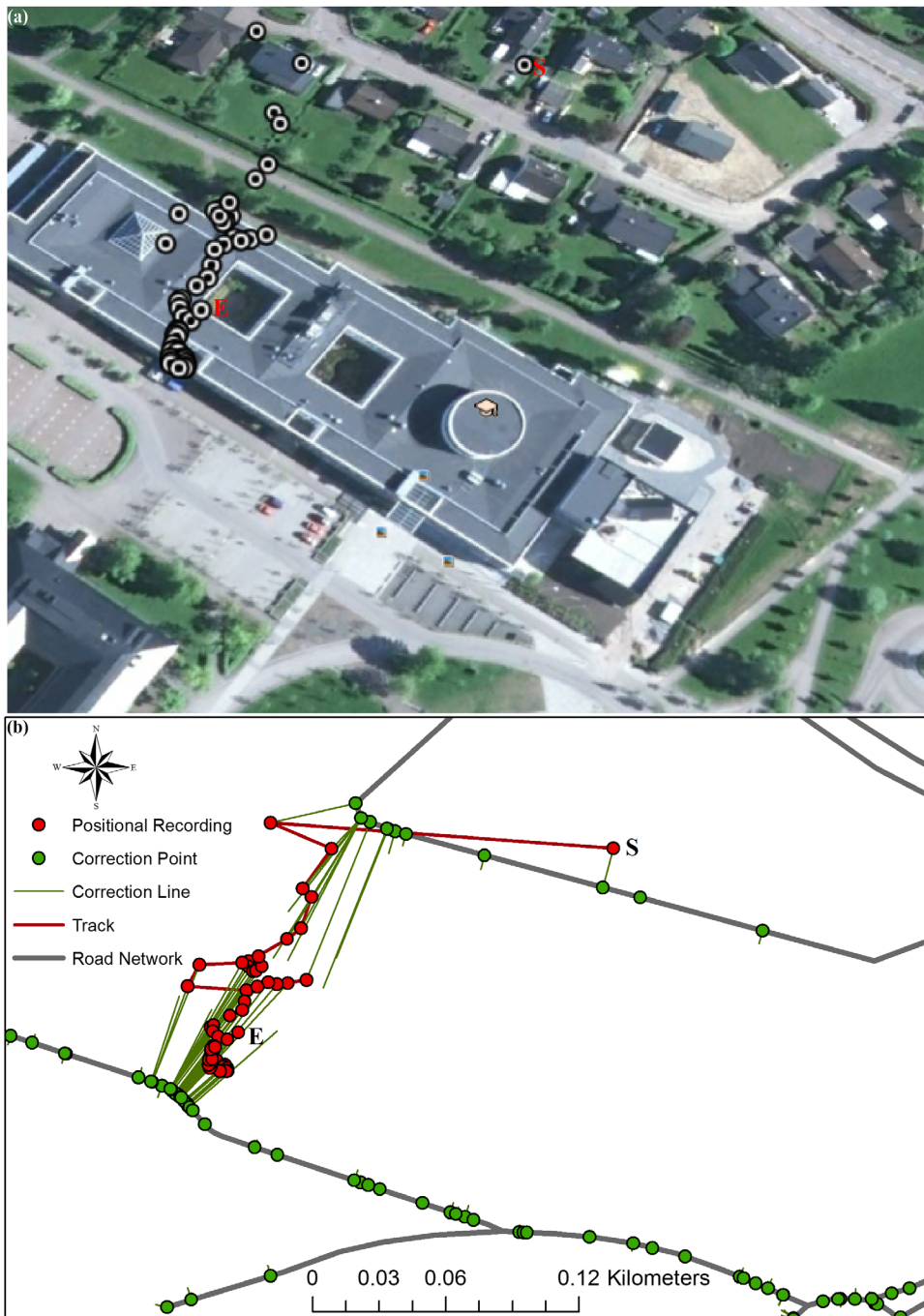


Figure 10. (a) A number of positional recordings in the earth image. (b) The map-matching gave big error in correction.

complexity; we therefore applied a global map-matching algorithm using a small subset of the whole data as an example for illustration. The subset consists of 285 GPS positions, a road network section of 1458 vertices and 677 road segments. Figure 9 illustrates an example

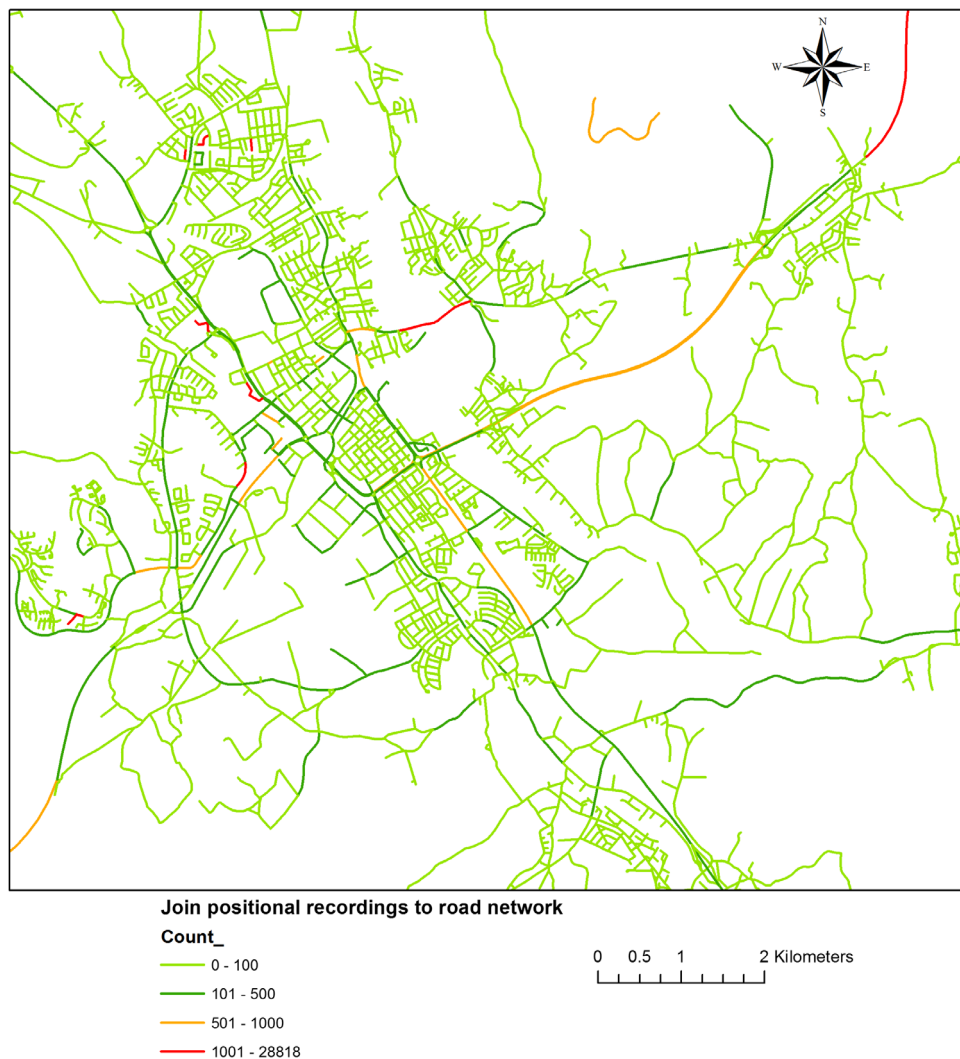


Figure 11. Frequency of positional recordings on the road network.

of matching an off-road track to the road network, in which Figure 9(a) shows that the off-road track is matched to the road and the correction is shown in Figure 9(b) and (c) shows the trip after the match.

However, map-matching cannot always provide good results (Quddus et al. 2007). As is shown in the real earth image in Figure 10(a), the positional recordings from the GPS data show that the volunteer drove through a school building from the start point S to the end point E, which cannot be true. These positional recordings are then matched to the two nearest roads with correction lines as is illustrated in Figure 10(b). However, this volunteer can never arrive to E from S if he just travelled on the two peculiar road parts as matched. The error of map-matching in such case could be due to the inaccuracy of the GPS data, the insufficiency of the road network map or the inappropriate use of the GPS device. This problem further exhibits the necessity of data processing to uncover the performance of the

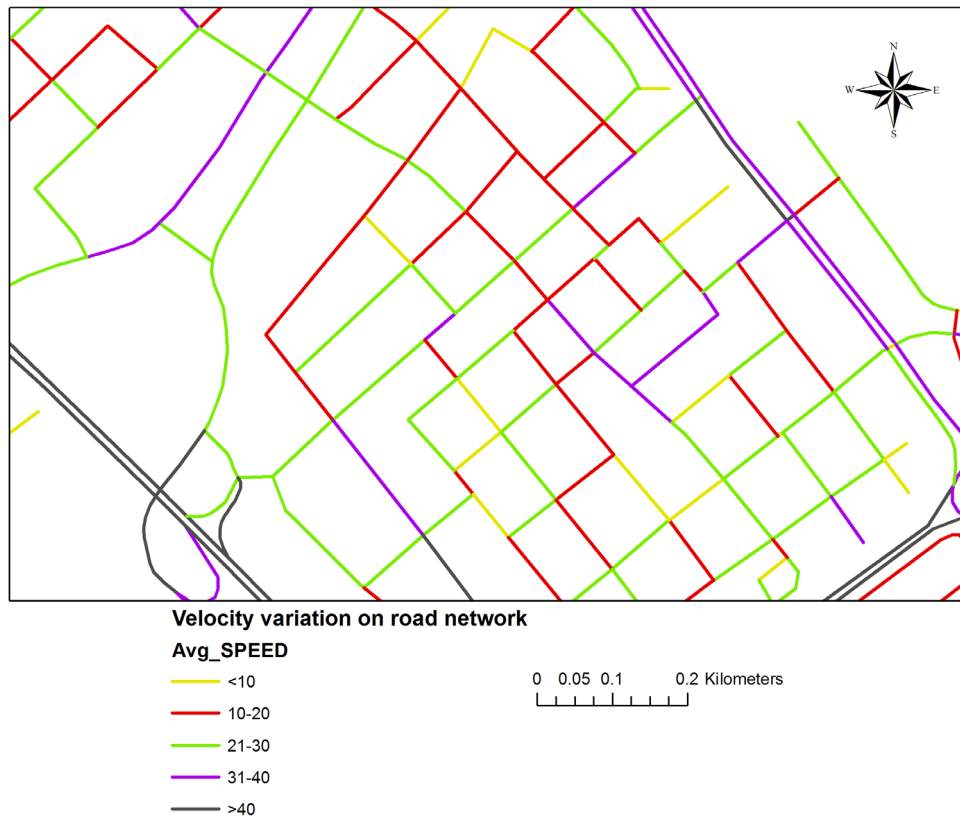


Figure 12. Variation of velocity on the road network in the centremost area of Borlänge.

data. Studies reviewed in this paper have investigated some possible causes and solutions to improve the data performance of reliability, accuracy, availability and integrity.

5.2.2. Spatial join

Another crucial part is to show the spatio-temporal relationships of the tracks, which can be achieved by spatial join (Orenstein 1986). Spatial join tool is a technical implementation to compile the functionalities of spatial join. The spatial join tool in Arc GIS is one of the spatial geo-processing tools that is recommended for showing the features of movements if the data-sets are large or complex or both. In this procedure, 5068 volunteers' tracks from 306,664 positional recordings are matched by 3521 road segments.

Figure 11 illustrates that most of the roads in the centremost area have less than 100 positional recordings, which happened primarily on the local roads or private streets. Roads that have between 101 and 500 positional recordings are the second most common, which take place mainly on the national roads. This is due to the usage and load capacity of the roads; the maximum number of joined positions to a road is 28,818.

Given a tolerance of 5 m, 90% of the trips on the plane match the road network. It captures the complexity of the real car movements in urban areas. We can further visualise the variation of average velocities on the road network by connecting velocities onto a map. Figure 12 illustrates the variation of speed when cars drive on the roads of the centremost

area, given a speed limit of 40 km/h. Most of the cars drive within 40 km/h due to the influence of the surroundings, road conditions, speed limit and other restrictions.

6. Conclusion

This paper aims to outline a general procedure for processing GPS tracking data and to discuss all issues related in the processing. The procedure is illustrated step by step by processing the real-world GPS data of 300 car movements that predominantly happened in a centre city of Sweden, Borlänge. The procedure provides a detailed understanding of the capability of GPS devices and the output of the data. In addition, post-processing techniques with auxiliary information are found necessary and important for solving the inaccuracy of GPS data. The procedure applies methods to match GPS data with the road network in order to improve the data performance based on a clear definition of movement.

The processed data and the generated maps from the procedure can be used on a broad range of researches and applications. Processing the same or similar data types can provide valuable information to discriminate mobility patterns, derive accurate inference for environmental control, urban planning, location-based services and transportation management. It can also provide useful reference information for adjusting and improving the accuracy of the current GPS tracking devices.

In the future, the time threshold for defining the stops within one track could be changed and differences could be compared. The tolerance used in the reduction of the single tracks may also be altered to minimise the induced bias when precise matching is required. Other sensor information like acceleration rate and dilution of precision could be useful in processing GPS data. The performance of the procedure could be evaluated by processing GPS data from other types of GPS devices and transportation modes.

Note

1. Requested by email to xzh@du.se or from URL <https://github.com/Stefangemfind/MapMatching>

Disclosure statement

No potential conflict of interest was reported by the author.

References

- Ashbrook, D., and T. Starner. 2003. "Using GPS to Learn Significant Locations and Predict Movement across Multiple Users." *Personal and Ubiquitous Computing* 7 (5):275–286.
- Auernhammer, H. 2001. "Precision Farming – The Environmental Challenge." *Computers and Electronics in Agriculture* 30 (1–3): 31–43.
- Aughey, R. J. 2011. "Applications of GPS Technologies to Field Sports." *International Journal of Sports Physiology and Performance* 6 (3): 295–310.
- Brakatsoulas, S., D. Pfoser, R. Salas, and C. Wenk. 2005. "On Map-matching Vehicle Tracking Data." In *Proceedings of the 31st International Conference on Very Large Data Bases*, 853–864. Trondheim: VLDB Endowment.
- Bruntrup, R., S. Edelkamp, S. Jabbar, and B. Scholz. 2005. "Incremental Map Generation with GPS Traces." In *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, 574–579. Piscataway, NJ: IEEE.

- Cagnacci, F., L. Boitani, R. A. Powell, and M. S. Boyce. 2010. "Animal Ecology Meets GPS-based Radiotelemetry: A Perfect Storm of Opportunities and Challenges." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1550):2157–2162.
- Casas, J., and C. H. Arce. 1999. "Trip Reporting in Household Travel Diaries: A Comparison to GPS-collected Data." 78th Annual Meeting of the Transportation Research Board. Vol. 428. Washington, DC, January.
- Coutts, A. J., and R. Duffield. 2010. "Validity and Reliability of GPS Devices for Measuring Movement Demands of Team Sports." *Journal of Science and Medicine in Sport* 13 (1): 133–135.
- Doherty, S. T., N. Noël, M. L. Gosselin, C. Sirois, and M. Ueno. 2001. *Moving beyond Observed Outcomes: Integrating Global Positioning Systems and Interactive Computer-based Travel Behavior Surveys* (No. E-C026). Washington, DC: Transportation Research Board.
- Draijer, G., N. Kalfs, and J. Perdok. 2000. "Global Positioning System as Data Collection Method for Travel Research." *Transportation Research Record: Journal of the Transportation Research Board* 1719 (1): 147–153.
- Etienne, L., T. Devogele, and A. Bouju. 2012. "Spatio-temporal Trajectory Analysis of Mobile Objects following the Same Itinerary." *Advances in Geo-Spatial Information Science* 10: 47–57.
- Giannotti, F., M. Nanni, D. Pedreschi, F. Pinelli, C. Renso, S. Rinzivillo, and R. Trasarti. 2011. "Unveiling the Complexity of Human Mobility by Querying and Mining Massive Trajectory Data." *The VLDB Journal – The International Journal on Very Large Data Bases* 20 (5): 695–719.
- Godha, S., and M. E. Cannon. 2007. "GPS/MEMS INS Integrated System for Navigation in Urban Areas." *GPS Solutions* 11 (3): 193–203.
- Greenfeld, J. S. 2002. "Matching GPS Observations to Locations on a Digital Map." Transportation Research Board 81st Annual Meeting, Washington, DC, January 13–17.
- Grengs, J., X. Wang, and L. Kostyniuk. 2008. "Using GPS Data to Understand Driving Behavior." *Journal of Urban Technology* 15 (2): 33–53.
- Gruteser, M., and D. Grunwald. 2003. "Anonymous Usage of Location-based Services through Spatial and Temporal Cloaking." In *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*, 31–42. New York: ACM.
- Huang, A., and D. Levinson. 2012. "Accessibility, Network Structure, and Consumers' Destination Choice: A GIS Analysis of GPS Travel Data." Proceedings of the 91st Annual Meeting of the Transportation Research Board. Transportation Research Board of the National Academies. Washington, DC, January 22–26.
- Huang J., and H. S. Tan. 2006. "A Low-order DGPS-based Vehicle Positioning System under Urban Environment." *IEEE/ASME Transactions on Mechatronics* 11 (5): 567–575.
- Jia, T., B. Jiang, K. Carling, M. Bolin, and Y. Ban. 2012. "An Empirical Study on Human Mobility and Its Agent-based Modeling." *Journal of Statistical Mechanics: Theory and Experiment*. 2012 (11): P11024–P11044.
- Kharrat, A., I. S. Popa, K. Zeitouni, and S. Faiz. 2008. "Clustering Algorithm for Network Constraint Trajectories." In *Headway in Spatial Data Handling*, edited by Anne Ruas, Christopher Gold, 631–647. Berlin Heidelberg: Springer.
- Krumm, J., and E. Horvitz. 2006. "Predestination: Inferring Destinations from Partial Trajectories." In *UbiComp 2006: Ubiquitous Computing*, edited by Paul Dourish, Adrian Friday, 243–260. Berlin Heidelberg: Springer.
- Leduc, G. 2008. "Road Traffic Data: Collection Methods and Applications." Working Papers on Energy, Transport and Climate Change, 1, 55.
- Li, H., R. Guensler, J. Ogle, and J. Wang. 2004. "Using Global Positioning System Data to Understand Day-to-day Dynamics of Morning Commute Behavior." *Transportation Research Record: Journal of the Transportation Research Board* 1895: 78–84.
- Liao, L., D. J. Patterson, D. Fox, and H. Kautz. 2007. "Learning and Inferring Transportation Routines." *Artificial Intelligence* 171 (5–6): 311–331.
- Lou, Y., C. Zhang, Y. Zheng, X. Xie, W. Wang, and Y. Huang. 2009. "Map-matching for low-sampling-rate GPS trajectories." In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 352–361. New York: ACM.
- Marais, J., M. Berbineau, and M. Heddebaut. 2005. "Land Mobile GNSS Availability and Multipath Evaluation Tool." *IEEE Transactions on Vehicular Technology* 54 (5): 1697–1704.

- Modsching, M., R. Kramer, and K. ten Hagen. 2006. "Field Trial on GPS Accuracy in a Medium Size City: The Influence of Built-up." 3rd Workshop on Positioning, Navigation and Communication, 209–218, Hannover, March.
- Mustière, S., T. Devogele. 2008. "Matching Networks with Different Levels of Detail." *Geoinformatica* 12 (4): 435–453.
- Ochieng, W. Y., M. Quddus, and R. B. Noland. 2009. "Map-matching in Complex Urban Road Networks." *Revista Brasileira De Cartografia* 2 (55): 1–14.
- Orenstein, J. A. 1986. "Spatial Query Processing in an Object-oriented Database System." In *ACM SIGMOD Record*, edited by Carlo Zaniolo, Vol. 15, No. 2, 326–336. Washington, DC: ACM, June.
- Patterson, D. J., L. Liao, D. Fox, and H. Kautz. 2003. "Inferring High-level Behavior from Low-Level Sensors." In *UbiComp 2003: Ubiquitous Computing*, edited by Anind K. Dey, Albrecht Schmidt, Joseph F. McCarthy, 73–89. Seattle, WA: Springer, January.
- Quddus, M. A., W. Y. Ochieng, and R. B. Noland. 2007. "Current Map-matching Algorithms for Transport Applications: State-of-the Art and Future Research Directions." *Transportation Research Part C: Emerging Technologies* 15 (5): 312–328.
- Schlingelhof, M., D. Bétaille, P. Bonnifait, and K. Demaseure. 2008. "Advanced Positioning Technologies for Co-operative Systems." *IET Intelligent Transport Systems* 2 (2): 81–91.
- Schönfelder, S., H. Li, R. Guensler, and J. Ogle. 2006. *Analysis of Commute Atlanta Instrumented Vehicle GPS Data: Destination Choice Behavior and Activity Spaces*. ETH, Eidgenössische Technische Hochschule Zürich: IVT, Institut für Verkehrsplanung und Transportsysteme.
- Skog, I., and P. Handel. 2009. "In-car Positioning and Navigation Technologies – A Survey." *IEEE Transactions on Intelligent Transportation Systems* 10 (1): 4–21.
- Stafford, J. V. 2000. "Implementing Precision Agriculture in the 21st Century." *Journal of Agricultural Engineering Research* 76 (3): 267–275.
- Steiner, I., C. Bürgi, S. Werffeli, G. Dell'Omo, P. Valenti, G. Tröster, D. P. Wolfer, and H. P. Lipp. 2000. "A GPS Logger and Software for Analysis of Homing in Pigeons and Small Mammals." *Physiology & Behavior* 71 (5): 589–596.
- Turner L. W., M. C. Udal, B. T. Larson, and S. A. Shearer. 2000. "Monitoring Cattle Behavior and Pasture Use with GPS and GIS." *Canadian Journal of Animal Science* 80 (3): 405–413.
- Van Schaick J. 2010. "Future Scenarios for the Relation between Advanced Tracking Research and Urban Design and Planning." *Journal of Location Based Services* 4 (2): 70–92.
- Wagner, D. P. 1997. "Lexington Area Travel Data Collection Test: GPS for Personal Travel Surveys." *Final Report, Office of Highway Policy Information and Office of Technology Applications, Federal Highway Administration, Battelle Transport Division, Columbus*.
- Wenk, C., R. Salas, and D. Pfoser. 2006, July. "Addressing the Need for Map-matching Speed: Localizing Global Curve-matching Algorithms." In *Scientific and Statistical Database Management, 2006. 18th International Conference on*, 379–388. Vienna: IEEE, July 3–5.
- White, C. E., D. Bernstein, and A. L. Kornhauser. 2000. "Some Map Matching Algorithms for Personal Navigation Assistants." *Transportation Research Part C: Emerging Technologies* 8 (1–6): 91–108.
- Wolf, J. 2000. "Using GPS Data Loggers to Replace Travel Diaries in the Collection of Travel Data." Doctoral diss., Georgia Institute of Technology.
- Wolf, J., R. Guensler, and W. Bachman. 2001. "Elimination of the Travel Diary: Experiment to Derive Trip Purpose from Global Positioning System Travel Data." *Transportation Research Record: Journal of the Transportation Research Board* 1768 (1): 125–134.
- Zhang, N., M. Wang, and N. Wang. 2002. "Precision Agriculture – A Worldwide Overview." *Computers and Electronics in Agriculture* 36 (2–3): 113–132.
- Zheng, Y., Y. Chen, Q. Li, X. Xie, and W. Y. Ma. 2010. "Understanding Transportation Modes Based on GPS Data for Web Applications." *ACM Transactions on the Web* 4 (1): 1–36.
- Zheng, Y., L. Zhang, X. Xie, and W. Y. Ma. 2009. "Mining Interesting Locations and Travel Sequences from GPS Trajectories." In *Proceedings of the 18th International Conference on World Wide Web*, 791–800. New York: ACM.

Paper IV

Paper IV

Residential planning, driver mobility and CO₂ emission

Xiaoyun Zhao[♦], Kenneth Carling, Johan Håkansson

In a city there are hotspots that attract the citizens, and most of the transportation arises when citizens move between their residence and the primary destinations (i.e. hotspots). However, an ex-ante evaluation of energy-efficient mobility and urban residential planning has seldom been conducted. Therefore, this paper proposes an ex-ante evaluation method to quantify the impacts, in terms of CO₂ emissions induced by intra-urban car mobility, of residential plans for various urban areas. The method is illustrated in a case study of a Swedish midsize city, which is presently preoccupied with urban planning of new residential areas in response to substantial population growth due to immigration. In general, the CO₂ emissions increase from the continued urban core area (CUCA), to the sub polycentric area (SPA), to the edge urbanisation area (EUA), where EUA is trice the CO₂ emissions of CUCA. The average travel distances also increase in the same pattern, though the relative increase is more than 4 times. Apartment buildings could be more effective in meeting residential needs and mitigating CO₂ emissions than dispersed single-family houses.

Keywords: Ex-ante evaluation; GPS-tracking data; Spatial distribution; Urban form

1. Introduction

Mobility is inevitable in meeting our multiple needs. However, at the same time, mobility is costly. For the individual, it costs time, energy, capital and so on, while, for society, costs are commonly external, such as congestions, emissions, resources consumptions and so on. The societal costs typically increase along with urban expansion (Fujita and Thisse, 2013). Under the broad definition of cities as urbanised areas (Angel et al., 2005; Satterthwaite, 2008), some 78% of all carbon emissions have been accredited to growing cities (Grimm et al., 2008).

At the end of 2014, the global urbanisation ratio was 54%, and is expected to be 66% by 2050 (United Nations, 2014). Under this urbanisation trend, transport energy use and CO₂ emissions are projected to increase by nearly 50% by 2030, and more than 80% by 2050 (International Energy Agency, 2009). On the other hand, adoption of new technologies and fuels, together with smart urban development, are two main courses for mitigating global energy-related CO₂ emissions by 50%, from the current level, by 2050.

[♦] Xiaoyun Zhao is a PhD student in Microdata Analysis, Kenneth Carling is a professor in Statistics, and Johan Håkansson is a professor in Human Geography. All are at the School of Technology and Business Studies, Dalarna University, SE-791 88 Falun, Sweden. *Corresponding author:* Xiaoyun Zhao, e-mail: xzh@du.se, phone: +46-23-778509.

A considerable amount of research has been conducted all around the world to understand the ecological and social consequences of urban development (Pickett et al., 2011; Seto et al., 2011; Wang et al., 2012). Urban form reveals the relationship between a single city and its rural hinterland (Grimm et al., 2008), as well as the impact of human actions on the environment in and around a city (Alberti, 2005; Weng et al., 2007). In particular, there has been an increasing interest in identifying and understanding the effects of urban transportation pattern, as this understanding is crucially important for the design of effective urban planning and management strategies (Dieleman and Wegener, 2004; Schwarz, 2010; Dubovyk et al., 2011; Long et al., 2012; Tavares et al., 2012; Thapa & Murayama, 2010; Li et al., 2013).

Makido et al. (2012) found that the per capita CO₂ emissions from residential and passenger transports in cities is significantly correlated with spatial variables of urban form: for example, the spatial distribution of the residents and their travelled destinations. Urban planning, in particular land use for residential locations, plays an important role in shaping the structure of a city, and it greatly influences the demand for mobility (particularly by cars). Although urban forms vary in different continents, countries and cities, due to the spatial distribution of residents, urban planning policies, landscape, cultural backgrounds and economic status, and so on, the relationship between urbanisation and greenhouse gas (GHG) emissions has been broadly studied in various contexts, and urban planning and management are found to be crucial in mitigating the adverse impacts on the environment (Stead and Marshall, 2001; Schwanen, 2002; Buxton and Scheuer, 2007; Dhakal, 2008; Poumanyvong and Kaneko, 2010; Hamin and Gurran 2009; O'Neill et al., 2010; Ewing & Cervero, 2010; Li, 2011; Li et al., 2013; Naess 2012; Iwata and Managi, 2016).

Various studies indicate that increasing urban density significantly reduces per capita vehicle travel (Kenworthy and Laube, 1999, Kenworthy, 2003; Mindali et al., 2004), and results in lower fuel consumption and transport-related GHG emissions. Athens (2008) purported that a general trend towards more compact urban structures of higher densities and restricted urban expansion should be adopted to enhance sustainable travel patterns. However, according to Holden & Norland (2005), there is an ongoing debate on which urban form and land use can promote a more sustainable development. Therefore, tailored and multi-dimensional urban planning is warranted for each city and country (Dodman, 2009; Corfee-Morlot et al., 2009; Madlener & Sunak, 2011; Hoornweg et al., 2011; Kennedy et al., 2012), and city policymakers must select and implement urban planning instruments that are appropriate to the situation at hand (Iwata and Managi, 2016).

Nonetheless, regardless of the urban form under review, in most urban planning studies, it is commonly assumed that the residential locations are already fixed, while the consequences of locating various types of facilities have been studied (Dieleman et al., 2002; Krizek, 2003; Norman

et al., 2006; Perkins et al., 2009; Lindsey et al., 2011; Hong and Goodchild 2014). These studies mainly focus on an analysis of ex-post evaluation of the current urban form and its influence on the environment, and the empirical analyses are mainly done on metropolitan cities. Bourguignon and Ferreira (2003) pointed out that while the ex-post analyses examine the current situation after urban plans have been implemented, it is important to evaluate the potential impacts of counterfactual urban plans. It would be helpful for urban planners to have some estimates of how much each alternative plan would cost, and which aspect would be affected, and by how much, under each of the alternatives.

Land use-transport interaction (LUTI) models are attractive for addressing such issues as they may predict the impact of new infrastructure on the transport system (Acheampong and Silva, 2015). However, Johansen et al. (2015) argued that a LUTI model would, on the one hand, be expensive and labour-intensive to develop, and on the other hand, either too complex to understand, or too simple to capture all the urban processes. Kii et al. (2016), based on their comprehensive review, pointed out that more flexible and simplified modelling on a local scale that would extract the essences of urban activities is needed to meet the goal of mitigating emissions in urban planning. This paper proposes an ex-ante evaluation method in the spirit of LUTI models, but mainly evaluates the induced CO₂ emissions of car mobility from counterfactual urban residential plans to the primary destinations (i.e. hotspots) on a local scale, where the difference in induced CO₂ emissions of car mobility between the plans is quantified. The method is illustrated by examining a specific urban residential planning case. To the best of our knowledge, this method has not been adopted in the literature thus far.

Most studies we have reviewed have focused on examining the influence of land use and emissions in major cities. However, midsize cities are equally faced with challenges in urban planning for adapting the substantial increase of residents due to immigration and to mitigate the pollutant emissions. This paper considers a midsize city for the implementation of the method. The paper is structured as follows. Section 2 introduces the methodology. Section 3 describes the data used in the empirical case study, and Section 4 presents the results of the analysis. Section 5 concludes the paper.

2. Methodology

2.1 Ex-ante evaluation

Mobility is necessary for people to meet multiple needs, whereby intra-urban mobility between residential locations and hotspots is a primary part. Meanwhile, mobility induces negative externalities, such as CO₂ emissions. The proposed ex-ante evaluation method focuses on quantifying the impacts in terms of CO₂ emissions induced by intra-urban car mobility expected to

follow as a consequence of the adoption of residential plans in various urban areas. Essentially, ex-ante evaluation is a “what-if” analysis (Bourguignon and Ferreira, 2003). In this paper, two main questions are in focus:

- (1) What could be the change, compared to the current situation regarding CO₂ emissions in the city, if the residential plans were adopted?
- (2) Which subset of residential plans would reduce the CO₂ emissions the most if adopted?

To implement this ex-ante evaluation, first, the current origins and counterfactual origins, current destinations and counterfactual destinations¹ of the residents in the city need to be identified. Second, the emissions induced by moving from the origins to the destinations need to be quantified, where a network that gives the induced emissions by the mobility of cars is necessary for the quantification. Having constructed this network of induced emissions (in what follows denoted NoIEs), the emissions between any origins to any destination could readily be computed. From this, we may compare the current CO₂ emissions of a city (the factual case) with the residential plans (the counterfactual), and thereby evaluate the expected change in CO₂ emissions due to the mobility in the city arising from the adoption of some, or all of, the residential plans in the counterfactual scenarios.

2.2 Network of induced emissions (NoIEs)

In constructing the NoIEs, a model is needed for estimating the emissions based on cars’ mobility. Stead (1999) suggested using travel length as a proxy for vehicle emissions due to the ease in collection and computation; but this estimate is too coarse in the intra-urban mobility (Carling et al., 2013). Wielenmann et al. (2005) used average speed to estimate emissions from road traffic. They, however, overlooked that different driving behaviour and vehicle dynamics may nonetheless yield the same average speed. Davis et al. (2005) applied the international vehicle emissions (IVE) model. But obtaining the comprehensive conditions of vehicles and circumstance factors is complicated, and commonly it is found in these models that factors like speed, time, distance, acceleration and deceleration are crucial in estimating emissions (Demir et al., 2014).

Oguchi et al. (2002) proposed a model that uses these crucial factors, and it is particularly suitable for GPS tracking data that is commonly collected in the transportation field nowadays. The model estimates CO₂ emissions by considering the instantaneous working conditions of an on-road vehicle, such as speed change in terms of acceleration or deceleration, total travel time τ , and total travel distance D . In their model, change in speed is an important component, denoted by the indicator function δ_t , being one when the observed speed is higher than at the previous time point,

¹ We regard the starting point of a trip as an origin, and the end point of the trip as a destination.

and zero otherwise. This means that the acceleration deduced by the travel speeds v_t and v_{t-1} is considered. The model is

$$E = k_c \times \left[0.3\tau + 0.028D + 0.056 \sum_{t=2}^T \delta_t (v_t^2 - v_{t-1}^2) \right]^2$$

E = CO_2 emission amount (kg)

k_c = Coefficient of CO_2 emission being 0.002322 CO_2 (kg)/gasoline (cc)

τ = Travel time (sec)

D = Travel distance (m)

t = Time point when speed is observed

$\delta_t = 1$ (when the observed speed is higher than at the previous time point), 0 otherwise

v_t = Travel speed at point t (m/sec)

There is a need to distinguish between inter-urban and intra-urban traveling when it comes to estimating CO_2 emissions induced by the mobility in cars. In inter-urban traveling, the speed is fairly constant and time and distance are highly related; thus distance is the crucial determinant for the emissions. In intra-urban travel, however, constant speed is difficult to attain and the relationship between time and distance is much weaker, making speed change the crucial determinant for the emission quantity. For this reason, a good estimation of speed profile is crucial for deriving the NoIEs in an urban context, and the issue of estimating the speed profile is the focus of next subsection.

2.3 Speed profile

In what follows, and for reasons discussed below, we approximate the speed profile by a symmetric triangle function.

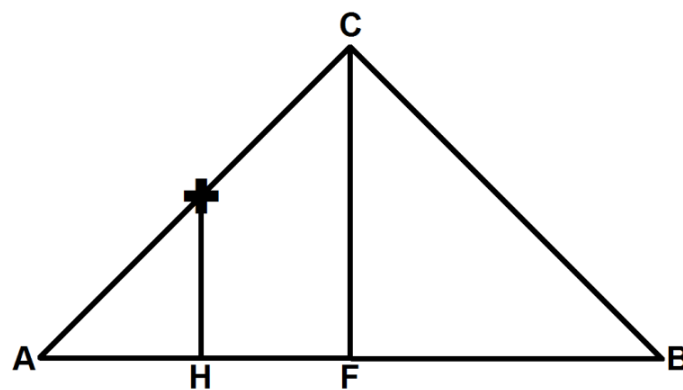


Figure 1. Illustration of the speed profile approximation.

² The three coefficients in the model were derived from experiments by testing real-time traffic conditions on urban roads which have been directly applied in various scenarios and found to be suitable for evaluation in estimating CO_2 emissions (Carling et al., 2013; Zhang and Zhu, 2015).

As is shown in Figure 1, line AB is a road segment, in which A and B represent the two intersections at the beginning and the end, while F represents the middle of this road segment. One car enters the road segment at A, with a low incoming speed, and then accelerates with a constant acceleration rate to attain the maximum speed in the middle of the road segment. Then, it decelerates at a constant rate to reach an outgoing speed at B, identical to the incoming speed at A. Having observed (e.g. from GPS recordings) the speed of a car at H during its travel on the road segment, the height of the triangle at the middle point (F) can be determined and thus an estimate of the maximum speed attained is provided. Upon knowing the length (distance) of the road segment, and having an estimate for the speed profile, the travel time for the road segment can be computed. Whereby, all the factors in the Oguchi et al. model (2002) have been quantified, providing for the estimation of the CO₂ emissions (E), produced by the car's travel on the road segment.

It goes without saying that the symmetric triangle function as an approximation of the car's speed profile is crucial for the computation of the NoEIs, and a close look at the validity of the approximation is warranted. We carried out a small experiment where roads in a residential area with infrequent car trips were chosen. Three drivers drove the same route and their speeds were measured every second by a GPS receiver (BT-338X). Figure 2(a) illustrates one such route of 350m, with 5 intersections. The drivers entered the arterial road from a collector road, and then drove out again. As is evident from Figure 2(b), the cars slowed down at the entry, speeded up, slowed down before the intersection, speeded up again, and then slowed down again at the exit, with some individual variation, but roughly in line with the triangle function.

The triangle function is an approximation to enable the estimation of the speed profile, but, admittedly, with a possible risk of underestimation (further, detailed examinations of the drivers' speed profile is given in Appendix A1). Considering that many (small) road segments in an urban area only have occasional car trips, there is a trade-off between bias and precision. The benefit of this triangle approximation is that the speed profile can be estimated (fairly accurately) by only one observation of the speed along the road segment, while any kind of elaborated, nonlinear specification requires multiple observations of the speed and would still run the risk of being highly imprecise.

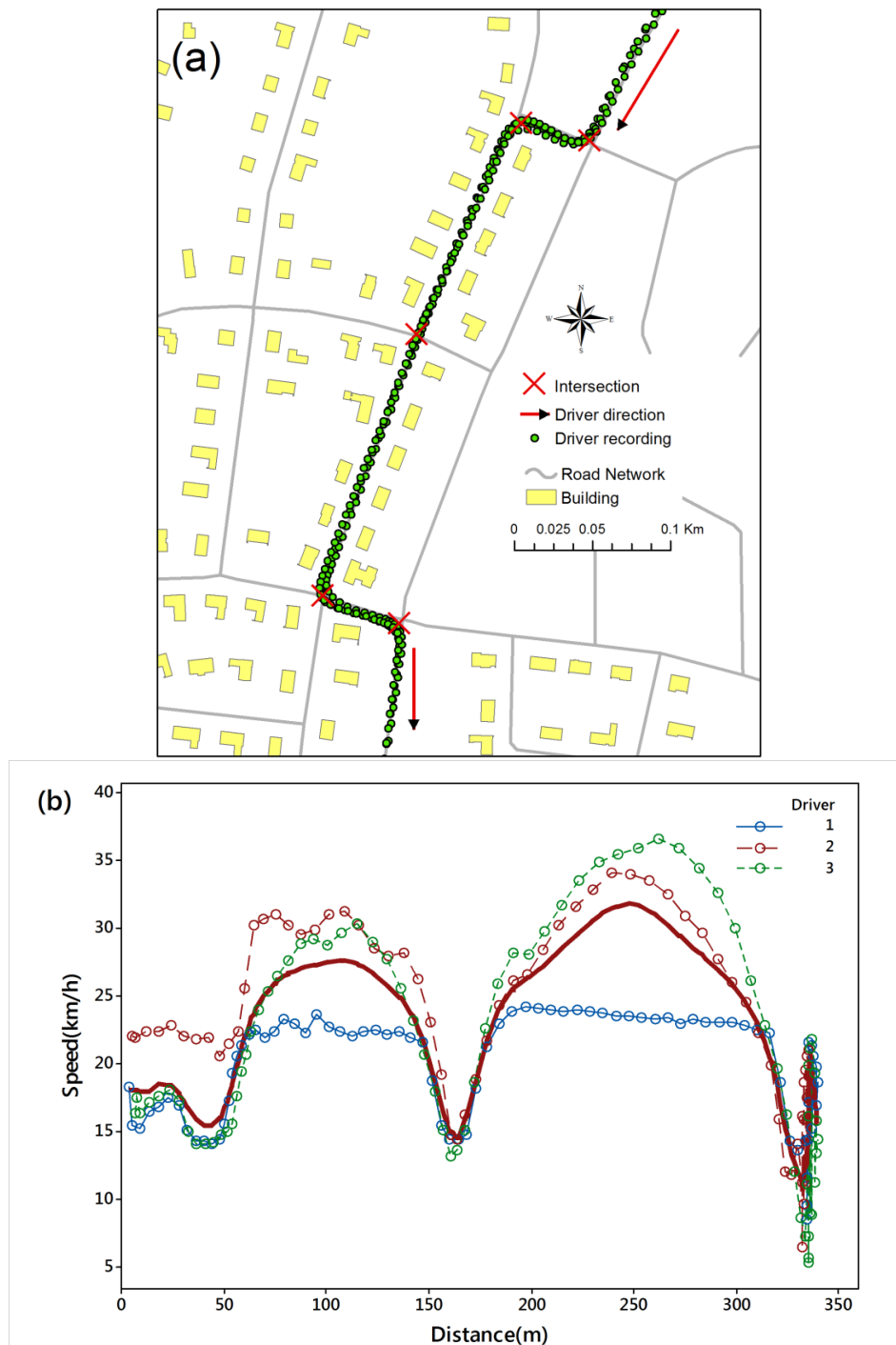


Figure 2. (a) Example of road segments travelled by three volunteers with recordings every second; (b) The observed speed profile for the three volunteers, where the thick-solid line is the LOWESS curve giving the general speed pattern.

3. Background of and data for the case city

In Sweden, 85% of the Swedish population lives in urban areas (Statistics Sweden, 2010). By 2050, the Swedish population is expected to grow by 15%, partly due to the dramatic increase of immigration (Statistics Sweden 2011). Midsize cities and the outskirts of bigger cities are expected to

accommodate the growth in population. Urban dwellers have moved away from the urban cores to the sub-centres (Anas et al., 1998), and to urban edge areas with rising settlements. This change has increased the dependency on car transportation, which leads to an increase in the number of cars per household and distances travelled (Behan et al. 2008), and the pollutant emissions due to car traffic have likely also increased. It is therefore challenging to reach the Swedish GHG target: a 40% reduction on 1990 levels by 2020, and a vision of 100% reduction (no net emissions of GHG to the atmosphere), by 2050 (Höjer et al., 2011).

Spatial planning in Sweden is largely the concern of the municipalities (in principle, the municipalities have a planning monopoly), which is regulated by law to accommodate certain land-use restrictions laid down in the Environmental Code, although they must coordinate their actions with other planning bodies (Persson, 2013). The two statutory plans at the municipal level are the comprehensive plan, and the detailed development plan. Since 1987, every municipality has adopted these two plans for spatial planning of sustainable land use, and coordinating the location and development of buildings, infrastructures, and various land use.

Borlänge is a midsize city in mid-Sweden that is situated 220 km northwest of Stockholm. The city is about 35 square kilometres, with approximately 50,000 residents. Furthermore, it is the labour market and commercial centre for the entire Dalarna region. Moreover, the continuous growth of immigrants and the economic growth in the city have led to an increased need for urban residential development and many residential plans have recently been presented.

Borlänge city is built up around a steel rolling mill on both sides of the Dalecarlia River, and a large railway yard in the city's south-western part, which act together as barriers for the development of residential areas in the city. Together with the old central business district along Stationgatan in the city core, the areas of Norra Backar, Domnarvet, Kvarnsvenden are sub-centres shaping the city in a heterogeneous polycentric pattern (Kloosterman & Musterd, 2001). The sub-centres have gradually been incorporated into an expanding, but coherent urban area, built around the original urban core. Other areas are newly spawned at nodes of the transportation network, far from the urban core, to sprawl the urbanisation to the edge area that is tangent to the city. We therefore categorise the city as follows: continued urban core area (CUCA), sub polycentric area (SPA) and edge urbanisation area (EUA).

Borlänge's development trends and housing styles tend to be quite typical of many other midsize cities in Sweden. The relatively compact central core and the polycentric sprawl pattern are also common to many midsize cities in Europe. The digital representation of residence and road network of Borlänge is detailed by the municipality and the National Road Data Base (NVDB). The current residents of the city are geo-coded into grids of 0.0625km^2 . We have identified the 1,291

centroid points of the grids, assigned the number of residents to each of the centroid points, as the grids' weight, and regard centroid points of the grids as the currently existing origins. Figure 3 illustrates how the residents of Borlänge are unevenly distributed in CUCA, SPA and EUA.

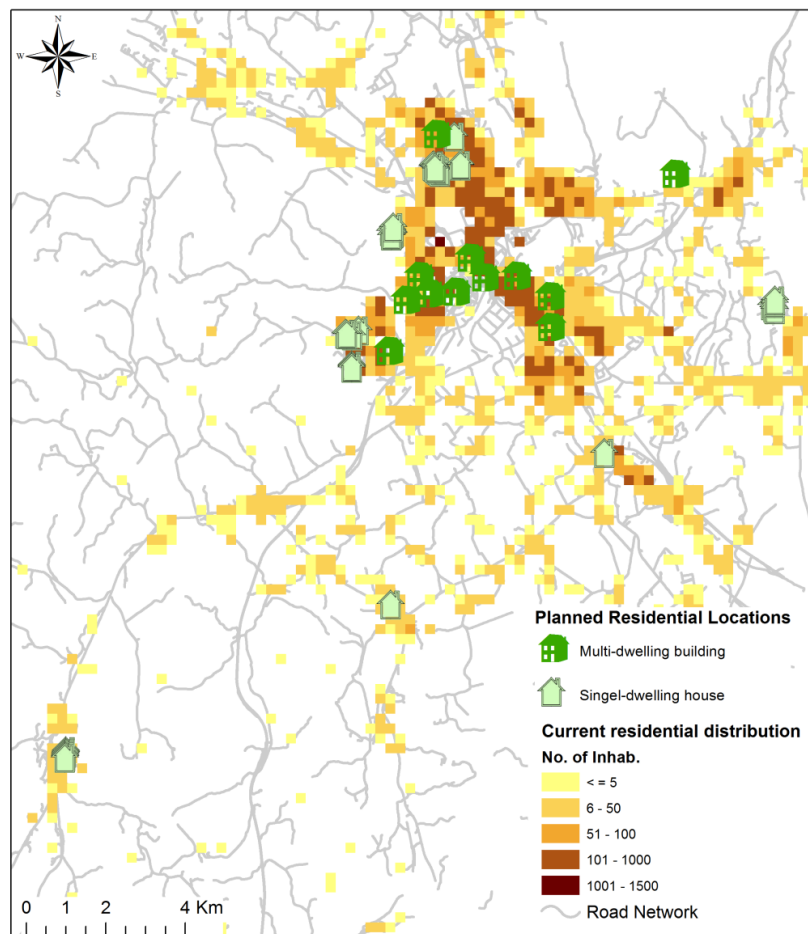


Figure 3. Planned new residential locations, according to the municipality's land planning office and the current residential distribution in Borlänge.

Figure 3 also illustrates the detailed residential development plan (expected to be adopted in the coming 5-15 years), according to the municipality's land plan office. Overall, the plan is to build new apartment buildings, single-family houses in the CUCA, SPA and EUA, by reusing abandoned places, parking lots and unexploited locations. 2,156 apartments are planned to be built at 12 locations to accommodate approximately 5,000 residents. 38 unexploited locations are planned to be developed into single-family houses for approximately 200 residents. In the CUCA, 4 apartment buildings that provide multiple residential units, and in a more compact pattern, are planned to be located. 7 apartment buildings and 25 single-family houses are planned in SPA, in a highly concentrated pattern adding to the current structure, while in EUA, there are plans for 1 apartment building and 13 single-family houses, dispersedly distributed to the north, south, and east of Borlänge.

In total, the plan is to meet the need for a population growth of 2% per year, which is a dramatic population increase compared to the average 0.4% annual population growth in the European Union³. The 50 planned residential locations, with a weight proportional to the planned number of residents, in the analysis are the counterfactual origins.

The mobility of cars in Borlänge is required for constructing the NoIEs for the city. We do this by making use of the observed mobility of 303 volunteers (residents of Borlänge), who were recruited via four sport clubs for mobility data collection from March 29 to May 15, 2011. Each volunteer was equipped with a GPS receiver (BT-338X) in the car to record their daily in-car mobility, where recordings were made every 30 seconds when the car was moving. The recordings provide information on geographical position, speed, date and time, and 154,579 valid recordings occurring in Borlänge are used for the data analysis in this paper (for details, see Zhao, 2015).

One question is: to what extent does the spatial distribution of the volunteers correspond to the spatial distribution of the population in Borlänge? Zhao (2015) demonstrated that the volunteers are spread throughout the city in a pattern similar to the overall population. However, in recruiting the volunteers, possession of a car was required. Presumably, people who live in SPA and EUA are more likely to own a car than those who live in CUCA, which explains why the volunteers have been found to be somewhat less concentrated in the CUCA area than the overall population.

Car driving must take place on the road network. The GPS recordings, however, are not always on the road network due to measurement errors, and data cleaning was therefore required before the analysis. According to the manufacturer, the BT-338X model should provide a geographical positioning within an error of 5 meters, and we therefore have applied a 5-meter tolerance in identifying valid recordings, and, as a consequence, have kept 129,353 recordings out of the original 154,579.

We use the speed information from the recordings to estimate the speed profile for each trip on a road segment in Borlänge. However, we first identified all recordings at intersections between road segments and those that connected to roundabout, for which the average speed was 15 and 20 km/h. These values were used as the incoming and outgoing speed, respectively. The estimated CO₂ emissions per trip for a road segment were derived by computing the average value of E over all trips that had been made by the volunteers on the road segment⁴.

³ http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_and_population_change_statistics

⁴ A truncation is applied for the estimate of the middle speed, if the speed profile implied that the maximum speed exceeds the speed limit plus 30km/h, then we trim this one as set to be the speed limit plus 30km/h. For road segments that have no recordings, the speed limit is assigned to the middle position.

We now turn to the identification of the primary destinations (i.e. hotspots) of the residents of Borlänge. The GPS receiver stops the recording once the car stops moving. 87% of the recordings indicated a moving car, while 13% of the recordings indicate that the car had stopped. There are two main explanations as to why a car stopped moving. One is for traffic reasons, such as traffic lights or queuing. The other is arrival at the destination. To identify destinations, we need to separate these two cases, and we have followed Jia et al. (2013) and classified any stop lasting less than 550 seconds as part of an ongoing trip, thus refraining from considering the location of the stop as a destination. For 7% of the recordings, the car starts moving again within 550 seconds, and thus the remaining 6% of the recordings are considered to identify the destination of the trip.

These 6% (or 7,932 recording positions) are identified as destinations and they are observed at multiple places, such as work places, kindergartens and schools, grocery stores and other types of stores, sports-arenas, as well as recreational areas on the edge of the city.

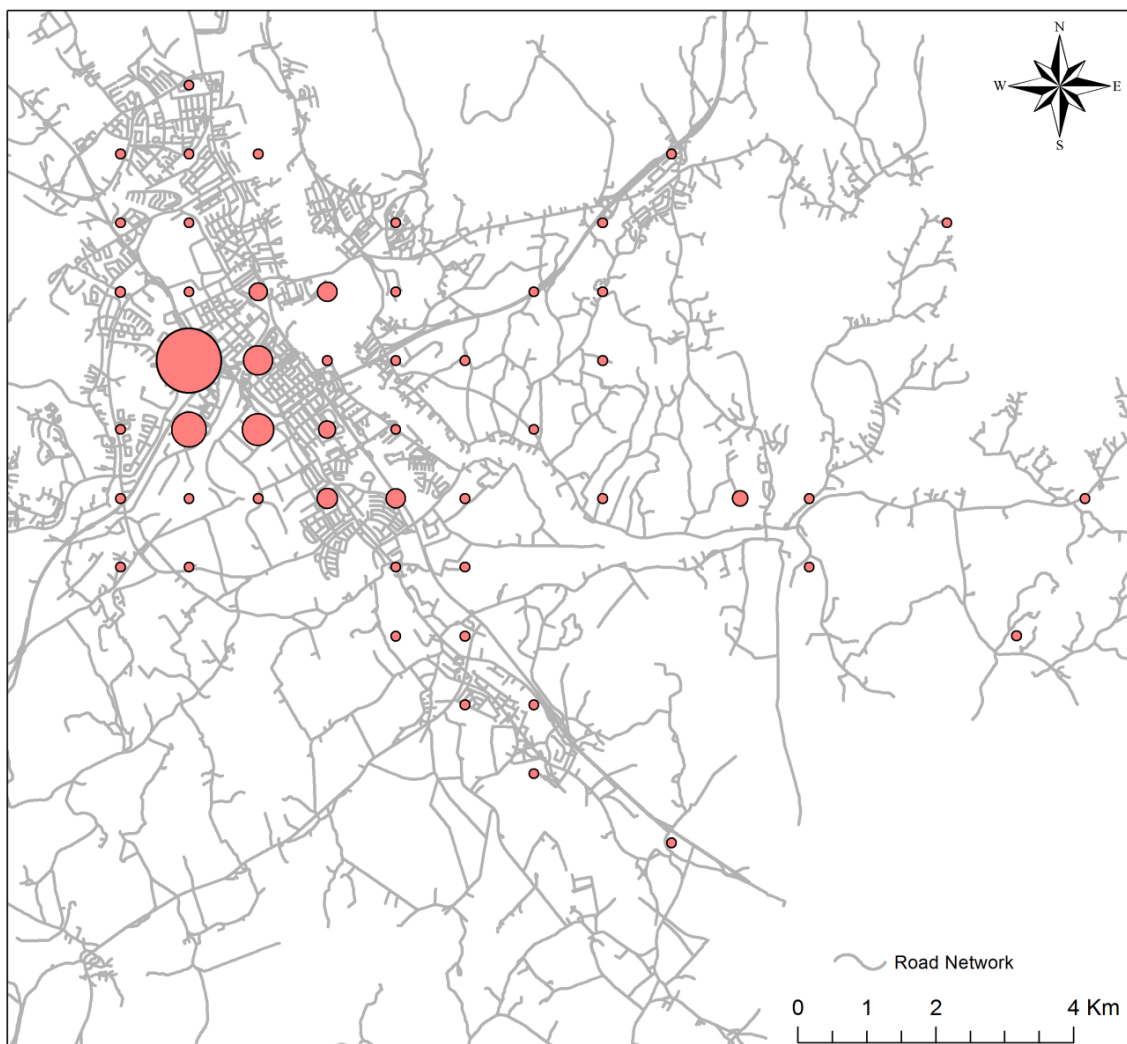


Figure 4. The spatial distribution of the 51 hotspots and their attractiveness.

The hotspots with the highest density are identified in the three main commercial areas: Kupolen in SPA, the city centre in CUCA, and Borlänge's largest grocery store, ICA Maxi, in SPA. These three destinations are hotspots, not only because they attract by a concentration of stores and restaurants, but also because they are important work places due to labour intensive work in the services sector⁵.

The number of observed destinations is very high, though most of these destinations are very seldom visited and therefore have little influence on the mobility in Borlänge. We have identified the 51 hotspots in Borlänge that attract more than 90% of all the trips in the city. These are depicted in Figure 4, and most of them are situated in the CUCA and SPA areas. In the counterfactual analysis, we assume that these factual destinations also are the counterfactual destinations, i.e. we expect future residents to live in the areas planned for residential areas to be attracted (to the same degree) by today's hotspots.

4. Results

Figure 5 shows the NoIEs of Borlänge standardised by the length of the road segment. "Low" indicates the 25 per cent of roads with the lowest emissions, "high" indicates the 25 per cent of roads with the highest emissions, while the remaining are the roads with medium-level emissions.

Emissions are high on some of the most frequently travelled roads, but also on the small streets in the residential areas. This heterogeneity implies that the residents' choice between the shortest, fastest or least emitting routes will affect the emissions generated by the mobility on the road network. In the ex-ante evaluation in this paper, we impose that residents opt for the shortest route as this has been found to be the best approximation to the residents' behaviour (Carling et al., 2013; Jia et al., 2013).

The origin-destination CO₂ emissions matrixes were constructed based on the NoIEs, weights were applied to each of the 51 hotspots proportional to the frequency in which the hotspots were visited. Currently, the average travel distance is 5.85 km, and the CO₂ emission per trip is 1.94 kg in Borlänge. The result of the counterfactual situation shows that 8 plans would imply lower CO₂ emissions than currently (detail information in Appendix, Table A2). These 8 locations amount to 47% of the planned residencies. Figure 6 shows that these 8 plans are apartment buildings to be located in CUCA and SPA. Relating to the current 1.94 kg of CO₂ emissions per trip, we classify the 50 residential plans into "lower", "similar" (within 15% of 1.94 kg) and "higher" (above 15% of

⁵ Unsurprisingly, a great deal of car stops takes place in the residential areas as people return to their homes. In the analysis, residential areas are not considered as destinations, albeit some mobility in a city could be attributed to residents visiting other residents.

1.94kg). There are 11 residential locations in the “similar” class planned to be 2 apartment buildings, and 7 single-family houses areas, all 11 to be located in SPA. The 31 residential locations in the class “above” are planning for 2 apartment buildings and 29 single-family houses areas and to be located in SPA and EUA.

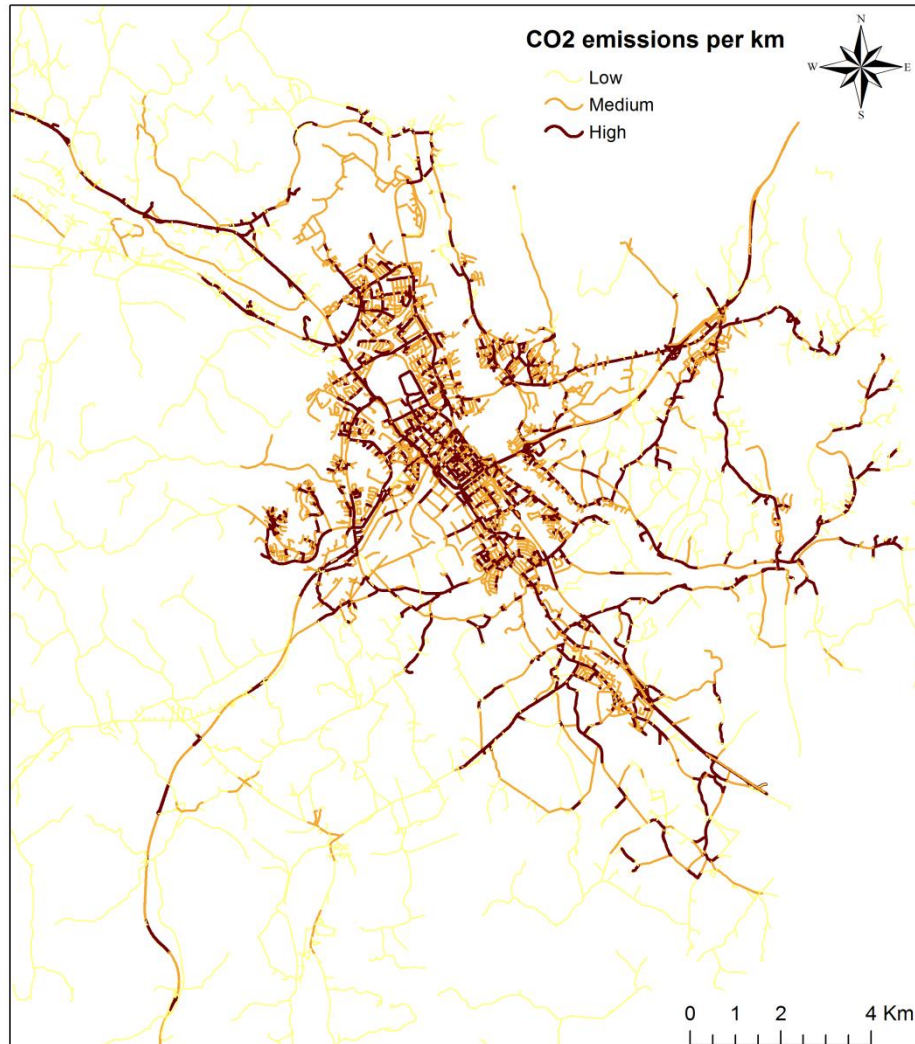


Figure 5. Road network of induced CO₂ emissions (NoIEs) standardised by the length of the roads.

In general, the plans deteriorate in terms of CO₂ emissions from CUCA to SPA to EUA. People who live in SPA will induce 1.95 times more CO₂ emissions per trip than people who live in CUCA. The induced CO₂ emissions per trip from people who live in EUA are the highest, and 4.32 times higher than in CUCA, and 2.57 times higher than SPA. The average travel distances for residents who will live in these three types of areas also differ to a large extent. People in SPA will travel 1.84 times longer than people in CUCA, while people in EUA will travel 4.11 and 2.23 times longer, respectively, than those in CUCA and SPA.

If none of the 14 plans in EUA were adopted, the remaining 36 plans would offer a residence for 54% of the new residents, and 35% of CO₂ emissions would be avoided. If all the 8 and 11 plans

that are in the classes “lower” and “similar” were realised, amounting to 51% of the planned residencies, 72% of CO₂ emissions would be avoided. If the optimisation objective is to minimise the CO₂ emissions while building 90% of the residencies, the 8 plans classified as “lower”, plus the one particular apartment building plan for 2,000 residents in EUA in Öster Barkargärdet, should be adopted. These 9 plans would provide homes for 92 % of the potential residents, while avoiding 89% of CO₂ emissions.

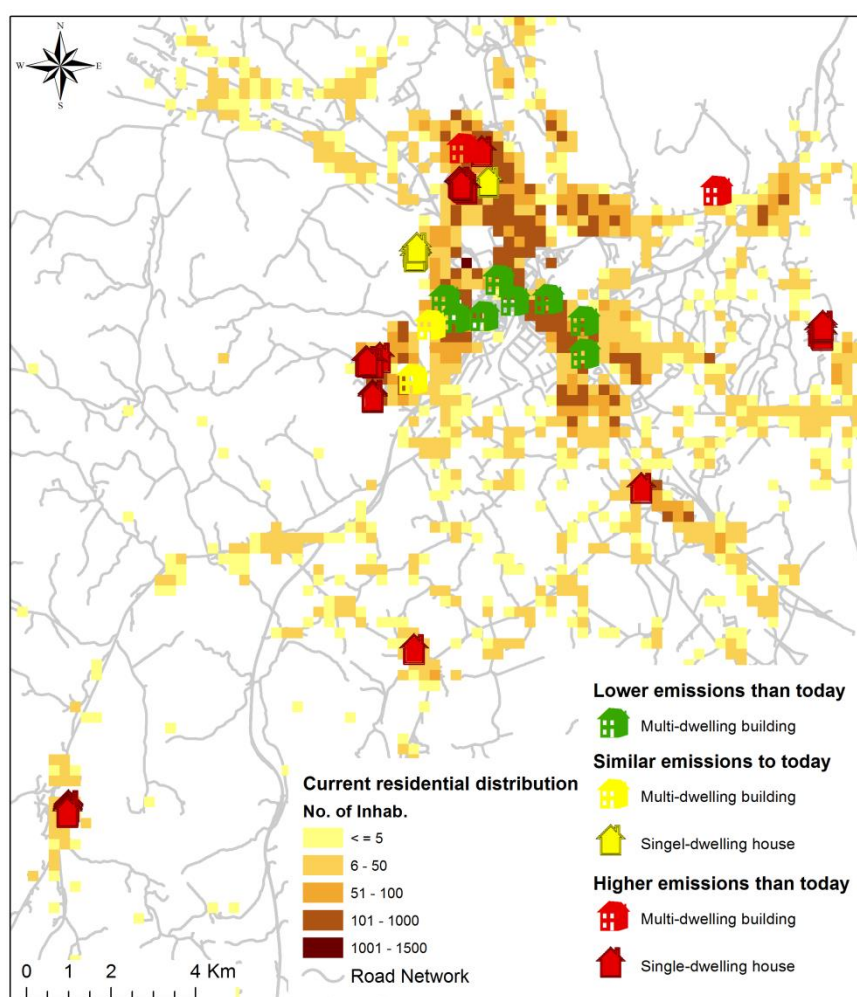


Figure 6. The 50 residential plans classified into whether they improve or worsen the mobility induced CO₂ emissions, compared to the present situation.

5. Concluding discussion

In most urban planning studies, the residential location has been considered fixed, and the consequences of where one sets up the various types of facilities has been debated. That is, the literature mainly focuses on the case of ex-post evaluation of relocating destinations, while few studies have considered an ex-ante evaluation of the location of the origins. This paper contributes by proposing an ex-ante evaluation method for quantifying the induced CO₂ emissions of car mobility in

a city at current and counterfactual situations. The results are compared in two dimensions, one between the current and the counterfactual situations, and the other between the counterfactual scenarios.

The empirical analysis was done for a case city (Borlänge), and considered 1,291 current residential locations, 50 counterfactual residential locations and 51 destinations in this midsize city. The urban plans were categorised into three areas, the continued urban core area (CUCA), sub polycentric area (SPA) and edge urbanisation area (EUA). The CO₂ emissions induced by car mobility of residents who counterfactually live in these three areas to the city's primary 51 destinations were quantified. In general, the CO₂ emissions increase from CUCA to SPA to EUA (1.17 kg vs. 2.28 kg vs. 3.00 kg). The average travel distances also increase from CUCA to SPA to EUA (3.40 km vs. 6.26 km vs. 13.96 km). The estimates may somewhat underestimate the benefits for people living in CUCA, as they may use other transportation modes than cars, as studied here. For this case, however, based on a check of which transportation mode is used by residents in Borlänge for their picking up of e-commerce packages, 83% were found to use cars rather than other transportation modes⁶, which suggests that the possible influence from other transportation mode to the result is minor.

The quantitative results indicate that urban residential plans can significantly influence CO₂ emissions induced by car mobility of residents. If the plans aim to increase the compactness of former urban cores and to increase the residential density of the developed sub centres, the CO₂ emissions can be mitigated effectively. An apartment building which can accumulate residents in a highly concentrated place would be preferable in reducing CO₂ emissions, which is in line with the advantages of a compact city (Burton, 2000; Holden & Norland, 2005). Urban expansion at the edge of a city could bring an increase in CO₂ emissions, due to long travel distance to hotspots, restricted accessibility and underdevelopment of public transport. On the other hand, in a polycentric city region, with several functional gravitation points that serve as destinations for the residents in the relatively peripheral parts, the scenario could be different, and the trade-offs between density and sustainability need to be adjusted (Naess, 2006; Reginster & Rounsevell, 2006; Westerink et al., 2013).

To be able to carry out the empirical analysis, certain simplifications, aggregations and assumptions were necessary. Firstly, we only had the residents in the city geo-coded with a precision of 250 m, and we could not identify whether a resident owned a car or not. Secondly, the speed profile used in deriving the network of the induced CO₂ emissions was based on data from 303

⁶ <http://www.itsdalarna.se/sv/sailor>, accessed at 2017-03-16.

volunteers in Borlänge, and a triangle function was used as an approximation of the speed profile. Further empirical checking of this approach would be welcomed to validate the ex-ante method. Furthermore, we follow Carling et al. (2013) and Jia et al. (2013) in supposing that the residents choose the shortest routes in their intra-urban mobility. However, in multiple-purpose trips and inter-urban trips, time and cost could possibly influence the choice of route.

It is also noted that, while this study has given specific consideration to empirical evidence in Borlänge, these findings are indicative of a more general relationship between residential planning and GHG emissions in many other European cities. In addition to residential planning and GHG emissions, city planners have multiple objectives to consider. Urban planning that is solely based on promoting one single goal is no longer the solution. Being able to identify the heterogeneous demands and combine strategies under various contexts is the way forward for sustainable urban development. Nonetheless, similar empirical studies in other cities, both within and outside Europe, would be helpful in moving toward a better quantitative understanding of urban planning effects.

References

- Acheampong, R. A., & Silva, E. (2015). Land use–transport interaction modeling: A review of the literature and future research directions. *Journal of Transport and Land Use*, 8(3).
- Alberti, M. (2005). The effects of urban patterns on ecosystem function. *International Regional Science Review*, 28(2), 168-192.
- Angel, S., Parent, J., & Civco, D. (2007, May). Urban sprawl metrics: an analysis of global urban expansion using GIS. In *Proceedings of ASPRS 2007 Annual Conference*, Tampa, Florida, May 7-11.
- Anas, A., Arnott, R., & Small, K. A. (1998). Urban spatial structure. *Journal of Economic Literature*, 36(3), 1426-1464.
- Athens, I. (2008). Relationships between urban form and travel behaviour in Athens, Greece. A comparison with Western European and North American results. *European Journal of Transport and Infrastructure Research*, 3(8).
- Behan, K., Maoh, H., & Kanaroglou, P. (2008). Smart growth strategies, transportation and urban sprawl: simulated futures for Hamilton, Ontario. *The Canadian Geographer/Le Géographe Canadien*, 52(3), 291-308.
- Bourguignon, F., & Ferreira, F. (2003). Ex ante evaluation of policy reforms using behavioral models. *The impact of economic policies on poverty and income distribution*, 123-142.
- Burton, E. (2000). The compact city: just or just compact? A preliminary analysis. *Urban Studies*, 37(11), 1969-2006.

- Buxton, M. and Scheurer, J., 2007. Density and outer urban development in Melbourne. *Urban Policy and Research*, 25(1), 91-111.
- Carling, K., Håkansson, J. and Jia, T., (2013), Out-of-town shopping and its induced CO₂-emissions, *Journal of Retailing and Consumer Services*, 20:4, 382-388.
- Corfee-Morlot, J., Kamal-Chaoui, L., Donovan, M. G., Cochran, I., Robert, A., & Teasdale, P. J. (2009). *Cities, climate change and multilevel governance*. OECD Environmental Working Papers 14. Paris: OECD Publishing.
- Davis, N., Lents, J., Osses, M., Nikkila, N., & Barth, M. (2005). Part 3: Developing countries: development and application of an international vehicle emissions model. *Transportation Research Record: Journal of the Transportation Research Board*, (1939), 155-165.
- Demir, E., Bektaş, T., & Laporte, G. (2014). A review of recent research on green road freight transportation. *European Journal of Operational Research*, 237(3), 775-793.
- Dhakal, S. (2008). Climate change and cities: The making of a climate friendly future. *Urban Energy Transition: From Fossil Fuels to Renewable Power*. Elsevier, Amsterdam, pp.173-193.
- Dieleman, F. M., Dijst, M., & Burghouwt, G. (2002). Urban form and travel behaviour: micro-level household attributes and residential context. *Urban Studies*, 39(3), 507-527.
- Dieleman, F., & Wegener, M. (2004). Compact city and urban sprawl. *Built Environment*, 30(4), 308-323.
- Dodman, D. (2009). Blaming cities for climate change? An analysis of urban greenhouse gas emissions inventories. *Environment and Urbanization*, 21(1), 185-201.
- Dubovyk, O., Sliuzas, R., & Flacke, J. (2011). Spatio-temporal modelling of informal settlement development in Sancaktepe district, Istanbul, Turkey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(2), 235-246.
- Ewing, R., & Cervero, R. (2010). Travel and the built environment: a meta-analysis. *Journal of the American Planning Association*, 76(3), 265-294.
- Fujita, M., & Thisse, J. F. (2013). *Economics of agglomeration: cities, industrial location, and globalization*. Cambridge University Press.
- Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., & Briggs, J. M. (2008). Global change and the ecology of cities. *Science*, 319(5864), 756-760.
- Hamin, E. M., & Gurran, N. (2009). Urban form and climate change: Balancing adaptation and mitigation in the US and Australia. *Habitat International*, 33(3), 238-245.

- Holden, E., & Norland, I. T. (2005). Three challenges for the compact city as a sustainable urban form: household consumption of energy and transport in eight residential areas in the greater Oslo region. *Urban Studies*, 42(12), 2145-2166.
- Hong, J., & Goodchild, A. (2014). Land use policies and transport emissions: Modelling the impact of trip speed, vehicle characteristics and residential location. *Transportation Research Part D: Transport and Environment*, 26, 47-51.
- Hoorweg, D., Sugar, L., & Gomez, C. L. T. (2011). Cities and greenhouse gas emissions: moving forward. *Environment and Urbanization*, 23(1):207–227.
- Hunt, A. and Watkiss, P. (2011) Climate change impacts and adaptation in cities: a review of the literature. *Climatic Change*, 104(1), pp. 13-49.
- Höjer, M., Gullberg, A., & Pettersson, R. (2011). Backcasting images of the future city—Time and space for sustainable development in Stockholm. *Technological Forecasting and Social Change*, 78(5), 819-834.
- International Energy Agency. (2009). *Transport Energy and CO2: Moving Towards Sustainability*. OECD Publishing.
- Iwata, K., & Managi, S. (2016). Can land use regulations and taxes help mitigate vehicular CO2 emissions? An empirical study of Japanese cities. *Urban Policy and Research*, 1-17.
- Jia, T., Carling, K., Håkansson, J.(2013). Trips and their CO2 emissions to and from a shopping center. *Journal of Transport Geography*. 33,135–145.
- Johansen, B. G., Hansen, W., & Tennøy, A. (2015). Evaluation of models and methods for analysing the interaction between land-use, infrastructure and traffic demand in urban areas. *TØI Report*, (1415/2015).
- Kennedy, C., Demoullin, S., & Mohareb, E. (2012). Cities reducing their greenhouse gas emissions. *Energy Policy*, 49, 774-777.
- Kenworthy, J. R., & Laube, F. B. (1999). Patterns of automobile dependence in cities: an international overview of key physical and economic dimensions with some implications for urban policy. *Transportation Research Part A: Policy and Practice*, 33(7), 691-723.
- Kenworthy, J. R. (2003). Transport energy use and greenhouse gases in urban passenger transport systems: a study of 84 global cities. In: *Proceedings of the Third Conference of the Regional Government Network for Sustainable Development*, Notre Dame University, Fremantle, Western Australia, September 17– 19, 2003.
- Kii, M., Nakanishi, H., Nakamura, K., & Doi, K. (2016). Transportation and spatial development: An overview and a future direction. *Transport Policy*, 49, 148-158.

- Kloosterman, R. C., & Musterd, S. (2001). The polycentric urban region: towards a research agenda. *Urban Studies*, 38(4), 623-634.
- Krizek, K. J. (2003). Residential relocation and changes in urban travel: does neighborhood-scale urban form matter? *Journal of the American Planning Association*, 69(3), 265-281.
- Li, J. (2011). Decoupling urban transport from GHG emissions in Indian cities—A critical review and perspectives. *Energy policy*, 39(6), 3503-3514.
- Li, X., Zhou, W., & Ouyang, Z. (2013). Forty years of urban expansion in Beijing: what is the relative importance of physical, socioeconomic, and neighborhood factors?. *Applied Geography*, 38, 1-10.
- Lindsey, M., Schofer, J. L., Durango-Cohen, P., & Gray, K. A. (2011). The effect of residential location on vehicle miles of travel, energy consumption and greenhouse gas emissions: Chicago case study. *Transportation Research Part D: Transport and Environment*, 16(1), 1-9.
- Long, Y., Gu, Y., & Han, H. (2012). Spatiotemporal heterogeneity of urban planning implementation effectiveness: Evidence from five urban master plans of Beijing. *Landscape and Urban Planning*, 108(2), 103-111.
- Madlener, R., & Sunak, Y. (2011). Impacts of urbanization on urban structures and energy demand: What can we learn for urban energy planning and urbanization management? *Sustainable Cities and Society*, 1(1), 45-53.
- Makido, Y., Dhakal, S., & Yamagata, Y. (2012). Relationship between urban form and CO 2 emissions: evidence from fifty Japanese cities. *Urban Climate*, 2, 55-67.
- Mindali, O., Raveh, A., & Salomon, I. (2004). Urban density and energy consumption: a new look at old statistics. *Transportation Research Part A: Policy and Practice*, 38(2), 143-162.
- Næss, P. (2006). *Urban structure matters: residential location, car dependence and travel behaviour*. Routledge, London and New York
- Naess, P. (2012). Urban form and travel behavior: Experience from a Nordic context. *Journal of Transport and Land Use*, 5(2), 21-45.
- Norman, J., MacLean, H. L., & Kennedy, C. A. (2006). Comparing high and low residential density: life-cycle analysis of energy use and greenhouse gas emissions. *Journal of Urban Planning and Development*, 132(1), 10-21.
- O'Neill, B. C., Dalton, M., Fuchs, R., Jiang, L., Pachauri, S., & Zigova, K. (2010). Global demographic trends and future carbon emissions. In *Proceedings of the National Academy of Sciences*, 107(41), 17521-17526.

- Oguchi, T., KATAKURA, M., & TANIGUCHI, M. (2002). Carbondioxide emission model in actual urban road vehicular traffic conditions. *Doboku Gakkai Ronbunshu*, 2002(695), 125-136.
- Perkins, A., Hamnett, S., Pullen, S., Zito, R., & Trebilcock, D. (2009). Transport, housing and urban form: the life cycle energy consumption and emissions of city centre apartments compared with suburban dwellings. *Urban Policy and Research*, 27(4), 377-396.
- Persson, C. (2013). Deliberation or doctrine? Land use and spatial planning for sustainable development in Sweden. *Land Use Policy*, 34, 301-313.
- Pickett, S. T. A., Cadenasso, M. L., Grove, J. M., Boone, C. G., Groffman, P. M., Irwin, E., et al. (2011). Urban ecological systems: scientific foundations and a decade of progress. *Journal of Environmental Management*, 92(3), 331-362.
- Poumanyong, P., & Kaneko, S. (2010). Does urbanization lead to less energy use and lower CO 2 emissions? A cross-country analysis. *Ecological Economics*, 70(2), 434-444.
- Reginster, I., & Rounsevell, M. (2006). Scenarios of future urban land use in Europe. *Environment and Planning B: Planning and Design*, 33(4), 619-636.
- Satterthwaite, D. (2008). Cities' contribution to global warming: notes on the allocation of greenhouse gas emissions. *Environment and Urbanization*, 20(2), 539-549.
- Schwanen, T. (2002). Urban form and commuting behaviour: a cross - European perspective. *Tijdschrift voor economische en sociale geografie*, 93(3), 336-343.
- Schwarz, N. (2010). Urban form revisited—Selecting indicators for characterising European cities. *Landscape and Urban Planning*, 96(1), 29-47.
- Seto, K. C., Fragkias, M., Güneralp, B., & Reilly, M. K. (2011). A meta-analysis of global urban land expansion. *PLoS ONE*, 6(8), e23777.
- Stead, D. (1999). Relationships between transport emissions and travel patterns in Britain. *Transport Policy*, 6(4), 247-258.
- Stead, D., & Marshall, S. (2001). The relationships between urban form and travel patterns. An international review and evaluation. *European Journal of Transport and Infrastructure Research*, 1(2), 113-141.
- Statistics Sweden (SCB):
http://www.scb.se/statistik/MI/MI0810/2010A01/MI0810_2010A01_SM_MI38SM1101.pdf
- Tavares, A. O., Pato, R. L., & Magalhães, M. C. (2012). Spatial and temporal land use change and occupation over the last half century in a peri-urban area. *Applied Geography*, 34, 432-444.
- Thapa, R. B., & Murayama, Y. (2010). Drivers of urban growth in the Kathmandu valley, Nepal: examining the efficacy of the analytic hierarchy process. *Applied Geography*, 30(1), 70-83.

- United Nations, (2014). *World urbanization prospects (2014 revision)*. Retrieved from: <https://esa.un.org/unpd/wup/publications/files/wup2014-highlights.Pdf>
- Wang, H., He, Q., Liu, X., Zhuang, Y., & Hong, S. (2012). Global urbanization research from 1991 to 2009: A systematic research review. *Landscape and Urban Planning*, 104(3), 299-309.
- Weng, Q., Liu, H., & Lu, D. (2007). Assessing the effects of land use and land cover patterns on thermal conditions using landscape metrics in city of Indianapolis, United States. *Urban Ecosystems*, 10(2), 203-219.
- Weilenmann, M., Soltic, P., Saxer, C., Forss, A. M., & Heeb, N. (2005). Regulated and nonregulated diesel and gasoline cold start emissions at different temperatures. *Atmospheric Environment*, 39(13), 2433-2441.
- Westerink, J., Haase, D., Bauer, A., Ravetz, J., Jarrige, F., & Aalbers, C. B. (2013). Dealing with sustainability trade-offs of the compact city in peri-urban planning across European city regions. *European Planning Studies*, 21(4), 473-497.
- Zhang, L. D., & Zhu, W. X. (2015). Delay-feedback control strategy for reducing emission of traffic flow system. *Physica A: Statistical Mechanics and its Applications*, 428, 481-492.
- Zhao, X. (2015). On processing GPS tracking data of spatio-temporal car movements: a case study. *Journal of Location Based Services*, 9(4), 235-253.

Appendix

A1: Detailed examination of speed profile

The following describes the examination of validity of the triangle approximation for speed profile in detail. Figure A1-1(a) illustrates an example road of 700m that lies between a gas station STATOIL and a hotspot ICA MAXi. 1, 225 recordings from several drivers who have repeatedly driven on this road segment were obtained.

Figure A1-2 shows the speed profile of the recorded speeds. The LOWESS curve indicates that the speed profile can be approximated by a symmetric triangle function. Compared to this road segment, most road segments have much less traffic flow, even occasional car trips. As is shown in Figure A1-1(a) and Figure A1-1(b), those road segments in the red circles only have 1 or 2, or even no, recording(s). Moreover, the volunteers' movements were measured every 30 seconds, which was insufficiently frequent to obtain recordings for the many short road segments in the city. This is problematic, as the Oguchi et al. (2002) model requires at least two speeds on a road segment to estimate the induced CO₂ emissions. Therefore, a post-test was conducted, where three drivers were equipped with the same GPS receivers (BT-338X) while taking the same route, and recordings were collected every second.



Figure A1-1. Example 1: (a) Positional recordings the volunteers have made on a main road; (b) Positional recordings that the volunteers have made on arterial roads and private roads.

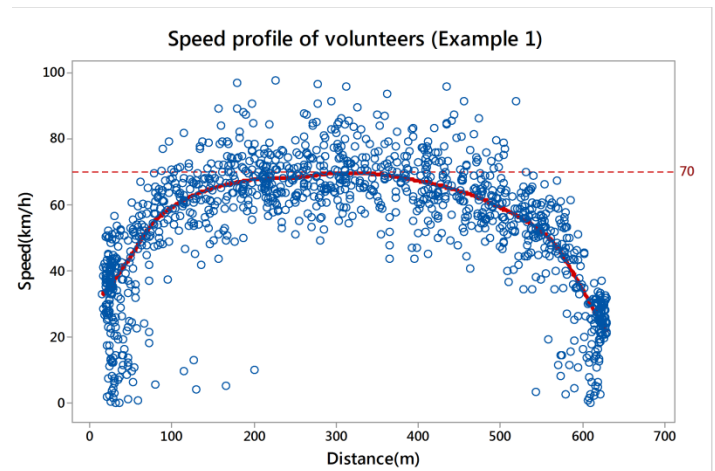


Figure A1-2. The speed profile for the recordings on road segments in Figure A1-1(a). The observed speed profile, where the thick solid line is the LOWESS curve giving the general speed pattern.

Figure A1-3(a) shows an example route that has both the volunteers and the post-test drivers' recordings. Along the route there is a traffic light, a speed bump, a left turn to a bigger road and a speed bump again. Hence, according to the triangle approximation, we should expect 4 occurrences of accelerations and decelerations. Figure A1-3(b) shows the speed profile from the three drivers. Although the same details for the volunteers are unavailable, we show the average speed profile for all the volunteers who travelled on this route (see Figure A1-3(c)). The post-test drivers and the volunteers behave similarly: they first speeded up then slowed down at the traffic light, then speeded up to about 40 and slowed down for the speeded bump, speeded up to about 40 again and slowed down for the left turn, again speeded up and slowed down for the speed bump, and finally speeded up to 60. Thus, both the volunteers and the post-test drivers' speed profiles approximate to a symmetric triangle function.

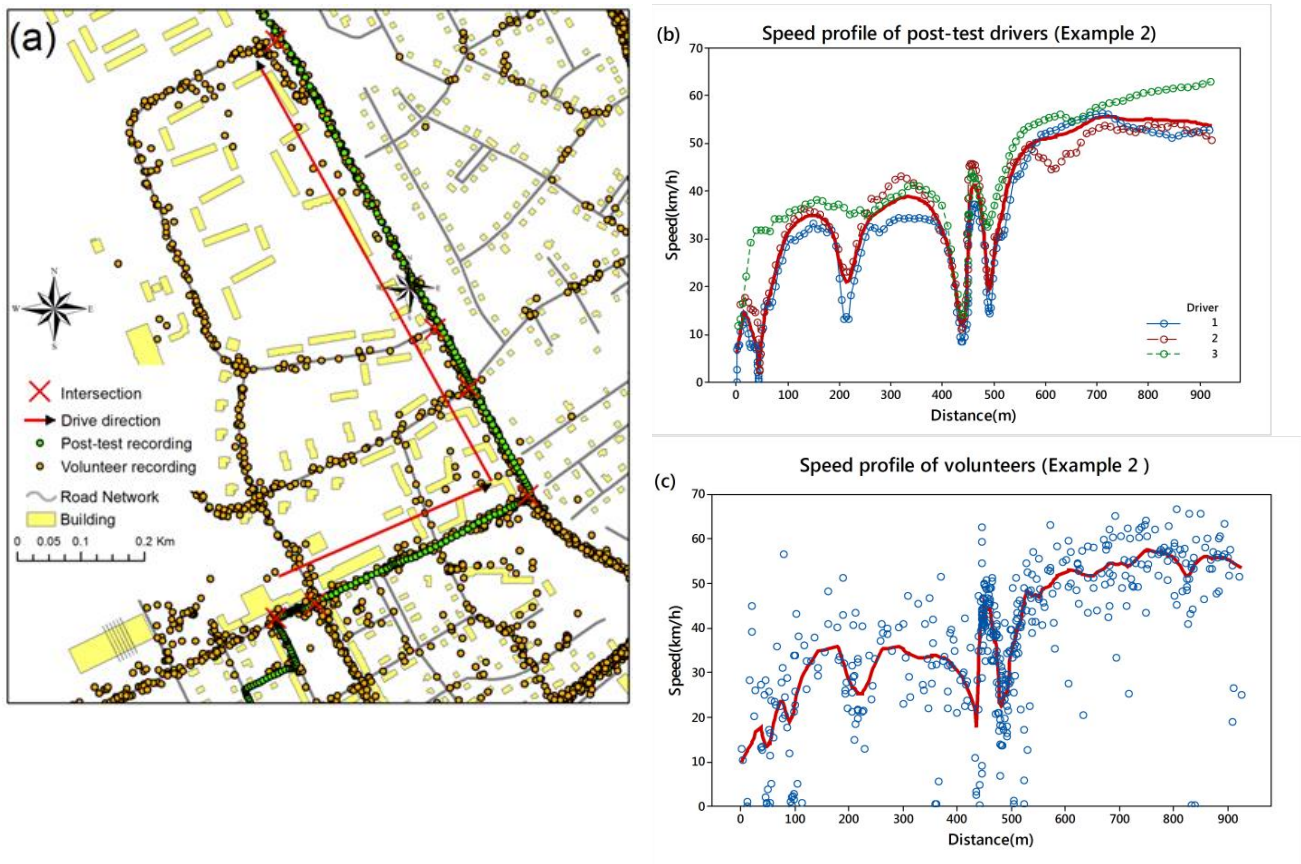


Figure A1-3. Example 2: (a) The road segments that had recordings from the volunteers and the post-test drivers; (b) The speed profile of the recordings from the post-test drivers; (c) The speed profile of the recordings from the volunteers. The thick solid line is the LOWESS curve giving the general speed pattern.

In another example as is illustrated in Figure A1-4(a), the driver drove out from an arterial road to a collector road. There is one walk/bike path, one speed bump and then the driver arrives at a roundabout, so we should expect 3 sets of accelerations and decelerations. Figure A1-4(b) and Figure A1-4(c) give the (average) speed profiles of the three post-test drivers and the volunteers, respectively. Once again, they illustrate that the symmetric triangle function is fairly accurate for estimating the speed profile.

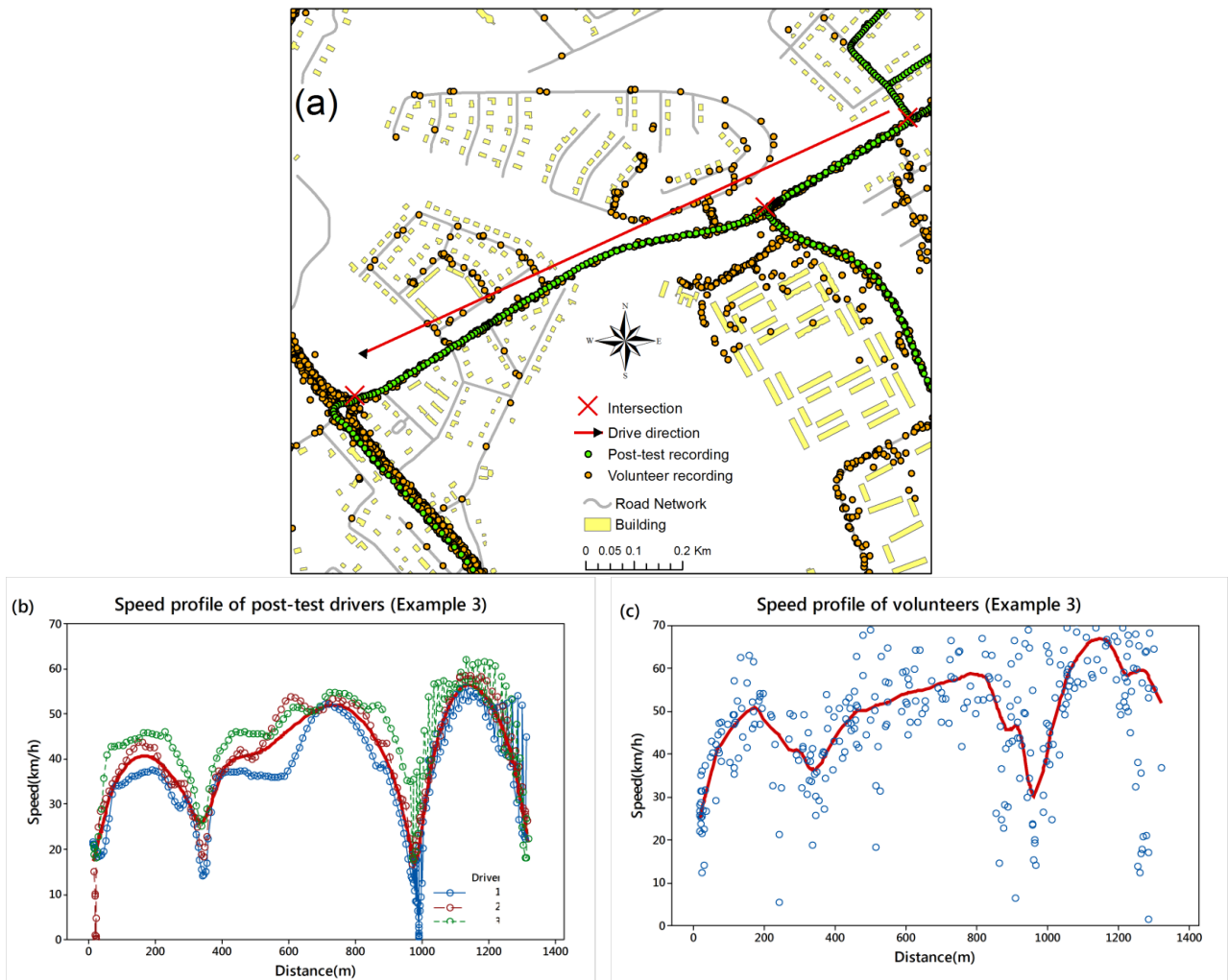


Figure A1-4. Example 3: (a) The road segments that had recordings from the volunteers and the post-test drivers; (b) The speed profile of the recordings from the post-test drivers; (c) The speed profile of the recordings from the volunteers. The thick solid line is the LOWESS curve giving the general speed pattern.

A2: The average travel distance and CO2 emissions in different scenarios

Table A2: The average travel distance and CO2 emissions per trip, per person, of current stage and the counterfactual stage.

	<i>Average travel distance per trip(km)</i>	<i>Average CO₂ emission per trip(kg)</i>	<i>Number of Residents</i>
Current situation	<u>5.853291 (0.02)</u>	<u>1.935696 (0.004)</u>	46, 942
Counterfactual situation (averaged based on the areas)			
CUCA	<u>3.400389 (0.07)</u>	<u>1.166777 (0.05)</u>	800
SPA	<u>6.261431 (0.19)</u>	<u>2.279009 (0.07)</u>	1570
EUA	<u>13.958923 (1.49)</u>	<u>3.001324 (0.11)</u>	2056

Note: Standard error is in the parentheses.

Counterfactual situation of each 50 residential plans				
ID of Residential Relocation Plan	Average travel distance per trip(km)	Average CO ₂ emission per trip(kg)	Planned number of residents	Type of area
1	3.322068	1.091739	400	CUCA
2	3.262528	1.097872	200	CUCA
3	3.419151	1.1894	100	CUCA
4	3.59821	1.288097	100	CUCA
5	3.764156	1.322576	40	SPA
6	3.819089	1.382415	80	SPA
7	4.405589	1.502445	400	SPA
8	4.594678	1.715279	800	SPA
9	5.763789	1.941382	4	SPA
10	5.460784	1.946403	30	SPA
11	5.778661	1.946876	4	SPA
12	5.616241	1.968728	80	SPA
13	5.341958	2.033146	4	SPA
14	6.033929	2.134565	4	SPA
15	6.0504	2.137013	4	SPA
16	6.0504	2.137013	4	SPA
17	6.0504	2.137013	4	SPA
18	6.184124	2.156885	4	SPA
19	6.214365	2.161379	4	SPA
20	6.819743	2.307555	4	EUA
21	11.71086	2.388115	8	EUA
22	6.898326	2.409262	2000	EUA
23	6.349138	2.470331	4	SPA
24	7.543796	2.528684	4	SPA
25	7.543894	2.528715	4	SPA
26	7.553595	2.544834	4	SPA
27	6.496725	2.571739	4	SPA
28	6.485287	2.572827	4	SPA
29	6.485074	2.572847	4	SPA

Paper IV

30	7.805994	2.629331	4	SPA
31	6.659685	2.63633	4	SPA
32	6.659685	2.63633	4	SPA
33	6.691149	2.641006	4	SPA
34	6.717806	2.644967	4	SPA
35	6.692213	2.658346	40	SPA
36	6.820574	2.66024	4	SPA
37	7.550281	2.665502	4	SPA
38	7.578937	2.669761	4	SPA
39	7.603381	2.673393	4	SPA
40	9.826835	2.884317	4	EUA
41	9.87237	2.888909	4	EUA
42	9.959335	2.897678	4	EUA
43	10.12383	2.914267	4	EUA
44	10.14795	2.916699	4	EUA
45	19.97554	3.397386	4	EUA
46	19.97433	3.39822	4	EUA
47	19.97977	3.398768	4	EUA
48	20.00999	3.40086	4	EUA
49	20.05814	3.406671	4	EUA
50	20.0679	3.409835	4	EUA

Paper V

5966.1

On assessing governmental sustainable residential planning and its alignment with residents' and estate investors' objectives

Xiaoyun Zhao[♦], Kenneth Carling, Johan Håkansson

There are three key actors in forming the sustainable spatial distribution of residency in an area, (local) government, the estate investor and the resident, each with its own objective. Most urban planning studies have mainly focused on the ex-post evaluation of residential development by considering the objective of each actor separately. This paper outlines a conceptual model where the three key actors and their unique objectives are integrated with the aim of providing an ex-ante evaluation of residential development for government to make policies operational on a micro level. The methodology is implemented on a Swedish city, where sustainable residential development is in high need due to the influx of immigrants. The case study demonstrates that the model can integrate the macro and micro actors well. The model can provide noteworthy insights for the government on where the objectives of sustainability, livability and profit can be met. A sensitivity check of the parameter settings shows that the implementation of the model is robust for replication in other cities.

Keywords: Sustainable urban development; CO₂ emissions; Multi-agent system model; Urban mobility

1. Introduction

Under the tide of sustainability and livability, governmental residential planning, estate investment and residential locations are gradually forming the spatial distribution of residency in a city. As Godschalk (2004) points out: “Land use planning faces both an opportunity and a threat. On the one hand, it is widely counted on and expected to deliver both sustainable development and livable communities. On the other hand, it must cope with serious conflicts in the values related to these two beguiling visions.” There are three key actors in forming the residential, spatial distribution, namely, (local) government, the estate investor and the resident, each with its own objective. One strand of the literature has focused on the actors' objectives and their determinants, while another strand has discussed how a government may form policies for sustainable, residential development. However,

[♦] Xiaoyun Zhao is a PhD student in Microdata Analysis, Kenneth Carling is a professor in Statistics, and Johan Håkansson is a professor in Human Geography. All are at the School of Technology and Business Studies, Dalarna University, SE-791 88 Falun, Sweden. *Corresponding author:* Xiaoyun Zhao, e-mail: xzh@du.se, phone: +46-23-778509.

to render such policies efficient on a micro level, the microscopic behavior of the other actors needs to be integrated into governmental policy. The aim of this paper is to develop such an integrated method for the identification of efficient, sustainable governmental policies.

In this paper, we set out to outline a conceptual model for the interaction of the three key actors in the forming of residential development – government, (residential) estate investors, and residents – each actor with their unique objective. Furthermore, we develop a methodology to ex-ante assess if, and where, the government’s sustainability objective may be in conflict with the objectives of the other market actors. The method is applied and demonstrated in a representative Swedish case study, the mid-sized city of Borlänge, which, similar to many other Swedish cities, is currently planning for massive residential development in response to population growth due to *inter alia* -the influx of immigrants.

In outlining the conceptual model, we define each actor’s objective (in utility terms), where we draw heavily on previous research that has scrutinized the actor’s objective separately. We regard the government as the macro actor in residential development, but purport that the government exerts the resident’s living preferences and the estate investor’s willingness to invest in its utility function. The objectives and behavior of the estate investor and the resident are formulated on a micro level, meaning that our approach belongs to the tradition of microscopic models frequently applied in the transportation science when it comes to model the behavior of the market actors (Waddell, 2002; Iacono et al., 2008; Haase, and Schwarz, 2009).

The implementation of the methodology requires access to some key variables or proxy-variables thereof: (1) the spatial distribution of market value of residencies and the built-up cost, (2) road network and its traffic flow and the corresponding emissions (or other outputs giving rise to negative externalities), (3) availability of data on fractions of green land and water in the surrounding area, and (4) residents’ consumption. In the case study, we demonstrate how these variables can be obtained, measured or derived. The implementation of the methodology also requires a number of parameters to be set. To ensure that the results for the case study are robust for these settings, we conduct a sensitivity analysis where the most reasonable setting (default) is replaced by an aberrant setting.

Accordingly, the paper is structured as follows: Section 2 provides a review of the literature upon which our conceptual model that integrates the objectives of the three actors in residential development is based. In Section 3, we outline the conceptual model. Section 4 is devoted to the case study we use for demonstrating the implementation of the methodology for assessing if, and where,

the government's sustainability objective may be in conflict with the objectives of the other market actors. Section 5 concludes the paper.

2. A review of previous research

In residential development literature, the government is typically regarded as the macro actor, whilst the estate investor and the resident act on a micro level. The relationship between these two levels can be structured as illustrated in Figure 1 (*cf* Sun et al., 2014). The government possesses the power of approving a certain location for residential development. The (estate) investors and residents behave in the market, based on the assessment of willingness to develop and willingness to live. This microscopic behavior may promote or hinder the fulfillment of the governmental objective.

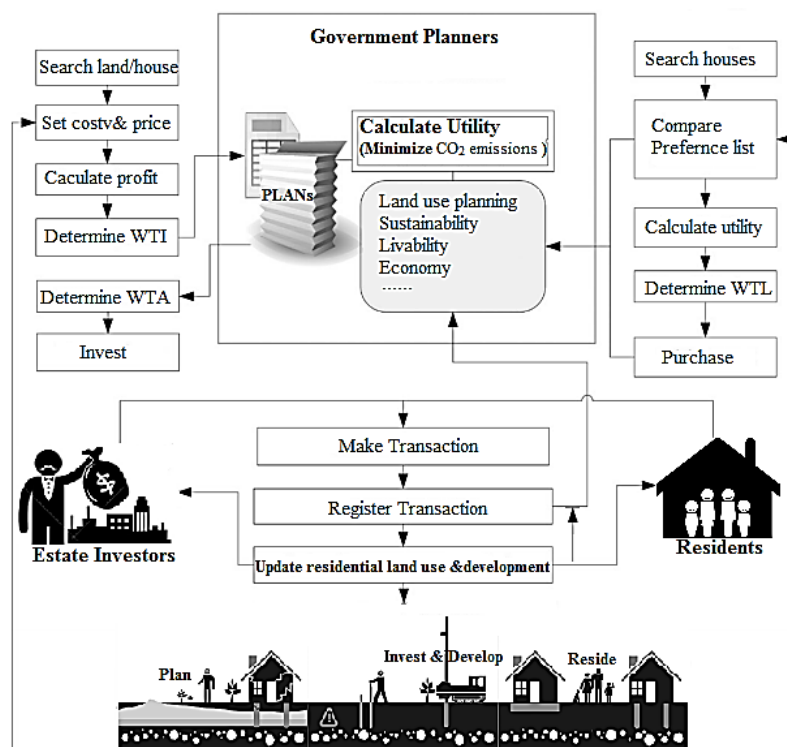


Figure 1: The relationship between government planners, residents and investors in residential land planning (WTI = Willingness to invest, WTA = Willingness to approve, WTL = Willingness to live).

Campbell and Cocco (2007) point out that estates are risky assets with volatile prices and that it is therefore important for residents to assess carefully the estate's location with regard to accessibility (via transportations), its surrounding as well as its relation to the residents' disposable income, and this, in turn, influences the strategy of the investor. Thus, on the estate market, the spatial distribution of market value and the built-up cost of estates are primary factors in influencing the decision of investors and residents. Given that residential development lies in a complex socio-ecological system (Cook, et al., 2012), the market value of estates is found to be dependent on the location (Alonso, 1964; Gelfand et al., 2004; Kiel and Zabel, 2008). As Kuethe (2012) states, "real

estate is tied to a particular location, and as a result, many of its characteristics are shaped by the surrounding physical environment. This notion has motivated a substantial body of research that addresses the price impact of various land use activities on neighboring properties.” For instance, Luttik (2000) found that natural elements, such as trees or water, increase property values. Moreover, transportation opportunities are also found influential in evaluating potential residences (Weisbrod, 1980; Srour, 2002; Chatman, 2009; Tillema, 2010).

Yiu and Tam (2004) conducted a review of empirical studies on property price gradients and identified two estimation methodologies and three assumptions on the spatial structure of an urban area that predominated. The two methods are: (1) the hedonic pricing model and (2) the repeat-sales model. In addition, the three assumptions on the spatial structure of an urban area are: (1) the monocentric assumption; (2) the non-monocentric/polycentric assumption; and (3) no *a priori* assumption about the urban spatial structure. Common to the literature, reviewed by Yiu and Tam (2004), is that the decision-making by one actor is unrelated to other actors’ decision-making, where the government as an actor is typically overlooked. For the government, sustainable development is a pillar in setting the macroplanning policy (Brugmann, 1996). This objective is usually evaluated by examining urban form strategies (Clark, 2001; Jabareen, 2006), or energy strategies (Dincer, 2000; Cooper et al., 2001; Lund 2007). In this strand of literature, it is almost taken for granted that the government and the market are incongruent in how residential development can achieve sustainability, livability and profit.

In another strand of the literature on residential development, however, a tie between the government and the other market actors is recognized. Bone et al. (2011) reviewed ‘top-down’ approaches like deterministic models, statistical models, and system dynamic models used for the purpose of identifying effective, governmental policies to achieve governments’ objective in residential development. There is also the ‘bottom-up’ approach, such as cellular automata (CA), but Zhang et al. (2010) explained that these models have limitations in properly simulating the land use while integrating with the complex social, economic and human system of urban dynamics. Alternatively, Ferrand (1996) regarded the multi-agent system (MAS) model as a set of agents interacting in a common environment, but modifying themselves according to their objectives. Further, Parker et al. (2003) suggested that MAS models of land-use are well suited for representing complex spatial interactions under heterogeneous and dynamic conditions. Furthermore, Li and Liu (2007) proposed integrating MAS and CA to simulate residential development under different planning scenarios.

However, residents and investors often overlook the macroscopic policy constraints exerted by the government in the land use decision-making process (Bone et al. 2011). Most studies regard the government as a macro actor (we use actor and agent interchangeably in this paper), with exclusive power to set the general land-use development, and these studies do not, explicitly, consider the microscopic choice behavior of residents and investors (Huang et al., 2014; Zhang et al., 2015). The conceptual model we develop in the next section is strongly rooted in the MAS microscopic modelling tradition of considering investors' investment strategy and residents' choice behavior, but it also explicitly integrates governmental objectives that have been downplayed in MAS modelling so far. Moreover, it has conventionally focused on a single hotspot and thereby neglected the fact that a city has multiple hotspots that meet the multiple needs of the residents, and that most of the transportation arises because of residents' travel between their residence and those hotspots (Zhao et al., 2016). The conceptual model developed allows for multiple hotspots.

3. A conceptual model of the interaction of government, estate investor, and resident and their objectives

As discussed in the previous section, Parker et al. (2003) suggested that multi agent (i.e. actor) system models of land-use are well-suited for representing complex spatial interactions under heterogeneous and dynamic conditions, while Li and Liu (2007) proposed integrating MAS and CA to simulate the residential development under different planning scenarios. Here, we purport that governmental residential planning plays a crucial role in shaping urban pattern, while the behavior of investors and residents plays an important role in promoting, or hindering, the fulfillment of the planning.

Governments are presumed to be careful in their planning to balance the environmental, political, personal and economic development, and integrate its plans with investors' investment strategy and residents' selection behavior of dwellings on a local level, bearing in mind the residents' desire to access all hotspots in the city. In the outline of the conceptual model we present the objective of each actor, one at the time.

3.1. The government's objective

In residential planning, governments are increasingly taking sustainable development as the primary objective to be balanced with personal, economic, political, cultural developments. Ultimately, governments are responsible for structuring residential locations to mitigate greenhouse gas (GHG), foremost CO₂ emissions. In what follows, we stipulate that the administrative area under

governmental responsibility can be portioned in H homogenous zones (or grids), indexed by h . Furthermore, we assume that the government is concerned about the negative externality generated by the CO₂ emission induced by intra-urban travel between the residential location h and the hotspots, and we denote by E , the CO₂ emission quantity. The government is also concerned with several other factors that we simply collect and denote by Q . In maximizing the societal utility U , the government makes decisions in maximizing the sub utility \tilde{U} , by developing a location h for residency based on mitigating the CO₂ emissions accordingly:

$$(1) \quad \max_{E,Q} U = f(E, WTL, WTI, Q) \propto \max_E \tilde{U} = \tilde{f}(E, WTL, WTI)$$

which implies that E and Q are independent factors, a necessary assumption. It should be noted that the objective of CO₂ emission mitigation could readily be replaced in the conceptual model by any other governmental objective, if so desired.

3.2 The resident's decision rule for choosing residential location

In estimating the willingness of a resident to live at h , land-use type, land price, surrounding environment, accessibility, general public facilities, and education opportunities have been identified as important factors (Li and Liu, 2007, as well as others reviewed in Section2). Note that the last three factors all can be combined and measured as a function of travel distance to the hotspots, implying that the WTL at a location h will decrease with the travel distance to the hotspots.

Thus, with a natural extension that encompasses hotspots, Li and Liu (2007), the resident's utility related to travel distance to hotspots from the location h is expressed as:

$$(2) \quad U_{travel}(h) = w_1 e^{-b_1 D_{road}} + w_2 e^{-b_2 D_{highway}} + w_3 e^{-b_3 D_{hotspots}}$$

where D_{road} and $D_{highway}$ are the Euclidian distances from h to the nearest road and highway, respectively. $D_{hotspots}$ is the weighted network travel distance from h to the hotspots, where the weights are determined by the frequency by which the hotspots are visited by the residents (see Zhao et al., 2016). b_1 , b_2 , and b_3 are decay coefficients and, w_1 , w_2 , and w_3 are the weights representing the resident's utility of being subject to the three types of distances.

Another factor identified in the literature review that a resident considers in its utility function and, therefore, in the choice of residential location is the surrounding environment:

$$(3) \quad U_{env}(h) = w_4 G(h) + w_5 W(h)$$

where $G(h)$ and $W(h)$ are the percentages of green land and water area at location h , where, once more, the weights express the resident's relative utility of green land and water area.

Constrained by income, price of the estate and the consumption of a composite of other goods are also factors influencing the preference of a residential location h , in order to maximize the utility:

$$(4) \quad \max_h U = w_P P(h) + w_Z Z + w_{env} U_{env}(h) + w_{travel} U_{travel}(h)$$

w_P , w_Z , w_{env} and w_{travel} are the weights that the resident assigns to each factor. $P(h)$ is the estate price in location h , while z denotes the consumption of a composite of other goods. Piecing eq. (2-4) together, the willingness to live at location h , rather than location h' of resident r can be formulated as (McFadden, 1978):

$$(5) \quad WTL_r(h) = \frac{\exp(U(r, h))}{\sum_{h' \in H} \exp(U(r, h'))}$$

3.3 The estate investor's decision on where to invest

Estate investors primarily consider the preferences of residents in home buying to set the investment strategies. The willingness of the investor to invest at location h is determined by the profit (π), where we abstract from uncertainty in the investment.

$$(6) \quad \max_{d, h} \pi = dP(h) - C(d)$$

where d is the development density, and C is the land and the construction cost in developing at location h . The willingness for investor i to carry out an investment at location h is:

$$(7) \quad WTI_i(h) = \frac{P(h)d}{C(d)}$$

where an investment takes place at h if $WTI_i(h) > \pi_t$ where π_t is the investor's required return on capital.

4. Residential development in Borlänge as a demonstrating case study¹

The municipality of Borlänge lies 220 km northwest of the capital Stockholm, and its size is about 637 km², with approximately 50,000 residents asymmetrically distributed in the area. According to Statistics Sweden², 93% of the residents in Borlänge live in either a single-dwelling house, or an apartment in a multi-dwelling building, which, in total, amounts to 15,768 residencies³. In what follows, it is realistically assumed that all the residencies are tradable on the estate market, although a few residencies may be subject to restrictions for trading. Further, for the analysis, we assume that all the land in Borlänge has the potential to be developed or redeveloped into residential areas, and apply a 1 km² grid to divide the municipality into 637 grids, where the centroid of each grid represents the location for possible development.

4.1 Variables needed to implement the methodology

The first variable that is needed to implement the methodology is the market value of an estate at location h . To estimate the market value of the estates, we use the transaction data of single-dwelling houses and multiple-dwelling buildings from January 2013 to December 2016, as reported on Hemnet⁴. There were 2209 transactions during that time period, and for the 347 estates that were transacted multiple times, we have used the average price. Figure 2(a) shows the distribution of all current single-dwelling houses and the multi-dwelling buildings in the municipality. Figure 2(b) illustrates the estates that have been transacted from 2013-2016.

Looking at Figures 2(a) and 2(b), simultaneously, the transacted estates are, as expected, distributed in a pattern similar to all the residential locations in Borlänge. To estimate the market value of the non-transacted residential locations, we have spatially interpolated the market value of the nearest transacted estate. In doing so, 90% of the non-transacted residential locations are within one kilometer of the nearest transacted estate. In some of the grids, there is no estate at all. For those grids, we have estimated the market value in the grid to be the market value at the adjoining grid with a market value.

¹ Data and the source code in *R* and ArcGIS are freely available upon the request to the first author of this paper.

² <http://www.scb.se/HE0111-en>, accessed at 2017-01-17.

³ Provide by Construction and Plan Office, Borlänge Municipality

⁴ The biggest website for estate transaction in Sweden, <http://www.hemnet.se/>, accessed on 2016-12-28.

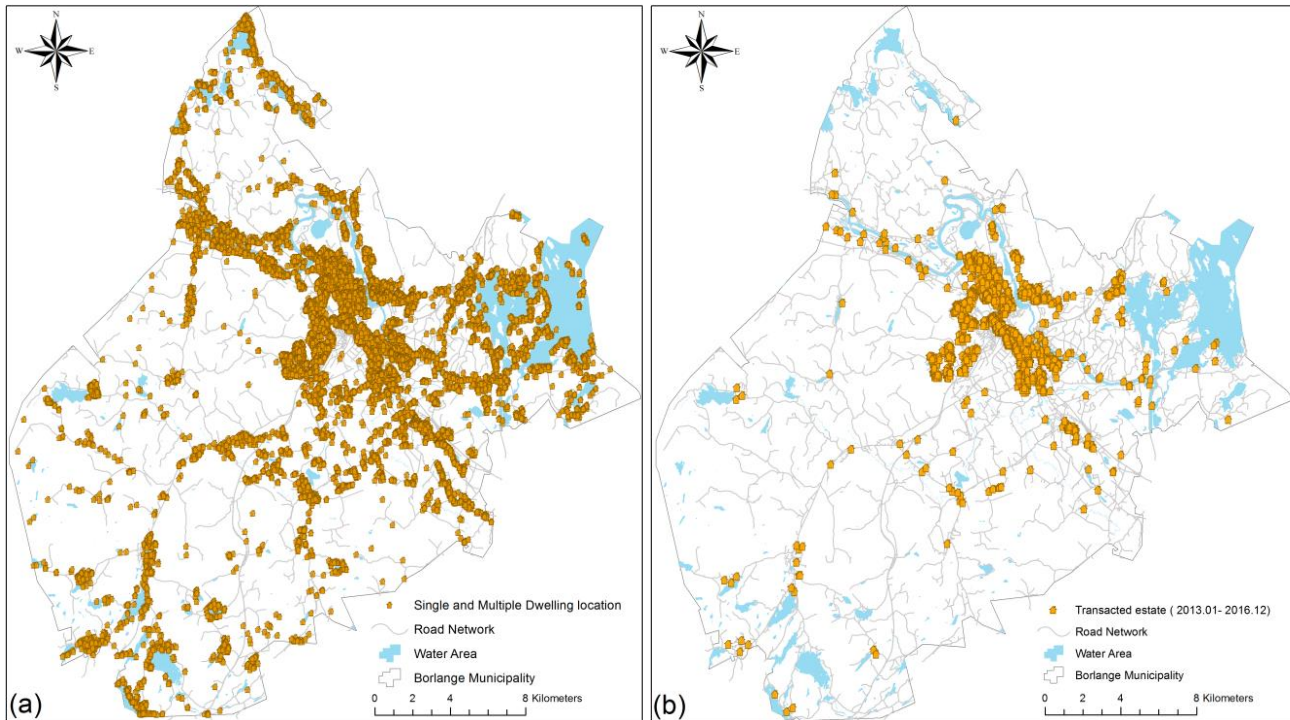


Figure 2: (a) the distribution of single-dwelling houses and multi-dwelling buildings in Borlänge
 (b) the distribution of transacted estates from January 2013 to December 2016.

As for the construction and land cost variable, we impose it to be uniform in Borlänge, based on the historical land selling price in the municipality (an approximation that may not apply elsewhere). However, the cost of construction may differ between investors, where we assume a variation in the cost between investors to be quantifiable by a random term $\varepsilon \sim N(0, \delta^2)$, as the market for housing construction is unlikely to be perfectly competitive. Operationally, we set the difference to amount to 5%, such that the construction and land cost C is on average 97.5% of the revenue with a standard deviation δ of 1.25% of C . Furthermore, we require at least one investor to be willing to invest at location h for the location to be of interest to investors.

The second variable that is needed to implement the methodology is the traffic flow on the road network. To estimate the decay coefficients b_1 , b_2 and b_3 , we follow Li and Liu (2007):

$$(8) \quad b_1/b_2 = f_{highway}/f_{road}$$

$$(9) \quad b_1/b_3 = (z_1 f_{road} + z_2 f_{highway})/f_{road}$$

where $f_{highway}$ and f_{road} are the average traffic densities (*number of vehicles per time unit by road length*). In Li and Liu (2007), z_1 and z_2 are assumed to be the number of roads and highways connected to the city center. Here, we take a more elaborated approach by acknowledging that travels

to multiple hotspots are required for fulfilling the needs of a resident. Zhao et al. (2016) found that 51 hotspots attracted more than 90% of the residents' movements in Borlänge. Therefore, to estimate z_1 and z_2 , we compute the weighted average number of roads and highways that connect to the 51 hotspots. From this estimation procedure, the decay coefficients of b_1 , b_2 and b_3 are set to 0.001, 0.0005 and 0.00025, respectively.

The third variable that is needed to implement the methodology is the percentage of green area and water area around location h . In each grid, the percentage of green land⁵ and water area⁶ was calculated. The fourth variable that is needed to implement the methodology is the CO₂ emissions induced by travelling from location h to the hotspots. We applied the network of induced emissions derived by Zhao et al. (2016), and calculated the CO₂ emission from each potential residence to each hotspot.

With the needed variables at hand, we can compute (emission) E , (willingness to live) WTL , and (willingness to invest) WTI for each location h , and check the utility for each of the three actors, compared to a threshold value of residential development at the location h . Zhao et al. (2016) found in Borlänge that the current, average CO₂ emission induced by the residents' trips by car to the hotspots was 1.94 kg. Residential locations that would induce CO₂ emission of less than 1.94 kg improve the governmental utility, and this value, therefore, serves as a natural threshold value for the local government in the city council.

The $WTL(h)$ is an index variable expressing the residents willingness (or unwillingness) to live at location h . We have set a location h to be attractive to residents if it exceeds the 3rd quartile of the distribution of WTL in Borlänge. Setting a low threshold value for this index would imply that almost all the locations in the municipality would be liked by the residents and, thereby, they would not exercise any influence on the regional development, which is unrealistic. Setting a high threshold value highlights the areas in Borlänge that are highly attractive to the residents, and the 3rd quartile is fairly high threshold value. As a threshold for the investors' willingness to invest, we use the return (0.63%) on buying a 10-year Swedish government bond.

Figure 3(a) shows the preference of the three actors for residential development in the 637 locations (grids) in Borlänge municipality. Locations that meet the governmental objective of lowering CO₂ emission, while at the same time appealing to the residents and the investors, are marked by the color green. White areas are those locations that are unattractive to all the three actors, or, in a few locations, only attractive to the investors and therefore unlikely to be developed.

⁵ Based on SLU Forest Map, Dept. of Forest Resource Management, Swedish University of Agricultural Sciences.

⁶ Provided by Construction and Plan Office, Borlänge Municipality

However, the locations marked with red color are attractive to both residents and investors, while undesirable for the government and, therefore, subject to tension between market forces and the government. Finally, yellow marked locations are ambiguous in that they appeal to investors and the government, but are not very appealing to the residents.

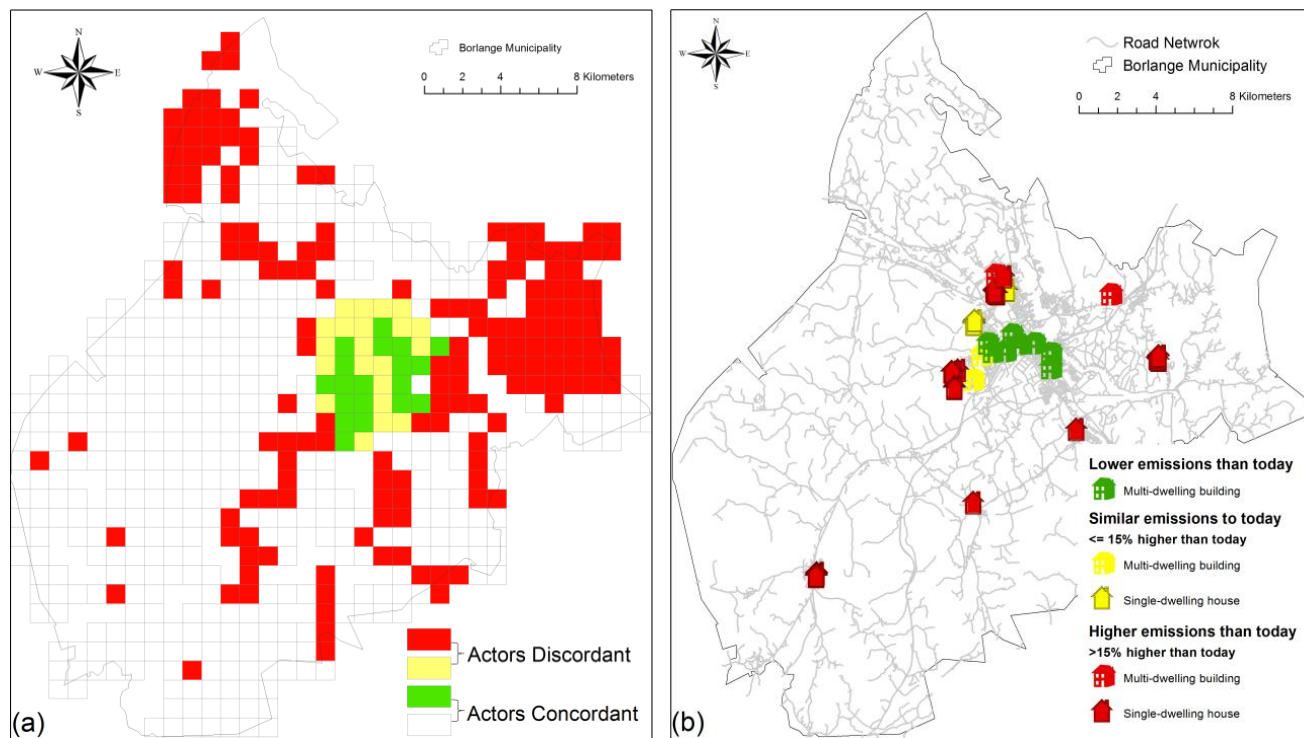


Figure 3: (a) The market's alignment with sustainable residential development in Borlänge

(b) Planned residential developments in Borlänge and their estimated induced CO₂ emissions.

Are market forces in Borlänge in alignment with the governmental objective of sustainable residential development, specifically with regard to CO₂ emissions? We are inclined to say yes. For most parts of the municipality, either the three actors have a common interest, or a common disinterest in residential development, with the exceptions being the north-western area and the east-most area of Borlänge, which appeal to residents and investors only. To get an insight into how the city council in Borlänge manages market forces jointly with the sustainability objective, Figure 3(b) depicts the presently existing 50 planned residential areas to be developed in Borlänge by the symbol of a house. A green, yellow and red colored house symbol indicates whether the induced CO₂ emissions would be below, on, or above the threshold value of 1.94 kg per trip, respectively. Green, planned residential developments amount to 95% of all future houses to reside in, meaning that the sustainability objective is being effectively fulfilled. However, as is also evident from Figure 3(b), the city council also compromises with the market forces, as there are some locations under development which are in conflict with the sustainability objective.

4.2 Sensitivity check

Apart from the variables needed to implement the methodology, some parameters also need to be set to make the methodology operational. Figure 3(a) was derived by applying a set of default values for these parameters. The proper choice of values is hard to nail down and to check the sensitivity of the results presented in the previous sub-section. We have re-done the analysis applying alternative settings, as listed in Table 1.

Table 1: Sensitivity checks of the default parameters settings in the implementation of the conceptual model.

Situation	Default	Alternative setting
1	$b_1 = 0.001, b_2 = 0.0005, b_3 = 0.00025$	$b_1 = 0.01, b_2 = 0.005, b_3 = 0.0025$ $b_1 = 0.0001, b_2 = 0.00005, b_3 = 0.000025$
2	$w_1 = w_2 = w_3 = 1/3$	$w_1 = 1/4, w_2 = 1/2, w_3 = 1/4$
3	$w_4 = w_5 = 1/2$	$w_4 = 1/4, w_5 = 3/4$
4	G is estimated based on total volume of all types of tree	G is estimated based on total volume of Birch, Deciduous, Oak and Beech
5	z is assumed identical for all residents	Z is differentiated in 50,000, 100,000, 300,000 in equal shares
6	Swedish government bond 10-year, return rate 0.63%	Swedish real total stock return rate 6.1%
7	$w_p = w_{env} = w_{travel} = 1/3$	$w_p = 1/9, w_{env} = 4/9, w_{travel} = 4/9$
8	C is assumed to be the cost of one house	C is assumed to be the cost of density d houses

In estimating the distance decay coefficients in equations (8-9), b_1 was normed to be 0.001, in line with Li and Liu (2007). We have also checked with b_1 set to 0.01 and 0.0001. Equal weights were set in equations (2-4) as default. As alternative settings, residents consider distance to highway more important than to normal roads and hotspots, value proximity to water more than green land, (as Kaplan and Kaplan (1989) have stressed: “Water is a highly prized element in the landscape”), and value surrounding environment and less travel more than house price in their choice of residential location. As a default setting, the residents have a common consumption of other goods, while we have checked the alternative that the residents are divided in three groups (low, middle and high incomers/consumers). For the investors, the required return on capital has also been set to the real total stock return, as well as letting the cost of construction decrease with d . The alteration of the parameter settings exercised little influence on the results shown in, and discussed, in relation to Figure 3(a).

5. Conclusion

In most urban planning studies, the formation of residential development has considered the objectives of social planners, investors and residents individually. Those studies also mainly focus on

the ex-post evaluation of residential development. This paper outlines a conceptual model where the three key actors – (local) government, estate investor, and resident – and their unique objective are integrated. The primary motivation for this contribution is to improve on governmental policies towards sustainable residential development.

The conceptual model developed draws heavily on previous modelling work in the literature that has had a more partial scope, and it is strongly rooted in the tradition of microscopic multi-agent system models. The conceptual model can be made operational to provide decision support for (local) governmental residential planning on a micro level.

As a demonstration of the conceptual model's ability to be operational, we have examined the case of the Borlänge municipality where residential planning is in a highly active state, as a consequence of substantial population growth due to the influx of immigrants. The application of the methodology in Borlänge provides several noteworthy insights. The first is that a city like Borlänge has multiple hotspots and that therefore the assumption of polycentric cities is the most plausible in urban planning of residential locations (Yiu and Tam, 2004). The second insight is that the market actors' behavior is, in general, aligned with the governmental objective of sustainable, residential development. However, on considering the present residential plans of Borlänge, it seems that the local government is aware of the market forces and compromises between its sustainability objective and the market actors' preference, in areas where the actors are not in alignment with the local government's objective.

The implementation of the methodology requires access to some key variables and the setting of a number of parameters. In relation to the demonstrating case study, we provide a detailed listing of these variables and indicate from where they may be retrieved, so that the conceptual model may be implemented elsewhere. In a detailed sensitivity check of the parameter settings for the case study, we found that the choice of settings has little to no impact on the model output, and therefore dares to claim that the implementation of the conceptual model is insensitive to the parameter settings, although this needs to be checked whenever the model is used in a specific case.

References

- Alonso, W. (1964). *Location and land use*. Harvard University Press, Cambridge.
- Bone, C., Dragicevic, S., & White, R. (2011). Modeling-in-the-middle: bridging the gap between agent-based modeling and multi-objective decision-making for land use change. *International Journal of Geographical Information Science*, 25(5), 717-737.

- Bourassa, S. C., Cantoni, E., & Hoesli, M. (2007). Spatial dependence, housing submarkets, and house price prediction. *The Journal of Real Estate Finance and Economics*, 35(2), 143-160.
- Brugmann, J. (1996). Planning for sustainability at the local government level. *Environmental Impact Assessment Review*, 16(4-6), 363-379.
- Campbell, J. Y., & Cocco, J. F. (2007). How do house prices affect consumption? Evidence from micro data. *Journal of monetary Economics*, 54(3), 591-621.
- Chatman, D. G. (2009). Residential choice, the built environment, and nonwork travel: evidence using new data and methods. *Environment and Planning A*, 41(5), 1072-1089.
- Clark, M. (2001). Domestic futures and sustainable residential development. *Futures*, 33(10), 817-836.
- Cook, E. M., Hall, S. J., & Larson, K. L. (2012). Residential landscapes as social-ecological systems: a synthesis of multi-scalar interactions between people and their home environment. *Urban Ecosystems*, 15(1), 19-52.
- Cooper, J., Ryley, T., & Smyth, A. (2001). Energy trade-offs and market responses in transport and residential land-use patterns: Promoting sustainable development policy. *Urban Studies*, 38(9), 1573-1588.
- Dincer, I. (2000). Renewable energy and sustainable development: a crucial review. *Renewable and Sustainable Energy Reviews*, 4(2), 157-175.
- Ferrand, N. (1996). Modeling and supporting multi-actor spatial planning using multi-agents systems. In *Proceedings of the Third NCGIA Conference on Integrating GIS and Environmental Modelling*, Santa Fe, New Mexico, USA.
- Gelfand, A. E., Ecker, M. D., Knight, J. R., & Sirmans, C. F. (2004). The dynamics of location in home price. *The journal of real estate finance and economics*, 29(2), 149-166.
- Godschalk, D. R. (2004). Land use planning challenges: Coping with conflicts in visions of sustainable development and livable communities. *Journal of the American Planning Association*, 70(1), 5-13.
- Haase, D., & Schwarz, N. (2009). Simulation models on human-nature interactions in urban landscapes: a review including spatial economics, system dynamics, cellular automata and agent-based approaches. *Living Reviews in Landscape Research*, 3(2), 1-45.

- Huang, Q., Parker, D. C., Filatova, T., & Sun, S. (2014). A review of urban residential choice models using agent-based modeling. *Environment and Planning B: Planning and Design*, 41(4), 661-689.
- Iacono, M., Levinson, D., & El-Geneidy, A. (2008). Models of transportation and land use change: a guide to the territory. *Journal of Planning Literature*, 22(4), 323-340.
- Jabareen, Y. R. (2006). Sustainable urban forms their typologies, models, and concepts. *Journal of planning education and research*, 26(1), 38-52.
- Kaplan, R., & Kaplan, S. (1989). *The experience of nature: A psychological perspective*. Cambridge University Press, Cambridge, UK.
- Kiel, K. A., & Zabel, J. E. (2008). Location, location, location: The 3L Approach to house price determination. *Journal of Housing Economics*, 17(2), 175-190.
- Kuethe, T. H. (2012). Spatial fragmentation and the value of residential housing. *Land Economics*, 88(1), 16-27.
- Li, X., & Liu, X. (2007). Defining agents' behaviors to simulate complex residential development using multicriteria evaluation. *Journal of Environmental Management*, 85(4), 1063-1075.
- Lund, H. (2007). Renewable energy strategies for sustainable development. *Energy*, 32(6), 912-919.
- Luttik, J. (2000). The value of trees, water and open space as reflected by house prices in the Netherlands. *Landscape and Urban Planning*, 48(3), 161-167.
- Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: a review. *Annals of the Association of American Geographers*, 93(2), 314-337.
- Srour, I., Kockelman, K., & Dunn, T. (2002). Accessibility indices: Connection to residential land prices and location choices. *Transportation Research Record: Journal of the Transportation Research Board*, (1805), 25-34.
- Sun, S., Parker, D. C., Huang, Q., Filatova, T., Robinson, D. T., Riolo, R. L., & Brown, D. G. (2014). Market impacts on land-use change: An agent-based experiment. *Annals of the Association of American Geographers*, 104(3), 460-484
- Tillema, T., van Wee, B., & Ettema, D. (2010). The influence of (toll-related) travel costs in residential location decisions of households: A stated choice approach. *Transportation Research Part A: Policy and Practice*, 44(10), 785-796.

- Waddell, P. (2002). UrbanSim: Modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association*, 68(3), 297-314.
- Weisbrod, G., Lerman, S. R., & Ben-Akiva, M. (1980). Tradeoffs in residential location decisions: Transportation versus other factors. *Transportation Policy and Decision-Making*, 1(1), 13-26.
- Yiu, C., & Tam, C. (2004). A review of recent empirical studies on property price gradients. *Journal of Real Estate Literature*, 12(3), 305-322.
- Zhang, H., Zeng, Y., Bian, L., & Yu, X. (2010). Modelling urban expansion using a multi agent-based model in the city of Changsha. *Journal of Geographical Sciences*, 20(4), 540-556.
- Zhang, H., Jin, X., Wang, L., Zhou, Y., & Shu, B. (2015). Multi-agent based modeling of spatiotemporal dynamical urban growth in developing countries: simulating future scenarios of Lianyungang city, China. *Stochastic environmental research and risk assessment*, 29(1), 63-78.
- Zhao, X., Carling, K., & Håkansson, J. (2016). *Residential planning, driver mobility and CO₂ emissions*. Working papers in transport, tourism, information technology and microdata analysis, (2016:5), Dalarna University, ISSN: 1650-5581.



Dalarna Doctoral Dissertations No. 6
ISBN: 978-91-85941-84-1
Microdata Analysis
School of Technology and Business Studies
Dalarna University
SE – 79 188 Falun Sweden
www.du.se