Individual accessibility and segregation on activity spaces: an agent-based modelling approach

Individual Accessibility and Segregation on Activity Spaces

an agent-based modelling approach

Marcus Vinicius Pereira Saraiva

A thesis submitted for the degree of Doctor of Philosophy

Department of Geography
Birkbeck, University of London
London, April 2020

Statement of Originality

I, Marcus Vinicius Pereira Saraiva confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

One of the main challenges of cities is the increasing social inequality imposed by the way population groups, jobs, amenities and services, as well as the transportation infrastructure, are distributed across urban space. In this thesis, the concepts of accessibility and segregation are used to study these inequalities. They can be defined as the interaction of individuals with urban opportunities and with individuals from other population groups, respectively. Interactions are made possible by people's activities and movement within a city, which characterise accessibility and segregation as inherently dynamic and individual-based concepts. Nevertheless, they are largely studied from a static and place-based perspective. This thesis proposes an analytical and exploratory framework for studying individual-based accessibility and segregation in cities using individuals' travel trajectories in space and time.

An agent-based simulation model was developed to generate individual trajectories dynamically, employing standard datasets such as census and OD matrices and allowing for multiple perspectives of analysis by grouping individuals based on their attributes. The model's ability to simulate people's trajectories realistically was validated through systematic sensitivity tests and statistical comparison with real-world trajectories from Rio de Janeiro, Brazil, and travel times from London, UK. The approach was applied to two exploratory studies: São Paulo, Brazil, and London, UK. The first revealed inequalities in accessibility by income, education and gender and also unveiled within-group differences beyond place-based patterns. The latter explored ethnic segregation, unveiling patterns of potential interaction among ethnic groups in the urban space beyond their residential and workplace locations. Those studies demonstrated how inequality in accessibility and segregation can be studied both at large metropolitan scales and at fine level of detail, using standard datasets, with modest computational requirements and ease of operationalisation. The proposed approach opens up avenues for the study of complex dynamics of interaction of urban populations in a variety of urban contexts

Acknowledgements

First of all, I wish to thank my supervisor Joana Barros for her support, advice and encouragement over the last five years. Her mentoring efforts went beyond PhD supervision, for which I am deeply grateful. I would also like to thank my friend and former supervisor Maurício Polidori, who set me on this path since my early undergrad years.

I am very grateful to my upgrade panel members, Dr Flávia Feitosa and Dr Paul Elsner, who gave me invaluable feedback and helped this thesis take the shape it currently has.

I am very thankful for being part of the incredibly diverse research environment of Birkbeck's Department of Geography. My thanks to everyone I have come across along these years in the department. Meeting you all has broadened my horizons beyond what I could have ever hoped for. Special thanks to Gemma Miller, Harriet Smith, and Hurmine Dormer-Dobson, for their unfailing support and assistance.

I would like to thank Vinicius Netto and his colleagues from UFF, who generously provided the data for the city of Rio de Janeiro. I also wish to thank the RESOLUTION research team, who welcomed me in their exciting and enriching research discussions, and also provided invaluable data for this research. Thanks to the entire open source community who provided all the software and open data that made this research possible.

I would like to acknowledge the financial support provided by the CAPES Foundation, Ministry of Education of Brazil, in the form of the PhD scholarship that has allowed me the opportunity to complete this research.

I am very grateful for the incredible friends I have made here: Cobus, Ash, Marit, David, Michael, Piero, and Lynsey. Special thanks to my Brazilian friends Sandro, Agnes, Fabiano, and Josi, whose companionship made us feel at home in London. I also thank my dear friends Rogério and Vanessa, who welcomed us back in Brazil so warmly. Thanks to the "Pelotinhas" group: Aline, Beta, Gabi, Gabriel, Gui, Jerê, Karol, Manô, and Nat, for welcoming us in the best reunion I could have hoped for. Very special thanks to Kellen, Matheus, and Zeni, who gave our furry friend Tinho lots of love and attention during our absence.

And finally, I would like to express my gratitude to my family. My brothers, José and Cristiano, and my sister Mariana, who inspire me every day with their focus and perseverance. I am forever grateful for the two sets of parents life has given me: my mothers Vera and Manuela (Ina), and my fathers José Alberto and Ramão (Titita). My final thanks goes to my loving wife Eliana, my partner in all moments, for accompanying me in this journey.

I dedicate this thesis to the memory of Ina, for all her care and love.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cities are arguably "the largest and most complex artefacts created by human activity" (Zamenopoulos and Alexiou 2012, 328), home to a large number of individuals who interact with each other and take advantage of the activities, amenities and services cities have to offer. As such, cities potentialise social and commercial exchanges, and accelerate the development and spreading of new ideas and technologies. The inherent complexity of cities, combined to the continuous growth of the urban population worldwide, brings many opportunities as well as challenges to urban planning, design, and management.

One of those challenges is the increasing social inequality imposed by the way population groups, jobs, amenities and services, as well as the transportation infrastructure, are distributed across urban space. It is common that citizens who belong to different groups, be it due to their wealth, social status, race, or ethnicity, live and spend time in areas spatially separated from each other. As a consequence, people have unequal access to urban opportunities, and have fewer chances to interact with individuals from different groups. Segregated minority groups are often particularly affected by reduced access to opportunities, which reinforces the effects of spatial segregation (Reardon 2006). In this context, accessibility and segregation studies have a central role in understanding and proposing solutions to those problems, which have a significant impact on the social and economic well-being of the urban population.

Although seldom studied together, accessibility and segregation are intrinsically linked subjects. Both concepts can be understood, in a broader sense, as measures of potential interaction. Hansen (1959, 73) defines accessibility as the "potential of opportunities for interaction", referring to interactions between people and urban opportunities and activities. Segregation, as defined by Freeman (1978, 413), is characterised by "any restrictions on interactions" among people perceived as belonging to different social groups. Despite their empirical and theoretical connections, both issues are usually studied as separate subjects, each with its own specific theoretical and methodological frameworks.

Yet, both subjects are studied geographically either from a *place-based* or an *individual-based* perspective. The place-based perspective is more popular and considers accessibility and segregation as properties of spatial units rather than of individuals. This is a useful perspective in studying both problems, as it informs real-world decisions such as policy-making and public investment allocation destined to improve accessibility and reduce segregation. However, the place-based approach has been criticised for being theoretically incomplete, as it does not account for individual and dynamic aspects of accessibility and segregation (Miller 2007; Kwan 2013). In other words, in the place-based approach, residents of each spatial unit are considered to have the same level of accessibility and experience the same level of segregation as all other residents of the same spatial units, so individual differences are not accounted for.

The individual-based approach aims to overcome the limitations of the place-based approach by acknowledging that different individuals, even when living in the same location and belonging to the same social group, can have significantly different experiences in terms of the opportunities they have access to and people they can interact with. To achieve a finer-grained level of representation, individual-based studies of accessibility and segregation are usually founded in Hagerstrand's (1970) time geographic theoretical framework, which is based on individuals' trajectories in space and time, rather than on artificially defined spatial units. As such, individual-based methods allow for the identification of intra-group and within-place differences in accessibility and segregation that are not possible with place-based methods.

The individual-based perspective is much less explored in the literature than the place-based perspective. The need for more individual-base studies on accessibility has been identified and discussed in the literature since the 1970s (Pirie 1979; Lenntorp 1976; Kwan 1998; Geurs and van Wee 2004). Researchers have called for more individual-based studies on segregation as well, although the discussion on this topic is more recent (Kwan 2013; Farber et al. 2015; Yip, Forrest, and Xian 2016).

The prevalence of place-based over individual-based studies of accessibility and segregation can be attributed to two main factors: the restricted availability of individual activity and trajectory data required for individual-based studies, and the difficulty of representing individual's spaces of activity in a computationally efficient way. These two factors make developing individual-based methods of studying accessibility and segregation challenging. As a result, a limited number of empirical studies using the individual-based approach have been conducted, and those are usually applied to small study areas or use data obtained from small samples of individuals.

The main aim of this thesis is *to develop an analytical and exploratory framework for studying accessibility and segregation for entire cities and large* *populations following an individual-based approach.* In order to overcome the challenges of data availability and computational complexity of existing methodologies, this thesis proposes an innovative method based on urban modelling techniques. An agent-based model was developed to artificially generate individual trajectories in space and time dynamically, employing standard and readily available datasets such as travel surveys and census origin-destination matrices. Agent-based modelling is a suitable approach for this kind of study due to its ability to simulate the behaviour of individual agents, as well as their interactions with each other and with the environment. Agent-based models provide a dynamic simulation environment which can incorporate individual mobility and the temporal dimension in accessibility and segregation studies, which are usually aggregated and static. Hence, this study aims to contribute to the body of literature by building methods that allow more theoretically accurate, scalable, and reproducible individual-based studies on accessibility and segregation.

1.1 Thesis Outline

This thesis is divided into two parts. The first part is composed by chapters 2 and 3, and presents a literature review on the topics of accessibility, segregation, and urban modelling. The second part (chapters 4 to 8), discusses the model development, its verification and validation, as well as two empirical applications. A brief summary of each chapter is provided in this section.

Chapter 2 introduces the theoretical foundation of the thesis. The chapter is divided into three sections. In the first section, the concepts of accessibility and segregation are discussed, and an overview of how those concepts are approached from the place-based perspective is presented. The second section discusses accessibility and segregation from the individual-based perspective, introducing the theoretical framework of time geography and activity spaces. The third section discusses the methodological challenges of individual-based studies of accessibility and segregation.

Chapter 3 presents an overview on the methodological foundation of this thesis. The chapter discusses the agent-based modelling technique used in this study, highlighting its advantages and limitations. The chapter briefly reviews agent-based models that incorporate accessibility, segregation, and individual movement.

Chapter 4 presents the proposed agent-based model and discusses its development. This chapter also introduces individual-based measures of accessibility and segregation proposed in this thesis.

Chapter 5 presents the model's verification and sensitivity analysis tests, carried out in a variety of abstract scenarios. The chapter presents the results of tests and discusses the effects of the model's parameters and initial conditions on agents' movement, as well as on accessibility and segregation metrics outputs.

Chapter 6 presents the model's validation exercises carried out in two real-world cities. The first exercise compares the agents' artificially generated trajectories to real world trajectories from Rio de Janeiro, Brazil. The second scenario compares agents' travel times to real-world travel times from London, UK. This chapter concludes with a discussion on the accuracy of the model's results, as well as its limitations.

Chapter 7 presents two simulation exercises that demonstrate how the model can be effectively used to explore accessibility and segregation issues in empirical applications. The chapter is divided into two sections, one dedicated to each empirical study. The first discusses inequalities in accessibility in São Paulo, Brazil, by income, education, and gender. The second focuses on ethnic segregation in London, UK, exploring patterns of potential interaction among ethnic groups based on individual movement across the urban space.

Chapter 8 concludes the thesis with a discussion of the contributions and limitations of the study. The chapter also discusses potential avenues for the further development of the methodological framework proposed in this thesis.

Chapter 2

Accessibility and Segregation

Accessibility and segregation have been the object of a large number of research studies and a main theme in urban planning practice since early 20^{th} century. This chapter presents a review of those studies, with two main objectives. The first objective is to introduce the concepts of accessibility and segregation, and discuss the main differences between the place-based and individual-based perspectives from which those subjects have been approached in the literature. The second objective is to detail the individual-based approach to accessibility and segregation studies, highlighting the theoretical advantages of this approach over the place-based approach, as well as its methodological challenges.

This chapter is organised in three sections. The first section introduces the concepts of accessibility and segregation, and how they have been traditionally studied from a place-based perspective. The section also includes a discussion of the limitations of the place-based approach. The second section focuses on individual-based studies of accessibility and segregation. This section discusses the concept of activity space and the time geography theoretical framework, which are the theoretical and methodological foundations of individual-based studies of accessibility and segregation, including this thesis. The third section discusses the implementation of individual-based measures of accessibility and segregation following time geographic principles, as well as the challenges of implementing such measures.

2.1 Background

This section introduces the definitions of accessibility and segregation used in this thesis. The section also presents an overview of place-based studies and metrics of accessibility and segregation, which form the main stream of studies on both subjects. The section concludes with a discussion on the advantages and limitations of the place-based approach.

Accessibility

Handy and Niemeier (1997) argue accessibility is the main reason people live in metropolitan areas: to enjoy the large number and variety of opportunities to engage in social and economic activities. Due to its importance, accessibility has been a central theme in transport and regional planning and research for many decades (Hansen 1959; Ingram 1971; Batty 2009a). However, despite this long tradition, there seems to be a consensus in the literature that the term accessibility can be easily misunderstood (Ingram 1971; Pirie 1979; Handy and Niemeier 1997; Geurs and van Wee 2004; El-Geneidy and Levinson 2006).

Accessibility is often mistaken for similar concepts such as mobility (El-Geneidy and Levinson 2011), for example. However, while mobility refers only to people's ability to move between places, accessibility refers to people's ability to reach places where they can participate on valuable activities (El-Geneidy and Levinson 2011). Hence, mobility is just one of two main components of accessibility most frequently referred to in the literature (Hansen 1959; Ingram 1971; El-Geneidy and Levinson 2011), which are *transportation* and *land-use*. The transportation component is determined by mobility-related factors such as travel distances, travel times, and costs associated to moving to a desired destination. The *land-use* component is characterised by the spatial distribution, quality, and variety of opportunities, activities and amenities in a city or region. In this sense, accessibility can be seen as a trade-off between the cost of going to some place and the benefits obtained by participating on the activities available at that place (Handy and Niemeier 1997; Batty 2009a).

Based on the interaction of land-use and transportation systems in society, El-Geneidy and Levinson (2006, 11) suggest that "[a]ccessibility indicates the collective performance of land-use and transportation systems and determines how well that complex system serves its residents". However, the land-use and transportation components alone do not adequately represent the complexity of the accessibility concept. Handy and Niemeier (1997) argue different people may evaluate their level of accessibility differently, even when located at the same place. That difference may stem from a number of factors, such as access to different means of transport, personal responsibilities which may impact an individual's time constraints, as well as personal preferences and necessities.

This view on the importance of individual factors that influence accessibility is shared by a number of authors (Lenntorp 1976; Pirie 1979; Miller 1991; Kwan 1998; Geurs and van Wee 2004). A more complete description of accessibility, provided by Geurs and van Wee (2004), includes four components: the aforementioned *land-use* and *transportation* components, plus *individual* and *temporal* components. This classification considers temporal constraints, such as opening times of places of activities and services, as well as individual preferences, possibilities, and needs that may affect an individual's access to such activities.

The conceptual differences discussed so far highlight the two perspectives from which accessibility is studied and measured: the *place-based* and the *individual-based* perspectives. In a broader sense, place-based studies focus on the land-use and transportation components of accessibility, while individual-based studies also take the individual and temporal components into consideration.

Segregation

In geographical and urban studies, segregation has been approached mainly from the residential perspective. This particular view refers to segregation as the spatial separation between different population groups in an area, where each group lives in distinct and often homogeneous neighbourhoods (Reardon 2006). Although different groups living in distinct neighbourhoods may not be a problem per se, this pattern usually stems from complex discriminatory processes against minority groups (Park and Kwan 2017). As argued by Young (2002), segregation is often associated with social issues such as minority groups' poor access to jobs, adequate housing, and public services, as well as lack of interaction and communication between different social groups which can help to perpetuate those problems.

Although the concept of segregation as different population groups living spatially separated from each other seems straightforward, it is seen as a complex and multidimensional phenomenon (Massey and Denton 1988; Reardon and O'Sullivan 2004). Reardon and O'Sullivan (2004) defined two spatial dimensions of segregation[1]: the *exposure/isolation* dimension, and the *evenness/clustering* dimension. Those dimensions are organised along two axes, shown in figure 2.1.

The *exposure/isolation* dimension refers to the probability of encounter between individuals of different groups (exposure) or the same group (isolation). The *evenness/clustering* dimension refers to the differential distribution of the members of population groups in the study area, who may be tightly grouped into clusters or evenly distributed in space. It may be argued those spatial dimensions of segregation represent different views on the phenomenon, similarly to the accessibility measures discussed previously.

It is also important to note the dimensions of segregation were conceived with residential segregation in mind, which is the dominant stream of studies in this subject. Although it can be argued they fit more naturally in place-based studies of segregation, individual-based studies of segregation have also relied on

1. Those dimensions are based on the five dimensions of segregations originally proposed by Massey and Denton (1988): *evenness*, *exposure*, *concentration*, *centralisation*, and *clustering*.

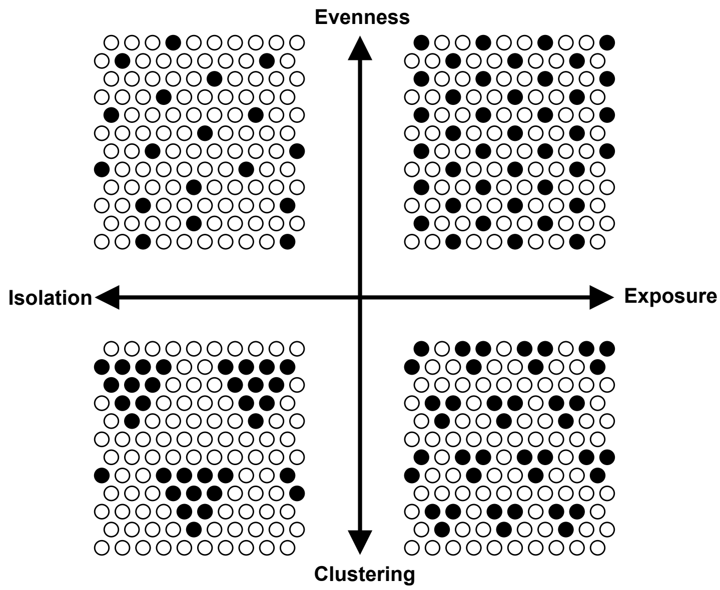

Figure 2.1: Dimensions of spatial segregation. Source: Reardon and O'Sullivan (2004, 126)

those dimensions, as discussed later in this chapter.

Place-based versus individual-based perspectives on accessibility and segregation

As previously discussed, accessibility is defined by Hansen (1959) as the potential for interaction between people and urban opportunities and activities, while Freeman (1978) defines segregation as the existence of restrictions on interactions between people from different groups. Those broader definitions of accessibility and segregation characterise both as inherently individual-based concepts. Also relevant for this thesis' argument is that such interactions presuppose individual mobility. Indeed, accessibility directly regards people's movement in order to reach desired activity locations. Interactions among different population groups are also dependent on places people visit while carrying out their daily activities, and also on how people move between such places. Furthermore, the temporal dimension is as important as the spatial dimension in accessibility and segregation studies, as interactions are only possible when people share the same location at the same time.

However, those individual and dynamic characteristics of accessibility and segregation are seldom recognised in studies on both subjects. Specifically, most quantitative studies rely on aggregate and static measures of accessibility and segregation, thus focusing on a partial view on the issues. In what follows, a brief overview of the most usual place-based (aggregate and static) measures of accessibility and segregation is presented, followed by a discussion on the main limitations of this approach[2].

Place-based measures of accessibility

Many different accessibility measures were developed with different aims and applications, but predominantly following the place-based approach. The measures of accessibility most widely used in research and planning applications are the ones in the group of place-based measures, referred to as *location-based* by Geurs and van Wee (2004), which include *distance*, *gravity*, and *cumulative-opportunities*.

Distance measures are the simplest class of accessibility measures. Those are based on the idea of "nearness" (Batty 2013), or the property of a place to be within easy reach of other places. *Gravity* measures consider, in addition to distance, the level of opportunities at each urban zone as the attractiveness of that zone. The level of opportunities at each place can be measured in different ways, according to the objectives of the study. Common examples are levels of employment, commercial activity, and number of urban facilities such as schools and hospitals. *Cumulative-opportunities* measures define accessibility as the number of opportunities that can be reached from an origin point within a specified distance or travel time (Geurs and van Wee 2004). These measures usually take forms like *the number of jobs within 30 minutes travel time*, or *the number of schools within 15 minutes walk*.

Cumulative opportunities measures are widely used in urban planning because they represent accessibility in meaningful units, which are easy to understand and compare. Owen (2014) argues cumulative-opportunities measures interpret accessibility as the *value* of living somewhere, value being derived from the opportunities made available by living at that location. This line of reasoning is opposite to the distance and gravity measures, which can be interpreted as a measure of the *cost* of living somewhere, or the impedance that needs to be overcome in order to reach desired opportunities from a particular place.

Place-based measures of segregation

Due to the multidimensional nature of segregation, it is commonly agreed that no single measure can portray the level of segregation in a city or region (Massey and Denton 1988). Hence, in empirical studies, a combination of indices is usually

2. Readers interested in more comprehensive literature reviews on accessibility are referred to Geurs and van Wee (2004) and El-Geneidy and Levinson (2006), while literature reviews on segregation were carried out by Reardon and Firebaugh (2002) and Reardon and O'Sullivan (2004)

applied to measure the evenness/clustering and exposure/isolation dimensions separately.

Among the most popular measures of evenness are the *dissimilarity index (D)* (Duncan and Duncan 1955) and the entropy-based *information theory index (H)* (Theil and Finizza 1971), even though a series of other indices were developed over the decades (Reardon and Firebaugh 2002). The *dissimilarity index (D)* measures the variation between the population composition in the local spatial unit (neighbourhood, census tract) and the composition of the whole study area. Its results can be interpreted as the proportion of the total population that would need to relocate to even out the population distribution in the study area (Reardon and O'Sullivan 2004). The *information theory index (H)* is a measure based on diversity, measuring the degree which local neighbourhoods are less diverse than the whole study area.

Systematic evaluations of segregation indices carried out by Reardon and Firebaugh (2002) and Reardon and O'Sullivan (2004) indicated that the *information theory index (H)* is a more robust and reliable metric of the evenness/clustering dimension of segregation in comparison to the *dissimilarity index (D)*. Further findings by Barros and Feitosa (2018) indicate the *information theory index* to be more informative at local scales than the *dissimilarity index*. While the *dissimilarity index* can only indicate if the population composition of any given local area is similar or different than the population composition of the entire study area, the *information theory index* also identifies the local areas with lower or higher diversity in comparison to that of the study area.

Measures of the exposure/isolation dimension of segregation were first conceived by Bell (1954) and reintroduced in the 1980s by Lieberson (1981), remaining popular ever since. The *exposure index (P)* measures the probability of members of one group meeting members of other group, while the *isolation index (Q)* measures the probability of encounter between members of the same group. Reardon and Firebaugh (2002) and Reardon and O'Sullivan (2004) also evaluated the *exposure (P)* and *isolation (Q)* indices and found them to be satisfactory measures of that dimension of segregation, thus justifying their popularity among the research community.

Limitations of place-based measures of accessibility and segregation

Place-based studies on accessibility and segregation, represented by the place-based metrics presented earlier among many others, are the main streams of decades of successful research on both subjects. However popular, the place-based perspective presents the aforementioned issues of ignoring individual differences in capabilities, constraints, and preferences, as well as their dynamic patterns of movement throughout the city. Specifically, three main issues can be identified with place-based measures of accessibility and segregation: a) the ecological fallacy, b) the Modifiable Areal Unit Problem (MAUP), and c) the Uncertain Geographic Context Problem (UGCoP).

The ecological fallacy problem occurs when characteristics of an individual are assumed to be the same as the individual's group or area where the individual is located. Neutens, Schwanen, and Witlox (2011) argue place-based measures of accessibility and segregation present this problem, as they assume all people who live in the same area have the same levels of accessibility and experience the same levels of segregation, which is not realistic. The Modifiable Areal Unit Problem (MAUP) also stems from the interpretation of results based on aggregated data and areal units. MAUP issues result from using areal units to aggregate individual observations, as the artificially defined shape and size of the areal units can affect the results of the analysis (Openshaw 1984).

The Uncertain Geographic Context Problem (UGCoP), as formulated by Kwan (2012), refers to the uncertainty regarding the geographic context of people's activities. Such context includes people's places of residence, work and recreation, as well as the open spaces used to travel between those places, which are rarely restricted to a single spatial unit of analysis. Considering only one of those spaces (which, in accessibility and segregation studies, is usually the place of residence) may lead to an incomplete view on people's experience, due to the significant amount of time people spend on everyday activities outside their home neighbourhoods.

All three aforementioned problems stem from the fact individuals are not restrained to their residential zones as defined by urban planning bodies and census offices, for which most geodemographic data are made available. People's spaces of activity are complex and hardly correspond to traditional spatial units of analysis. This is partly a result of the increased mobility of urban populations and the emergence of the so called mobilities paradigm in the social sciences (Urry 2007). This paradigm suggests static representation of cities in traditional studies may not be enough to capture human experiences of segregation and diversity, and that the "urban experience may be qualitatively different now compared to when many of the core concepts of segregation were initially developed" (Yip, Forrest, and Xian 2016, 156).

The complexities of contemporary life call for more dynamic representations of space and social interactions, which may be achieved by approaching accessibility and segregation issues from an individual-based perspective. The following section will discuss the theoretical foundations and empirical applications of individual-based approaches to accessibility and segregation found in the literature.

2.2 Individual-Based Approaches to Accessibility and Segregation

This section introduces the concept of activity space and the time geographic theoretical framework, which form the theoretical foundation of individual-based accessibility and segregation studies. This is followed by a discussion on how accessibility and segregation can be interpreted as time geographic concepts.

2.2.1 Activity Spaces and Time Geography

According to Golledge (1997), the concept of *activity space* originated and developed during the 1960s and 1970s by behavioural geographers in the USA, although Dijst (1999) identified a series of earlier studies referring to different terms with similar meaning dating back to the 1930s. Activity spaces are part of the broader concept of *action space*, which describes the subset of places "with which the individual has direct contact as the result of day-to-day activities" (Horton and Reynolds 1970). The concept of action space encompasses two main components: the *movement* component, which is called *activity space*; and the *communicating over space* component, which refers more specifically to communication channels such as telephone, newspaper, television and, more recently, internet. According to Golledge (1997, 279), "activity spaces represent direct contact between individuals and their social and physical environments".

Trajectories of individuals are an essential part of activity spaces. Those trajectories can be hierarchically classified according to the focal point around which they occur (Jakle, Brunn, and Roseman 1976). For instance, the main focal point of an individual is their place of residence, from where a number of trips of different lengths are made. Longer trips are usually those to other activity locations, which often act as secondary focus points around which other shorter trips occur, and so on (figure 2.2).

Dijst (1999) classifies action spaces in three groups: *actual action space*, the places an individual has actually visited; *potential action space*, the places within reach of an individual; and *perceptual action space*, the places known by an individual. Thus, the actual action space is a subset of the potential action space. The perceptual action space can hypothetically cover the whole potential action space, in a situation where an individual has complete knowledge of their environment. However, in practice, the perceptual action space also tends to be much smaller than the potential action space.

Despite the definitions discussed above, there is no specifically defined set of methods to identify individuals' activity spaces in the literature, and most studies rely on Hägerstrand's (1970) time geographic constructs for that purpose

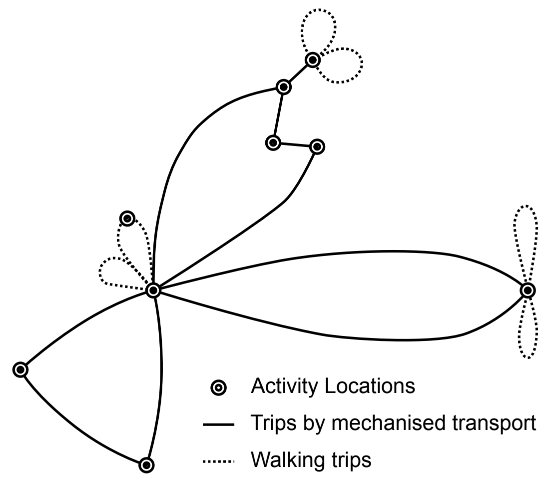

Figure 2.2: Individual activity space based on hierarchical movement. Source: (Jakle, Brunn, and Roseman 1976, 101)

(Patterson and Farber 2015). The time geography approach was developed during the 1960s at the Department of Geography of the Lund University in Sweden (Golledge 1997). This approach was made famous after Torsten Hägerstrand's (1970) presidential address at the ninth European Congress of the Regional Science Association, entitled "What about People in Regional Science?". Lund University geographers' objective was to add individuals' time and space constraints to a field of research which was, and still is, mainly focused on spatial and aggregate statistics. Hägerstrand argued there was a gap in research to be explored, between the domains of *biography* (or a detailed account of an individual's life) and *aggregate statistics* of whole populations, which time geography aims to fill.

In time geography, particular emphasis is given to space and time constraints, which limit the freedom of individuals. Space and time are both limited resources, and are inseparable from each other (Hägerstrand 1970), so people's choices and actions are bound by a set of constraints of varying levels of strength. Hägerstrand (1970) classifies those constraints into three groups: *capability* constraints, referring to people's transportation and communication capabilities, as well as basic needs (such as eating and sleeping); *coupling* constraints, referring to the need to join other individuals, and access tools and materials, in order to carry out some activity; and *authority* constraints, referring to rules, laws and customs that may allow or deny an individual access to certain spaces at certain times. Those key constraints, combined, account for most of the factors that stop people from participating in activities. For this reason, Golledge (1997) argues time geography shifts the point of view from what people do and why, to what prevents people from doing something or interacting with someone else.

The principal concept in time geography is the *space-time path*, which represents the trajectory of an individual (or any material object) in time and space during his/her/its whole existence. The path connects all movements and stationary activities of an individual through a series of control points and the links between them. Control points represent stops, turns, and changes in velocity in the trajectory. The path can be plotted in a three-dimensional graph, where the x and y axes represent space and the z axis represents time (figure 2.3a). Several characteristics of the path are condensed in that graph, such as movement and velocity (the slope of the line, the shallower the slope, the faster the movement), stationary activities and their durations (vertical lines, the longer the line, longer the activity duration), and the geographic area covered (the projection of the line on the xy plane).

Locations in space and time where many paths converge are called *bundles* (Hägerstrand 1970), and are necessary for carrying out most (if not all) activities. For example, participating on a work meeting requires the bundling of paths of several coworkers. Catching a bus requires the bundling of the paths of the passengers, driver and vehicle, as well as materials such as the fuel that powers the bus. Advances in communication technology can make the formation of bundles easier, for example when meetings are held remotely (Hägerstrand 1970; Kwan and Weber 2003).

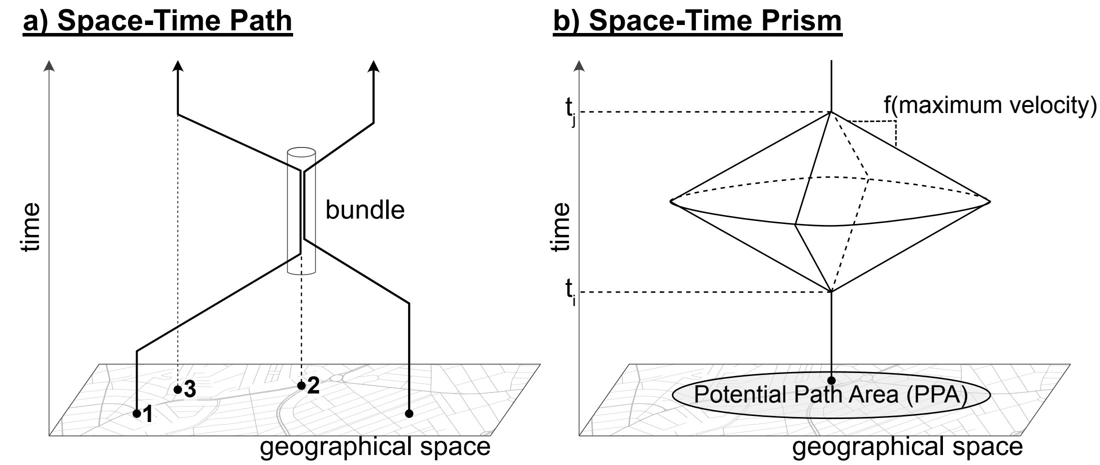

Figure 2.3: Space-Time Path and Space-Time Prism. After Miller (2005).

The concept of *space-time prism* expands the path, and represents the areas an individual can reach given space and time constraints (figure 2.3b). While the path represents areas actually visited by an individual, the prism represents the possible future trajectories "branching off" the main path in the future. The prism represents periods of time when the individual has freedom to choose activities to carry out, delimited by the ending time and location of one fixed activity (such as work or school) and the beginning of the next fixed activity. This free time is called the individual's *time budget*, which can be used at the individual's discretion (Neutens, Schwanen, and Witlox 2011). The prism encapsulates the individual's existence in space and time, as "[i]t is impossible for individuals appear outside the walls of the prism" (Hägerstrand 1970, page 14). The smaller the space-time constraints acting upon an individual, larger is that individual's space-time prism. For example, the higher the individual's travel velocity (due to access to fast transportation methods), or the more flexible the start and end times of regular activities, the larger the individual's prism will be.

Although the space-time prism fully encompasses time-geographical concepts, the *potential-path area* (PPA) is more frequently used in practical applications (Miller 1991, 1999; Patterson and Farber 2015). The PPA is defined as the projection of the prism in two-dimensional space (figure 2.3b), delimiting the spatial extent accessible by an individual (Golledge 1997), and is often associated to the aforementioned concept of potential activity space (Dijst 1999). Two-dimensional PPAs have methodological and operational advantages over three-dimensional space-time prisms, mainly in terms of computational and data requirements. Although the PPA is conceptually derived from the space-time prism, the former can be calculated without reference to the latter, thus reducing the computational burden (Miller 1991). A derived construct (and perhaps more useful in practice) is the Daily Potential Path Area (DPPA), which is the subset of the PPA that includes the activities and movements carried out by an individual during the course of one day (Weber and Kwan 2002).

All those time geographic concepts, together, form a cohesive theory of time and space, representing individuals's possibilities of interaction according to their personal constraints. Despite its advantages, the time geographic framework only gained popularity in the 1990s, mainly due to an increase in computing power available for research. This allowed the development of computational methods to deal with datasets with the required level of detail to derive space-time concepts, such as the pioneer attempts by Miller (1991) and Kwan (1998).

2.2.2 Accessibility and Segregation as Time-Geographic Concepts

As discussed above, time geography presents a comprehensive theoretical and methodological framework from which individuals' interactions with urban opportunities and with other individuals can be studied. According to that framework, *actual* interactions are represented by bundles of space-time paths, while *potential* interactions are represented by the intersections of space-time prisms in three-dimensional space. Hence, individual-based accessibility and segregation measures can be obtained from those intersections, as follows.

An individual's accessibility can be interpreted, in time geographic terms, as the intersection of the individual's space-time prism and that of the urban opportunities and activities in the city, as shown in figure 2.4. In that diagram, H and W represent the individual's places of residence and work, respectively. The individual has a *time budget* between times t_i, when they finish their responsibilities at home, and t_j, when they must be available for work. During that time budget, they need to commute to work but can also carry out any discretionary activities.

The urban area the individual can reach (the potential path area - PPA)

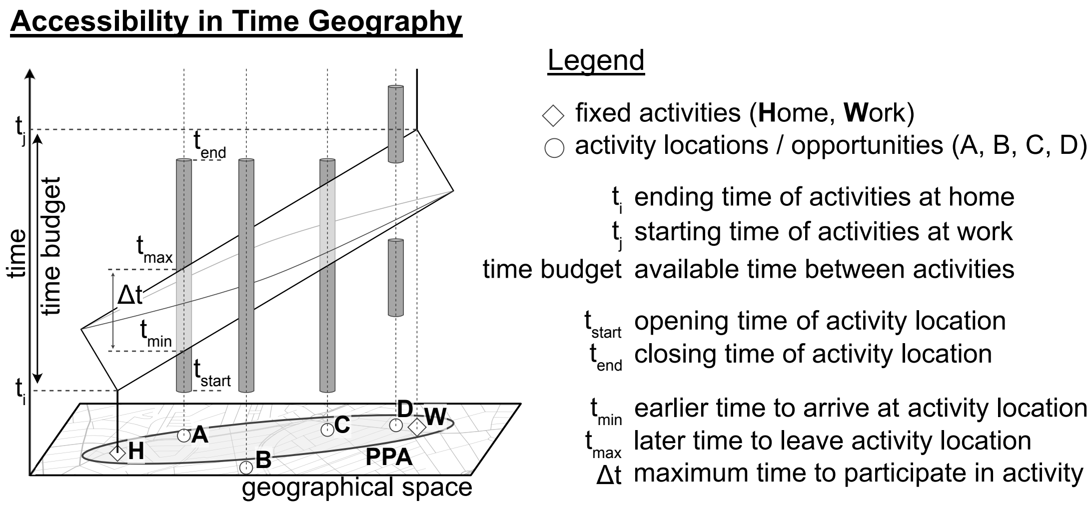

Figure 2.4: Accessibility as the intersection of space-time prisms.

is marked in the diagram as the projection of the individual's space-time prism in two-dimensional space. A sample of the opportunities available at the study area are marked as A, B, C, and D in the diagram. Those opportunities' space-time prisms are represented by intermittent vertical cylinders, since their locations are fixed in space. The interruptions in the cylinders represent closing hours of establishments. In that example, opportunities A and C are accessible to the hypothetical individual, as their space-time prisms intersect. The volume of the intersections represents the amount of time the individual can spend at those locations. Opportunities B and D are not accessible to the individual, for different reasons: opportunity B is outside the geographical area of the individual's PPA, while opportunity C's opening hours do not coincide with the individual's available time.

Individual-based measures of segregation tend to be based on the exposure of people from one population group to people from other groups (or from the same group, in the case of isolation). This measure fits well within the time-geographical paradigm, as exposure can be seen as the intersection between space-time prisms of different individuals. Figure 2.5a shows the space-time prisms of two individuals with their own places of residence, places of work, and time budgets. Following a similar approach to the previous example, the exposure of one individual to another is represented by the intersection between their prisms. This intersection merely represents the possibility of both individuals being co-present in space and time. According to Pred (1977, page 209), the intersection represents the "necessary (but not sufficient) conditions for virtually all forms of interaction [...] involving human beings". As previously discussed, actual interactions are represented by the bundle of space-time paths, as shown in figure 2.5b.

Accessibility and segregation, when viewed from the theoretical perspective discussed here, are clearly related concepts. Moreover, time geography allows for the same methodological framework to be used in studies of both issues, fur-

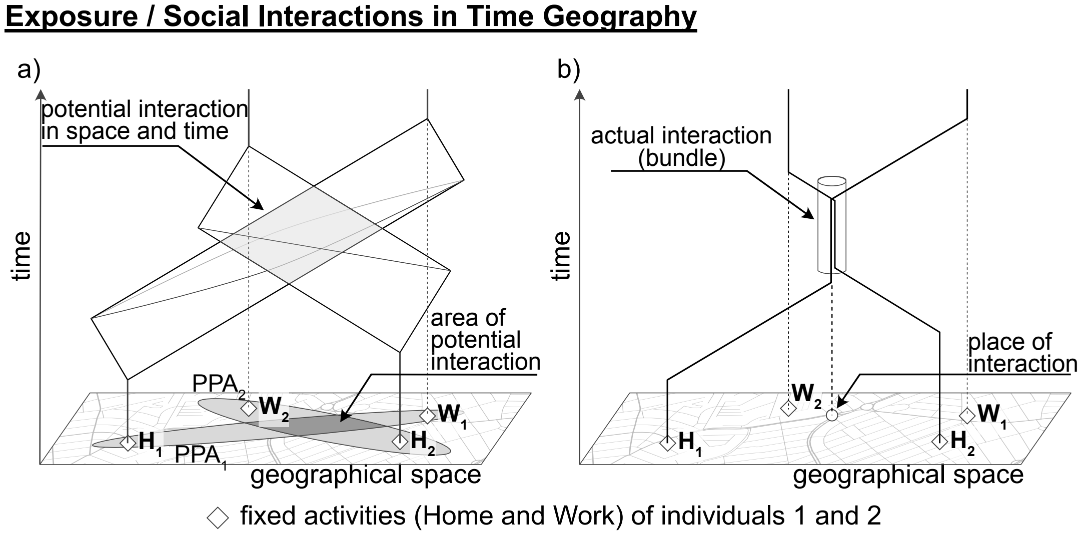

Figure 2.5: Exposure as the intersection of space-time prisms.

ther integrating these study areas. The remainder of this chapter will discuss how the theoretical representations of individual-based accessibility and segregation have been implemented in the literature, as well as the technical challenges faced by those studies.

2.3 Methodology and Challenges of Individual-Based Approaches

This section discusses the methods used in individual-based accessibility and segregation studies. The methodology of such studies can be summarised into three steps: 1) acquiring the necessary data, 2) delimiting activity spaces, and 3) measuring accessibility and segregation on activity spaces. Each step presents its own challenges, which are also discussed in this section. The following subsections detail those steps, starting from step 2 (delimiting activity spaces), because data acquisition is highly dependent on the choice of activity space representation made for each study.

2.3.1 Building Potential Path Areas and Activity Spaces

A series of techniques to define people's activity spaces can be found in the literature. Those range from complex three-dimensional prisms such as the traditional time geographic space-time prisms previously introduced, to simpler two-dimensional representations of potential path areas. Those techniques are discussed in this section, after the summary shown in table 2.1.

Table 2.1: Summary of methods for defining activity spaces.

Category	Method
Geometric - 3D	Space-Time Prism Network-Time Prism
Geometric - 2D	Standard Deviation Ellipse (SDE) Minimum Convex-hull Polygon (MCP) Standard Deviation Circle (SDC) Road Network Buffer Spatial Units (Census Tracts)
Continuous (Raster)	Kernel Density Estimator (KDE)
Discrete (Point-based)	Feasible Opportunity Set (FOS)
Discrete (Network-based)	Potential Path Tree (PPT) Potential Network Area (PNA)

Geometric - 3D Prism

Three-dimensional representations of activity spaces include the *space-time prism* and the *network-time prism*, which is a road network-based version of the former. The time geographic space-time prism was introduced in the previous section, and its visual representation was shown in figure 2.3b. Network-time prisms were developed to improve space-time prisms by considering the constraints of the road network on people's movement (Neutens et al. 2008). This advancement was made possible by new CAD and GIS technologies, which allowed the development of complex three-dimensional and network-based representations of space-time prisms, such as the ones proposed by Neutens et al. (2008), Kuijpers and Othman (2009), and Kuijpers et al. (2010). An example of network-time prism can be seen in figure 2.6.

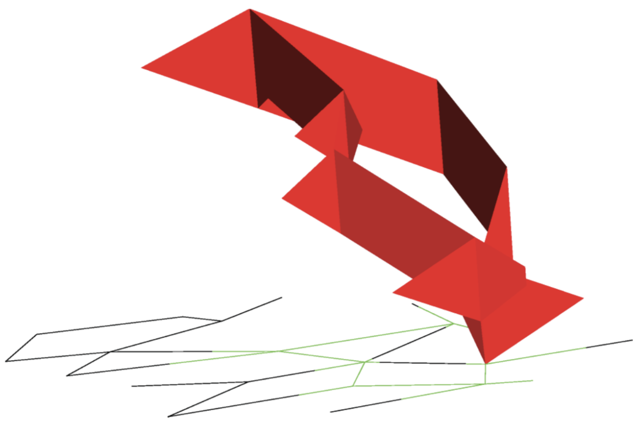

Figure 2.6: Three-dimensional representation of a Network-Time Prism. Source: Kuijpers et al. (2010, 1228)

Such three-dimensional and network-based techniques have a considerable computational cost, particularly when when large populations are considered. Furthermore, the geometric complexity of an individual's prism can increase very fast depending on factors such as the locations of trips' origins and destinations, number of necessary stops along the way, and varying travel velocities (Miller 1991). Network-time prisms have the additional burden of calculating a large number of shortest paths (Kuijpers et al. 2010). The inherently visual nature of those techniques is also a problem, as a large number of intersecting three-dimensional prisms would be very hard to visualise and interpret. Due to those limitations, none of the aforementioned studies attempted to calculate space-time prisms and network-time prisms for a large population.

Geometric - 2D Areas

Methods for building potential-path areas in two-dimensional space from individuals' activity locations can be based on circular, elliptical, and polygonal areas, as well as buffers along the road network and the simple adaptation of traditional spatial units such as census tracts. Those techniques are discussed below.

A simple elliptical PPA can be derived from a minimum of two activity locations and an amount of travel time, otherwise a circular PPA can be derived when only one activity location is defined (Lenntorp 1976; Stopher, Hartgen, and Li 1996). For higher numbers of known locations, standard deviation ellipses (SDE, shown in figure 2.7a) and 95% confidence ellipses (Schönfelder and Axhausen 2003) are popular alternatives, as they can capture the orientation and dispersion of a point pattern. Rai et al. (2007) tested other elliptical forms such as the Cassini oval, the bean curve and the superellipse, finding the best representation for the person's activity space depends on the person's specific travelling patterns.

The minimum convex-hull polygon (MCP), which is the minimum polygon that encompasses a collection of points (see figure 2.7b), is also a popular technique for representing individual PPAs. This method consists in finding the MCP of all the places visited by an individual (Buliung and Kanaroglou 2006; Kamruzzaman and Hine 2012). Some extensions to the MCP were proposed to account for differential travel distances and speeds along the road network, as well as accounting for reachable locations instead of only visited ones. For example, identifying all points reachable by an individual given some travel time threshold (Sherman et al. 2005), or using the person's longest distance travelled in a day as maximum reachable distance (Casas, Horner, and Weber 2009). Standard deviation circles (SDC, shown in figure 2.7c) can be an alternative to SDEs and MCPs, as they can be obtained with at least one place of activity recorded, unlike SDEs and MCPs which require at leat three locations (Kamruzzaman and Hine 2012).

An individual's PPA can also be defined by a buffer along the trajec-

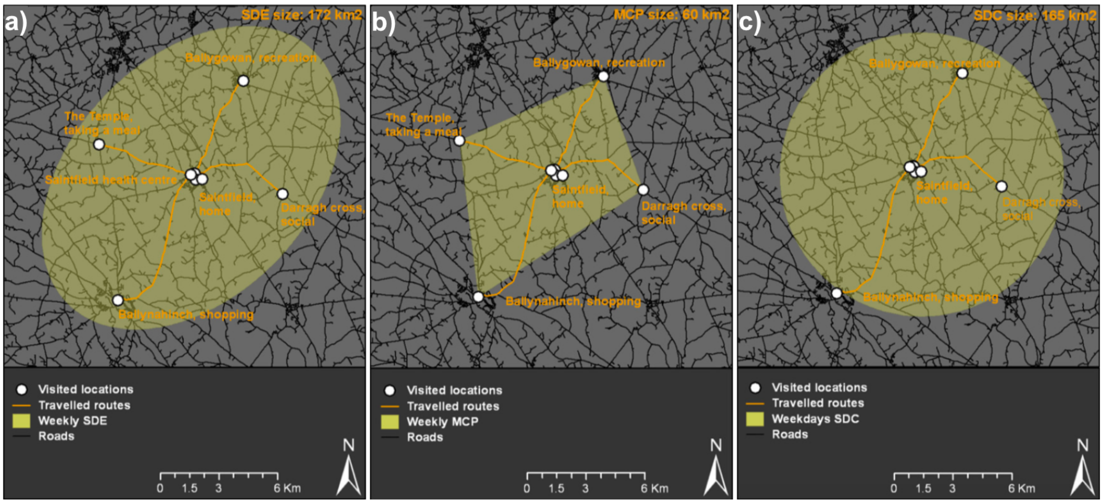

Figure 2.7: Two-dimensional representations of activity spaces: a) Standard Deviation Ellipse (SDE), b) Minimum Convex-hull Polygon (MCP), and c) Standard Deviation Circle (SDC). Source: Kamruzzaman and Hine (2012, 111)

tories connecting the individual's sequence of activity locations. For example, this technique was used in studies by O'Sullivan, Morrison, and Shearer (2000), Schönfelder and Axhausen (2003), Sherman et al. (2005), and Zenk et al. (2011). Sections of those buffers can be suppressed to account for places along the trajectory that the individual may not be able to access, or may not be familiar with, as in Chaix et al. (2012). For example, when commuting by high-speed or underground rail, only areas near stations are really accessible.

The last two-dimensional technique to be discussed here is the use of administrative boundaries or census tracts visited by an individual during their daily activities as a representation of their activity space (Wong and Shaw 2011). This is sometimes extended to the concept of intervening opportunities (Stouffer 1940), when not only the areas visited are counted as part of the activity space, but also the areas the individual has travelled through (Wong and Shaw 2011). One problem with this technique is that it can bring problems from aggregated measures of accessibility and segregation, such as the MAUP, back to individual-based approaches.

Continuous Surfaces - Raster

Continuous raster surfaces can also be found in the literature as representations of individual's activity spaces (Kwan 2000b; Schönfelder and Axhausen 2003; Kamruzzaman et al. 2011; Wang, Li, and Chai 2012). Those studies use Kernel Density Estimators (KDE) to interpolate continuous raster surfaces from point data, which usually represent the places visited by individuals. Those raster surfaces are used as indicators of individuals's activity density. Raster surfaces can be visualised in 3D, as shown in figure 2.8.

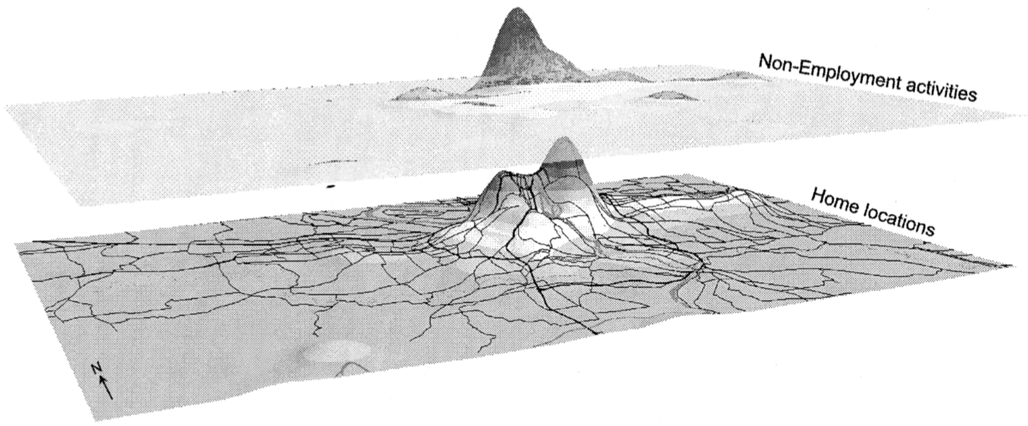

Figure 2.8: Raster surface representing the activity density pattern of a group of individuals. Source: Kwan (2000b, 193)

Discrete / Point-based

Discrete space PPAs, also known as punctiform or point-based, can be used to represent places people can actually reach and interact with. Discrete space PPAs are considered more realistic representations of people's activity spaces than two-dimensional and continuous spaces PPAs, because the former include: a) empty or unreachable spaces (Miller 1991; Kwan and Hong 1998), and b) areas which individuals have no direct contact with (Patterson and Farber 2015).

Miller (1991) proposed two types of discrete space PPAs. The first one is defining an individuals's PPA as the subset of activity locations reachable by that individual. That subset is often called the Feasible Opportunity Set - FOS (Golledge, Kwan, and Gärling 1994). The second type of discrete space PPAs is based on the street segments and intersections reachable by an individual given their movement constraints. In this context, the PPA is defined in terms of the size of the transportation structure that both allows and constraints people's movement (see figure 2.9).

The simplest technique to build a road network PPA is the potential path tree (PPT), which consists in calculating shortest path trees from the individual's fixed activity locations until a time budget threshold is reached (Miller 1991), as shown in figure 2.9a. This technique, however, leave gaps in the network due to edges unvisited by the algorithm (Miller 2007). To fill those gaps, Miller (1999) proposed the use of the extended shortest path tree algorithm from Okabe and Kitamura (1996), generating a Potential Network Area - PNA (see figure 2.9b).

Road network PPAs can account for the topological structure and directionality of the road network, which can significantly decrease mobility due to turn restrictions, presence of one-way streets, and crosses on non-planar space such as overpasses and underpasses (Kim and Kwan 2003). Road network PPAs

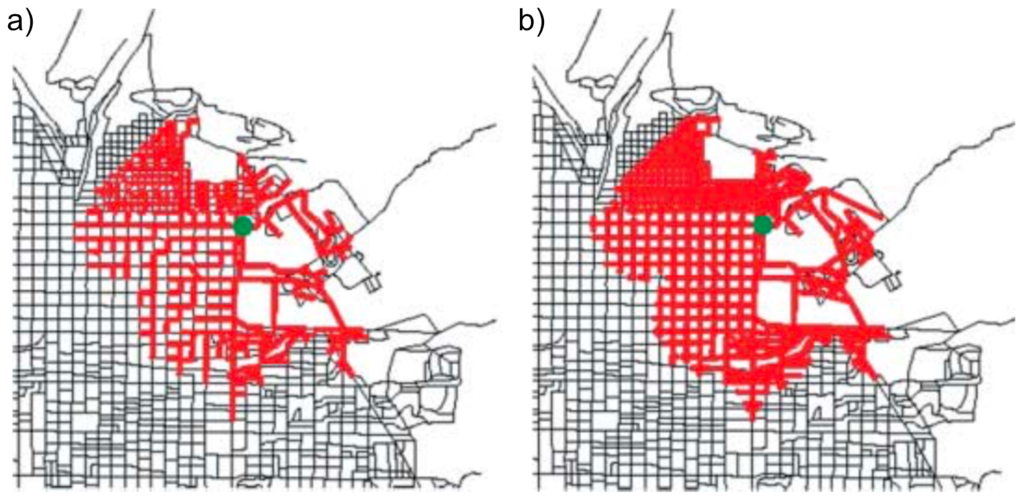

Figure 2.9: Street network representations of potential path areas: a) potential path tree (PPT), b) potential network area (PNA). Source: Miller (2007, 515)

can also consider different travel velocities and traffic times when calculating the areas individuals can reach during their time budgets. This can be done both statically, by assigning traffic times directly to street segments (Miller 1991; Kwan and Hong 1998), and dynamically, by considering varying travel times according to congestion levels in peak and off-peak hours (Wu and Miller 2001).

Discussion

The overview presented in this section demonstrates the variety of techniques researchers have developed to deal with the problem of identifying individual activity spaces. The variety of solutions stem from a trade-off between computational efficiency and theoretical consistency. Simpler activity space representations are easier to compute and require less data, but can be considered less theoretically accurate and may include areas inaccessible to the individual in reality. More complex three-dimensional space-time prisms may better represent an individual's actual activity space, but those are harder to compute, visualise, and interpret. Ultimately, the choice of activity space representation depends on the research objectives, computational resources, and data available.

2.3.2 Acquiring Data

All the methods for building activity spaces previously discussed rely on data on people's trajectories, activities, and time constraints in different levels of detail. This section will present an overview of the data collection methods found in the literature. A summary of those techniques is presented in table 2.2.

Table 2.2: Summary of data sources used for defining activity spaces.

Method	Characteristics
Travel diary surveys	**High Level of Detail** Detailed socioeconomic and travel behaviour data on each participant can be collected, as much as needed. **Low Scalability** Usually small samples, due to manual processes of data collection and processing.
GPS	**High Level of Detail** Detailed socioeconomic data on each participant can be collected. GPS data is a more reliable source of travel behaviour information than travel diaries. **Low/Medium Scalability** Providing GPS tracking devices to research subjects can be costly and limited to small scale studies. Providing a GPS tracking app that participants can install on their own smartphones increases the scalability of the research.
Participatory GIS	**Medium/High Level of Detail** High flexibility regarding the kind of information that can be collected. Potentially, detailed socioeconomic data on each participant can be obtained. **Low/Medium Scalability** Scalability depends on recruiting participants.
Social Media	**Low Level of Detail** Socioeconomic data on users are not available and may be inferred from census data when needed. **High Scalability** Social media APIs (such as Twitter) provides direct access to the activities of many users who decide to share their activity and location data. Issues in self-sampling of users, which may not represent the whole population.
Mobile phone	**Low Level of Detail** Socioeconomic data on users are not available or very limited and may be inferred from census data when needed. Spatial resolution may be limited. **High Scalability** Access to a significant proportion of the population can be achieved with this method.
Flow Data (OD)	**Medium Level of Detail** Socioeconomic and flow data on a large (if not the whole) parcel of the population can be obtained. However, there is no temporal information, nor trajectories. **High Scalability** Access to the entire population, in the case of Census OD data.

Travel diaries are the most common approach for collecting individual activity data for time-geographic studies. By using this technique, researchers are able to collect a vast array of socioeconomic and travel behaviour data on each participant, in a level of detail specifically tailored to the research needs. However, such studies are difficult to replicate or to expand to a larger population beyond the original sample. Kamruzzaman and Hine (2012) reviewed the sample sizes of several travel diary surveys used in studies carried out between 1999 and 2011. They found significant variation between the sample sizes, which ranged from 19 to 755 diaries collected. While most studies collected data for 1 to two days, up to a week, one particular study by Schönfelder and Axhausen (2003) collected travel activity information for 6 weeks. Although not comprehensive, the evidence seems to indicate there's no consensus regarding the size requirements of travel diary surveys. Although new software tools (Safi et al. 2017; Prelipcean, Gidófalvi, and Susilo 2018) can help automate travel diary collection and reduce issues such as low response rates and incomplete trip declaration by participants, those are still very new, rarely used, and their impacts on the field are yet to be seen.

One alternative to travel diaries is the use of large scale travel surveys carried out by government agencies and planning bodies. Examples are the U.S. National Household Travel Survey (Santos et al. 2011), Mobidrive for German cities (Schönfelder and Axhausen 2003), and the SIRS survey in France (Vallée and Chauvin 2012). A shortcoming of this approach is that the data may have been collected with focus on different objectives, which may lead to issues as some important information be missing, or the sample not being representative of the research's target population.

Origin-destination matrices (OD, also known as flow data) from censuses can be a useful datasource in time-geographic studies. Farber and colleagues (Farber et al. 2013; Farber et al. 2015) use flow data to calculate potential opportunities for social interaction among different population groups for metropolitan regions. The advantage of this technique is that socioeconomic data and trips' origins and destinations of a large parcel of the population can be obtained, in many cases, from the census. However, OD matrices usually contain only origins and destinations of trips, without temporal information nor the actual trajectories of trips.

Data collection can be partially automatised with the use of GPS technologies, both in dedicated devices or in apps installed in modern smartphones. Yip, Forrest, and Xian (2016) used an Android app to track the location of participants and to identify places of activities (defined as places where the participant stayed longer than 30 minutes). Greenberg Raanan and Shoval (2014) used this technique to compare perceived territorial boundaries, in the form of mental maps sketched by the participants, to the participants' actual trajectories collected via GPS receivers, finding a strong relationship. This technique brings many advantages over travel diaries and interviews, such as the elimination of the burden on the participant of remembering past activities and filling the diary, as their locations are recorded automatically by the app. However, this technique is still limited to voluntary participation, and reaching out to possible participants can be time consuming.

Participatory GIS (PGIS) techniques allow surveying individuals remotely through the combination of online mapping and questionnaires. Huck et al. (2019) collected data in Belfast, Northern Ireland, using an online PGIS platform for a study on religion segregation. Participants identified as Catholics, Protestants, and Other, were asked to divide the city in areas perceived as used mainly by a single group, or shared by more than one group. The authors used a *spray-can* technique so users could demarcate regions with fuzzy limits rather than the hard boundaries of traditional areal units such as census tracts. The sample used was limited to 33 recruited participants, but the technique has the potential to be scaled to larger samples depending on recruiting efforts.

Passive mobile positioning data, made available by mobile operators, allow the collection of data in very large scales. Silm and Ahas (2014a, 2014b), for example, were able to reach half the population of Estonia in their study of segregation among ethnic Russians and Estonians in the country. The information supplied by the operator can vary, and in their specific case the researchers were able to obtain, together with the call detail record, the sex, birth year and language the mobile phone user preferred to communicate with the service provider. With this information, the authors could identify the ethnicity of the users, assuming the language chosen is the individual's first language. Among the shortcoming of this technique are the spatial resolution of the data, which is restricted to the location of the cellular antenna closest to the user, and limited demographic information due to data privacy issues.

Another alternative for mass data collection is from social media feeds. Netto et al. (2018) collected data from Twitter users in the city of Rio de Janeiro, Brazil, generating a database and 20,029 geotagged tweets belonging to 2,543 users over the course of 56 hours. The place of residence of each user was estimated based on the location of their first tweet in the morning, and the income of the user was estimated by the characteristics of the census tract they live in. This technique allows gathering data of a large number of subjects when no other sources are available. The data collected are also more recent and frequently updated. As a shortcoming, the user base may not be representative of the whole population, and the socioeconomic characteristics of the user has to be estimated.

2.3.3 Measuring Accessibility and Segregation

This section discusses how the concepts and techniques discussed in this chapter were translated into measures of accessibility and segregation in the literature.

Measuring Accessibility

Several approaches for quantifying accessibility in people's activity spaces can be found in the literature. Those can be classified into four categories: *geometric* measures, *cardinal* measures, *temporal* measures, and *utility* measures.

Geometric measures are the most traditional in time geography. The volume of an individual's space-time prism can be considered a direct proxy for that individual's accessibility (Lenntorp 1976; Burns 1979; Miller 1991). However, due to the inherent difficulty of working with three-dimensional prisms, the area of two-dimensional potential path areas are more frequently used as measures of individual accessibility in practice, such as in Newsome, Walcott, and Smith (1998) and Kamruzzaman and Hine (2012). Such geometrical measures are not considered good measures of accessibility, though, as they may contain too many empty and unreachable spaces which provide no value to the individual (Miller 1991; Kim and Kwan 2003).

Cardinal measures are derived from the number of feasible opportunities in the individual's PPA, stemming from Lenntorp's (1976) work. These measures tend to be considered more adequate representations of an individual's accessibility than their PPA's geographic extent, due to the aforementioned issues with geometric measures. Following Kwan (1998) and Kim and Kwan (2003), a generic accessibility measure based on the number of reachable opportunities can be defined by equation 2.1.

$$A_s = \sum W_k I(k) \tag{2.1}$$

In equation 2.1, the function $I(k)$ indicates whether the activity k is part of the individual's feasible opportunity set (FOS), as per equation 2.2.

$$I(k) = \begin{cases} 1, & \text{if } k \in FOS, \\ 0, & \text{otherwise;} \end{cases} \tag{2.2}$$

In equation 2.1, W_k accounts for the weight of opportunity k. This parameter can be used to differentiate each opportunity's contribution to the individual's total accessibility, or else simple set to 1 to use the size of the choice set as an accessibility measure. Kwan (1998), for example, sets this parameter to the land parcel's area multiplied by a building-height factor, as a proxy for real parcel's properties such as retail floor space and employment. Cardinal measures can be interpreted as measures of freedom of choice (Neutens et al. 2010), since an individual is more likely to find suitable locations to carry out desired activities in larger opportunity sets.

Temporal measures of individual accessibility, according to Neutens et al. (2010), account for the amount of time available for carrying out each activity. For example, an opportunity has no practical value to an individual if its opening hours are not coincident with the individual's free time, or if the time available is too short to carry out the desired activity. Temporal measures are defined as per equation 2.3:

$$A_s = \max_{k \in FOS} \left[(t_k^e - t_k^s) I(k) \right] \tag{2.3}$$

In equation 2.3, t_k^e and t_k^s are the earliest ending time and latest starting time of activity k, respectively. While cardinal accessibility measures count all opportunities, temporal measures only consider the opportunity with the maximum benefit in terms of the time available to participate on the activity. This is a measure of an individual's temporal freedom (Neutens et al. 2010), or their freedom of choosing when and for how long to carry out a particular activity.

Utility measures were developed originally by Burns (1979) and extended by Miller (1999) based on concepts of random utility theory. In these measures, opportunities are differentiated by their utility to an individual, accounting for the benefits obtained from participating in that activity by weighing in factors such as attractiveness, possible activity duration and proximity (Neutens et al. 2010). A locational benefit can be defined as in equation 2.4.

$$B_{ik} = a_k T_{ij} e^{-\lambda t_{ik}} \tag{2.4}$$

In equation 2.4, B_{ik} is the locational benefit individual i obtains from participating in activity k, a_k is the attractiveness of activity k, T_{ij} is the maximum duration of activity k considering individual i time-budget constraints, t_{ik} is the combined travel time from previous activity to activity k and from activity k to the next activity, and λ is a travel time/distance decay parameter.

According to Neutens, Versichele, and Schwanen (2010), two versions of individual accessibility measures can be derived from the locational benefit concept: a) the *additive*, considering all opportunities in the FOS contribute to the individual's accessibility; and b) the *maximative*, assuming the utility an individual obtains from the opportunities available is equal to the opportunity with the largest benefit in the FOS. The additive and maximative measures are defined below, in equations 2.5 and 2.6, respectively:

$$A_{ik}^{add} = \sum_{k \in FOS} B_{ik} \tag{2.5}$$

$$A_{ik}^{max} = \max_{k \in FOS} B_{ik} \tag{2.6}$$

Measuring Segregation

The approach to measuring segregation on activity spaces usually follows similar techniques to the place-based segregation indices discussed earlier. The main difference is that activity space segregation studies use representations of space beyond the residential location, which are derived from people's activity and mobility patterns. For example, the population composition may present significant variation in different areas of the city throughout the day, even when fixed spatial units such as census tracts are considered, thus affecting the evenness/clustering dimension of segregation. People visiting different places have different probabilities of meeting members of other population groups, thus changing their levels of isolation and exposure. Once an activity space is defined for an individual, or even a group of individuals, applying the same place-based segregation indices to the activity spaces is usually a trivial task. A few techniques and possibilities are worth mentioning here, though.

The extent of an individual's activity space can give insights in their mobility levels and overall experience of the city. Some travel behaviour metrics, such as number of trips (Schönfelder and Axhausen 2003), number of places of activity visited (Yip, Forrest, and Xian 2016; Aksyonov 2011; Silm and Ahas 2014a), and geographical extent of activity locations (Palmer et al. 2013; Wang, Li, and Chai 2012; Wang and Li 2016) can be used to assess differences in activity spaces among groups. Some studies (Lee and Kwan 2011; Jang and Yao 2014; Huck et al. 2019) focus on techniques to visualise activity spaces, highlighting the extent of each group's reach over the urban area and identifying patterns of occupation of each group.

The dimension of segregation most explored in activity space segregation studies seems to be the exposure/isolation dimension. People living apart may interact with each other by visiting neighbourhoods mainly inhabited by other population groups (Yip, Forrest, and Xian 2016; Palmer et al. 2013) or in their trajectories on the road network when inbetween activities (Netto 2017; Netto et al. 2018). The amount of time spent on one's own territory rather than in territories perceived as belonging to other groups can be seen as an important indicator of ethnic isolation, such as the case of ultra-ortodox Jewish and Palestinian Muslim women in Jerusalem, which avoid each others' territory but share spaces inside secular Jewish areas (Greenberg Raanan and Shoval 2014).

Another method worth mentioning here was developed by Farber et al. (2015), who use origin-destination (OD) matrices from the census to calculate the Social Interaction Potential (SIP) of a region based on the concept of joint-accessibility (Farber et al. 2013). Joint-accessibility measures the amount of time available for two individuals to participate in the same activity together. The SIP represents the average volume of the intersection between the space-time prims of all pairs of individuals in a region by aggregating their joint-accessibilities. The measure can be decomposed by social group, to estimate the exposure or isolation of each group. However, flow data contains no temporal information nor real trajectories, which need to be estimated to calculate the study region's SIP.

Discussion

It is clear from the discussion above that accessibility and segregation are approached rather differently in the literature, when it comes to actually measuring both phenomena. Individual-based measures of accessibility seem to be more mathematically well defined and applicable to more diverse situations. Segregation studies, however, tend to use more ad-hoc approaches to quantifying the problem, which depend on the definition of activity space used and available data. As previously mentioned, individual-based segregation studies are more recent than individual-based accessibility studies, which may partially explain this difference.

2.4 Summary

This chapter discussed accessibility and segregation in light of the two main approaches both issues have been studied in the literature: place-based and individual-based. Although the place-based approach is more popular and less challenging to operationalise and interpret, the individual-based approach is considered theoretically more advanced and sensitive to fine-grained patterns.

More importantly for this thesis, accessibility and segregation are much related concepts when looked at from an individual-based perspective, sharing a common theoretical background and methodological framework. However, those similarities are seldom explored in the literature and the challenges of measuring accessibility and segregation at the individual level are significant.

This thesis proposes an alternative solution to those challenges, based on modelling. The following chapter will discuss the modelling techniques available to simulate individual trajectories and flows of people, which will be used to derive people's activity spaces and accessibility and segregation measures in this study.

Chapter 3

Urban Modelling

The previous chapter introduced the theoretical framework of individual-based studies of accessibility and segregation, including this thesis, which is based on Hägerstrand's (1970) time geography. This chapter will introduce the methodological framework used in this thesis, which is agent-based modelling.

This chapter is organised in three sections. The first section presents an overview, definitions, and concepts regarding agent-based modelling. The second section discusses the process, advantages, and challenges of agent-based modelling as a research method. The third section presents an overview on how the themes of accessibility, segregation, and individual movement have, so far, been approached using agent-based models.

3.1 Agent-Based Modelling

Models can be defined as simplified representations of reality (Batty 2009b). As argued by Gilbert and Troitzsch (2005), in the broadest meaning of the term, models are built as a means to understand the complex world we live in. Within the scope of this research, the term model is used specifically to refer to simulation models, as defined by O'Sullivan and Perry (2013):

> "In a simulation model, a computer is programmed to iteratively recalculate the modelled system state as it changes over time in accordance with the relationships represented by the mathematical and other relationships that describe the system." (9)

One of the main advantages of using simulation models in urban studies as well as other social sciences is they allow computers to be used as virtual laboratories where theories and hypotheses can be tested. This ability to conduct experiments in an artificial environment is one of the main reasons for the popularity of simulation models, in particular in social sciences, where it is “impossible, impractical or unethical” (O’Sullivan and Perry 2013, 16) to close a system (or part of it), such as a city, for experimentation. By using models, social scientists can build models of “artificial societies” focusing on specific aspects of the real world they are studying and use them to run experiments, test hypotheses, and build theories (Gilbert and Troitzsch 2005).

In spatial sciences, such as geography and urban planning, models are often used to “develop and test theories, ideas and hypotheses and to convey those ideas and concepts in teaching and education” (Torrens 2010, 428). The use of computational models to aid the development of theories and ideas by requiring their translation from natural language to a formal language which the computer can understand (Gilbert 2008) can be seen as a fundamental advantage of modelling for the development of social sciences. As computers require a complete and exact set of instructions to run a model, which is difficult to achieve using natural language, the actual process of describing the problem in a structured way is valuable as a theory building method, and can be compared to the role of mathematics in the physical sciences (Gilbert and Troitzsch 2005). According to Van der Leeuw (2004), the formalisation of the problem via computer programming is even more important in interdisciplinary research, because it provides a neutral view on the problem and makes communication among researchers from different fields easier and more precise.

Models of urban systems were first developed in a planning context, mainly with the aim of predicting future urban development (Batty 2008). Predictive modelling assumes the model replicates processes with a level of accuracy which is deemed good enough so the models can then be used to simulate the passage of time and predict the state of the system at some point in the future (Gilbert and Troitzsch 2005). Those models of first generation were mainly static and large scale representations of urban states[1]. Predictive models were based on systems approach, following principles which, according to Batty and Torrens (2005), were later found to be unsuitable to cities. Hence, in predictive models, cities were: treated as closed systems, containing a well defined boundary with the wider environment; assumed to eventually reach an equilibrium state; and modelled as if composed of homogeneous elements. It is now known that cities are open systems with no clear boundaries, usually in far-from-equilibrium states, and composed of many heterogeneous agents and objects (Batty and Torrens 2005), thus contradicting those early assumptions. This shift in understanding has marked a change in paradigm in urban science towards a complexity theory approach, as it became clear cities present many characteristics inherent to complex systems.

There is no universally agreed definition of complexity theory, in part

1. See Batty (2008) for an overview on the evolution of urban models.

due to it being a relatively recent field of study (Mitchell 2009). Despite that, there seems to be a consensus on the overall characteristics of complex systems. It is often agreed, for example, that in complex systems the whole is more (and different) than the sum of its parts (Batty 2000). Essentially, the complexity theory approach shifts the focus from systems structure to dynamics, highlighting the need for understanding the drivers of (continuous) change and disequilibrium taking place in complex systems such as cities. As such, complex systems present high level of unpredictability, not only because a complex system can respond to the same problem in many different and equally valid ways (Allen 2012), but also because of path-dependency, in which future states depend on initial conditions and previous responses which are impossible to ascertain (Batty 2008).

This unpredictability of complex systems is on the root of the concept of emergence (Batty and Torrens 2005). Emergence refers to macroscopic behaviours or structures that originate from the actions of individuals at the local level. Usually, the individuals' actions follow simple rules, or rules seemingly unrelated to the system's macro-behaviour (Mitchell 2009), which makes the resulting pattern seem surprising to the observer (Miller and Page 2007). These characteristics of complex systems, including cities, prompted a shift to more disaggregated and dynamic modelling techniques able to simulate emergent phenomena from the bottom-up (Batty 2008). According to Batty and Torrens (2005), it also marked a change in the main aim of modelling cities, from prediction-oriented to more theoretical discussions and production of what-if scenarios.

It is in this context the use of agent-based modelling techniques to simulate urban dynamics has become popular. Agent-based models aim to simulate macroscopic patterns that originate from the actions and behaviours of individual agents at the local level, which is in line with the cities' emergent characteristics. In agent-based models, agents are created within an environment, and are able to interact among themselves and with said environment (Gilbert 2008).

According to Sycara (1998), in agent-based models: individual agents have *incomplete knowledge* about the environment and other agents, and have to make decisions based on the information available; there is *no central control*, meaning no single entity is controlling the behaviour of the whole system; there is *no central data storage*, meaning information is distributed among agents and environment; computation is *asynchronous*, meaning each agent acts on their own time.

Agents can be defined as "computer systems that are capable of autonomous action in some environment in order to meet objectives that are delegated to them by us" (Wooldridge 2013, 8). Agents can represent individual entities such as "people, buildings, cars, land parcels, water droplets or insects" (Crooks and Heppenstall 2012, 88), or collective organisations such as firms and nation-states (Gilbert 2008). Although agents can have many properties such as intelligence, mobility, communication, perception, and vision, those attributes do not need to be present in every agent-based model since their importance vary from domain to domain (Wooldridge 2009). According to Crooks and Heppenstall (2012), at least three characteristics seem to be common to most agents: a) agents are active, meaning they impact the simulation they are part of; b) agents are autonomous, meaning they act independently of central control; and c) agents are heterogeneous, meaning they represent individuals with different attributes.

Wooldridge (2013) suggests *intelligent agents* have an extra set of properties: *proactiveness*, or the ability to pursue goals by their own initiative; *reactivity*, or the ability to perceive and adapt to changes in their environment; and *social ability*, which is the ability to interact with other agents when pursuing their goals. The combined effect of those properties generate more complex behaviour. Although an agent pursues a goal, it has to be able to react to changes in the environment, for reasons such as: a) the environment may change, rendering the initial assumptions about it obsolete; and b) the goal may not remain valid, so there is no reason to continue pursuing it. When modelling geographic phenomena, which are complex and dynamic, these reasons are usually present. Finally, social ability is not just communicating, but also cooperating and sharing goals, negotiating, and competing.

The *environment* is the space where agents operate and with which they interact. Usual environmental representations are continuous spaces, cellular grids and social networks (Crooks and Heppenstall 2012). The choice of environmental representation has an effect on practical decisions during modelling. For instance, the proximity function depends on the environmental representation: spatial distance for continuous space, adjacency for grid cells, and connectivity for social networks. The environment can represent a geographical space, in which case it is called spatially explicit (Gilbert 2008). The environment's complexity can vary significantly according to the needs of the model, sometimes being as elaborate as the agents (Gilbert 2008). Urban environments, which comprise social and geographical phenomena, present high levels of complexity, although part of that complexity is usually abstracted for modelling purposes.

Agent's behaviours and relationships to each other are governed by a set of rules defined in the model. Such rules are usually based on the domain knowledge of the field and existent literature, but can also be derived from data analysis (Crooks and Heppenstall 2012). The rules can be applied to the entire set of agents or to each agent individually, which is one of the advantages of agent-based models (Crooks and Heppenstall 2012). Even when the behaviour rules assigned to the agents are simple, their interactions can generate complex behaviour. This is particularly true in the social sciences, where most relationships between the involved variables are not linear (Gilbert and Troitzsch 2005).

Interactions among agents can be an important aspect of agent-based models. For example, agents can sense each other's presence, avoid collision, move in groups and relate to each other through social networks (Batty 2012). These interactions can vary from simply reacting to external stimuli to being goal-oriented, or take place synchronously at discrete time steps or asynchronously

(Crooks and Heppenstall 2012). The ability to simulate agent-to-agent interactions is one of the main advantages of ABM, setting them apart from more traditional modelling methods (Gilbert 2008). But agent-based modelling also allow agent-to-environment interactions, which are in the core of the definition of such models, and also space-to-space interactions, making them flexible tools for spatial simulation (Batty 2012).

Crooks and Heppenstall (2012) sum up the main three advantages of agent-based modelling over traditional modelling techniques. The authors argue that a) agent-based models act from the bottom-up, so they are specially fit to capture emergent phenomena; b) agent-based models provide a natural metaphor for representing real world complex systems in a software environment, as agents can represent individuals and organisations in a society, and the environment can represent the geographical space where this society exists; and, finally, c) agent-based models are flexible, allowing the development of simple yet extendable models, where complexity can be added over time in the form of agent's behaviour, intelligence and rules of interaction, as well as different aggregation levels and environmental complexity.

However, building agent-based models is not a trivial task. Deciding what aspects of the real world should be included or not in the model is arguably the most difficult step in model design (Gilbert and Troitzsch 2005). One principle often mentioned in model building is the *Occam's razor*, or principle of parsimony, which states that among many possible solutions, the one that is simpler and relies on fewer assumptions is the better one (Van der Leeuw 2004). What scientists should aim at, according to Gilbert and Troitzsch (2005, 32), is "a model that embodies the minimum number of assumptions, but which applies as generally as possible to many different circumstances". Hence, choosing a level of abstraction at the beginning and keeping it throughout the model building process is not considered a good strategy. A better approach is treating simulation as a research process, and not as an end per se. Often, starting with a simpler model, which is easier to specify, to implement, and to understand, and that can be incrementally extended is considered more effective (Gilbert and Troitzsch 2005). Following this approach, it is easier to know when the model achieved the level of detail necessary to replicate the target phenomenon to an adequate degree, thus avoiding building a model more complex than necessary. Model design and building, as well as the challenges involved in the modelling process, are the subject of the next section of this review.

3.2 Challenges in the Modelling Process

There are many challenges throughout the modelling process. In what follows, the different stages of the modelling process will be presented and their challenges discussed. Those stages, according to Gilbert and Troitzsch (2005), are: a) model design; b) model building; c) verification and validation; and d) publication.

3.2.1 Model Design

Model design is a key stage of the modelling process. It starts by the identification of the research problem, in the form of a real world phenomenon the research aims to explore (Gilbert and Troitzsch 2005). From the overall research object, a specific research target is defined. This step also comprises the selection of which aspects of the real world are to be included in the model and which are to be left out. Observing the system in the real world is essential to acquire an initial understanding of its structure and of the processes regulating its dynamics. Techniques such as the Pattern Oriented Modelling (POM) framework, proposed by Grimm and colleagues for ecological modelling research (Grimm and Railsback 2012; Grimm et al. 2005) are useful at this stage of model development, as well as the later evaluation stage. POM's main reasoning is that patterns encode information on systems' structure and organisation, hence using such patterns in model design is an effective way of tying the model to the real world system.

Crooks, Castle, and Batty (2008) identified seven key challenges in agent-based model development for spatial simulation, spanning all stages of model building. One of such challenges is the defining the purpose of the model, which can focus on policy application, theory building, or any point in between. While the first generation of urban models were mostly focused on policy and prediction, agent-based models are usually more speculative and focused on exploring what-if scenarios (Batty 2008). The decision regarding the model's purpose is crucial as it dictates subsequent development and outcomes. For instance, when the aim is prediction, the requirements regarding the amount and accuracy of data, as well as on the accuracy of the assumptions built into the model's logic, are much higher than when the aim is exploratory (Gilbert and Troitzsch 2005).

Other two challenges related to model design are the role of theory and agent representation. Models can be useful as virtual laboratories where experiments can be carried out, but for that they need to be grounded in theory and domain knowledge. However, the theory is many times just implicit in some models and obfuscated by ad hoc assumptions. Defining which real-world entity agents should represent is, perhaps, the decision that has the most significant impact on the results of the simulation. By definition, agents can represent any kind of object from the real world at the individual level. However, this kind of individual representation sometimes is not the most adequate and some sort of aggregation is required. Hence, agents could represent from small groups of individuals to larger entities such as a firm or a city. The problem with such aggregations is that the processes that act at the individual scale are inherently different than processes acting at the aggregated scale, so the model's design and rules must be adapted accordingly (Crooks, Castle, and Batty 2008).

Also related to model design and agent specification is the decision-making process implemented in the model. The ability of agents to make decisions at the individual level is a fundamental advantage of agent-based models when compared to other forms of modelling (Wooldridge 2013). While traditionally models have been built based on the assumption that humans always make rational decisions and have complete information about the problem at hand, such assumptions are now known to be false (Heppenstall, Malleson, and Crooks 2016; Manley and Cheng 2018; Portugali 2011). A wide range of techniques were developed aiming to replicate human cognition traits into agents' behaviour, stemming from the principle of bounded rationality (Simon 1957). Bounded rationality suggests that humans are unable to take perfectly rational decisions both due to the lack of cognitive ability and incomplete information about the problem. Decisions, then, are taken despite those limitations, aiming to satisfy the current needs in the best possible way, but never perfectly.

3.2.2 Model Building (Programming)

Model building consists of translating the conceptual model into computer language so it can be executed and the planned experiments can be carried out. There are basically two ways in which this task can be done: the first one is by programming the model from scratch, preferably in an object-oriented language; the second is by using one of the many agent-based modelling platforms and libraries available (Gilbert 2008; Crooks and Heppenstall 2012).

The main advantage of programming a model from scratch is that it gives the modeller complete control over all stages of the modelling development cycle (Crooks and Heppenstall 2012). However, this can be time-consuming because parts of the model which are non-central to the research problem being addressed but essential to a functioning model need to be programmed as well. These parts include a graphical user interface (GUI), data import and export, and visualisation (Crooks and Heppenstall 2012). In the case of spatially explicit models, such as urban models, this may include dealing with GIS-specific data formats which require specific knowledge and experience from the researcher. However, some libraries not specific to agent-based modelling may be used to aid in some specific tasks such as GIS data input and output, data visualisation, statistics, and so on. The availability of such libraries is heavily dependent on the programming language of choice.

Agent-based modelling programming platforms are useful for building models at any scale, from simple prototypes to massive models containing millions of individual agents (Kravari and Bassiliades 2015). The main advantage of using one of such platforms is the improved efficiency in terms of programming time, debugging and deploying the application. Part of this efficiency comes from pre-built non-specific aspects of the software, such as the GUI and visualisation tools mentioned earlier. Yet another feature of agent-based modelling platforms that increases efficiency is the available building blocks of ABM that are readily available for reuse instead of building from scratch. Features such as dynamic schedule of agents' actions, randomisation of call orders, synchronisation and time-keeping, and tools to keep track of agents' states are all useful, necessary in many models, and yet may be difficult to implement and thoroughly test. Agent-based modelling platforms help keeping the program free of bugs because the application building blocks were developed by professional programmers and tested by many users (Gilbert and Troitzsch 2005). One disadvantage of modelling platforms is they are limited in their capabilities, possibly hindering development if they are not open for extension (Gilbert and Troitzsch 2005).

A large number of agent-based modelling platforms are available, and comprehensive reviews were conducted by Abar et al. (2017), Kravari and Bassiliades (2015), and Railsback, Lytinen, and Jackson (2006). Popular programming languages in the agent-based modelling community are Java, C++ and C#. NetLogo (Wilensky 1999) is a popular platform that provides an entry-level programming language integrated into a complete modelling environment, which is easier to learn than more generic programming languages. Besides its simplicity, NetLogo contains extensions that enable advanced features such as GIS capability, and the integration with other programming languages with additional capabilities such as R and Python.

One common recommendation regarding the choice of programming language, either when building a model from scratch or using a modelling platform, is that the language should be object-oriented (OO). In the OO paradigm, objects are individual entities which contain actions and data. Thus, OO provides a clear metaphor for programming agent-based models.

One of the key challenges in agent-based modelling presented by Crooks, Castle, and Batty (2008) can be related to the programming stage of model development: making agent-based models operational. This challenge refers to building a model that is robust and user-friendly enough to be used outside academia or the environment in which the model was developed. This requirement adds a considerable amount of programming hours and expertise from the part of the research team, to the point that most operational models used in planning are built and maintained by professional teams of software developers.

3.2.3 Evaluation

Model evaluation is technically the step following model programming. However, in practice, both activities are carried out simultaneously. Some of the model evaluation tasks usually start as soon as the model starts producing observable results (Gilbert and Troitzsch 2005). However, there is no standard or consensus in the urban modelling community on how the validation process should be carried out, or even regarding the terminology to be used (Ngo and See 2012; Heppenstall, Malleson, and Crooks 2016). The whole process of model evaluation is highlighted by Crooks, Castle, and Batty (2008) as one of the key challenges of agent-based modelling, because the specifics of agent-based models make them more difficult to evaluate than traditional static and aggregated urban models. There is general agreement in the literature that many activities are part of model evaluation, such as verification, calibration, validation, sensitivity analysis, and robustness analysis (Gilbert and Troitzsch 2005; Ngo and See 2012; Crooks and Heppenstall 2012; Batty and Torrens 2005; Manson, Sun, and Bonsal 2012; Grimm and Berger 2016). These steps will be discussed in this section.

Verification and validation are key steps in model evaluation. A brief description of those concepts is provided by North and Macal (2007, 30-31) as: "Verification is the process of making sure that an implemented model matches its design. Validation is the process of making sure that an implemented model matches the real-world". So both processes together aim to make sure the model is correctly implemented and adequately replicates the real world phenomenon it is supposed to. Large part of the verification process is debugging (Gilbert and Troitzsch 2005). The usual strategy is building the model iteratively, adding few features at each time and testing the model frequently. In the social sciences modelling literature, this process is sometimes called unit testing (Gilbert and Troitzsch 2005; Crooks and Heppenstall 2012). However, in computer science, unit testing refers to the more specific process of writing automated tests to the code, which are rerun every time a new version of the software is built. Failing to pass in one of those tests means a new feature has introduced some error into a routine which was previously working. This formal procedure is rarely discussed and implemented in the literature on models of social systems. However, the larger the model and the team working on it, the more necessary it is.

Ngo and See (2012) suggest two other processes are also part of verification: face validation and sensitivity analysis. Ngo and See (2012) define face validation as the qualitative and visual interpretation of the model's outputs. This process is done in the early stages of the model's development, and is less rigorous than the validation of the model's outputs carried out at later stages. Face validation includes checking if the model's variables are kept consistent with agents' and environment's states, if the agents' behaviours correspond to input parameters' values, and if the results appear realistic (Ngo and See 2012).

Sensitivity analysis consists in varying the model's initial conditions and parameters in a systematic way aiming to verify which changes have the most impact on the simulations outcomes. A clear issue in this task is that, even a small number of parameters and input data variations can present a large number of possible combinations that must be tested. Additionally, the model may contain stochastic processes that lead to different possible outcomes for the same set of inputs, increasing the number of necessary model runs. Thus, sensitivity analysis can be a time consuming and resource intensive process. This process can be made more efficiently by computational tools such as NetLogo's BehaviourSpace (Wilensky 1999), which provides a way of automating many runs of the same model and recording the results. Also, an experienced modeller will have an idea on which parameters have the strongest effect on the simulation, and what range of values make more sense, narrowing down the number of necessary model runs (Gilbert and Troitzsch 2005).

Another useful method of model evaluation is the process of robustness analysis (Grimm and Berger 2016), which fits between model verification and its application to real-world scenarios. The key aspect of robustness analysis is to assess the resistance of the model's results to major changes in the model's structure and parameters. In the authors' words, the process is conducted by intentionally "trying to break the model", running it in extreme conditions in order to identify the model's limits of operation and also its main structural weak points. This technique is similar to sensitivity analysis, and may be carried out simultaneously depending on the situation.

Once the model is verified and its implementation matches the conceptual design, it can be applied to the scenarios it was intended for. This is done in two steps: calibration and validation. Calibration is the process of fitting the model to a specific scenario by finding the set of parameters that generate the most optimal results. This is often compared to validation, as both processes may involve fitting the model to actual data (Crooks, Castle, and Batty 2008). However, calibration is focused on optimising the model's input parameters (Ngo and See 2012) while validation focuses on finding the goodness of fit between the outputs and the real world (Manson, Sun, and Bonsal 2012). Hence, validation produces a measure of the degree to which the model accurately reproduces some patterns present in the modelled phenomenon, and not a simple binary indicator of valid/invalid (Crooks, Castle, and Batty 2008). It may also produce an indication on which situations the model performs better or worse, and which elements impact the model's performance. In this process, the pattern-oriented modelling (POM) method mentioned earlier is also useful in helping to select which aspects of the real world phenomenon are more likely or not to be reproduced by the model, or which aspects can be more easily compared to the model's outputs (Grimm and Railsback 2012; Heppenstall, Malleson, and Crooks 2016).

The process of evaluating agent-based models presents many challenges (Crooks, Castle, and Batty 2008; Gilbert and Troitzsch 2005), and many published models only present partial attempts on validation (Batty and Torrens 2005; Heppenstall, Malleson, and Crooks 2016). It can be argued that the actual complexity and uncertainty of the world prevents a model being fully validated (Batty and Torrens 2005). For instance, real world processes are often stochastic, and models are designed with stochasticity built-in. Hence, the state of the world measured and the model's results trying to replicate that state are one of many possibilities and, given the same conditions, different outcomes could be as likely to emerge (Gilbert and Troitzsch 2005). Other possible issue is path-dependence, or situations where the simulation may be sensitive to precise initial conditions. In these situations, a small change on those conditions (e.g. due to measurement imprecision) may have a large effect on the final results (Gilbert and Troitzsch

2005). Finally, Batty and Torrens (2005) highlight the relationship between validity and replicability. In traditional laboratory science, a hypothesis must be replicated and verified many times before it is considered valid. Even though urban modelling tries to emulate laboratory science in the sense that computers are 'virtual laboratories', modelling experiments are very rarely replicated in more than one occasion in order to test their validity in different scenarios and study areas. In the occasions when this happens, the model usually has been modified enough for the subsequent case studies that it cannot be considered the same model anymore.

3.2.4 Publication

Although publication is common practice in all research fields, communicating agent-based models presents its own specific challenges (Grimm and Railsback 2005; Crooks, Castle, and Batty 2008). One of such challenges stems from one of the advantages of agent-based modelling mentioned earlier in this chapter: the translation of the problem into a formal computer language (Gilbert 2008), which may be an issue because of space restrictions in traditional paper-based publications (Gilbert and Troitzsch 2005). While other forms of modelling may be translated into more concise forms such as equations, agent-based models require more lengthy and detailed explanations if they are to be replicated (Grimm and Railsback 2005). This issue is being solved recently with the aid of technology. For instance, scientific journals are migrating to digital formats, which allow the publication of additional content such as source code, and the raw data used to obtain the results. Migrating to digital formats also improves model communication by allowing more interactive and dynamic forms of visualisation. For instance, agent-based models are inherently dynamic, potentially producing results in the form of animated graphics which can be share to communicate results more effectively than static images with textual description.

Fulfilling similar role than publication is making the model available to the wider public. Sharing model and data, as well as making simulation software more user-friendly, increased the chance of such models being used by stakeholders as actual decision support tools (Waldrop 2018). For example, models may allow the user to change its parameters *on-the-fly*, immediately changing the model's behaviour and the simulation's outcomes. Rich visualisations can attract the interest of the wider public and decision-makers, increasing the impact of the model and research findings outside academia.

As a final note on this section, it is important to consider that all the processes discussed above do not produce a complete and definitive model, but rather a useful, well tested and validated version of the model. Modelling is usually an iterative process, where one model is used as a starting point for new models, and so on. Also, as new data become available, new experiments may be conducted, which may identify limitations in the model which were previously undetected. At each point along the model development process, the researcher may find it necessary to return to a previous point to improve or correct some specific aspect of the model. Communicating the model frequently during its development is also useful in this regard, by gathering feedback on the model's theoretical background, assumptions and partial results from the wider community.

3.3 Segregation, Accessibility and Movement in Agent-Based Models

A wide range of agent-based models applied to accessibility and segregation studies can be found in the literature. While a comprehensive review of models that incorporate aspects of segregation or accessibility was deemed to be out of the scope of this thesis, it is important to have an overview on how agent-based models have dealt with those topics. As such, this review will present an overview of models which are relevant to this research. Three categories of agent-based models have been reviewed. The first two concern models which directly deal with accessibility and segregation, respectively. The third category concerns models that simulate people's trajectories in space and time, which here are referred to as 'models of movement'. Below each of those categories will be discussed.

3.3.1 Segregation in Agent-Based Models

Dynamic simulation models of residential segregation often stem from the work of Thomas Schelling (1971, 1978) on segregation as an emergent phenomenon. Schelling studied how the formation of segregated neighbourhoods could be explained by the preferences of individuals for living among their own kind. He proposed a simple model where households (akin to agents) are divided into two groups and attributed a level of preference towards individuals of the same group. These agents are then randomly positioned into a grid. Agents are considered satisfied with their location if the presence of agents of their group in their neighbourhood is equal or higher than the agents's stated preference. If this condition is not met, the agent then is relocated to a satisfactory empty space. This simple abstract model revealed, rather counterintuitively, that even low preferential biases could lead to the emergence of highly segregated residential patterns.

Schelling's work is considered the first application of agent-based modelling in the social sciences, even though he conceived his model before the advent of agent-based modelling as a research technique. His description fits well into an agent-based modelling framework as it includes individual agents interacting within an environment, and who make decisions based on their preferences and partial knowledge of the world. This led to many researchers replicating and extending his work using computational agent-based modelling techniques. Such applications include testing different preference functions (Bruch and Mare 2006; Clark 1991; Pancs and Vriend 2007), different definitions of neighbourhood (Fossett and Waren 2005; Laurie and Jaggi 2003; O'Sullivan, MacGill, and Yu 2003), and a vector-based spatial representation instead of the usual regular grid (Crooks 2010).

The aforementioned studies tend to follow Schelling in simulating racial or ethnical segregation in abstract or synthetic environments. Other studies, such as the segregation model by Benard and Willer (2007), applies Schelling's rules to wealth and status of agents rather than race or ethnicity. When segregation by income and social class are considered together, studies tend to use other theoretical frameworks rather than Schelling's. Examples are the simulation of urban expansion and formation of informal settlements and low-income areas in the peripheral areas of Latin American cities by Barros (2012), the formation of slums in India (Patel, Crooks, and Koizumi 2012) and South Africa (Shoko and Smit 2013), and the MASUS model of income segregation by Feitosa, Le, and Vlek (2011). There are also a number of models of ethnic segregation which are not based on Schelling's framework such as the model of ethnic segregation in Tel Aviv by Benenson, Omer, and Hatna (2002), which is based on the concept of spatial cognitive dissonance by Portugali, Benenson, and Omer (1997).

Despite following a range of different approaches, all of reviewed studies focused on the same objective of modelling the formation of residential segregation. During this review, no models were found that approached the problem of segregation from an analytical perspective, using agent-based modelling techniques to study and measure segregation in cities.

3.3.2 Accessibility in Agent-Based Models

Accessibility has been considered an important component on models of urban structure since the seminal work of Hansen (1959) on the influence of accessibility on land use distribution. Traditional large scale and operational models in the group of Land Use Transportation Interaction (LUTI) models also have accessibility as one of their main components.

In LUTI models, urban structure is considered as a result of a two-way interaction between transport and land use (Chang 2006), referred to as the "land use transport feedback cycle" (Wegener 2004, 129). This cycle consists on a loop in which land use distribution determines origins, destinations and the amount of trips that take place in cities, which in turn affect the general traffic distribution and levels of accessibility, which influence residential and commercial locations, thus closing the cycle (Ettema, Arentze, and Timmermans 2007).

Agent-based models of urban processes often employ a similar logic to those earlier models. However, the way in which agent-based models deal with accessibility tends to be different due to its bottom-up approach. In agent-based models, accessibility is usually treated as a variable in the agents' decision-making process. Broadly, its implementation can be classified into two approaches: place-based and people-based. Place-based measures of accessibility in dynamic models are usually treated as static layers, which is usually one of the model's inputs or it is calculated at the start of the simulation. This is the case in residential location choice modules inside larger LUTI models, such as the model by Salvini and Miller (2005), as well as the independent residential location models by Fontaine and Rounsevell (2009) and Jordan, Birkin, and Evans (2014).

The person-based approach treats accessibility as a variable of the agent, considering individual place of residence and work as well as mode of transport. For example, in the SelfSim model by Zhuge and colleagues (Zhuge et al. 2016), agents evaluate potential new places of residence based on the distance to their individual workplace and other places of activity, in addition to house characteristics such as size and price. Other models combine both approaches, such as the ones developed by Lee et al. (2010) and Babakan and Alimohammadi (2016). In those models, agents consider accessibility to opportunities such as schools, transport, and shops, as well as the distance of the new home to the agent's own workplace when selecting a new place of residence.

One important development regarding people-based accessibility in agent-based location choice models is the integration with transport models in order to obtain more realistic travel times and distances. For example Zhuge et al. (2016) integrate the MATSim (Horni, Nagel, and Axhausen 2016) activity-based model to the Residential Location Choice (RLC) and Real Estate Price (REP) components of the aforementioned SelfSim model. MATSim generates daily activity plans and trips for agents based on their current and potential new residences, which are then scored using a utility function. These scores are used as accessibility measures in the agents' decision of bidding for a new residence.

Accessibility is also important in travel demand agent-based models. Models in this category usually follow the Activity-Based Approach (ABA), which is based on the idea that travel is a derived demand from individuals' activities (McNally and Rindt 2007). Among the theoretical roots of ABA is Hägerstrand's (1970) time geography and the system of constraints that limits individuals' freedom to participate in activities (Axhausen 2000; Kurani and Lee-Gosselin 1997). Activity-based models aim to simulate activity schedules for households, considering details such as car ownership and sharing, division of tasks among household members, and actual feasibility of activities considering time constraints (McNally and Rindt 2007). The focus on individuals and households allows transport modelling in a much more disaggregated level than conventional transport modelling methods (Rasouli and Timmermans 2014), which makes the activity-based approach a good fit for agent-based modelling implementation. Examples of agent-based activity-based models are ALBATROSS (Arentze and Timmermans 2004; Timmermans and Arentze 2011), MATSim (Horni et al. 2009; Horni, Nagel, and Axhausen 2016), and SIMMOBILITY (Azevedo et al. 2017; Lu et al. 2015).

A common trend found in the agent-based models reviewed in this section is that accessibility is used as one variable among others in the agent's decision-making process. Differently from the segregation models previously discussed, accessibility has been approached using Hägerstrand's (1970) time geographic framework in agent-based models that aim to simulate people's activity schedules. However, those models are not primarily used on the analysis of individuals' accessibility. Instead, those models seem to be mainly predictive and exploratory, using accessibility as a means to the end of simulating processes such as residential location choice and travel behaviour.

3.3.3 Agent-Based Models of Movement

This section covers models which simulate individuals' movement in an environment, here referred to as 'models of movement'. These models do not deal specifically with either accessibility or segregation, but are included here because they use ABM to simulate individuals' trajectories and travel behaviour, which are relevant to the approach for building individual activity spaces being proposed here.

The problem of modelling individual movement through an environment is well suited for agent-based modelling, and it has been a prolific research area. Agent-based modelling has been applied to many types of individual movement, mostly focusing on specific types of movement or modes of transport. Below a brief overview of models that deal specifically with movement of people, namely pedestrian, cycling, traffic and passenger transport simulation, is presented.

Pedestrian models

Agent-based models of pedestrian movement are used to simulate detailed pedestrian navigation and interactions at the micro scale. Those models can be broadly categorised into two groups: a) models of normal pedestrian flow situations; and b) models of evacuation in emergency situations. It is widely agreed that both situations lead to very different behavioural responses from people.

Agent-based modelling is a useful technique to study the pedestrian navigation process. Besides the relatively simple problem of finding the shortest path to the destination, agent-based models allow the simulation of more complex dynamics. For example, Turner and Penn (2002) and Antonini, Bierlaire, and

Weber (2006) include agents' visual perception in the path planning process, while Stubenschrott et al. (2014) propose a dynamic path-finding process where agents continuously update their plans based on new information gathered. In a different perspective, Vahidi and Yan (2016) use an agent-based model to explore how pedestrians actively change the environment by creating new paths and trails. Pedestrian navigation modelling also focuses on techniques to simulate obstacle avoidance (Liu et al. 2014) and moving through bottlenecks (Dai, Li, and Liu 2013; Torrens 2012). These methods are important in the simulation of pedestrian flows in situations of high pedestrian traffic, such as train stations (Rindsfüser and Klügl 2007; Stubenschrott et al. 2014), and street parades (Batty, Desyllas, and Duxbury 2003). Those models have been proven useful in improving the design of those stations and the organisation of street events.

Particular focus has been given to agent-based simulation of evacuation from dangerous situations. The simulation of building evacuation due to fire or other hazards is a frequent topic of research (Korhonen et al. 2007; Okazaki and Matsushita 1993; Pelechano and Badler 2006; Tsai et al. 2011; Batty and Hudson-Smith 2014; Shi, Ren, and Chen 2009). Agent-based models also have been applied to larger scale evacuation situations, such as stadiums (Samuelson et al. 2008), concert venues (Wagner and Agrawal 2014), and large religious pilgrimages (Basak and Gupta 2017). A particularly large application has been developed to the U.S. government to simulate a nuclear attack and subsequent spread of radioactive fallout in Washington D.C., described by Waldrop (2018). The model includes 730.000 agents that try to reunite with family and friends, look for shelter and health-care, and evacuate the area while facing a changing environment. In these conditions, counterintuitive situations emerge such as people's tendency to flee towards the disaster area, instead of away, looking for missing family members and friends.

Cycling models

Concerns about the effects of increasing traffic congestion, air pollution, and greenhouse gas emissions on public health and quality of life in cities have led to increased interest in the active modes of transport: cycling and walking. As a result, cities around the world are increasing investment in pedestrian and cycling infrastructure in an effort to move away from costly and often ineffective investment in vehicle-oriented infrastructure (Cavill et al. 2008; Rissel 2009; Rissel et al. 2013).

Although simulating cyclists' movement is not very different from other types of movement, there are particularities. Beyond giving preference to shorter travel times and distances, cyclists have been found to avoid steep slopes and prefer riding over smooth surfaces (Li et al. 2012; Menghini et al. 2010; Milakis and Athanasopoulos 2014). Concern about road safety also seems to be a major factor, as cyclists tend to avoid routes with high traffic volumes and dangerous intersections and junctions (Li et al. 2012; Sener, Eluru, and Bhat 2009), while there is a preference towards continuous cycling routes (Sener, Eluru, and Bhat 2009; Menghini et al. 2010).

Cycling modelling has been done both by adapting large scale transportation models and by developing simpler specific models. For instance, Ziemke, Metzler, and Nagel (2017) demonstrate the use of MATSim transport model to simulate bicycle traffic, relying on OpenStreetMap for cycling infrastructure data. Examples of ABMs specifically developed for cycling modelling are the space syntax based wayfinding model by Rybarczyk (2014) and the model of cycling safety by Thompson, Savino, and Stevenson (2015).

Urban traffic and passenger transport models

Modelling traffic dynamics is the last stage (route choice) in the traditional four-step model (FSM) of transport modelling (McNally 2007). The first three steps (trip generation, trip distribution, and mode choice) deal with simulating the volume of travel between pairs of urban zones, in an aggregated manner. The aforementioned activity-based approach (ABA) was developed to simulate travel demand in a disaggregated way, and to be more consistent with people's real travel behaviour than the FSM (McNally and Rindt 2007). Although FSM and ABA differ significantly regarding travel demand modelling, both use similar methods when it comes to on-road traffic simulation (Rasouli and Timmermans 2014). Agent-based models of traffic dynamics aim to simulate the collective effects that emerge from individual's driving behaviour.

Simulating a large number of individual agents interacting can be computationally expensive, so traffic and transportation models face a trade-off between complexity and scalability. Small-scale agent-based models are used to study micro-scale behaviours and interactions among individual drivers. These include complex driving behaviour at intersections (Doniec et al. 2008), car following (Hao, Ma, and Xu 2016), lane-changing (Dailisan and Lim 2016), and parking (Benenson, Martens, and Birfir 2008). Despite the fast increase in the computational power available to simulations, models at this level of detail are very difficult to scale to large study areas. Hence, their use tends to be restricted to evaluate the performance of traffic signal, intersection design, and road infrastructure in small sections of the urban road network.

Large-scale, and relatively simpler, agent-based models aim to simulate traffic dynamics of large spatial areas. These models can simulate people's response to changing traffic conditions such as congestion levels and road closures. Techniques developed to enable these models to scale to large urban areas include cellular and queue methods. For instance, the TRANSIMS (Smith et al. 1995) model uses a cellular-based technique, where each cell is approximately the size of one vehicle and driver agents flow through the cells one at a time in the direction and speed of traffic. The queue system is implemented in the popular MATSim model (Horni, Nagel, and Axhausen 2016). In this approach, each street segment is a waiting queue where agents added to the tail and leave in the other end in the same order they got in, and only after an amount of time has passed. The capacity of the queue (maximum number of cars in it at any given time) and the time spent in it is dependent on the length and maximum driving speed of the segment. Although these models represent a simplification of drivers' behaviour, they are highly scalable and allow the simulation of whole cities and countries.

Simulation of passenger trips via public transport is usually included in larger transportation models. Examples of popular models that include multimodal transportation capabilities are TRANSIMS (Serras 2005; Smith et al. 1995), MATSim (Horni, Nagel, and Axhausen 2016) and SIMMOBILITY (Adnan 2015). These models are able to simulate multimodal transportation networks and calculate travel demand for specific modes. Following the aforementioned activity-based approach, individuals make choices regarding mode of transport based on their households' daily activity plan, space and time constraints, as well as car ownership and preferences. These choices affect and are affected by factors such as public transport's quality, extent and reliability.

3.4 Summary

This chapter introduced agent-based modelling as the methodological framework used in this thesis. The characteristics of agent-based models seem to fit well with the requirements of this research, which is to simulate people's trajectories in space and time at the individual level in order to derive their activity spaces. This style of modelling is inherently dynamic, providing a fitting simulation framework for the problem at hand, in contrast with static and aggregated urban models of previous generations (Batty 2008). Furthermore, the heterogeneity of urban populations can be easily represented in agent-based models, as each individual can have different capabilities, constraints and preferences (Crooks and Heppenstall 2012).

Although many existent models can simulate people's trajectories, it was shown in this review that such models are usually embedded into larger transport modelling contexts or have very specific aims. Using or adapting an existing model to study accessibility and segregation from a time geographic perspective would bring an undesirable level of detail and complexity to the process. Hence, this thesis proposes a novel agent-based model to simulate movement of people in an urban environment, and to generate individual-based metrics of accessibility and segregation from the simulated movement patterns. The model is described in the following chapter.

Chapter 4

The AxS Model

This chapter introduces the **AxS Model** (Accessibility × Segregation, pronounced *access*), aimed at studying accessibility and segregation from an individual-based perspective. The AxS Model is based on the theory on accessibility and segregation from a time geographical perspective discussed in chapter 2, and is built using the agent-based modelling technique discussed in chapter 3.

This chapter is organised in three sections. Section 4.1 presents the model's objectives and an overview of its logic and simulation process. Section 4.2 presents the model's components and algorithms in detail. Section 4.3 presents the outputs produced by the model, including dynamic movement patterns and the individual-based accessibility and segregation metrics proposed in this thesis.

4.1 Model Overview

As discussed in chapter 2, individual-based studies of accessibility and segregation are founded on Hägerstrand's (1970) time geographic theoretical framework. Time geography is based on the idea that an individual's trajectory in space and time determines the activities they can participate in, as well as the people they can interact with. This is in line with the classical definitions of accessibility as the "potential of opportunities for interaction" (Hansen 1959, 73), and of segregation as "any restrictions on interactions" among people from different social groups (Freeman 1978, 413), adopted in this thesis.

Two main challenges in applying the time geographic framework to individual-based studies of accessibility and segregation were previously discussed in this thesis. The first challenge is the restricted availability of data on individuals' trajectories in space and time. The second challenge is the computational complexity of representing large populations at the level of detail and disaggregation required for such studies. Those two challenges restrict most individual-based studies of accessibility and segregation to local neighbourhoods and small population samples, making it difficult to scale such studies to entire cities and large populations.

Efforts to overcome the data availability and computational complexity challenges have been reviewed and discussed in chapter 2. Those efforts rely on analytical methods for representing individual activity spaces more efficiently and innovative data collection techniques. Differently from previous studies, this thesis addresses those challenges through the agent-based simulation methodology introduced in chapter 3. The data availability challenge was overcome by artificially generating the unavailable individual data (trajectories in space and time) from readily available aggregated data (such as census origin-destination matrices). The computational complexity challenge was addressed by simulating individuals in an agent-based model, which provided a direct and efficient representation of the real world system the model is representing, which is the city and its inhabitants. Agent-based models also are dynamic, making it easier to represent the time dimension which is an integral part of time geography.

The methodology developed in this thesis was implemented in the AxS model. The model's main goal is to simulate individuals moving through an urban environment between activity locations. During movement, spatial patterns and metrics are produced to allow the study of individuals' access to opportunities and services. Furthermore, patterns of potential interaction among individuals and population groups along their trajectories are visualised and quantified, allowing the study of segregation between groups. Additionally, this thesis aims to make the model easy to operate and scalable to large metropolitan areas, allowing its application in different case studies. That was done mainly by keeping the model's data and computational power requirements low.

The overall simulation process is presented in the diagram of figure 4.1. A brief overview of the *input*, *process*, and *output* stages of the simulation is presented below. The model's components and processes are further detailed in the following section, and the model's source code is available at www.mvpsaraiva.com/thesis.

(a) **Input.** The model requires two sets of input data. The first is a dataset containing the trips to be simulated, which is built outside of the model from aggregated data sources such as origin/destination matrices. The trips dataset is used to create agents during the model's execution, and each agent is responsible for carrying out a single trip. The second input dataset consists of GIS data describing the study area. The GIS input data is used to build the simulation environment during the model's initialisation.

(b) **Process.** During the model's execution, agent's move iteratively towards their destination following a custom pathfinding algorithm to navigate the environment. As agents move, their individual trajectories are plotted on-screen, providing a dynamic visualisation of the population's movement patterns. The model also keeps track of encounters between agents during their trajectories, as well as travel time and distance statistics, which are used to calculate the model's output accessibility and segregation metrics. The process is repeated until all trips from the input dataset are simulated.

(c) **Output.** At the end of the simulation, the model produces two types of outputs. The first type comprises the population's movement patterns, in the form of a series of snapshots or videos. The second type consists of individual-based accessibility and segregation metrics. Those metrics are calculated based on the aforementioned movement patterns, on individual agents' travel time and travel distance statistics, and on the number and diversity of encounters between agents during the simulation.

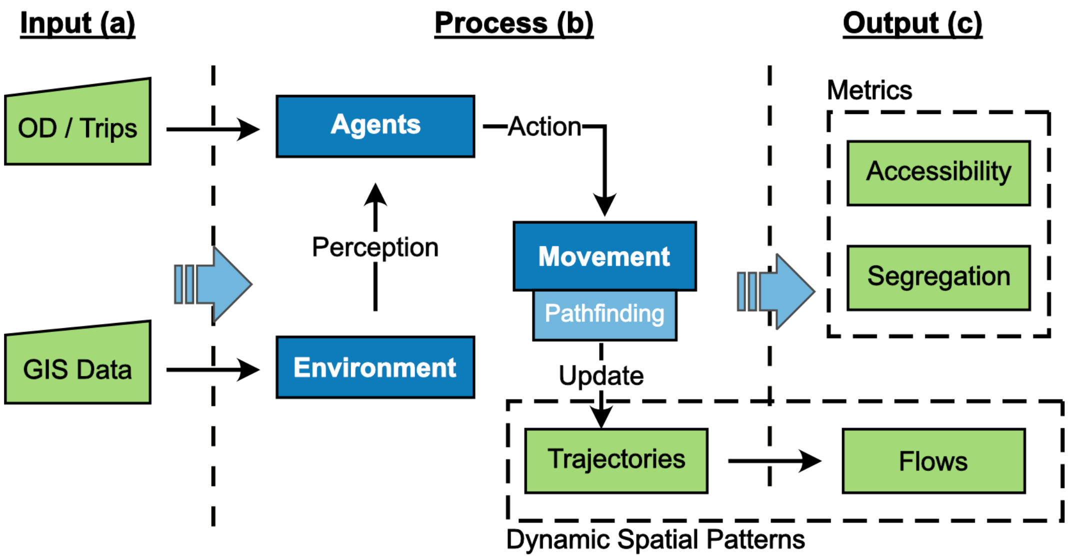

Figure 4.1: AxS model's conceptual diagram.

4.2 Model Details

This section discusses in detail all the components and processes of the AxS model. The model's representation of the urban environment and population is described, followed by the simulation process and pathfinding algorithm. This section also presents the input parameters and datasets required by the model, as well as the model's user interface.

4.2.1 Environment

The model's environment represents the urban area where people live, work and travel through. Most agent-based models of transport and movement (section 3.3.3) use a network-based representation of the urban road system, which allows for a more detailed and accurate representation that system. However, a raster-based environmental representation was chosen for the AxS model due to simplicity and scalability. The difference between the two forms of environmental representation can be seen in figure 4.2.

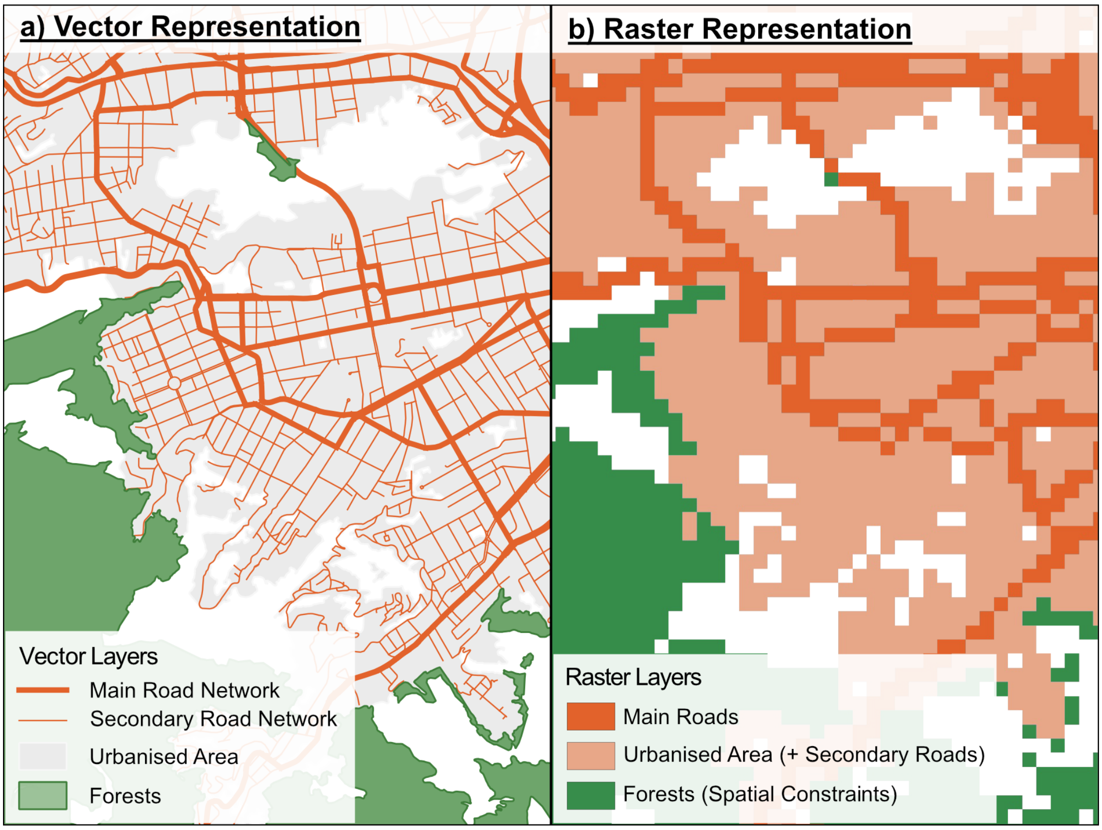

Figure 4.2: Vector versus raster environmental representations.

Representing the urban road network as a grid instead of a network brings important advantages for simulating individual movement at large scales. The raster representation is less resource intensive when compared to a network representation. The difference is even more significant for large metropolitan areas, which can be simulated with the AxS model on standard personal computers. This form of representation is also simpler and easier to build, which is important for areas where the necessary datasets are incomplete or hard to obtain.

The model's environment consists of a regular grid of square cells. Each cell has five input properties (or variables) that are defined at the beginning of the simulation based on the input data. Those properties are: the cell's *urbanisation state*, which is a boolean value indicating whether the cell is urbanised or not; the *code* of the OD zone or census areal unit where the cell is located; the presence of a *main road*; the existence of *spatial constraints* on the cell's location, representing areas agents cannot move through; and, the number of *urban opportunities* (such as commerce, services, and amenities) available in the cell. Each of those properties is read from a separate input raster file during the model's initialisation. The urbanisation state, main roads, and spatial constraints rasters are represented in figure 4.2b.

Movement in a Raster Environment

Simulating movement on a raster environment is significantly different than on a network environment. The main limitation of the raster representation is the lack of the road network's topological structure, which has an important impact on agent's navigation. A specific pathfinding algorithm was developed for the model to account for that, which is detailed in section 4.2.3. An overview of the environment's characteristics that impact agent navigation is presented here, as follows.

All cells in the model's environment are traversable even if they are not represented as the road network, with the exception of cells marked as a spatial constraint. Those spatial constraints represent obstacles such as forests and large bodies of water, which agents cannot access or traverse. However, only the main road network is explicitly represented in the model's environment, while the local road network is assumed to be embedded into the regular urban cells (see figure 4.2). Agents use the main road network during the pathfinding process whenever possible, in search for an efficient route to their destinations. This means agent movement in the model is only facilitated by the main road network, but not restricted to it. Hence, agents are able to move more freely in the AxS model's raster environment than they would in a network environment, where restrictions such as one-way routes can be implemented. Additionally, it is important to note space is continuous in the model, rather than discrete. This means that, regardless of chosen cell size, agents can be located at any point inside a cell, and can move distances smaller than an entire cell in each time step.

The model's environment also lacks an explicit representation of the public transportation system. Features such as the shape of railway networks and bus lines, as well as their access points (train stations and bus stops) and timetables, are not represented. The model uses agents' movement speeds as a simple proxy for different transportation modes. The process of setting movement speeds which are representative of real world transport modes is discussed in the Agents section of this chapter (4.2.2).

During the model's execution, agent's trajectories are dynamically plotted onscreen and the number of agents that pass through each cell is updated. The model keeps track of the total flow of agents through each cell (the cell's *flow* output variable), as well as the flow of agents of different groups (the cell's *flow-by-group* output variable). The flow properties are used to produce the model's spatial patterns visualisation and metrics outputs, discussed in section 4.3.

The simplifications made in the design of the AxS model allow for an efficient representation of urban areas, but arguably reduce the model's accuracy. Hence, those simplifications represent a trade-off between a more scalable model and a more detailed representation of the real world. A discussion of the effects of these design choices in the model's results is presented in chapters 5 (Verification and Sensitivity Analysis) and 6 (Validation).

Spatial and Temporal Scales

The AxS model's spatial and temporal scales are flexible, so the model can be adapted to study areas of varying size and complexity. Spatial and temporal scales are also connected, in order to assure distances travelled by agents in a set amount of time, determined by the agents' movement speeds, accurately replicate real world travel distances, times, and speeds.

Setting the spatial scale is done by choosing grid and cell sizes to appropriately represent the study area. There is no set restriction regarding the extent of the study area, apart from computational limits. While smaller cell sizes allow for a more detailed spatial representation, they also increase the grid size in terms of number of cells and, consequently, increase the computational power required to run the simulation. However, the model's design imposes certain limits to cell sizes. In the model's spatial representation, local roads are assumed to be embedded into regular urban cells, which is not possible with very small cells (such as 100m and below). Conversely, large cells (such as 500m and above) implicate in loss of detail in the study area's spatial representation, possibly leading to issues such as main roads merging with each other. During the model's development, a size of 200m was found to be suitable for simulating very large metropolitan areas with an acceptable level of detail.

The temporal scale refers to the amount of time each iteration in the model corresponds in the real world. Setting the temporal scale is done by choosing a time scale factor, which indicates how much time, in minutes, one model's iteration represents. By default, one iteration in the model corresponds to one minute in the real world, but this factor can be changed depending on each study's requirements. Time is discrete in the model, meaning the chosen time scale factor is the minimum amount of time that can be represented in the model.

Abstract Environments

The model has built in functionality for generating abstract environments, which are useful for testing the model in controlled situations. The abstract scenarios are composed of pre-defined building blocks representing elements of a city, which can be used in any combination. These elements are shown in figure 4.3, where the urban areas are represented in brown and the road network is represented in orange.

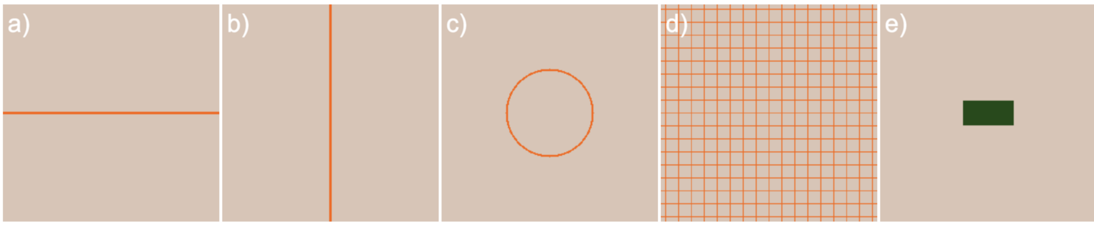

Figure 4.3: Building blocks of abstract environments: a) main horizontal road, 3 cells wide, crossing the centre of the grid; b) main vertical road, 3 cells wide, crossing the centre of the grid; c) main circular road around the city centre, with 50 cells radius; d) regular grid of secondary roads, spaced at 15 cells; e) park in the city centre, 60 cells wide by 30 cells tall, spatial constraint.

Three examples of possible combinations of those elements can be seen in figure 4.4. The size of the abstract study area is 250 by 250 cells. The urbanised area can be set to occupy the whole grid (as in figure 4.4a), or only a circular area of 100 cells radius at the grid's centre (as in figures 4.4b and 4.4c).

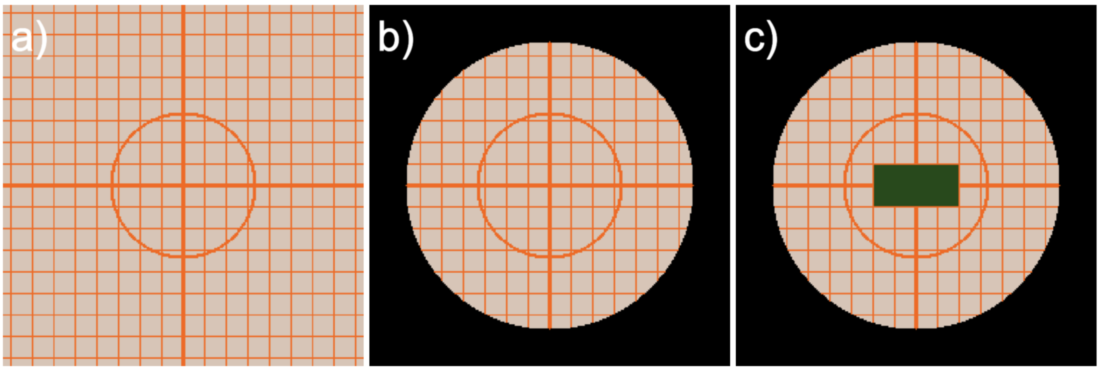

Figure 4.4: Examples of abstract environments built into the AxS model.

These options of abstract environments are built into the model for convenience, and were used for the verification and sensitivity tests presented in chapter 5. More sophisticated abstract environments can be created outside the model using any GIS tool, and then imported into the AxS model as GIS data.

4.2.2 Agents

Agents, in the model, represent people who move through an urban environment to reach desired activity locations. The model does not have specific restrictions regarding the purpose of their travel, so agents can represent people commuting to work, children going to school, shoppers travelling to commercial areas, or individuals doing multi-purpose trips combining some of those objectives. Each agent can be assigned a group according to the study area's demographic characteristics, such as race, income, or gender. Different transport modes are also supported, although through a simplified movement speed proxy, which is further detailed later in this section. Those decisions on travel purpose, travel mode, and definition of population groups depend on the objectives of each study and the available data.

The agent population in the simulation can match the entire population of the study area or, most usually, a sample of it. There are no specific requirements regarding the size of the population sample to be input into the model. As a general rule, the sample must be large enough to accurately represent the characteristics of the study area. Those characteristics include the distribution of origins and destinations, the proportion of each group in the area's population, as well as the proportion of travellers using each transport mode. If the sample is too small, smaller groups and sparsely populated areas may be underrepresented. Conversely, a large sample requires additional computing power to run, which can increase run times significantly and unnecessarily. The effects of the number of agents in the simulation are explored in the verification and sensitivity analysis tests in chapter 5.

Agents have two sets of state variables: *input* properties, and *output* variables. Agents' input properties are set during the model initialisation based on the trips dataset. Agents' *group*, *transport mode*, *origin zone* and *destination zone* are obtained directly from the trips dataset. The value of the *speed* property is assigned by matching each agent's *transport mode* to the appropriate speed parameter value (*speed pedestrian*, *speed bicycle*, *speed bus*, *speed car*, *speed motorcycle*, and *speed train*). Specific origin and destination cells are also assigned to the agents by the model. This is done by randomly selecting one of the cells with matching origin and destination zone codes and assigning those to each agent as origin and destination of its trip. Those properties can be seen in table 4.1.

Basic travel statistics are calculated at the end of each agent's trip and saved as agents' output variables. The iteration the agent was created is stored in the *starting iteration* variable, and the agent's total *travel time* is measured by the number of iterations elapsed since then. The *travel distance* is measured by the number of cells (or fractions of cells) traversed by each agent during its trip. The *net time budget* is the amount of time the agent has to participate in discretionary activities. It is calculated by subtracting the agent's *travel time* from the *gross time budget* parameter (which is the same for all agents). Those statistics are used to calculate the agents' accessibility metrics, detailed later in this chapter (section 4.3.2). The agents' output variables can be seen in table 4.2.

Table 4.1: Agent's input properties.

Variable Name	Description	Possible Values	Units / Data type
group	Numeric identifier of agent's population group	1 to 20	Integer
transport-mode	Mode of transport	pedestrian, bicycle, bus, car, motorcycle, train	String
speed	Movement speed	0.1 to 10	Cells per Iteration
origin-zone	Origin zone code	Any in the list of OD zones	String
origin-cell	Starting point of the trip	Set of cells in origin zone	Cells
destination-zone	Destination zone code	Any in the list of OD zones	String
destination-cell	Ending point of the trip	Set of cells in destination zone	Cells

Table 4.2: Agent's output variables.

Variable Name	Description	Possible Values	Units / Data type
starting-iteration	Iteration when the agent was created	1 to simulation time	Integer
travel-time	Time since the start of the trip	1 to simulation time	Integer
net-time-budget	Time budget after travel time is discounted	0 to 120	Iterations
travel-distance	Distance travelled by the agent	1 to grid size	Floating point

Transport Modes and Movement Speed

As previously mentioned in the Environment section of this chapter (4.2.1), transport modes are not explicitly represented in the AxS model. Instead, agents' movement speeds are used as a simple proxy for different modes of transport. Agents' movement speed depend on the model's spatial and temporal scales chosen for the simulation. Hence, real world movement speeds, in kilometres per hour (km/h), must be converted into model speed units (cells per iteration, cpi), according to chosen *cell size* (spatial scale) and *time scale factor* (temporal scale). Speed in cpi indicates how many cells (or fractions of cells) an agent can travel at a single unit of time.

The conversion of speed from real world to model units is done according to equation 4.1, where s_{cpi} is the model speed in cells per iteration, s_{kmh} is the speed in km/h, cs is the cell size and ts is the time scale factor.

$$s_{cpi} = \left[\frac{1000 \times s_{kmh}}{60 \times cs}\right] \times ts \quad (4.1)$$

As an example, a normal walking speed of 5 km/h was converted to cells per iteration based on different cell sizes (50m, 100m, 200m, and 500m) and time scale factors (1 min, 5 min, and 10 min), and the results can be seen in table 4.3.

Table 4.3: Comparison between movement speeds in real world units (*kilometres per hour*, *km/h*) and model units (*cells per iteration*, *cpi*).

Walking Speed	5 *km/h*		
	time scale (minutes)		
cell size (m)	1	5	10
50	1.7	8.3	16.7
100	0.8	4.2	8.3
200	0.4	2.1	4.2
500	0.2	0.8	1.7
	speed (*cpi*)		

Travelling speeds for each transport mode must be estimated according to the study area, accounting for factors such as quality of infrastructure, reliability and frequency of public transportation options, and levels of congestion. Accurate travelling speeds are key to the accuracy of the model's results, particularly for time sensitive outputs such as accessibility metrics and travel time statistics.

Once transport modes' travelling speeds are calculated for the study area, they are input into the model through a set of mode-specific parameters (*speed pedestrian*, *speed bicycle*, *speed bus*, *speed car*, *speed motorcycle*, and *speed train*). Each agent is assigned a movement speed at the start of their journey, according to their transport mode and corresponding input parameter value.

4.2.3 Simulation Process

The AxS model's simulation process is illustrated by the flowchart presented in figure 4.5. The model starts the simulation by reading the input trips dataset, which contains the population sample to be used in the experiment. Since loading the entire population at once would require too many computational resources, the *number of agents* parameter is used to control the maximum number of active agents in the simulation at any given time. Hence, the model gradually generates agents, assigning them trips from the input dataset, until the limit imposed by the *number of agents* parameter is reached. After that point, the model waits for current agents to complete their trips before resuming generating agents from the population sample.

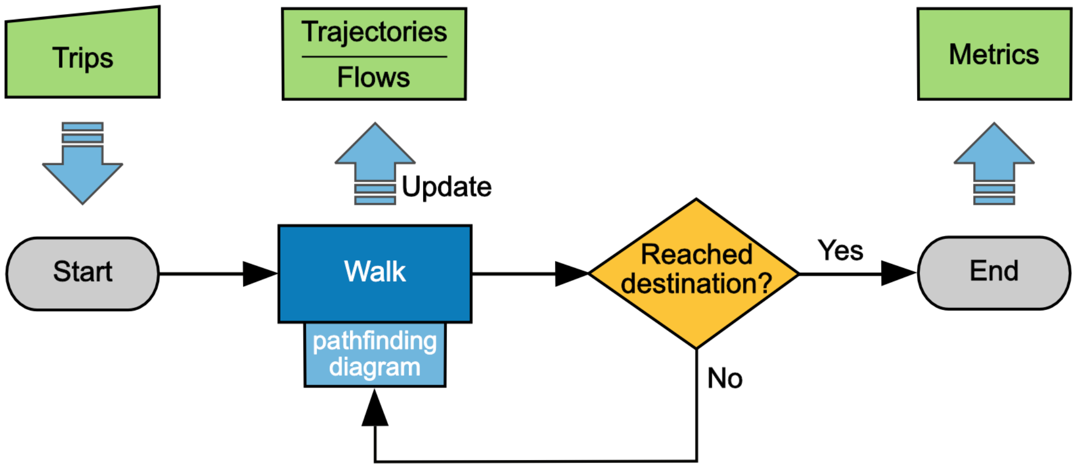

Figure 4.5: Flowchart of the AxS model's simulation process.

Agents move iteratively towards their destinations, following the pathfinding algorithm detailed later in this section. During movement, agents' trajectories are dynamically generated and visualised onscreen. Agents keep track of travel statistics and encounters with other agents during their trajectories, producing data that will be used to calculate accessibility and segregation metrics at the end of the simulation. Each agent that reaches its destination is removed from the simulation and replaced by a new one, who carries out another trip. The model is executed until the entire list of trips is read and all trajectories are simulated. At that point, the output accessibility and segregation metrics, as well as aggregated flow patterns, are produced and can be exported for further analysis.

Pathfinding Algorithm

A custom, raster-based, pathfinding algorithm was developed for the AxS model. This algorithm draws on concepts from the literature on human cognition in order to emulate, in a very simplified manner, human decision-making when navigating through an environment. Cognition-based route finding is a relatively novel development in the fields of transport modelling and traffic simulation, with recent advances on network-based algorithms (Kazagli, Bierlaire, and Flötteröd 2016; Manley and Cheng 2018). However, since the AxS model's simplified raster environment does not support the use of network-based methods, a tailored pathfinding approach had to be developed for the model, as detailed below.

The literature on human cognition and navigation process indicates that route finding is an imperfect process that is highly dependent on individual characteristics. Individuals have their personal mental representations of space (Lynch 1960), due to the way the human brain stores and organises spatial information. As suggested by Downs and Stea (1973, 315), such mental representations are "incomplete, distorted, schematised and augmented", and present "both group similarities and idiosyncratic individual differences". As individuals' spatial knowledge is incomplete and imperfect, people are "unlikely and unable to select an optimal route between origin and destination" (Manley, Addison, and Cheng 2015, 124).

The literature also presents evidence that individual spatial knowledge tends to be hierarchical, where major urban features such as main roads and landmarks are likely to be known by more people than local roads and regular buildings (Couclelis et al. 1987; Kuipers, Tecuci, and Stankiewicz 2003; Jiang and Jia 2011). It is understood that the navigation process is made stepwise (Lynch 1960; Couclelis et al. 1987), and decisions are made at important locations along the way such as major street intersections and well-recognised landmarks. This suggests that the hierarchy of the urban landscape has a significant effect on people's navigation decisions. As each decision has a cascading effect on the remainder of the trip (Manley, Addison, and Cheng 2015), the longer the route the more unpredictable it tends to be.

The AxS pathfinding algorithm is based on the evidence discussed above: the **hierarchical** nature of spatial decisions in regard to the street network, the **incompleteness** and **imperfection** of individuals' spatial knowledge, and the **stepwise** nature of the navigation decision-making process. Those principles suggest that it is highly unlikely that individuals always use optimal routes. Thus, in the AxS pathfinding algorithm, agents have a limited perception of the environment, indicated by their fields of view, which represents the incompleteness of human spatial knowledge. Agents need to estimate the current distance to the destination and the cost of each possible path, hence they are unable to find and follow the optimal path towards the destination. No complete path is planned in advance of the trip. Rather, the pathfinding process occurs iteratively during the entire walk. Finally, the rationale for the agents' movement behaviour is to minimise the cost of moving through the environment. This is achieved by giving preference to movement through the main road network whenever possible.

Those principles were the basis for the development of a custom version of the A* pathfinding algorithm (Hart, Nilsson, and Raphael 1968), which is a graph-based algorithm that aims to find the shortest path between two nodes of a graph. A* adds an heuristic function to the Dijkstra's (1959) algorithm, in which nodes closer to the destination are checked first, reducing the number of nodes that need to be checked before the calculation is completed. The AxS model's algorithm adapted A* to a raster environment and adopted a stepwise approach to the calculation. Hence, calculation is carried out in stages for each stretch within the agent's current field of vision. In addition, randomness is added at each step, replicating the uncertainties of the navigation process.

Figure 4.6 details the full AxS pathfinding algorithm and illustrates each of the algorithm's four steps.

- **Initial State** (figure 4.6a): at the beginning of the pathfinding process, the agent is created at the origin of its trip and is assigned a destination cell.

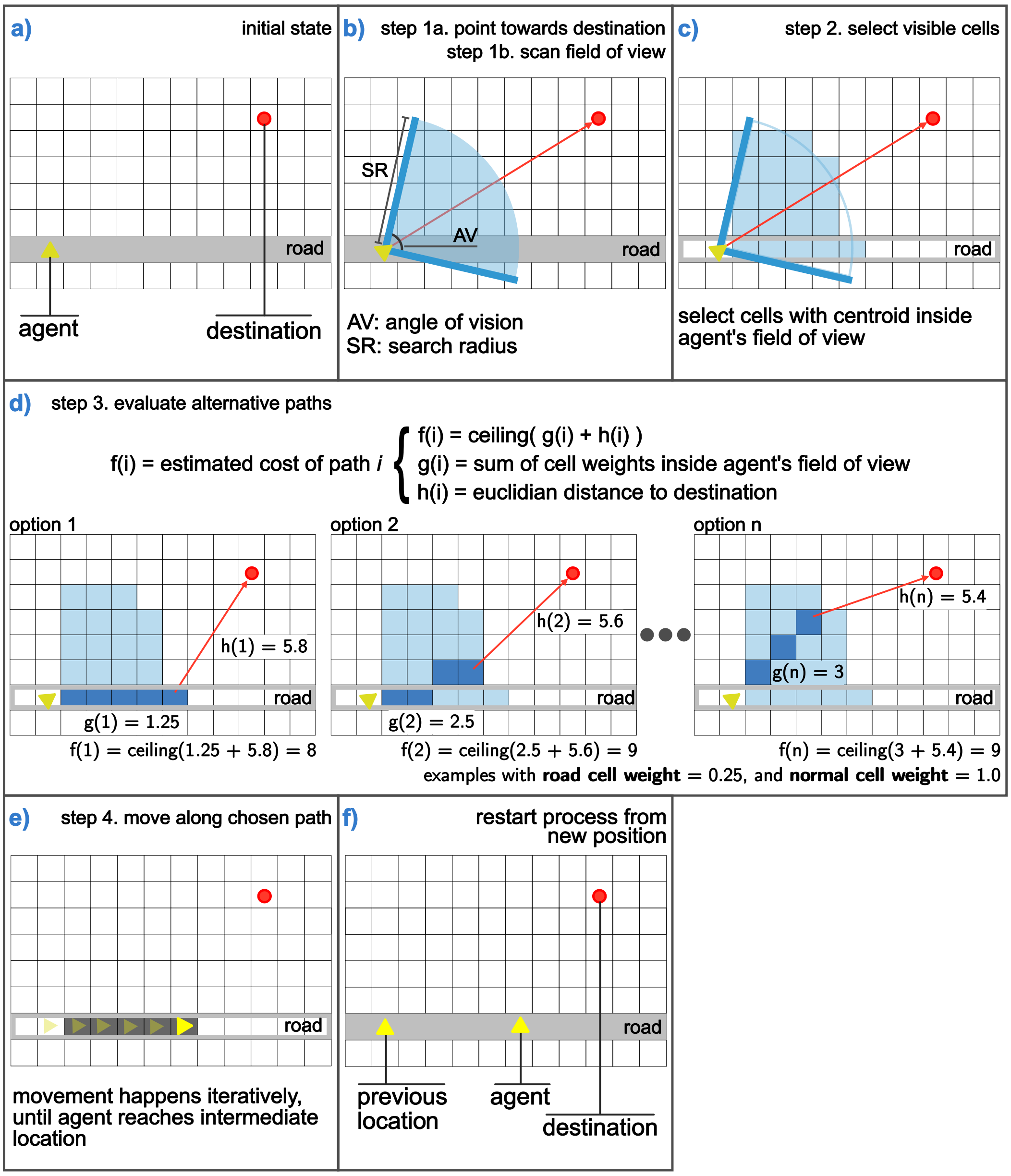

Figure 4.6: Scheme of pathfinding algorithm.

- **Step 1** (figure 4.6b): the agent points towards the destination cell and scans the neighbourhood inside its field of view. The agent's field of view consists of the cone of cells directly in front of the agent, as defined by the *angle of vision* and *search radius* parameters.

- **Step 2** (figure 4.6c): the agent selects the visible cells, which are the ones whose centroids are located inside the agent's field of view (represented in light blue).

- **Step 3** (figure 4.6d): the agent evaluates the alternative subpaths inside the set of visible cells previously selected, randomly choosing one of the least cost subpaths to follow.

- **Step 4** (figure 4.6e): the agent moves along the subpath selected in step 3.

Step 4 may take several iterations to complete, depending on the length of the subpath and the agent's movement speed.

Once the agent reaches the end of the current subpath (figure 4.6f), the four-step pathfinding process is restarted from the agent's current location. The pathfinding and movement processes are repeated until the agent reaches the destination cell.

The most complex stage of the pathfinding process is step 3, which is further detailed here. In this step, agents evaluate the set of subpaths currently visible to them, in order to chose one of those subpaths to follow. The subpaths are evaluated through equation 4.2. In that equation, the estimated total cost $f(i)$ of subpath i is the sum of the costs of the known $g(i)$ and unknown $h(i)$ segments of the total path, rounded up to the next integer.

$$f(i) = ceiling(g(i) + h(i)) \tag{4.2}$$

The known segment of subpath i contains the cells inside the agent's field of view, and its cost $g(i)$ is determined by the sum of the weights W of all cells c in the segment S, as per equation 4.3. The weight of each cell represents the cost of moving through the cell. Urban cells have $weight = 1$, while road network cells' weight is set by the *road weight* parameter. In the example presented in figure 4.6, $road\,weight = 0.25$.

$$g(i) = \sum_{c \in S} W(c) \tag{4.3}$$

The cost $h(i)$ of the unknown segment of subpath i is estimated by the agent using an Euclidean distance heuristic. As per equation 4.4, the estimated cost $h(i)$ is the straight-line distance $d()$ between the centroid of the n^{th} (last) cell of subpath i and the destination cell D.

$$h(i) = d(i_n, D) \tag{4.4}$$

The reasons for rounding up the subpath's estimated cost $f(i)$ are twofold. First, rounding the estimated cost represents the uncertainty and lack of precision in the distance evaluation process made by humans. Second, rounding allows for subpaths with similar costs to be considered as equivalent alternatives for moving forward, in recognition that minor differences are not considered when people navigate through urban environments in reality. At the end of step 3, the agent assesses all subpaths and randomly chooses one of the least cost subpaths to follow. Through this process, agents can take alternative path choices even when origins and destinations are similar.

Occasionally, agents get blocked by spatial constraints along their path and are unable to move forward. Under those circumstances, agents follow an *extended search* procedure trying to circumvent the obstacle, illustrated in the flowchart of figure 4.7. During the extended search, the agent restarts the pathfinding process from step 1 in the next iteration, but with their *angle of vision* increased to 360º, and their *search radius* set to the double of the original value. If a path is found in those conditions, the agent follows that path and continues navigating following the normal pathfinding procedure. However, if the agent is unable to find an unobstructed path even under those relaxed conditions, the agent is removed from the simulation and their attempt is recorded as a failure in the model's output statistics.

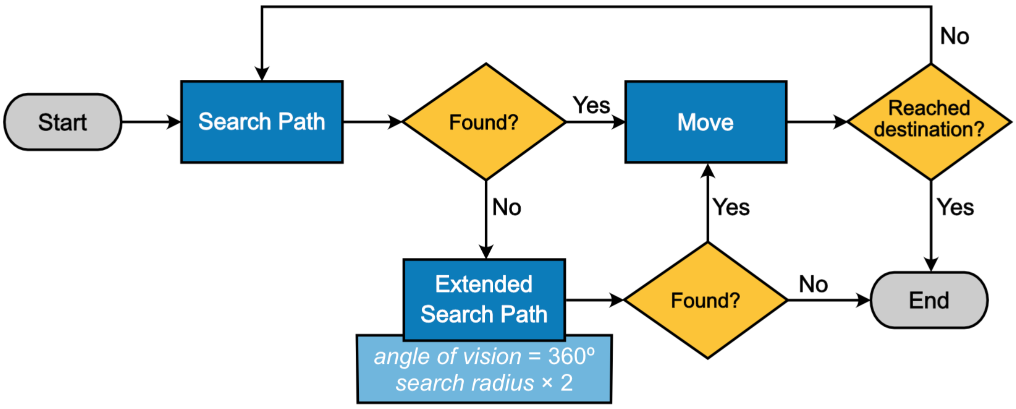

Figure 4.7: Flowchart of agent's pathfinding when facing obstacles.

4.2.4 Input Parameters

The model's can be controlled via a set of configurable parameters available on the program's user interface. Some of these parameters were already mentioned in the previous sections, when relevant. For the sake of completion, parameters discussed in previous sections will be listed here as well, and additional information will be included when relevant.

The *number of agents* parameter defines the number of simultaneously active agents in the simulation, which is kept constant during the simulation process until the entire trips dataset is read through. Changing the number of agents mainly affects the model's performance, as more active agents require more computational power to simulate. Another effect of increasing this parameter is the resulting aggregated flow patterns form faster, as a larger number of trips are simulated at once. The effect of the number of agents in the model's results will be explored further in the verification and sensitivity analysis tests presented in chapter 5.

The simulation flow is controlled through three parameters, as follows. The *stop condition* defines if the simulation should stop at a specific iteration (defined by the *stop at iteration* parameter), after a set number of agents is simulated (defined by the *stop at agents* parameter), or after the whole input trips dataset is simulated.

Two parameters control agents' navigation behaviour: *search radius* and *angle of vision*. Those parameters, combined, define the area of the environment that the agent can perceive during the pathfinding process. In an abstract way, *search radius* and *angle of vision* represent the agents' partial knowledge of the environment.

A set of parameters control the agents' movement speed per transport mode: *speed pedestrian*, *speed bicycle*, *speed bus*, *speed car*, *speed motorcycle*, and *speed train*. Speeds are input in model speed units: cells per iteration. The value of the generic *speed* parameter is attributed to the agents' movement speed when no transport mode information is available. The *speed* parameter is also useful for simulations in abstract environments, and for verification and sensitivity analysis purposes.

The *road weight* parameter defines how preferable (or least costly) it is for the agents to move through the main road network than through the local road network.

The parameter *calculate accessibility* is used to activate or deactivate the calculation of accessibility measures by the model. Similarly, *calculate copresence* parameter defines if the model should check for encounters between agents during the simulation and calculate the output segregation metrics. These parameters exist mainly for performance reasons, as deactivating those procedures when they are not necessary significantly reduces the model's run times.

Finally, two parameters adjust the export settings. Results are exported automatically at the end of the simulation when *export on finishing = true*. Similarly, when *export video = true*, the model creates and exports an animation of the simulation process by capturing every iteration into a frame.

4.2.5 Input Data

The AxS model requires two groups of data as input: *GIS data*, for building the model's environment; and *trips* data, containing origins and destinations of trips. Those data requirements were already briefly mentioned in this chapter, and are further detailed here.

The set of *GIS* input data is composed of raster and vector files containing information about the study area, as described in table 4.4. Datasets must be prepared using any GIS software package before being input into the model, following two basic requirements so that they can be correctly loaded: all data must use the same projection system; and raster files must have the same geographical extent and cell size, matching the spatial scale defined for the study area.

Table 4.4: GIS input data.

File	Description
urban_areas.asc	Raster file in the ASCII grid format where urban cells have value 1 and non-urban cells have value 0.
outline.shp	Optional polygon shapefile containing the boundary of the study area, used only for visualisation.
road_network.asc	ASCII grid file containing the main road network of the study area.
od_zones.asc or od_zones.shp	Polygon shapefile or ASCII grid file containing the OD zones of the study area.
od_codes.csv	CSV file containing the codes of the OD zones.

The *trips dataset* must be built outside of the model from aggregated OD data, and then input during the simulation setup. Common sources of aggregated OD data are national censuses and travel behaviour surveys conducted by transport and urban planning bodies. The set of trips of a study area, or a sample of the entire set, must be extracted from the OD matrix and converted into a list containing the origin and destination zones of each trip, alongside the agent's group and transport mode. This conversion from a matrix into a list is relatively straightforward and can be done using freely available statistical packages such as the R statistical programming language (R Core Team 2016) used in the case studies of this thesis. The trips dataset consists of a single table, in CSV file format. The structure of the trips table can be seen in table 4.5.

Table 4.5: Trips input data.

Field	Description	Data Type
origin_zone	Code of trip's origin zone.	String
destination_zone	Code of trip's destination zone.	String
group	Agent's group identifier.	Integer
transport_mode	Agent's mode of transport.	Float

4.2.6 User Interface

The model was developed using the NetLogo agent-based modelling environment (Wilensky 1999). NetLogo is a simple, yet powerful, flexible, and user friendly agent-based modelling environment. It has an extension that allows access to GIS data, making it a suitable tool for geographic agent-based modelling. NetLogo makes it possible to build simple user interfaces for input and control of the model.

It also provides tools to build plots and visual outputs to monitor the model's behaviour. Most importantly for modelling geographical phenomena, NetLogo provides a spatially explicit environment which can be visualised dynamically on the main window.

The AxS model's user interface can be seen in figure 4.8. The study area is shown at the centre of the user interface, and it is updated dynamically as the simulation unfolds. The input parameters and simulation controls are located on the left-hand side of the interface. The controls for building abstract environments and running model verification, sensitivity analysis, and validation tests are located on the right-hand side of the interface.

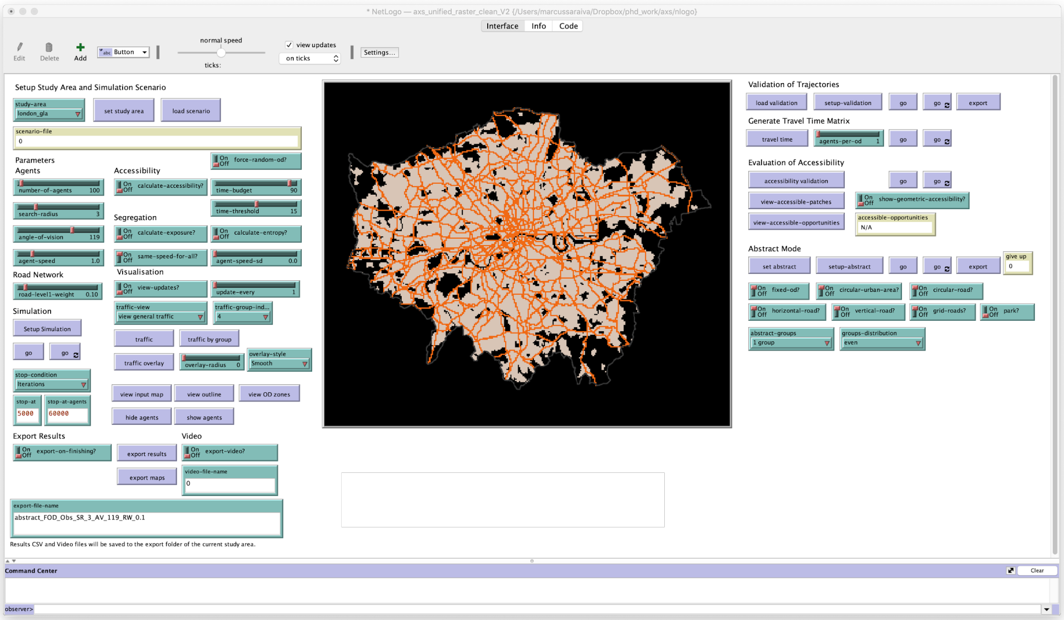

Figure 4.8: AxS Model interface in NetLogo.

4.3 Output Spatial Patterns and Metrics

The AxS model produces a series of outputs that allow the study of accessibility and segregation in cities from an individual-based and dynamic perspective. The outputs are of two types. The first type comprises spatial patterns of movement of individuals and groups. The second type includes accessibility and segregation metrics based on those movement patterns, which were specifically created or adapted from the literature for this study. Those outputs and metrics are detailed in this section.

4.3.1 Movement Patterns and Flow Metrics

The simulated trajectories of many individual agents can be visualised dynamically in the model's graphical interface, providing a rich visualisation of large scale movement patterns of individuals and groups in cities. In the context of this study, those spatial patterns represent the collective activity spaces of groups of individuals, and are also referred to as aggregated flows. Such patterns are useful to provide an overall view of the areas of the city where individuals of each group travel through or visit more often and, thus, are more likely to interact with each other and with the local environment.

The model produces aggregated flow patterns by keeping track of the number of agents that move through each cell, as well as the population group they belong to, during the course of the simulation. Those metrics are stored in the cells' *flow* and *flow-by-group* variables, and are updated dynamically every time an agent steps onto a cell. Visual representations of those patterns can be exported both as static images and as animations, allowing the model's results to be communicated more effectively to the research community and stakeholders. Those values can be exported as raster files as well, that can be open in GIS software for further analysis and visualisation.

Examples of flow patterns produced by the model at different time steps can be seen as a series of snapshots in figure 4.9. Those maps were produced for the city of São Paulo, in Brazil. They show aggregated flows of the entire population (column a), as well as aggregated flows per income class (high income and low income individuals in columns b and c, respectively). The different patterns produced by each group are strikingly distinct, showing significant differences in the patterns of occupation of the urban environment by each group. Specifically, it is noticeable how high income individuals concentrate in the city centre, while low income individuals travel mostly through São Paulo's peripheral areas.

As time passes, it can be seen the patterns gradually consolidate, demonstrating flow patterns emerge from the bottom up as many individuals complete their assigned trips. As expected, patterns produced by larger populations consolidate faster than the ones produced by smaller groups of individuals. For example, flow maps of the entire population (column a) consolidate faster than the maps of separate groups. The formation of these patterns will be further explored in the sensitivity analysis tests of chapter 5.

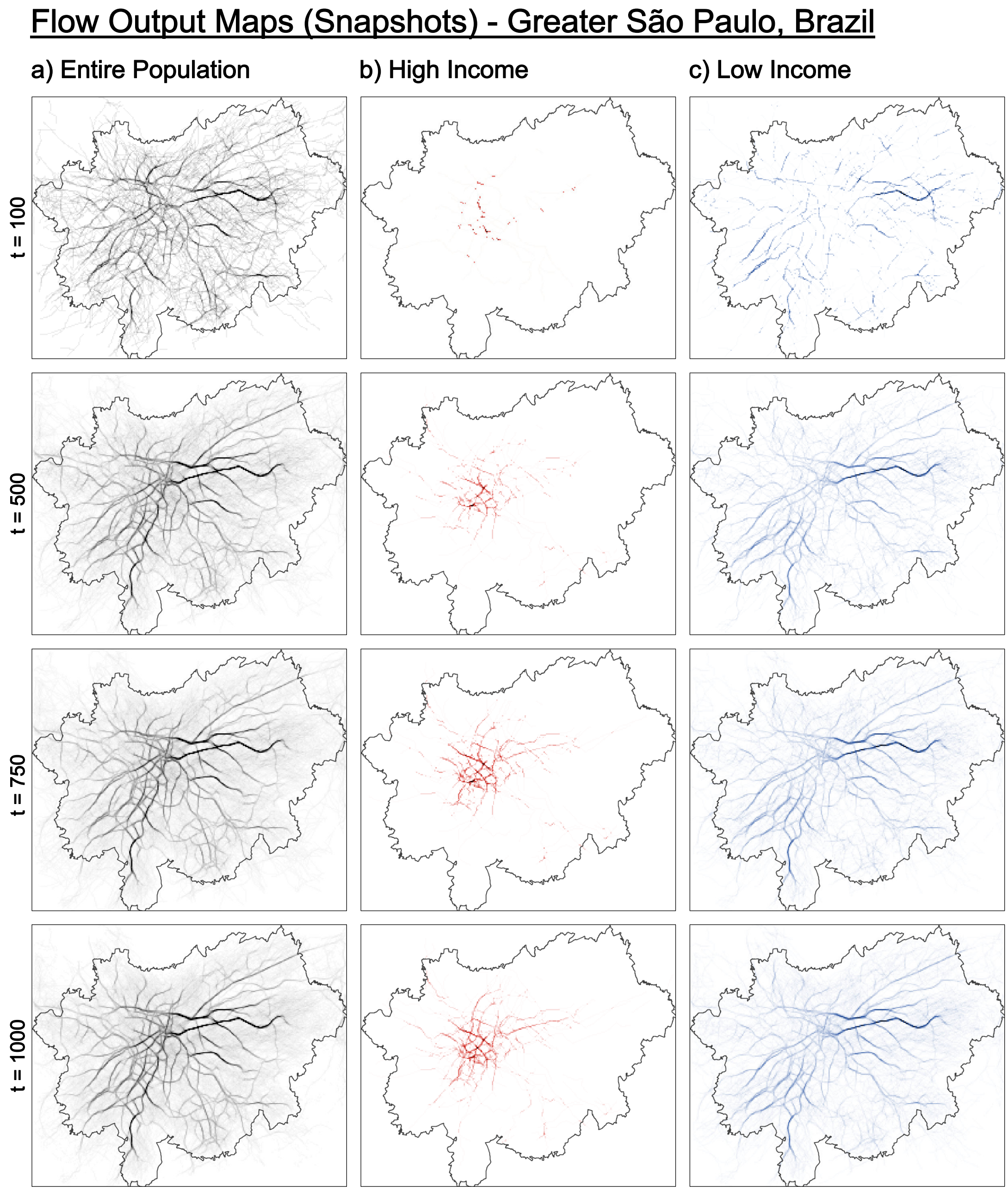

Figure 4.9: Sample of the AxS model's flow output rasters for the São Paulo and adjacent municipalities, in Brazil, by income groups: a) flows of the entire population; b) flows of high income individuals; c) flows of low income individuals.

4.3.2 Accessibility Metrics

The individual-based accessibility metrics proposed in this thesis are calculated based on agents' simulated trajectories and travel times. A visual representation of the logic behind those metrics is shown in figure 4.10. The figure illustrates the hypothetical commuting trajectories of two individuals (*blue* and *red*), whose home and work locations are marked as H and W in the map. At specific times and following their own schedules, individuals *blue* and *red* travel from home to work following their chosen paths, marked on the map as the blue and red lines connecting those locations.

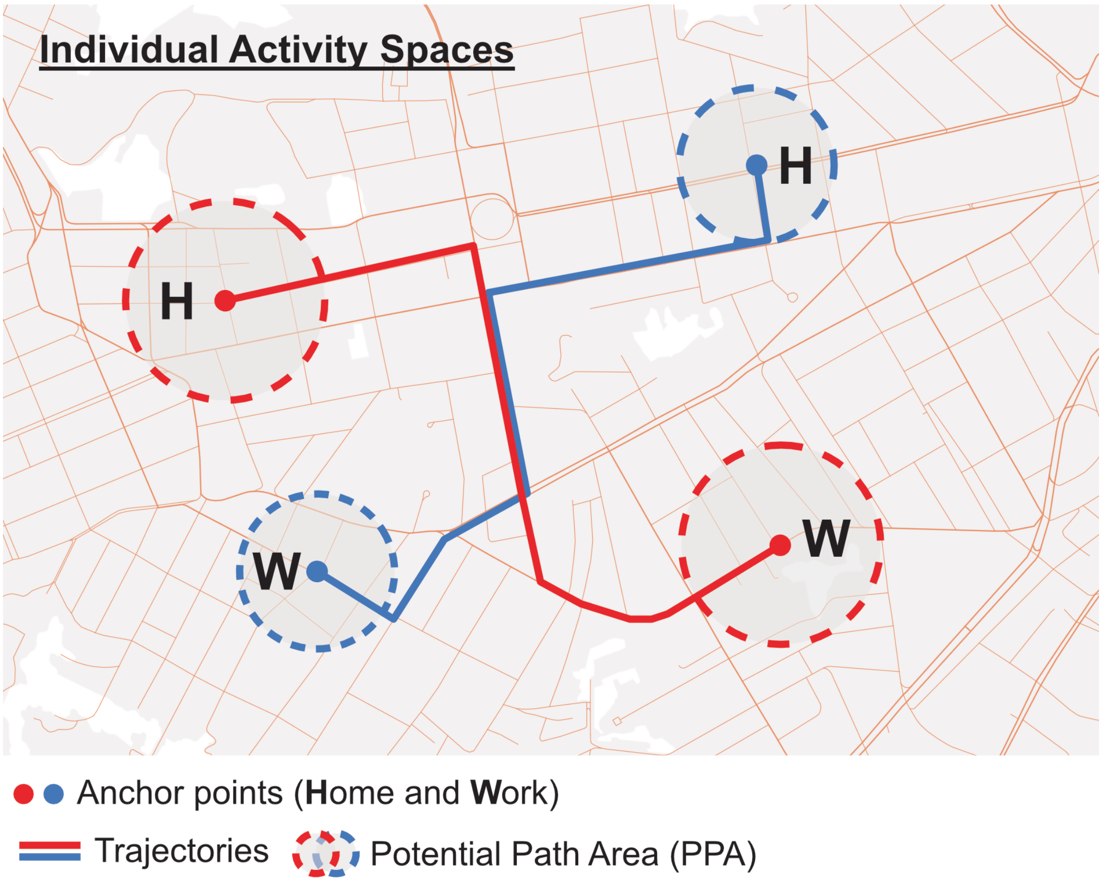

Figure 4.10: Hypothetical activity spaces of two individuals.

Once the main activities carried out at home or work are completed, the individual can use their free time to travel to other locations in order to participate on different activities. Each individual has a *time budget*, which is a set amount of time that can be used to participate on any discretionary activity. Time spent commuting is discounted from the individual's time budget, so individuals with longer travel distances and/or slower travel speeds have less available time for participating on other activities. The area the individuals can reach on their free time, considering their specific time budgets and movement speeds, is the Potential Path Area (PPA). In this study, for simplicity, the PPA is delimited by circular areas around the individuals' home and work locations. Hence, the PPA indicates the maximum area an individual could theoretically reach according to their speed, but without considering the street network's morphology. The two hypothetical individuals' PPAs are represented by the blue and red dashed circles in figure 4.10. Based on this process, mobility and accessibility metrics are calculated, as follows.

Mobility Metrics

Simple mobility metrics, such as travel time and distance, are useful for studying individual travel behaviour, and also serve as input for calculating accessibility metrics. The total travel time tt_a of agent a is calculated by subtracting the agent's departure time t_i from the agent's arrival time t_f, as per equation 4.5. Since departure and arrival times are measured in *iterations*, travel times must be converted to real world time units according to the temporal scale ts defined during model setup. The temporal scale indicates how much time a model iteration represents, in minutes.

$$tt_a = (t_f - t_i) \times ts \tag{4.5}$$

The distance travelled d_a of agent a is defined by the number of cells c the agent has traversed during its trajectory, as per equation 4.5. Similarly to travel time, travel distance must be converted to real world units according to the spatial scale defined during the model setup, which is done by multiplying the number of cells traversed by the cell size cs, in metres.

$$d_a = count(c, a) \times cs \tag{4.6}$$

Potential Path Area

The first step to calculate individual accessibility is defining the agent's PPA. The radius r_a of the agent's circular PPA (figure 4.10) is defined by the agent's available time for participating in discretionary activities (net time budget - tb_n), and the agent's movement speed s_a, as per equation 4.7.

$$r_a = \frac{tb_n \times s_a}{2} \tag{4.7}$$

The net time budget tb_n used in equation 4.7 is calculated by subtracting the agent's travel time tt_a and the minimum time needed to meaningfully engage in an activity t_m from the gross time budget tb_g, as per equation 4.8. While travel time is calculated individually per agent, t_m and tb_g are exogenous parameters that are input by the model's user and applied uniformly to all agents.

$$tb_n = tb_g - tt_a - t_m \tag{4.8}$$

All cells and opportunities located within the PPA radius from the origin and destination points of the agent's trip are included in the agent's PPA set. The measures of geometric and cardinal accessibility derived from those sets and employed by the model are detailed below.

Geometric and Cardinal Accessibility Measures

The geometric and cardinal accessibility metrics and equations proposed here are inspired by Lenntorp's (1976) work, and were adapted to raster environment in this thesis.

The geometric accessibility A_a^{geo} of agent a is calculated by equations 4.9 and 4.10. In those equations, $I(c)$ is a square function that identifies wether cell c belongs to the agents's PPA set or not. Geometric accessibility is simply a measure of the size of an individual's PPA, which is represented by the number of cells an agent can access in the model's environment given the agent's mobility and time constraints.

$$A^{geo} = \sum_{c \in C} I(c) \tag{4.9}$$

$$I(c) = \begin{cases} 1, & \text{if cell } c \in \text{PPA} \\ 0, & \text{otherwise} \end{cases} \tag{4.10}$$

Similarly, the cardinal accessibility A_a^{card} of agent a is calculated by equations 4.11 and 4.12. In those equations, $W(o)$ is a square function that identifies if opportunity o belongs to the agents's PPA set. Cardinal accessibility is a measure of the number of opportunities and individual can access given their mobility and time constraints.

$$A^{card} = \sum_{o \in O} W(o) \tag{4.11}$$

$$W(o) = \begin{cases} 1, & \text{if opportunity } o \in \text{PPA} \\ 0, & \text{otherwise} \end{cases} \tag{4.12}$$

The main difference between cardinal and geometric accessibility is that the cardinal measure considers the distribution of opportunities in the study area as well as individual mobility patterns. In this sense, an individual with low mobility will necessarily have low geometric accessibility, but they can have high cardinal accessibility depending on the density of opportunities in the areas they visit.

4.3.3 Segregation Metrics

The segregation metrics proposed in this thesis are based on the aggregated trajectories of groups of individual agents, which are here referred to as collective activity spaces. Figure 4.11 shows trajectories from home to work of a sample of individuals belonging to hypothetical groups *Red* and *Blue*. In real-world studies, individuals can be grouped according to socioeconomic characteristics such as income, ethnicity, or level of education.

The activity and movement patterns of many individuals, such as the ones illustrated in figure 4.11, highlight sections of the urban space that are shared among many groups, as well as areas where a single group is predominant. Those areas are marked on the map as single-colour dashed circles, representing encounters and possible interactions between individuals of the same group, and dual-colour dashed circles representing possible interactions between individuals of both groups.

Two types of segregation measures are proposed in this thesis, based on the groups' collective activity spaces: measures of *diversity* at street level, and measures of *copresence* in space and time. Those measures are detailed in this section.

Diversity

This thesis proposes adapted versions of Theil's (1971) *information theory index H* as indicators of diversity at street level. In this study, Theil's index H is calculated on the AxS model's raster environment, quantifying the diversity of agents that moved through each cell based on the cell's aggregated flow counts.

The information theory index is based on *entropy*, which is a measure of

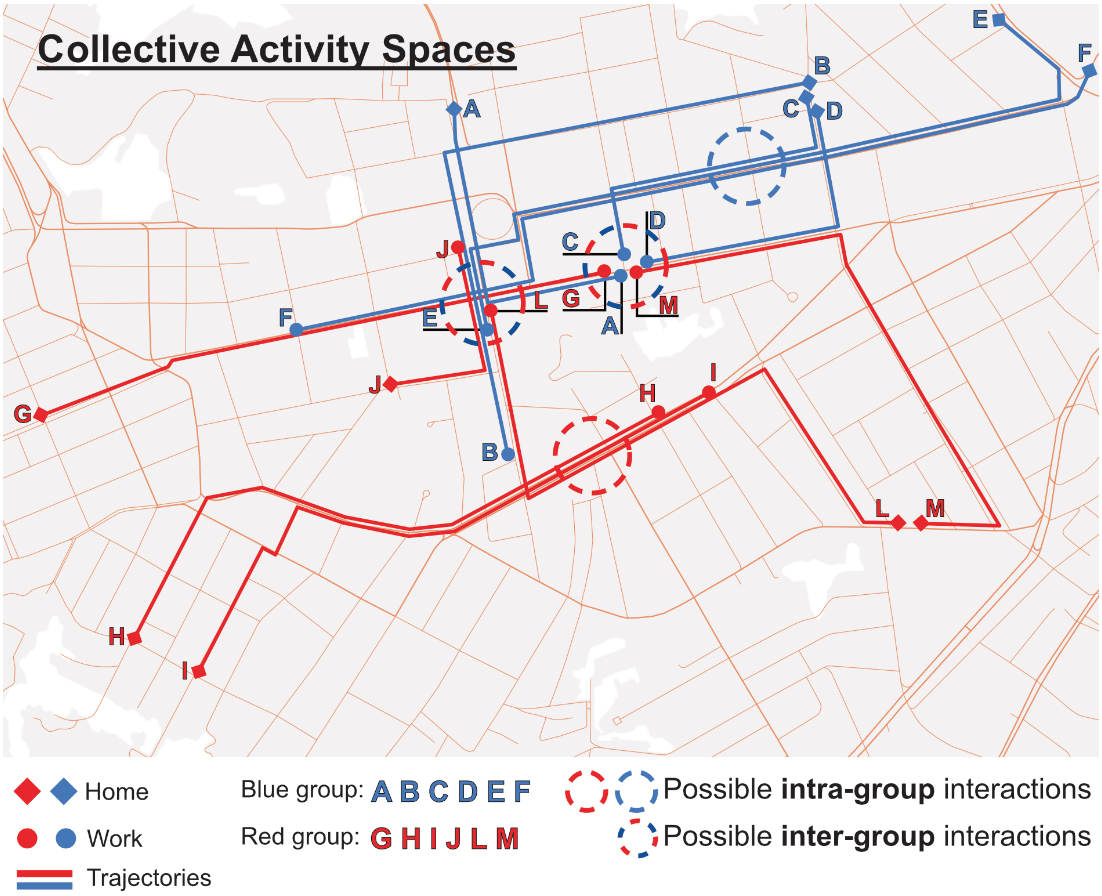

Figure 4.11: Hypothetical activity spaces of two groups of individuals.

diversity. Hence, local and global entropy indices were also adapted from Theil and Finizza (1971) to measure diversity on the AxS model's aggregated flows. The global entropy index E summarises the diversity of the entire study area's flows, as per equation 4.13. In that equation, τ_m is the proportion of flows of population group m in the study area, and M denotes the number of population groups.

$$E = \sum_{m=1}^{M} (\tau_m) ln \left(\frac{1}{\tau_m} \right) \tag{4.13}$$

Entropy can also be calculated locally, measuring the diversity of flows on each cell of the model's environment. The local entropy index E_c of cell c is calculated as per equation 4.14. In that equation, τ_{cm} is the proportion of flows of population group m in cell c, and M denotes the number of population groups.

$$E_c = \sum_{m=1}^{M} (\tau_{cm}) ln \left(\frac{1}{\tau_{cm}} \right) \tag{4.14}$$

Global and local entropy values range from 0, indicating all flows in the study area (for global entropy) or the cell (for local entropy) belong to a single group, to $log(M)$, which indicates aggregated flows of all groups are evenly distributed in the urban area (or in the cell, in the case of local entropy).

The information theory index H is calculated by comparing the entropy of the entire study area to the entropy of the local areas (in the case of this study, individual cells). The global index H measures the average deviation between each cell's entropy and the grid's global entropy, as per equation 4.15. In that equation, H denotes the global index H of the study area, the total grid's flow count is represented by F, while the local flow count of cell c is denoted by F_c, and C indicates the number of cells in the grid. Global and local entropy values are represented by E and E_c, respectively.

$$H = \sum_{c=1}^{C} \left[\frac{F_c(E - E_c)}{EF} \right] \tag{4.15}$$

Global index H values vary between 0, when each cell has the same entropy as the entire grid (maximum integration), and 1, when the city is totally segregated and each cell contains flows of only one group.

Local index H measures how much each cell is more or less diverse than the study area, as per equation 4.16. The notation used in equation 4.16 is the same as the previous one (4.15), apart from h_c that denotes the local index H of cell c.

$$h_c = \frac{F_j(E - E_c)}{EF} \tag{4.16}$$

Negative values of the local index H indicate the cell has higher entropy than the grid, hence it is more diverse. Positive values indicate the opposite: the grid (city) is more diverse than that particular cell.

The segregation indices introduced here are indicators of the evenness/clustering dimension of segregation. They measure how evenly distributed are the aggregated flows of different population groups in the study area. They can also be interpreted as measuring segregation from the perspective of a shop (or any other activity location) on a street. In this sense, entropy and index H can measure the diversity of that shop's potential customers, who are the people who walk in front of that shop on a daily basis.

Copresence

As previously discussed, in time geography, the temporal dimension is as important as the spatial dimension in determining possibilities of interaction. This means that actual or potential interactions between individuals require those individuals to be present at the same *place* at the same *time*. In this study, encounters and possible interactions between individual agents in space and time are referred to as *copresence*, and a series of copresence metrics were proposed to quantify those interactions.

During the simulation, the model keeps track of all encounters between agents in the model's environment. In a dynamic and spatially explicit agent-based model such as the AxS model, identifying when and where agents encounter each other is relatively straightforward. Encounters, in the model, happen when two or more agents share the same cell during the same iteration. The total number of encounters that took place during a simulation is named *absolute copresence*. In the following equations, C^{abs} denotes the absolute copresence, while C^{abs}_{mn} denotes the absolute copresence between agents of groups m and n.

Proportional and relative copresence measures are derived from the absolute copresence, as follows. The *proportional copresence* C^{prop}_{mn} (equation 4.17) measures the proportion of encounters between agents of groups m and n relatively to the absolute copresence. In a situation where all population groups are perfectly integrated, the values of relative copresence are expected to match the proportion of the populations of each pair of groups in the study area.

$$C^{prop}_{mn} = \frac{C^{abs}_{mn}}{C^{abs}} \tag{4.17}$$

The proportional copresence is easier to interpret than the absolute copresence, because the latter is highly dependent on a series of factors such as the size of the population sample and the number of active agents in the simulation.

The *relative copresence* C^{rel}_{mn} (equation 4.18) measures how much the proportional copresence between groups m and n deviates from the proportion τ_{mn} of the same groups in the study area's population. Negative values of relative copresence mean the probability of encounter is lower than expected, given the proportion of the groups in the study area.

$$C^{rel}_{mn} = \frac{C^{prop}_{mn} - \tau_{mn}}{\tau_{mn}} \tag{4.18}$$

The lower limit of relative copresence is -1, meaning no encounter happened between agents of the two groups in question. Positive values indicate the number of encounters is higher than expected given the groups' presence in the study area, indicating such groups are more integrated.

Local copresence indices are also proposed in this thesis. In the model's context, those indices measure the number of encounters among agents that happened on each cell. The total number of encounters that took place during the simulation at cell c is named *local absolute copresence*, and is denoted by C_c^{abs} in the following equations. Similarly, C_{cmn}^{abs} denotes the local absolute copresence between agents of groups m and n in cell c. In order to better interpret the model's results, measures of local expected and relative copresence were derived, as follows.

The *local expected copresence* C_{cmn}^{exp} indicates the expected absolute number of encounters between agents of groups m and n at cell c, given the proportion τ_{mn} of the groups' populations in the study area and the cell's local absolute copresence C_c^{abs}, as per equation 4.19.

$$C_{cmn}^{exp} = C_c^{abs} \times \tau_{mn} \tag{4.19}$$

The *local relative copresence* C_{cmn}^{rel} indicates the difference between the absolute and expected copresence indices of groups m and n at cell c, as per equation 4.20.

$$C_{cmn}^{rel} = C_{cmn}^{abs} - C_{cmn}^{exp} \tag{4.20}$$

Positive values of local relative copresence indicate higher than expected number of encounters has taken place at the cell, while negative values indicate the opposite trend. This makes the local relative copresence the easier to interpret of the local copresence indices proposed here.

The copresence indicators proposed in this thesis can be considered measures of the exposure/isolation dimension of segregation, as they measure the probabilities of interaction between individuals. In fact, copresence measures actual interactions in the model, which represent potential interactions in the real world. Obviously, it cannot be guaranteed an encounter between two individuals in the model's environment will be translated into meaningful interactions in the real world. This is even more relevant when the relatively large cell sizes used in the model are considered, as agents can be hundreds of metres apart from each other and still occupy the same cell. However, it can be said the minimal time geographic conditions for such interactions were met: at some point during their trajectories in space and time, two individuals were in close proximity to each other.

4.3.4 Exporting Outputs

Spatial patterns, accessibility and segregation metrics produced by the model can be exported for communication purposes and further analysis on specialised statistical and GIS software packages. Simple and open file formats were chosen for that purpose, described as follows.

Flow Metrics

The spatial patterns can be exported in the form of videos or of a series of snapshots showing the simulation dynamically unfolding. Those visual outputs can be visualised and exported directly from the model's interface. The flow metrics containing the aggregated flow of agents on each cell, both total and by population group, can be exported as rasters in ASCII grid format. Those files are simple text files which can be opened in most GIS software packages, where further analyses can be carried out. The artificial trajectories simulated by the model can also be exported to the *routes.csv* file. The structure of that file can be seen in table 4.6.

Table 4.6: Routes table structure (routes.csv).

Field	Description	Data type
agent_id	Agent's unique identifier	Integer
route_id	Route's unique identifier	Integer
x	Cell's x coordinate	Integer
y	Cell's y coordinate	Integer
order	Cell's order in route	Integer

Accessibility

Accessibility and travel statistics of each individual agent can be exported to the *accessibility.csv* file, detailed in table 4.7.

Table 4.7: Individual accessibility table structure (accessibility.csv).

Field	Description	Data type
agent_id	Agent's unique identifier	Integer
agent_group	Agent's group identifier	Integer
transport_mode	Transport mode	String
movement_speed	Agent's speed	Floating point
origin_zone	Origin zone's identifier	String
origin_x	Origin cell's x coordinate	Integer
origin_y	Origin cell's y coordinate	Integer
destination_zone	Destination zone's identifier	String
destination_x	Destination cell's x coordinate	Integer
destination_y	Destination cell's x coordinate	Integer
travel_distance	Total distance travelled (cells)	Floating point
travel_time	Total travel time (iterations)	Integer
gross_time_budget	Agent's gross time budget	Integer
net_time_budget	Remaining time budget after trip	Integer
geometric_accessibility	Agent's geometric accessibility value	Integer
cardinal_accessibility	Agent's cardinal accessibility value	Integer

Copresence

The encounters between agents that took place during the simulation can be exported to the *copresence.csv* file, detailed in table 4.8.

Table 4.8: Copresence table structure (copresence.csv).

Field	Description	Data type
time	Iteration of encounter	Integer
x	Cell's x coordinate	Integer
y	Cell's y coordinate	Integer
agent_id	Agent's unique identifier	Integer
agent_group	Agent group's identifier	Integer
other_id	Second agent's identifier	Integer
other_group	Second agent's group identifier	Integer

4.4 Summary

This chapter has presented the AxS Model, which implements accessibility and segregation measures based on time geographic concepts. The model's logic, which represents the translation of concepts discussed in the theoretical framework, and its implementation was presented and detailed. The next chapter will explore the model's behaviour and the effects of the parameters introduced here through a series of verification and sensitivity analysis tests.

Chapter 5

Verification and Sensitivity Analysis

This chapter presents a set of tests aimed at evaluating the AxS model's behaviour and outputs under different circumstances, covering the steps of verification and sensitivity analysis of the model building process. As discussed in chapter 3, verification can be understood as the process of making sure a model's implementation matches its design, while sensitivity analysis consists in systematically evaluating the effects of initial conditions and parameters on the model's outputs.

The analyses discussed in this chapter were carried out in abstract scenarios representing hypothetical cities. Those scenarios were designed to isolate the factors influencing the simulation and facilitate the analysis of the model's outcomes, by allowing each model's aspect to be analysed separately. The objective of the analysis is to provide a better understanding on how the model works and how the macro-scale results relate to agents' individual behaviour and to environmental conditions.

In what follows, the verification and sensitivity tests are presented in three sections. Section 5.1 focuses on agents' navigation process and resulting aggregated flow patterns, section 5.2 focuses on the model's accessibility outputs, and section 5.3 focuses the model's segregation outputs.

5.1 Agents' Navigation and Flow Patterns

The objective of the analyses discussed in this section is to evaluate the agents' navigation algorithm and resulting movement patterns. The basic abstract environment used in this set of tests is presented in figure 5.1. It consists of a circular urban area, represented by the light brown cells in the figure, served by a main road network represented by the orange cells. The main road network is structured by two primary roads, one horizontal and one vertical, crossing at the centre of the grid, complemented by a ring road around the city centre and a regular grid spanning the entire urban area. Agents can also move outside the main road network, on regular urban cells, assuming those cells are served by a local road network.

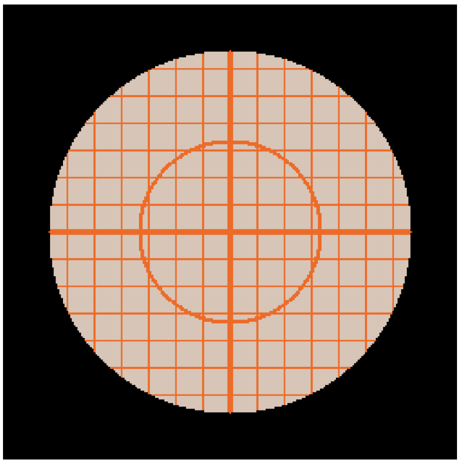

Figure 5.1: Basic abstract environment used in the agents' navigation sensitivity analysis.

Five exercises aimed at testing the effects of the parameters and environmental factors that influence agents' navigation were carried out, and are discussed in the following sections. Section 5.1.1 explores the effect of the road network on agents' movement patterns. Section 5.1.2 focuses on agent's environmental perception. Section 5.1.3 explores how the model's stochasticity manifests during agents' decision making process. Section 5.1.4 tests how agents manage to avoid obstacles in the model. Section 5.1.5 tests the effects of the number of active agents in the simulation on the model's results. Finally, section 5.1.6 summarises the results.

5.1.1 Road Network Weight

The *road weight* parameter represents how likely agents are to move along the main road network instead of outside it. The road weight parameter is a multiplication factor. Thus, *road weight = 1* indicates there is no difference between moving inside or outside the road network, while *road weight = 0.1* indicates the cost of moving through the road network is one tenth of moving outside it.

For this set of sensitivity analysis tests, the road weight parameter was assigned values between 0.1 and 1 at increments of 0.1. All other parameters were kept constant: *number of agents = 200*, *angle of vision = 180º* and *search radius*

= *10 cells*. Start and end points of each agent's trip were randomly located in the urban area.

Figure 5.2 shows the aggregated flow patterns for different values of road weight. It is clear, from the resulting patterns, that agents respond to lower road weights by changing their trajectories to coincide to the main road network. As the road weight parameter is increased and approximates the maximum value of 1, there is visibly less structure in the aggregated flow patterns, which become fuzzier as agents' use the local road network more frequently.

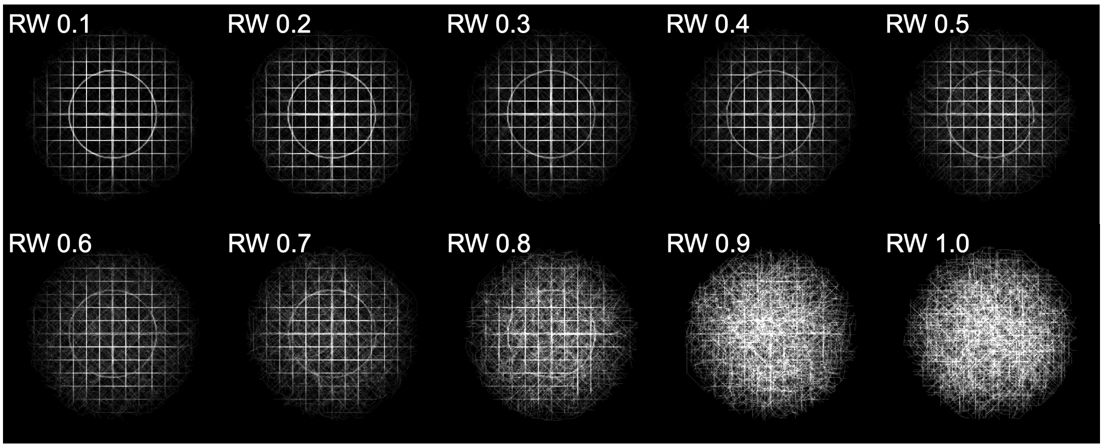

Figure 5.2: Aggregated flows patterns for different values of *road weight* (RW).

The results demonstrate the road weight parameter allows the representation of the road network in a raster environment that reproduces human movement in urban settings that resemble reality, where main roads are more frequently used than local roads. This is despite the lack of a topologically accurate representation of the road network, that could only be implemented in vector space.

5.1.2 Agents' Field of View

This analysis aims to test the impact of the parameters that define agents' field of view (*search radius* and *angle of vision*) in their navigation. Those parameters represent agents' local perception of the environment.

The search radius parameter was set to 3, 5, 7, and 10 cells, and the angle of vision parameter was set to 45, 60, 90, 120, and 180 degrees. The road weight parameter was kept constant at 0.1, and the number of active agents was set to 200. This set of values will also be used in the subsequent tests in this section.

The resulting flow patterns can be seen figure 5.3. As expected, larger fields of view produce more structured patterns. When values of search radius or angle of vision are too small, many agents deviate from the main road network as they cannot evaluate their surroundings effectively, thus generating fuzzier patterns.

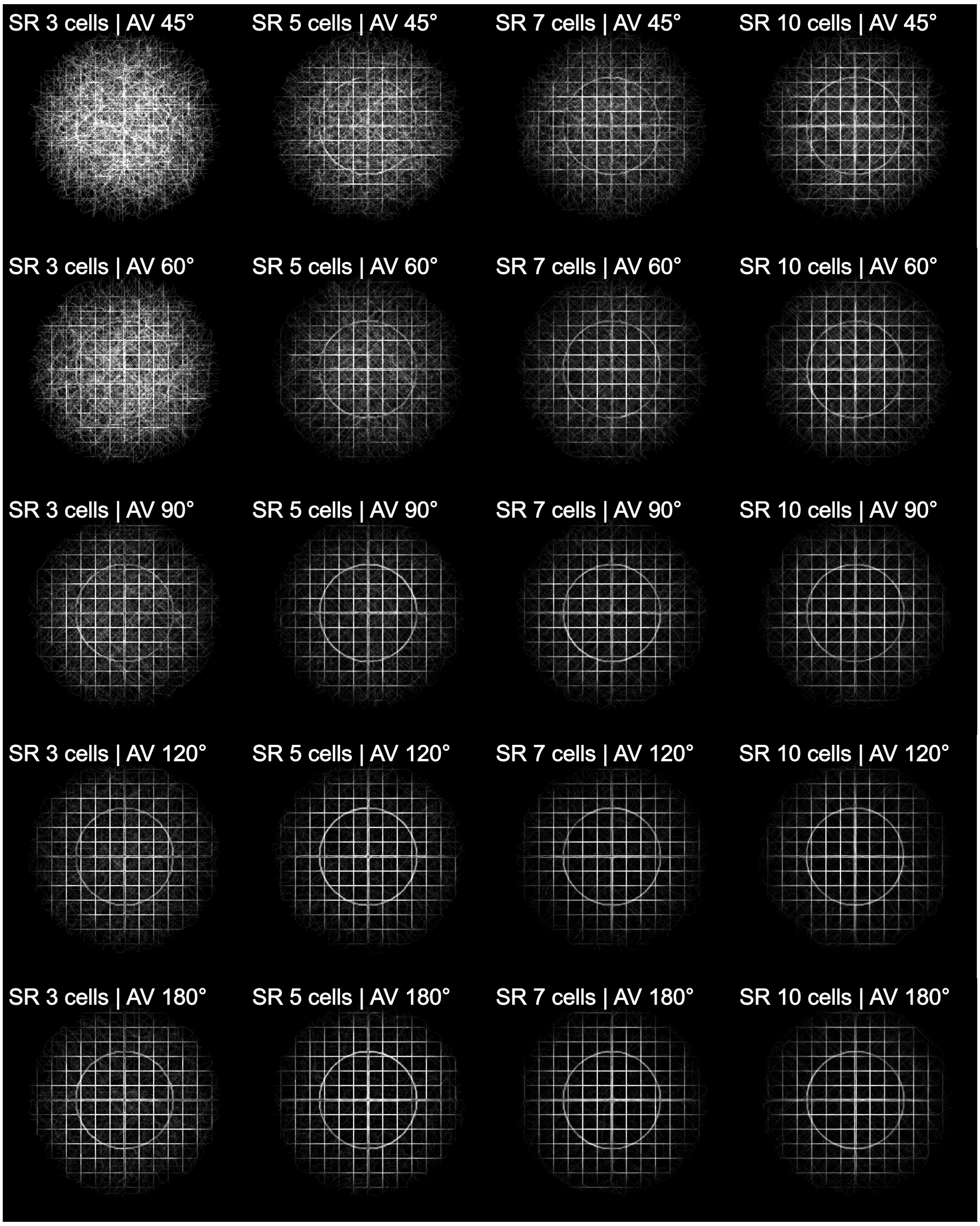

Figure 5.3: Aggregated flows patterns for different values of *search radius* (SR) and *angle of vision* (AV).

5.1.3 Stochasticity

This test was designed to check how stochasticity manifests in the AxS model's pathfinding algorithm by verifying how different agents behave when faced with alternative paths ahead of their trajectories. In this scenario, 500 agents were positioned at the same origin point in the bottom-left quadrant of the abstract study area, and were assigned the same destination point in the top-right quadrant (see figure 5.4).

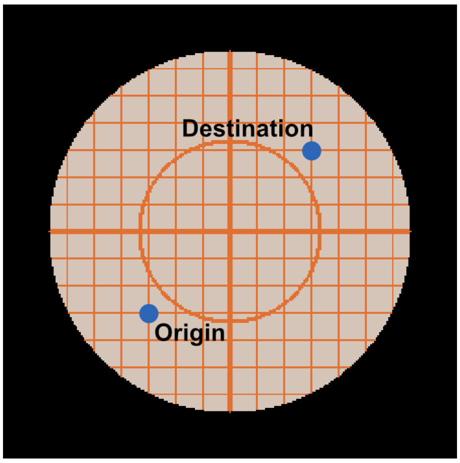

Figure 5.4: Origin and destination points for stochasticity test.

The results can be seen in figure 5.5, where brighter cells represent higher traffic of agents. It is noticeable the algorithm works as expected as, even with the same input parameters, many agents chose alternative routes. Hence, the algorithm is able to capture the diversity of choices present in the real world.

The results also demonstrate how larger fields of view make agents' navigation easier. When the field of view is too small, agents move almost on a straight line towards the destination, basically ignoring the road network. Larger fields of view, such as the ones towards the bottom rows of figure 5.5, allow agents to identify more of the alternative routes to the destination.

Another effect of the size of the field of view is noticeable. Taking for example the results for *search radius = 5* and *angle of vision = 120°*, agents made many small turns during the trajectory. When *search radius = 10* and *angle of vision = 180°*, agents more frequently chose wider routes with fewer turns.

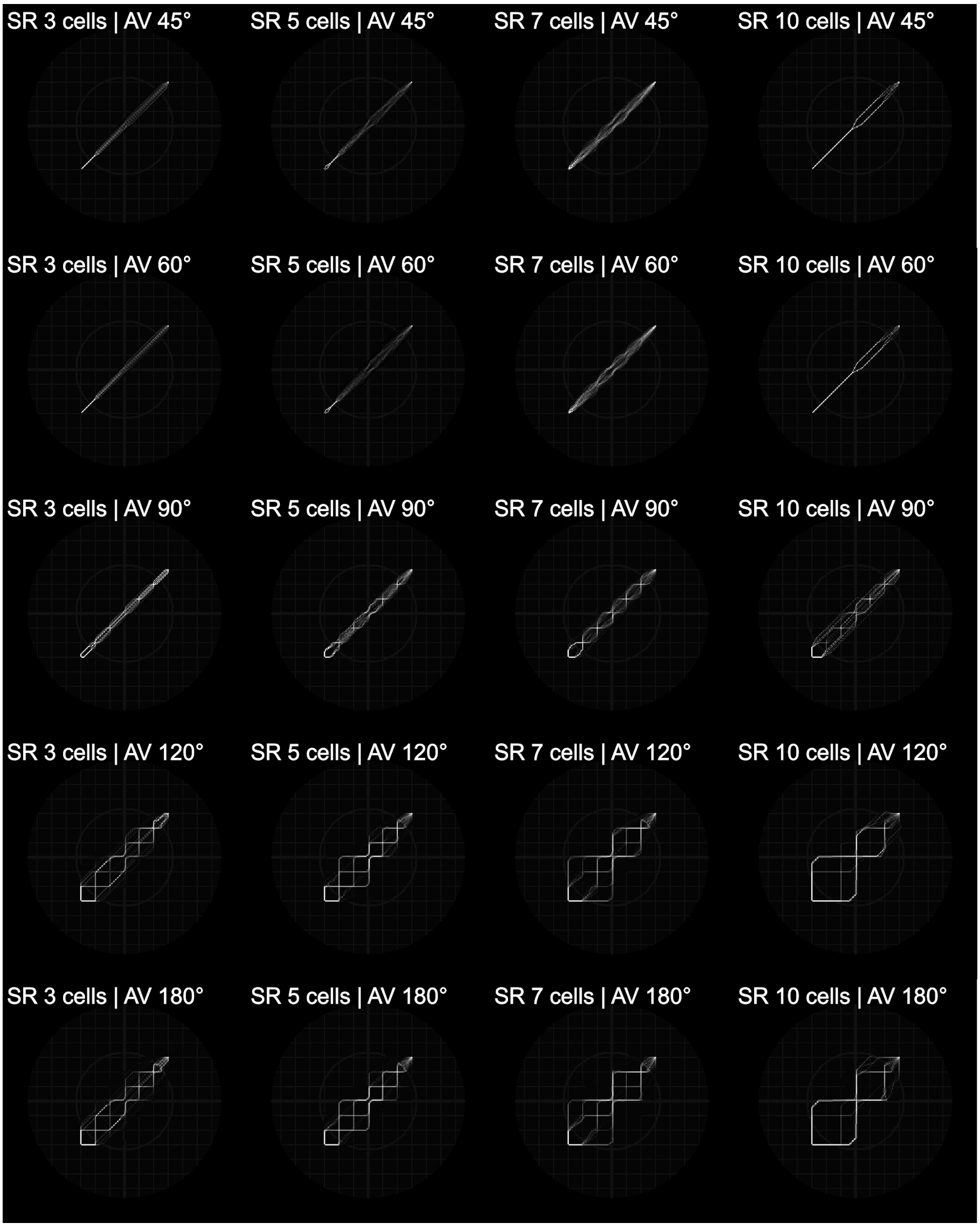

Figure 5.5: Trajectories chosen be agents when moving from the origin on the bottom-left to the destination on the top-right of the grid. Brighter shades of grey represent routes more frequently chosen.

5.1.4 Obstacle Avoidance

This scenario was designed to test the agents' capability of avoiding obstacles (cells marked as spatial constraints) in their route, such as large parks and bodies of water. A park was added to the centre of the hypothetical city, interrupting the main road network (figure 5.6). In a graph-based representation of the road network, avoiding this obstacle would be a trivial task for the agents, although that would require a wider knowledge of the urban environment. This test aims to verify if agents are able to circumvent the obstacle based only on their field of view, as defined by the *search radius* and *angle of vision* parameters.

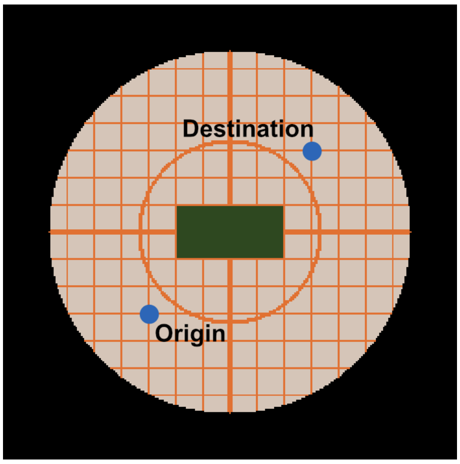

Figure 5.6: Origin and destination points, and central city park for obstacle avoidance test.

The results shown in figure 5.7 indicate agents are able to navigate around the park even when their field of view is relatively small. With the exception of three cases, when *search radius = 3 cells* and *angle of vision* equals to 45°, 60°or 90°, all agents managed to avoid the obstacle. The average failure rates for those three sets of simulations within 10 model runs can be seen in table 5.1. Interestingly, the average failure rate when *angle of vision = 60°* is slightly higher than when *angle of vision = 45°* (67% and 63%, respectively), despite the larger field of view generated by the former value. The failure rate drops to 52% when *angle of vision = 90°*, which is more in line with the expected result.

Table 5.1: Average failure rate in 10 runs.

Search Radius	Angle of Vision	Failure Rate
3	45°	63%
3	60°	67%
3	90°	52%

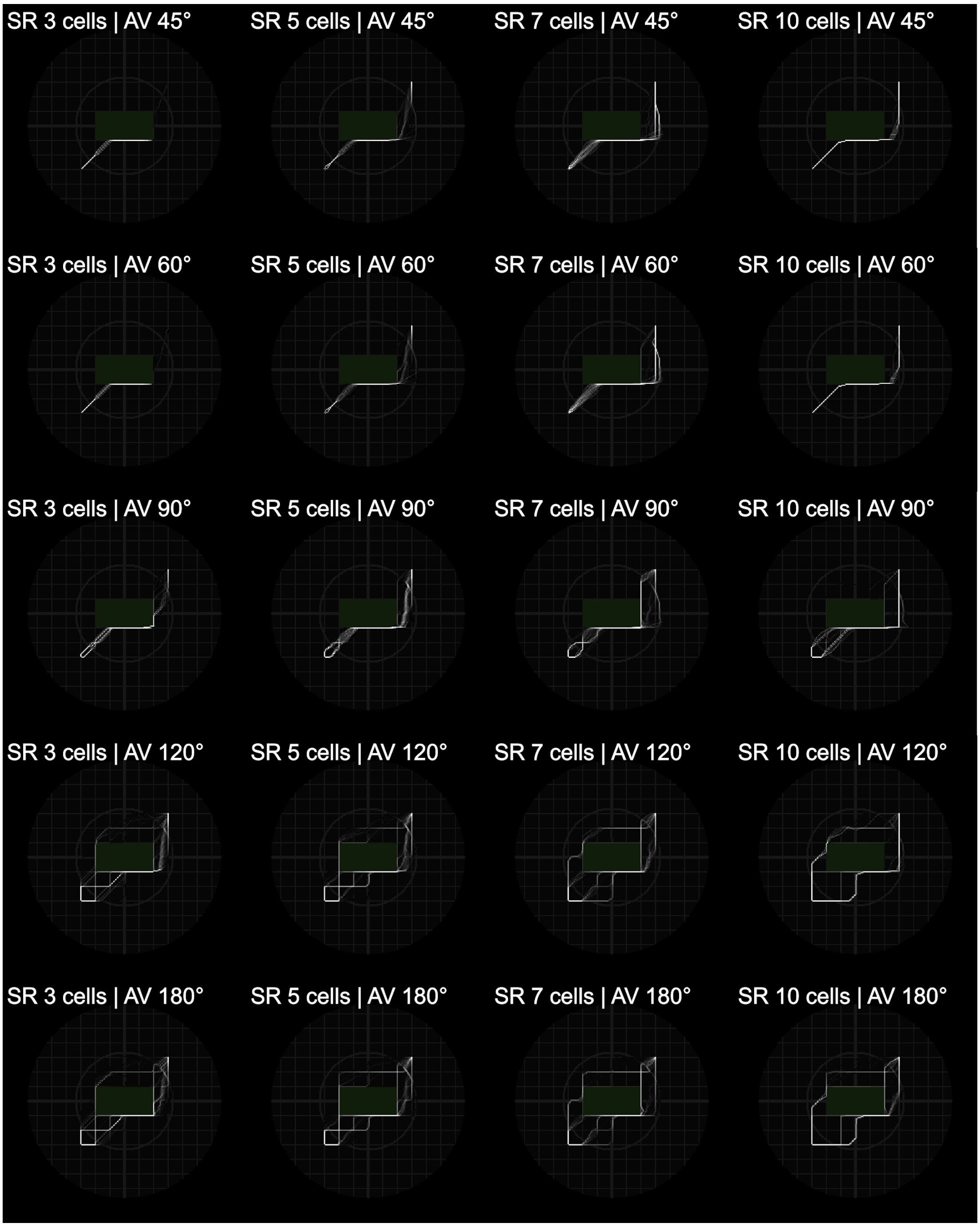

Figure 5.7: Trajectories chosen be agents when moving from the origin on the bottom-left to the destination on the top-right of the grid, while avoiding park on the city centre. Brighter shades of grey represent routes more frequently chosen.

Additional model runs were carried out to further explore those results. For the additional runs, the search radius was set to 3 cells, and the angle of vision was varied between 1°and 120°, at 1°intervals. Five simulations were ran for each parameter variation. The resulting failure rates can be seen in the plot of figure 5.8. Two tipping points are evident in the plot: the first at around 30°of angle of vision, when agents start being able to circumvent the park, and the second at around 90°, above which failure rates drop sharply. Those two transitions peak at 30°and 94°, as can be seen in the graph. Between those points, failure rates remain fairly stable around the median value of 61%. Another milestone is reached at 97°, when 100% of the agents are able to avoid the park and successfully get to the destination.

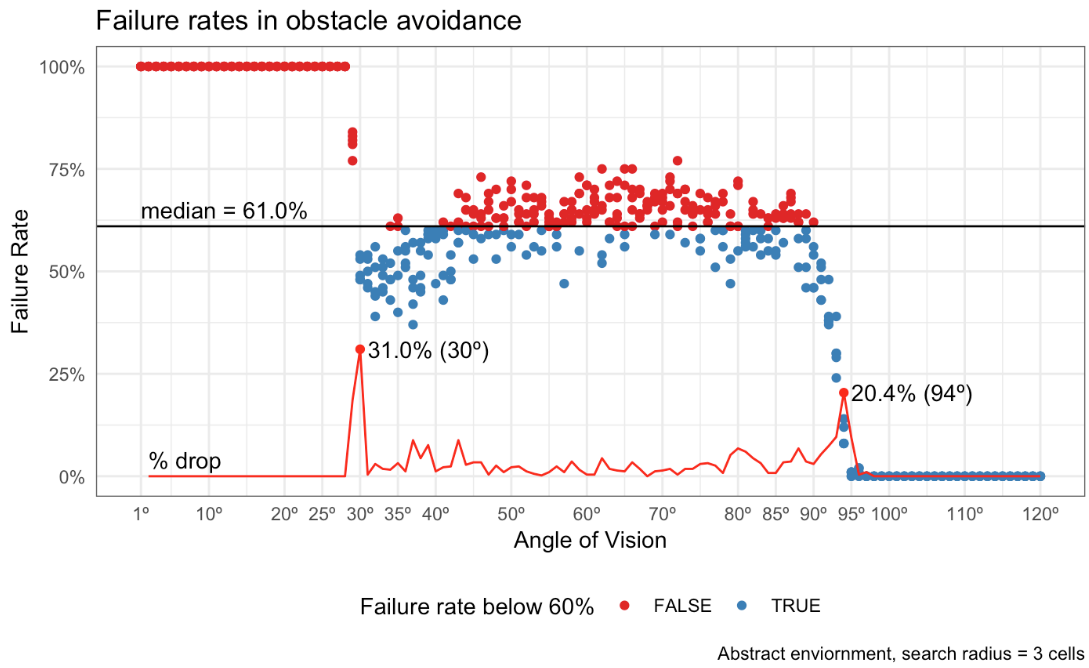

Figure 5.8: Failure rates in obstacle avoidance.

5.1.5 Number of Active Agents

This experiment was designed to verify the effects of the number of simultaneously active agents in the simulation in the formation of the model's aggregated flow patterns. Simulating an entire city's population (in the order of millions of travellers) in a model, especially one developed in NetLogo and designed to run on desktop computers, would not be viable. Hence, this test aims to verify if a relatively small number of agents is enough to generate visually meaningful patterns.

Five simulations were run for this test, setting the number of active agents to 10, 50, 100, 200, and 400. The resulting patterns are shown in figure 5.9, in snapshots taken at 100, 200, 300, 400, and 500 iterations. The results demonstrate the larger the number of active agents in the simulation, the faster the aggregated flows patterns form. Even with few agents, as in the first two rows of figure 5.9, the aggregated pattern eventually forms, albeit slower.

It is important to note that, when the population is divided into groups, the number of agents is split proportionally among the groups, which may slow down the formation of each group's flow pattern. Hence, the decision regarding the actual number of active agents in the simulations must be made considering population and group sizes in the study area.

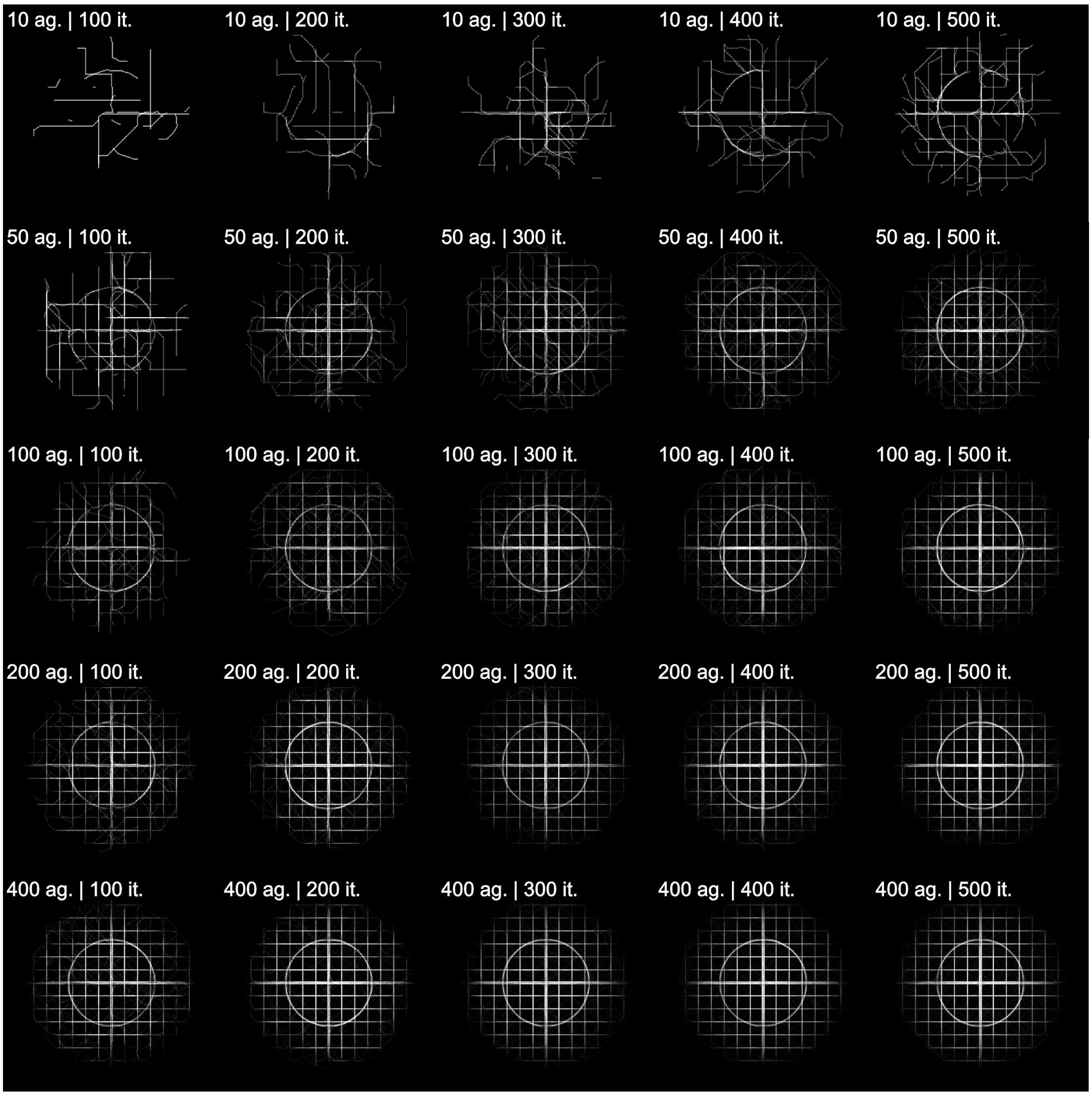

Figure 5.9: Resulting flow patterns with different agents' populations and at different stages of the simulation.

5.1.6 Discussion of Agents' Navigation Analysis

The agents navigation process and the flow patterns that result from individual agent's decisions are an integral part of the model's logic. The resulting agents' trajectories are the foundation for the definition of activity space developed in this study, from which accessibility and segregation measures are derived. Hence, it is important to understand the effects of different parameters and environmental conditions on agents' behaviour in the model.

Regarding the pathfinding algorithm developed for the AxS model, the tests demonstrated that agents are able to find their destination and navigate the urban environment even with limited fields of view and without the aid of a topological network. The tests helped establish guidelines on the size of the area observable by the agents required for efficient navigation and obstacle avoidance. These served as useful guidelines in the subsequent simulations carried out during the development of this thesis.

The tests also demonstrated how the stochasticity introduced in the model's rules manifests in the results. In reality, people's navigation is not a deterministic process, and the tests demonstrated the model is capable of capturing uncertainties existent in the process. Finally, the tests demonstrated that only a small number of agents is necessary for the simulation to produce meaningful patterns on the macro-scale. That is important because simulating the entire population of a large city would require significantly more computational resources than used by the AxS model.

5.2 Accessibility Outputs

This section focuses on testing the AxS model's accessibility outputs. In the model, accessibility is directly affected by two parameters: the *time budget*, which represents the amount of time individuals have available for travelling and participating in discretionary activities; and the *movement speed*, which depends on the agent's transport mode and affects the time spent travelling. Environmental factors, such as travel distance and the spatial distribution of opportunities in the city, also affect the model's accessibility results. The simulation scenarios used in this section were designed to test the effects of those parameters and environmental factors on accessibility.

Simulations were carried out in the abstract environment shown in figure 5.10. The commuting population is represented by three agents (A, B, and C), whose trips' origins and destinations are illustrated in figure 5.10a. This scenario was designed so that agents have clear straight paths to their destinations, which are indicated by the black arrows in figure 5.10a, thus removing the effects of stochasticity in pathfinding from the test results. Places of activity were randomly distributed in the study area, following a normal distribution from the central cell of the grid with standard deviation of 50 cells (figure 5.10b).

Spatial and temporal scales were not set in real world units (such as metres and minutes) for this exercise. Abstract units were used instead, thus time

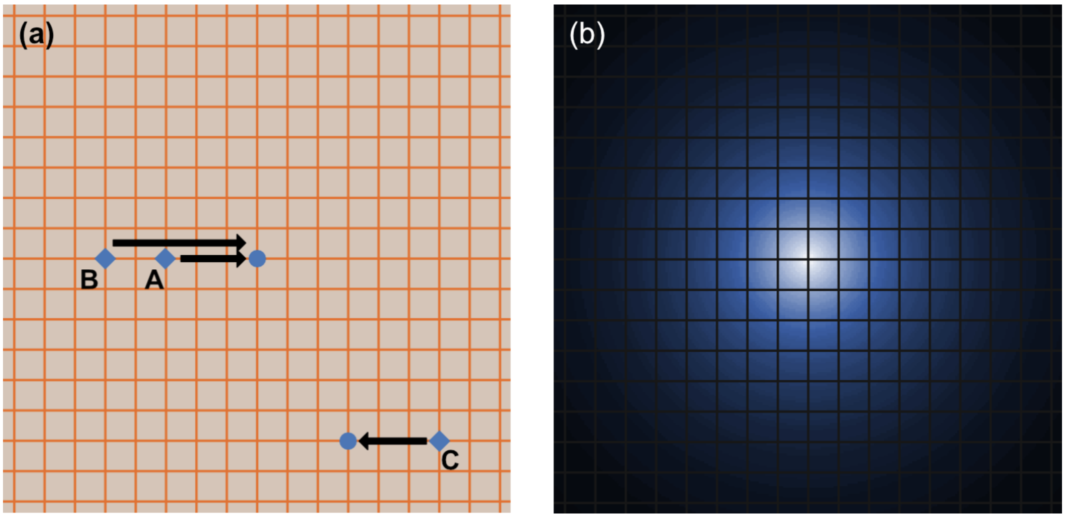

Figure 5.10: Abstract scenario for accessibility output tests: (a) origin and destination of the agents' trips; (b) spatial distribution of activity locations (opportunities) in the study area, with brighter colours representing higher density of opportunities.

was measured by the number of *iterations* elapsed since the start of the agent's trip, and speed was measured by the number of cells an agent can traverse during one time step (*cells per iteration*)

The effects of movement speed and time budget on agents' potential path area (PPA) are discussed in section 5.2.1, while section 5.2.2 compares the results of the geometric and cardinal accessibility metrics.

5.2.1 Potential Path Area

The sensitivity analysis exercise discussed here consists of 1083 model runs in total. The *time budget* parameter values tested range between 30 and 120 iterations, at intervals of 5 iterations. The agents' movement speeds tested range between 1 and 10 cells per iteration, at intervals of 0.5. For each combination of parameters, a single trip was simulated corresponding to one of agents A, B, or C.

As discussed in chapter 4, the PPA is the area an individual can reach using their remaining time budget after travel time is discounted. A sample of the agents' PPAs resulting from the parameter values discussed above is shown in figure 5.11. The figure shows the PPA's of agents A (left-hand side column), B (middle column), and C (right-hand side column) in different model runs. The rows in the figure show results for different movement speeds, ranging from 1 to 5 cells per iteration. The time budget was set to 60 iterations in all cases presented.

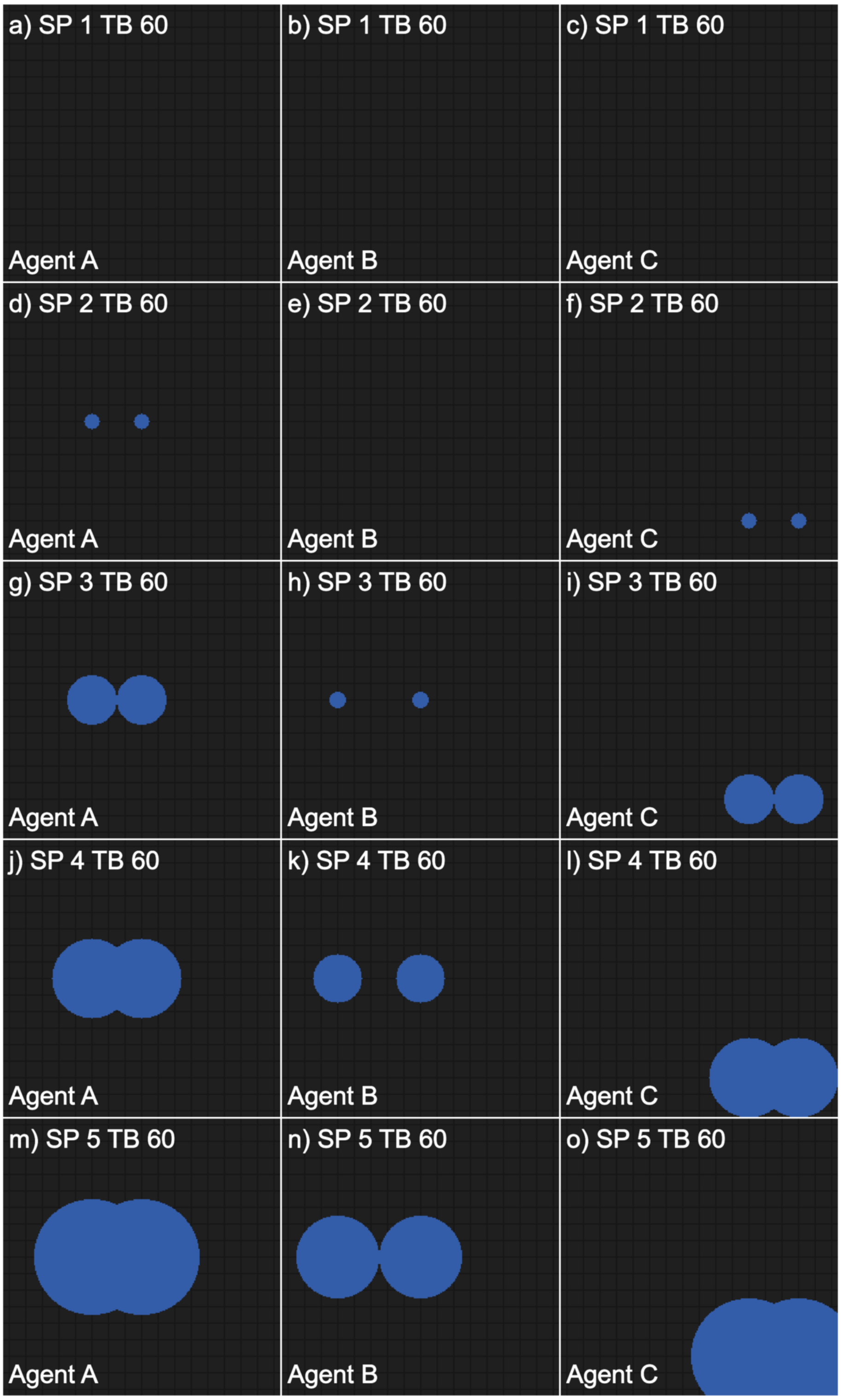

Figure 5.11: Sample of PPA results for agent's speed (SP) ranging from 1 to 5 cells per iteration and time budget (TB) equals to 60 iterations.

It is noticeable, from figure 5.11, that the larger the original time budget and movement speeds, the larger the agent's PPA. In this sense, when movement speeds are too slow relatively to the distance travelled (figures 5.11a, 5.11b, 5.11c, and 5.11e), there is no time budget left for agents to participate in any activity, thus their PPAs are empty. As movement speeds increase, so do the PPAs, until a limit is reached: either the PPAs around the origin and destination points merge together (e.g. figure 5.11j), or the PPA reaches the study area's limits (e.g. figure 5.11o).

5.2.2 Geometric and Cardinal Accessibility

In the AxS model, the *geometric* accessibility measure is equal to the number of urban cells in each agent's PPA. The variations in geometric accessibility according to movement speed and time budget can be seen in figures 5.12 and 5.13. Figure 5.12 shows movement speed on the x axis, with panels representing different time budgets (60, 75 and 90 iterations), while figure 5.13 shows time budgets on the x axis, with panels representing movement speeds (3, 5, and 8 cells per iteration). The similarity between the patterns shown in both graphs is striking. The main difference between them is the curve in figure 5.13 (time budget in the x axis) is slightly smother than the curve in figure 5.12 (speed in the x axis). This means both parameters have the same effect on geometric accessibility: faster speeds and longer time budgets allow agents to access larger areas.

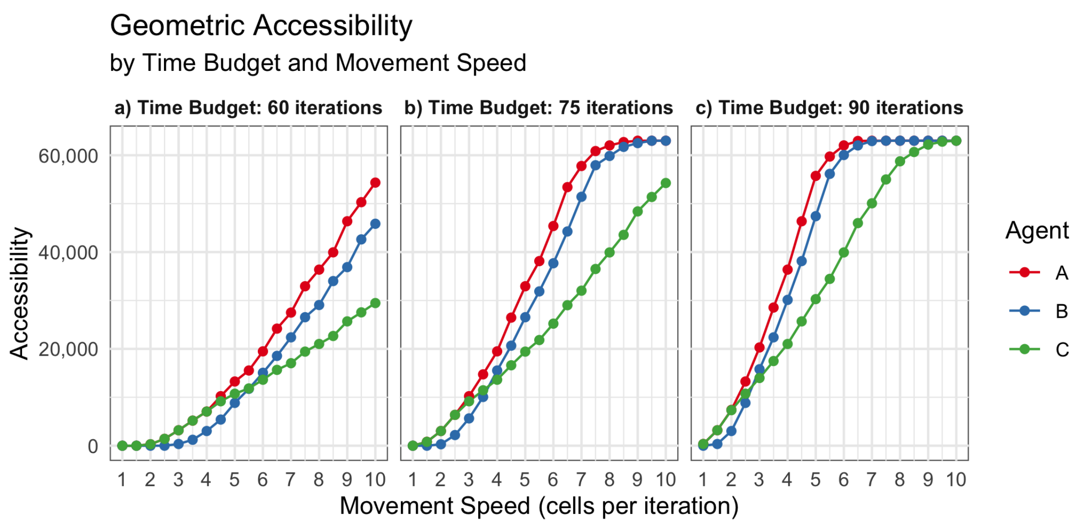

Figure 5.12: Geometric accessibility results for agents A, B, and C, with movement speed ranging from 1 to 10 cells per iteration and time budgets of 60, 75, and 90 iterations.

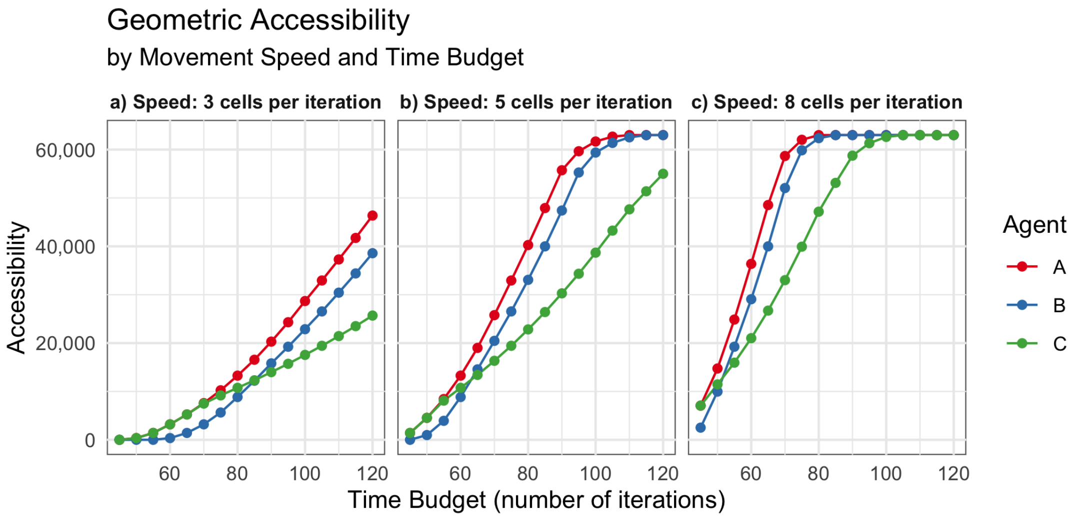

Figure 5.13: Geometric accessibility results for agents A, B, and C, with time budget ranging from 50 to 120 iterations and movement speeds of 3, 5, and 8 cells per iteration.

There are two more patterns worth discussing regarding the geometric accessibility results shown in figures 5.12 and 5.13. The first concerns agents A and C accessibility curves. When slower speeds and smaller time budgets are considered, agents A (red line) and C (green line), who travel the same distance in the simulation, have the same geometric accessibility. As speed and time budget are increased, agent C's accessibility curve changes slope and increases more slowly than agent A's. This happens because agent C's PPA reaches the study area's limits earlier than agent A's, as shown in figure 5.11. The second pattern concerns the accessibility curves of agents B (blue line) and C (green line). Agent B's accessibility curve starts lower than agent C's, but the situation quickly inverts. This is an effect of agent B's longer travel distance, which is a liability with slow movement speed and short time budgets. However, when those parameter values are increased, the larger PPAs around agent B's origin and destination points (figure 5.11n) compensate the extra distance the agent has to travel. Conversely, agent C's shorter travel distance causes the PPAs around agent C's origin and destination points to merge (figure 5.11l), thus cancelling accessibility gains obtained from faster speeds and longer time budgets.

While *geometric* accessibility simply measures the size of an individual's PPA, *cardinal* accessibility measures the number of opportunities available within and individual's PPA. The importance of the cardinal accessibility measure stems from the unequal spatial distribution of opportunities in cities. Hence, an individual with low geometric accessibility may actually have access to more opportunities than an individual with high geometric accessibility, depending on the locations they live, work, and visit. The origins and destinations of agents A, B, and C (figure 5.10a) were chosen to test the model's sensitivity to the spatial distribution of opportunities (figure 5.10b) in the abstract study area of this exercise.

The plots of figure 5.14 show the comparison between agents A, B, and C geometric (5.14a) and cardinal (5.14b) accessibilities at different movement speeds, at a constant time budget of 60 iterations. It is noticeable agent A's cardinal accessibility is higher than agent C's from the start, since agent A is located in an area with more opportunities than agent C. It is also noticeable agent B also has higher cardinal accessibility than C, despite agent B's longer travel distance and lower geometric accessibility. Cardinal accessibility values of agents A and B also increases much faster than agent C's.

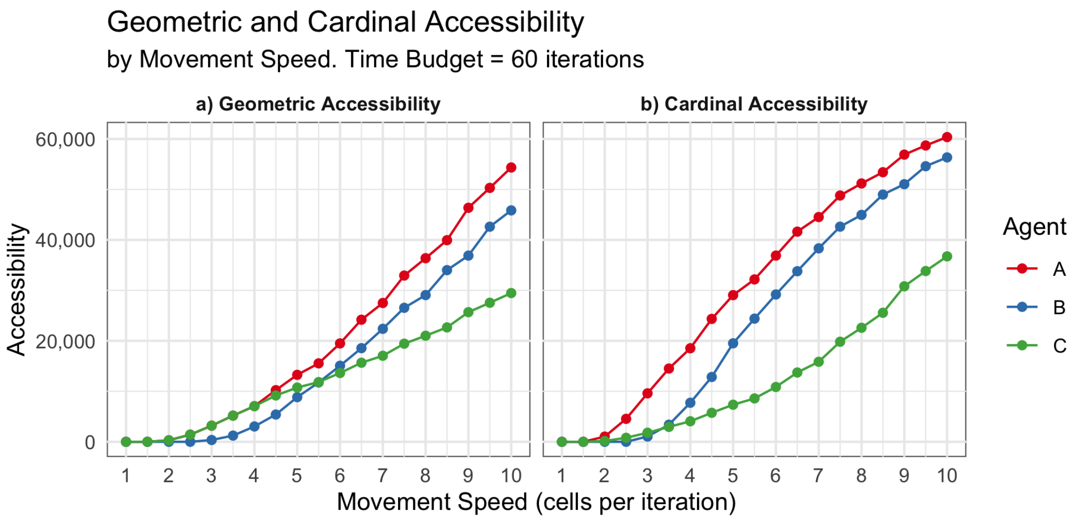

Figure 5.14: Geometric (a) and cardinal (b) accessibility results for agents A, B, and C, with movement speed ranging from 1 to 10 cells per iteration, and time budget of 60 iterations.

This test summarises the range of possibilities and trade-offs in people's residential, workplace, and other activity locations, movement speed (as a proxy for mode of transport), and free time available for carrying out discretionary activities. For example, agent A reaches a level of cardinal accessibility at a speed of 3 cells per iteration that is only reached by agent C when moving at 6 cells per iteration. In this example, agent A could represent an individual who lives and works near the city centre while the individual represented by agent C lives in the suburbs, so agent C needs to use a faster (and probably more expensive) means of transportation and travel longer distances to enjoy the same level of accessibility as agent A. Such level of flexibility and detail is impossible to achieve with traditional place-based accessibility measures.

5.3 Segregation Outputs

The tests presented in this section aim to analyse how different patterns of residential and workplace location of population groups in a study area affect the segregation output maps and metrics produced by the AxS model.

Four simulation scenarios, presented in section 5.3.1, were designed to illustrate hypothetical locational patterns of three population groups in an abstract study area. Those scenarios were used as inputs for the exercises presented in this section. All the simulations were carried out with parameters *road weight = 0.1*, *angle of vision = 180°*, *search radius = 10 cells*. The number of simultaneously active agents in each simulation was set to 500. Thus, the focus of this section is on the effects of the variation of the population distribution in the study area (initial conditions) rather than on parameter values.

The results are presented according to the dimensions of segregation (as discussed in chapter 2): section 5.3.2 discusses the evenness/clustering dimension, while section 5.3.3 focuses on the exposure/isolation dimension.

5.3.1 Initial Conditions

The sensitivity tests in this section were conducted on the abstract environment shown in figure 5.15. In this environment, the urban area occupies the entire grid space, and the road network is organised in a regular grid. This setup provides an environment as featureless as possible, which allows the main focus of the analysis to be set on agent-to-agent interactions rather than on the environment.

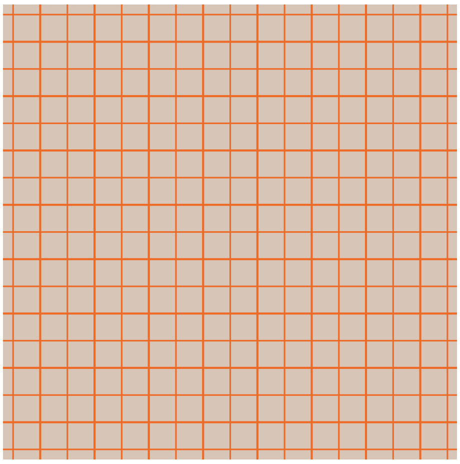

Figure 5.15: Abstract environment for the segregation tests.

Four input scenarios were created for the sensitivity tests in this section. In all four scenarios, the total commuting population is set to 6,000 individuals, distributed among 3 distinct groups: Group 1 (G1) has 4,200 individuals (70% of the population), Groups 2 (G2) and 3 (G3) have 900 individuals each (15% of the population). The population composition is kept constant across the four scenarios, and the only variation is on the spatial distribution of the groups in the study area. In all tests presented in this section, residential and workplace locations are used as origins and destinations of trips, respectively. The four scenarios are described as follows.

Scenario 1 represents a situation of integration between the population groups, serving as the baseline for this analysis. In this scenario, the individuals' places of residence and work are uniformly distributed throughout the study area. This means trips' origin and destination cells are randomly selected at the moment of agent's creation.

Scenario 2 (figure 5.16) also represents a situation of integration between the population groups, but not completely random as in scenario 1. It is inspired by real world cities, where residential and workplace densities tend to be higher in the city centre and lower near the urban edges. Residential and workplace densities in this scenario follow normal distributions from the grid's central cell. Group G1 is more widespread in the study area due to its larger size, and its residential location follows a normal distribution with standard deviation of 50 cells from the city centre. Groups G2 and G3 are smaller and both concentrated around the city centre, following normal distributions with standard deviations of 15 cells. Workplaces for the three groups follow a normal distribution with standard deviation of 15 cells from the grid's centre.

Scenario 3 (figure 5.17) aims to represent a situation where one of the three population groups is segregated from the others. In this scenario, groups G1 and G2 follow the same spatial distribution of scenario 2, both for residential and workplace locations. Population in G3, however, live and work at the upper-right quadrant of the study area. Their residential and workplace locations follow a normal distribution centred at coordinates $x = 185$ and $y = 185$, with standard deviation of 15 cells.

Scenario 4 (figure 5.18) aims to represent a situation where individuals of group G3 live in a segregated area of the city, but work at the city centre together with the other groups, thus being less segregated than in scenario 3.

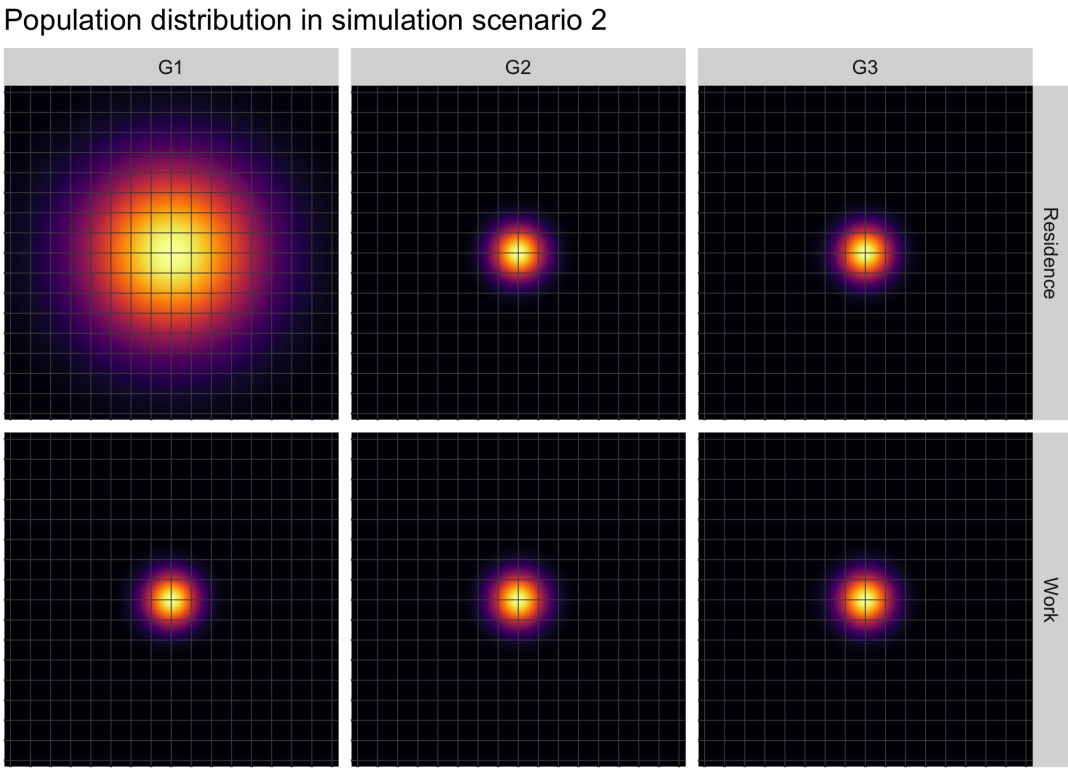

Figure 5.16: Residential and workplace distributions of population groups G1, G2 and G3 in scenario 2.

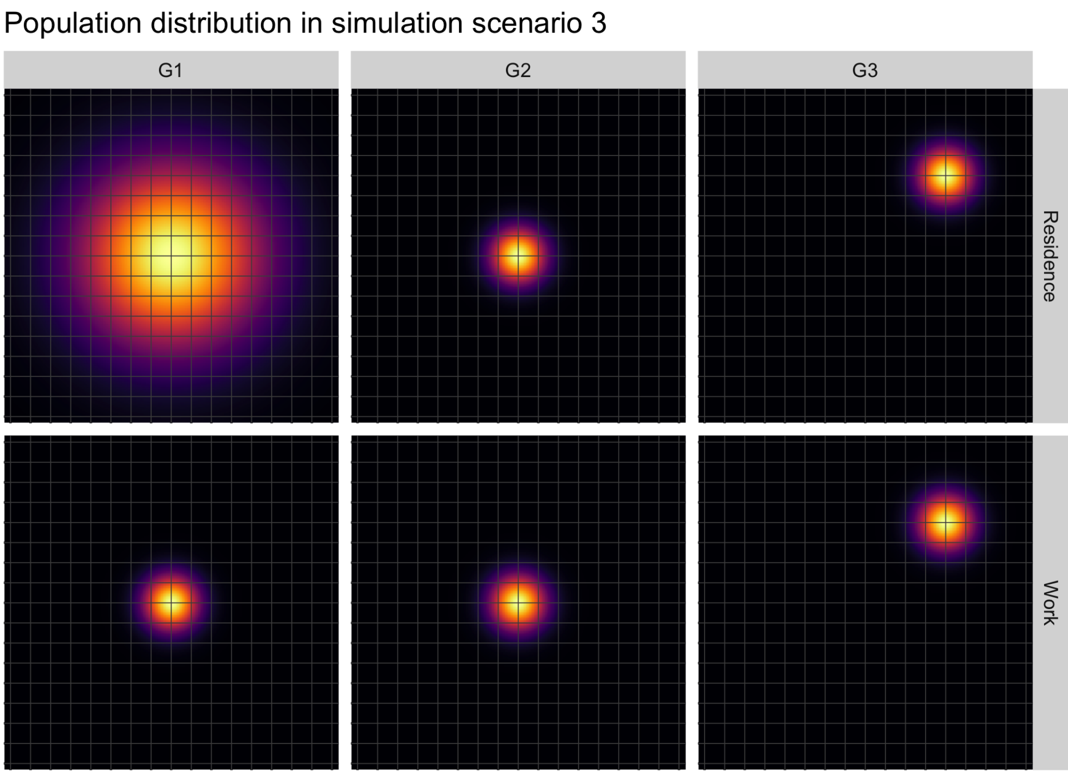

Figure 5.17: Residential and workplace distributions of population groups G1, G2 and G3 in scenario 3.

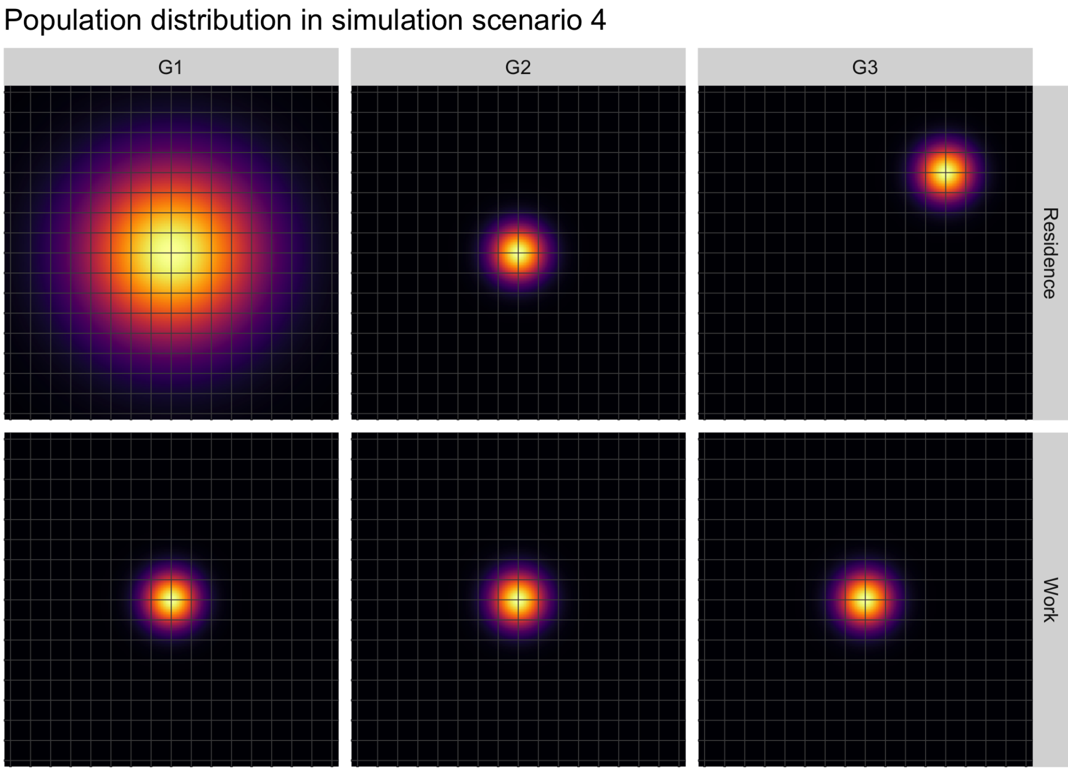

Figure 5.18: Residential and workplace distributions of population groups G1, G2 and G3 in scenario 4.

5.3.2 Evenness/Clustering Dimension

The evenness/clustering dimension of segregation can be analysed using the AxS model's outputs in two ways. The first is through the visual analysis of the aggregated flow patterns of each population group. The second is through the global and local versions of the information theory index (H), that provide quantitative indicators of segregation based on diversity. Both outputs were produced for the four simulation scenarios described above and the results are discussed in this section, as follows.

Visual / Flow Patterns

The model's aggregated flow patterns serve as visual representations of the areas individuals of each group travel through more frequently. Those patterns can provide indicators of potential interaction between groups that are made possible by their travel behaviour rather than their residential and workplace locations alone. Hence, those patterns represent, in the context of this thesis, the collective activity spaces of the study area's population groups. The resulting flow patterns of groups G1, G2, and G3 in the four simulation scenarios are shown in figure 5.19.

In the simplified conditions of this experiment, it is noticeable each group's flow pattern presents the expected outcomes for each simulation scenario, mainly linking agents' places of residence and work. However, those patterns are produced from the bottom up, from the decisions of many individual agents, and not simply by linking origins and destinations via the shortest paths between them. AxS individual trajectories present more variety than the shortest paths, thus the flow patterns spread over larger areas instead of concentrating on fewer routes.

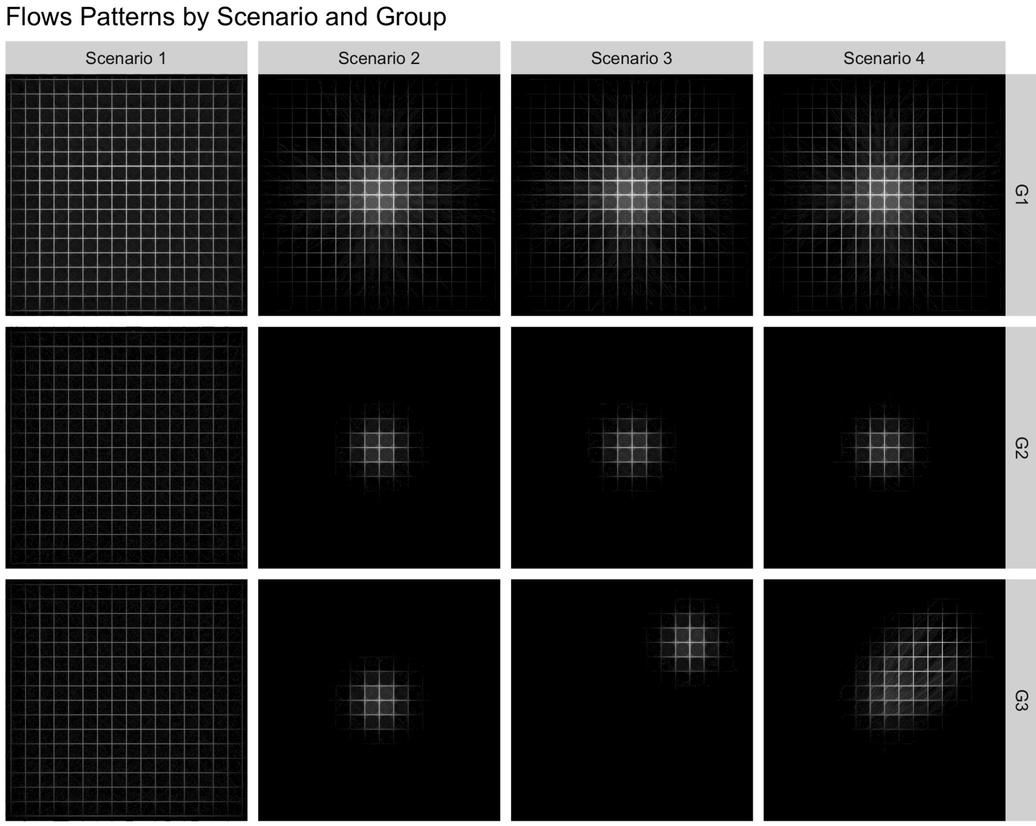

Figure 5.19: Aggregated flow patterns (collective activity spaces) of groups G1, G2, and G3 in four simulation scenarios.

Global index H

Global index H values for the four scenarios are presented in table 5.2. The values of index H range from 0 (indicating maximum integration) to 1 (indicating maximum segregation). The global index H for scenario 1 is very low (H = 0.03), reflecting how the population is evenly distributed in this scenario. Scenario 2 presents a slightly higher global index H (H = 0.17), indicating segregation is higher in this scenario when compared to scenario 1. The global index H is highest in scenario 3 (H = 0.47), indicating the measure identifies correctly the intentional increase in segregation built into this scenario. Scenario 4 is slightly less segregated than scenario 3 (H = 0.30). This is expected, since G3's individuals

Table 5.2: Global index H for four simulation scenarios.

Scenario	H
1	0.03
2	0.17
3	0.47
4	0.30

in this scenario have to travel to the city centre for work, which increases the overall evenness of flows in the study area.

Local index H

The maps of the local index H of the four simulation scenarios are shown in figure 5.20. Areas where the local diversity is similar to the overall diversity of the study area are represented in light grey in the figure. Areas of higher diversity have negative values of H, and are shown in shades of red in the figure, while areas of lower diversity have positive values and are shown in shades of blue.

The local index H map for scenario 1 is completely light grey, which indicates flows are evenly distributed in this scenario and segregation is low in the entire study area. The local index H map for scenario 2 clearly indicates areas of higher diversity (red areas in the centre of the map) and lower diversity (blue areas around the city centre), both concentrated on the main road network. Scenario 3 presents low diversity areas throughout the city, with very small pockets of higher diversity (in light shades of red) in the city centre and close to G3's area in the grid's upper-right quadrant. Scenario 4 presents a few more light-red areas than scenario 3, indicating more opportunities for interaction are made possible by G3's travel pattern in this scenario.

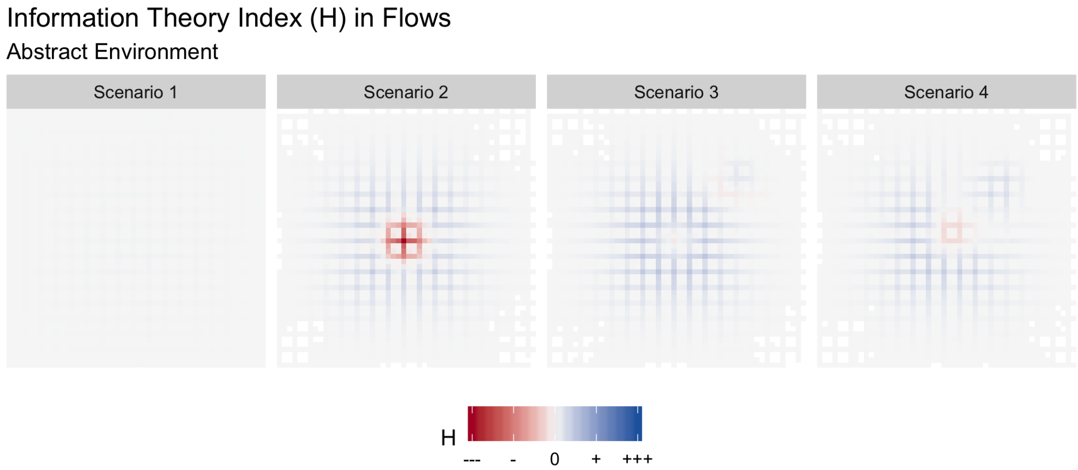

Figure 5.20: Local index H maps for the four simulation scenarios.

5.3.3 Exposure/Isolation Dimension

The AxS model provides direct measurement of the number of encounters between agents during their trajectories. Hence, measures of isolation (when agents belong to the same group) and exposure (when agents belong to different groups) can be generated, providing indicators of potential interaction considering spatial and temporal dimensions (*copresence*, in the model, as introduced in chapter 4). A set of simulations was run to test the model's copresence results, as follows.

Global Copresence

The relative copresence between each pair of groups in the four scenarios was calculated and can be seen in figures 5.21 and 5.22. The size of the bars in the plots indicate how much the proportion of encounters between agents of each pair of groups deviate from the proportion of the groups in the study area's population. Blue bars, to the right-hand side, indicate higher probability of encounter, while red bars to the left-hand side indicate lower probability. The solid red and blue bars represent the average copresence of 20 representative model runs in each scenario, while the error bars indicate the variation (upper and lower bounds) between model runs.

The graph of figure 5.21 shows intragroup copresence values, based on encounters between agents of the same group, thus indicating isolation. It can be seen in the graph that copresence in scenario 1 is close to 0 for all groups, which is the expected outcome due to the evenly distributed population in this scenario. Scenario 2 presents higher relative copresence between groups G2 and G3 than for G1, which can be explained by the larger dispersion of G1 in the study area, while G2 and G3 are spatially concentrated in the city centre. In scenario 3, the spatial separation of G3 increases intragroup copresence (isolation) of all groups: as fewer encounters between agents of different groups are possible in this scenario, the proportion of encounters between agents of the same group increases. In scenario 4, isolation decreases for all groups, but the difference is less significant for G3 who still is the most isolated group in this scenario.

The graph of figure 5.22 shows intergroup copresence values, based on encounters between agents of different groups, which indicates exposure. Similarly to the previous plot, copresence in scenario 1 is close to 0 for all pairs of groups, as expected. Scenario 2 presents higher copresence between G2 and G3, similar to G2 and G3 intragroup copresence values, due to their residential and workplace locations at the city centre. Scenario 3 presents the most striking results, as copresence values involving G3 are much lower due to the spatial isolation of this group in the study area. Scenario 4 presents similar trend of low copresence, but copresence values in this scenario are not as extremely low as in scenario 3.

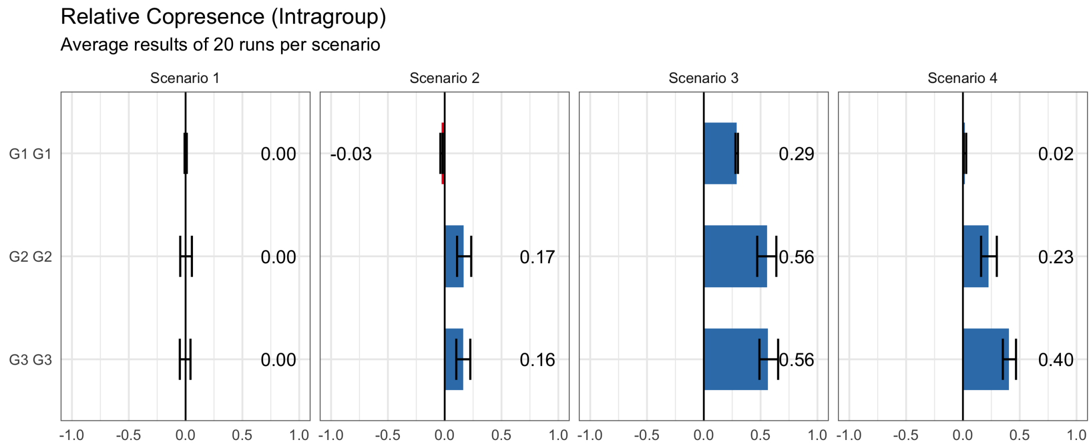

Figure 5.21: Relative intragroup copresence values by simulation scenario. Solid bars represent the average of 20 model runs, and error bars indicate lower and upper limits.

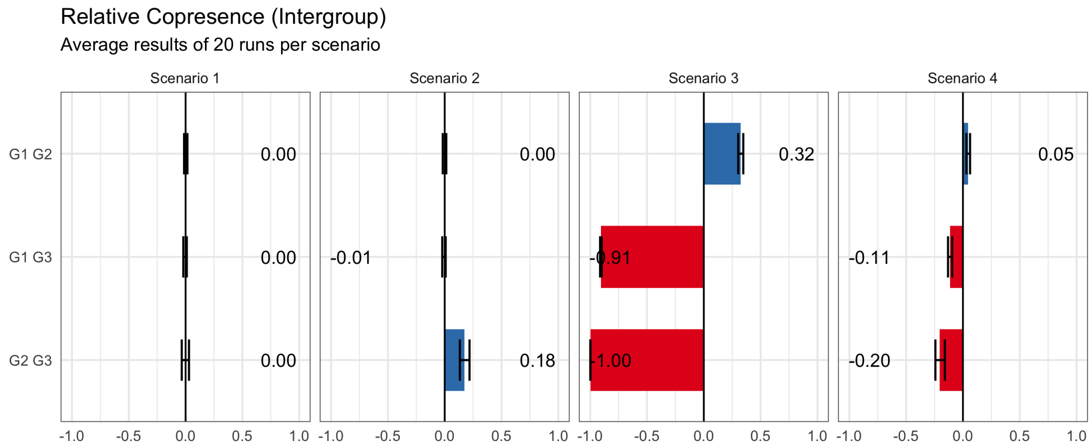

Figure 5.22: Relative intergroup copresence values by simulation scenario. Solid bars represent the average of 20 model runs, and error bars indicate lower and upper limits.

Local Copresence

The model's local copresence results indicate where encounters between agents happen more often in the simulation environment, denoting where encounters between individuals are more likely to happen in reality. The local relative copresence maps for the four simulation scenarios are presented in figures 5.23 and 5.24. Those maps show areas with higher copresence than expected in shades of red, and areas with lower copresence than expected in shades of blue.

Figure 5.23 presents intragroup local copresence maps, depicting encounters between agents of the same group. Copresence maps for scenario 1 show that, in that scenario, encounters happen randomly throughout the study area for all groups. The light shades of red in those maps, that coincide in space with the main road network, indicate slightly higher chance of encounter between agents along the main roads. This can be explained by agents using the main road network more frequently than the local roads, thus increasing the number of encounters on cells served by the main road network. The maps for scenario 2 present higher copresence in the city centre and lower towards the urban edges, showing a wider pattern for G1 and narrower patterns for G2 and G3, following the groups' distribution in the study area.

Copresence results for scenario 3 reflect the differences in residential and workplace locations of G3 in that scenario. A small area of lower copresence of G1 can be seen in the upper-right quadrant of the study area, where G3 is located. The same area can be seen in red in the copresence map of G3, indicating higher copresence at that location. Scenario 4 presents similar patterns to scenario 3, but stronger. The blue low copresence area of G1 is much larger in this scenario than in scenario 3, while the red high copresence area of G3 covers a larger area between G3's places of residence and work. Since G2's locational patterns do not change between scenarios 2, 3 and 4, its copresence patterns are the same in those scenarios.

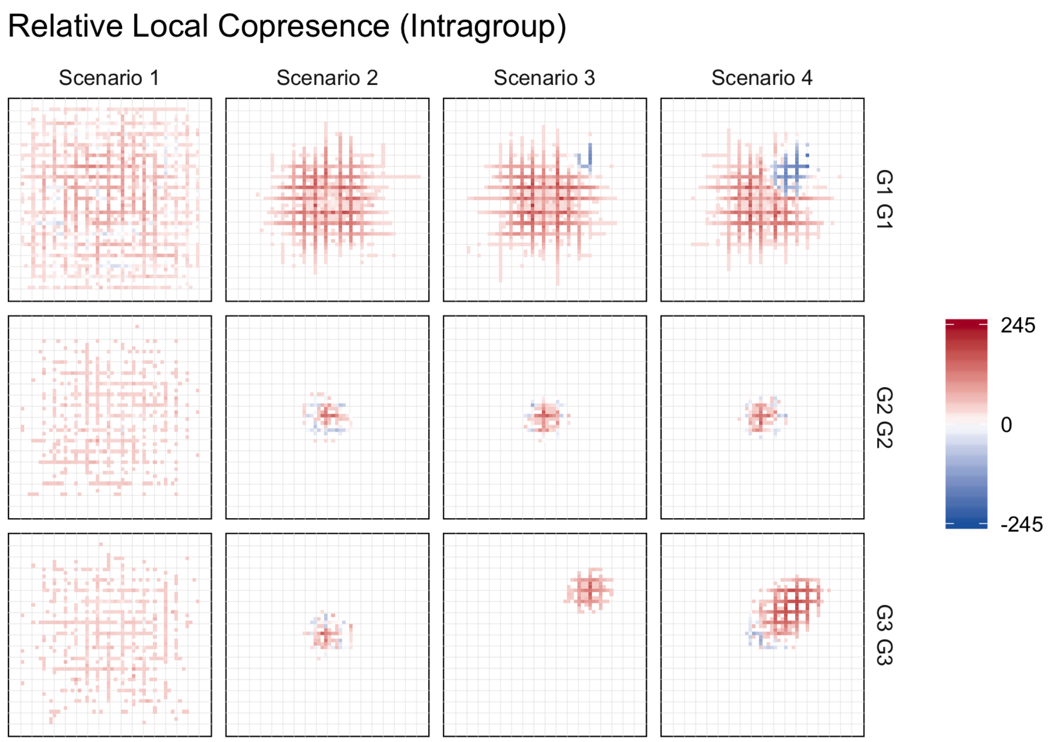

Figure 5.23: Relative local copresence (intragroup) of the four scenarios.

Figure 5.24 presents intergroup local copresence maps, depicting encounters between agents of different population groups. Scenario 1 presents random intergroup copresence values throughout the study area's main road network, which is expected and similar to the intragroup copresence results previously discussed. Scenario 2 presents areas with high intergroup copresence in the city centre (red cells) surrounded by areas of low copresence (blue cells), for all pairs of groups. Scenario 3 presents distinct patterns due to G3's location. Encounters between G1 and G3, in this scenario, concentrate on the upper-right quadrant of the study area, while chances of encounter between G2 and G3 are basically non-existent in this scenario, as demonstrated by the mostly empty copresence map of that pair (apart from a single light-blue pixel). The maps of scenario 4 are similar to the ones of scenario 3, but with stronger shades of red and blue. This indicates two opposing trends in the results. On the one hand, the longer travel trajectories of G3 make more encounters possible, which are represented by the red cells on the maps. On the other hand, in some situations, those possible encounters do not actually happen, hence the lower relative copresence in the dark blue areas of the maps.

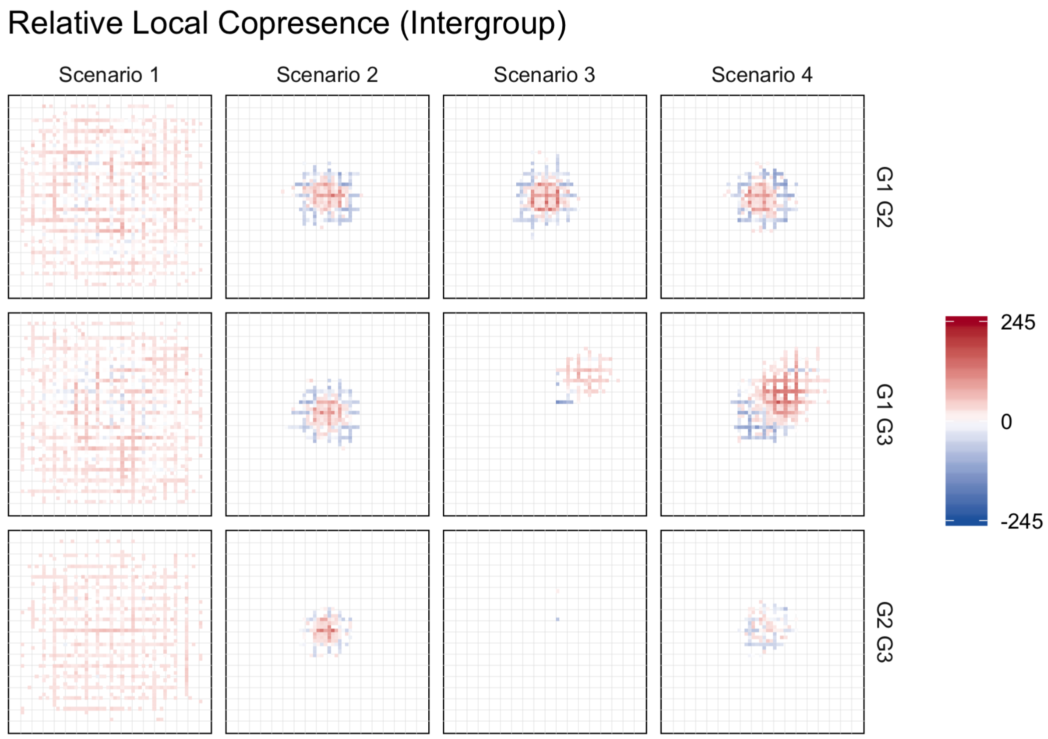

Figure 5.24: Relative local copresence (intergroup) of the four scenarios.

5.4 Discussion

The tests presented in this section were designed to evaluate different input settings of the AxS model, both isolated from each other as well as in combination. Those input settings include model's rules, parameters, environmental conditions, and agents' behaviour. Tests were carried out in abstract simulation environments and controlled scenarios, so the results could be interpreted in light of the expected outcomes for each situation.

The analysis of the basic procedures implemented in the model, related to agent's movement and pathfinding processes, demonstrated the agents behave in line to the rules described in the conceptual model, presented in chapter 4.

As such, the model has been through a verification process, as described in the model building stages introduced in chapter 3.

The evaluation of the individual accessibility measures generated by the model were reported in the second section of this chapter. Sensitivity tests were designed to test the impact of the time budget, movement speed, and distance to work on individuals' level of accessibility. Simulations covering a wide range of parameters were run, and the results demonstrated subtle differences in individual accessibility can be captured by the model.

The segregation measures produced by the model were evaluated through sensitivity analysis. Four sets of initial conditions were tested and their effects on segregation analysed. Measures of diversity and copresence were tested on scenarios of different levels of segregation. The results demonstrated that the measures are sensitive to the spatial distribution of the population in the study area, capturing patterns of segregation and interaction between groups based on dynamic movement patterns rather than on residential or workplace locations alone.

The sensitivity tests reported in this chapter, together, demonstrate through practical exercises how the theoretical concepts discussed in the literature review were translated and implemented in computational form as an agent-based model. The analyses reported so far also illustrate how the model behaves under a wide range of controlled situations. This was important also from a practical standpoint, as the controlled tests allowed the model to be debugged and programming errors to be fixed before the model was applied to real world scenarios. The next section will present the validation exercises, were the model's outputs were compared to real world data.

Chapter 6

Validation

This chapter presents a series of tests carried out with the objective of comparing the AxS model's results to real world data in an attempt to validate the model. As discussed in chapter 3, validation is one of the many challenges in agent-based modelling, due to both the complexity of the modelled phenomena and data requirements.

In fact, data availability is a major challenge for ABM validation, as rarely a single dataset can cover all aspects of a model. This was the case of the AxS model, which required two sets of validation tests each focusing on different output of the model. The first set of tests compared the artificial trajectories produced by the model's agents to trajectories of people in the real world using a dataset from Rio de Janeiro, Brazil. The second set of tests compared agents' travel times to driving and public transport travel times' datasets from London, UK. In both cases, the choice of study areas was determined mainly by data availability.

A choice was made to validate agents' behaviour rather than the model's output accessibility and segregation metrics. This is mainly due to the innovative approach proposed here, which allows measuring accessibility and segregation both at the individual level and at the scale of whole urban and metropolitan areas. Thus, comparing AxS generated metrics to existing metrics of accessibility and segregation, both place- and individual-based, would not make sense as a validation exercise. Agents' trajectories and travel times, though, are the building blocks on which AxS accessibility and segregation metrics are calculated and, thus, it is important to ensure the former are simulated with accuracy levels deemed sufficient for the purposes of this thesis.

This chapter is comprised of two main sections. Section 6.1 presents the trajectories validation, while travel times validation is discussed in section 6.2. The chapter concludes with a discussion on the combined results of the two datasets and a discussion on the validity of the model as a whole.

6.1 Trajectories Validation

This section evaluates the accuracy of the artificial trajectories generated by the AxS model by comparing them to trajectories of individuals in the real world. The following sections detail the study area and data sources used, the model's setup for the simulations, the methodology for measuring route similarity, and the results.

6.1.1 Study Area and Data Sources

The study area for this validation exercise is the municipality of Rio de Janeiro, Brazil, shown in figure 6.1. The city's geography contains natural features such as large hilly areas covered in dense forests inside the city's boundaries, which in the model are implemented as obstacles that agents cannot move through. The pathfinding algorithm implemented in the AxS model relies on local rather than global knowledge of the environment, so the aforementioned environmental constraints can be challenging for agents to navigate around. The effects of those environmental constraints in the model's performance will be detailed in the following analysis.

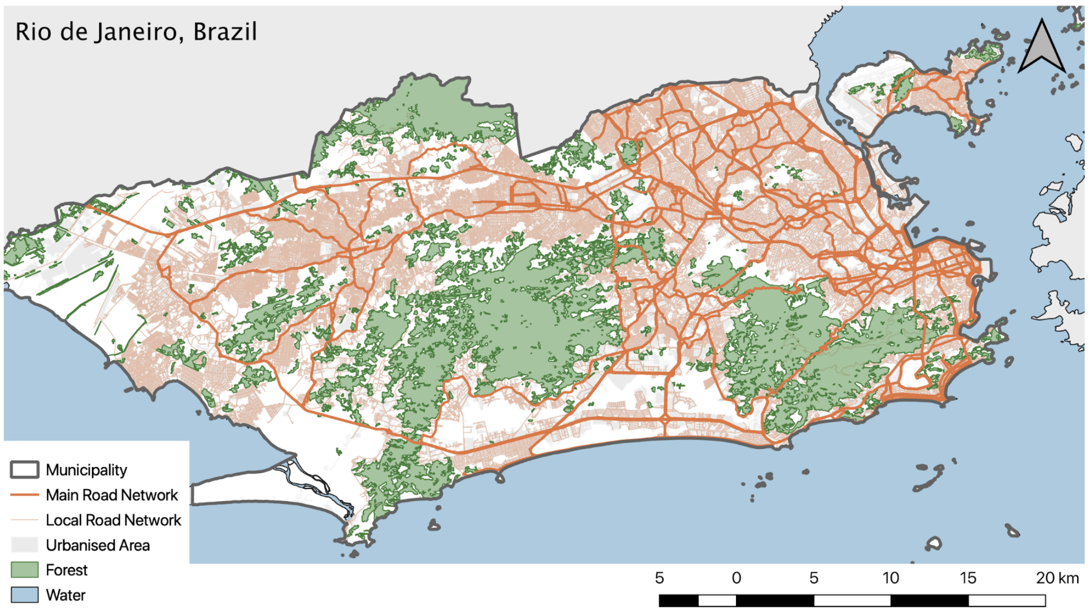

Figure 6.1: Map of the municipality of Rio de Janeiro with main road network and environmental features.

The trajectories dataset used for this set of tests was collected by researchers from the Federal Fluminense University[1] as part of the research project entitled "Mobilities and social performance of residents in affordable housing complexes in Rio de Janeiro"[2]. The dataset contains 1222 trajectories collected via interviews with residents of social housing projects located in the municipality of Rio de Janeiro. Information such as the purpose of each trip, the transport mode used, and the actual route followed during the trip were collected in the interviews and stored in GIS. Each trajectory uses one of three modes of transport: a) walking; b) public transport; and c) private vehicle. Any multimodal trajectories were split into single mode trajectories in the dataset. For example, a home-to-work trajectory composed of three legs – walking to bus stop, bus ride, and then walking to work address – counts as three distinct trajectories in the data, each with its own mode of transport assigned.

The remaining data necessary for this analysis were collected from openly available data sources. Road network information for the city of Rio de Janeiro was downloaded from OpenStreetMap, while political boundaries, urbanised areas, and environmental features (forests, mountains, and bodies of water) were obtained from the city of Rio de Janeiro's open data portal (www.data.rio).

6.1.2 Experiment Setup

Rio de Janeiro's spatial information was converted into a grid of 375 by 200 square cells of 200m size in order to be used as inputs to the model. OpenStreetMap road segments marked as motorways, primary, and secondary were included as representation of the city's main road network. Environmental features such as mountains, forests, and water bodies were included as areas agents cannot move through (or spatial constraints). The municipality's limits also were included as obstacles, thus preventing agents moving to neighbouring cities.

Of the 1222 trajectories in the dataset, 1146 are completely within the limits of Rio de Janeiro's municipality, while 76 are partially or totally outside those limits. 183 trajectories are contained inside one or two adjacent 200m by 200m cells, hence deemed too short to be used in this analysis. The problem of using very short routes for validation purposes is they are always simulated with 100% accuracy, due to the spatial resolution of the model's environment. More specifically, when simulating routes contained inside a single cell, the agent starts its trip at the right destination point, hence the 100% accuracy in the simulation. The same happens when a route spans only two adjacent cells, as there is no meaningful decision to be made by the agent other than moving directly to the trip's destination. In order to avoid bias in the results, those trajectories were

1. UFF - Universidade Federal Fluminense, in Portuguese.
2. Research project supported by the Ministry of Cities (MCIDADES), Ministry of Science and Technology (MCTI), and the National Council of Research (CNPq), Brazil, under Grant 550271/2012.

removed from the analysis.

Considering the conditions discussed above, 963 trajectories were selected for this validation exercise, shown in figure 6.2. It is noticeable from the map the trajectories in the dataset span the entire municipality, although the highest density of trajectories can be found in Rio de Janeiro's city centre, located in the East of the study area.

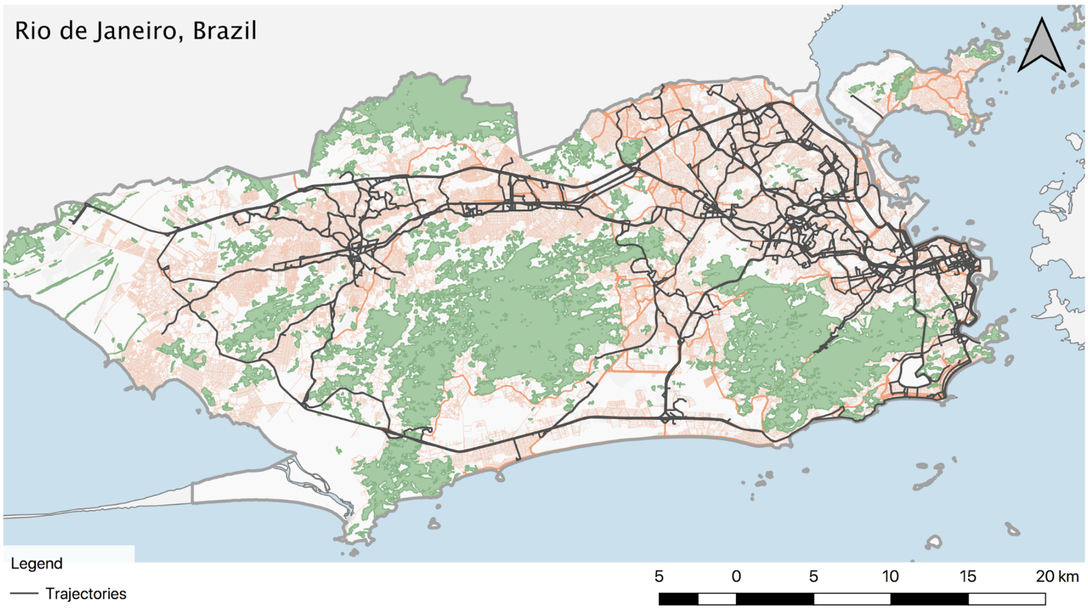

Figure 6.2: Map of the municipality of Rio de Janeiro, with the trajectories of the validation dataset in black.

The input values of the *search radius*, *angle of vision*, and *road weight* parameters used in the simulations are presented in table 6.1. To account for the stochasticity of the model, as the same parameters can yield different results, 20 simulations were run for each combination of parameters. Since there are 100 possible combinations of parameters, a total of 2000 simulations were ran for this analysis. The experiment was carried out using NetLogo's BehaviorSpace tool, which makes it easy to run large number of simulations automatically.

Table 6.1: Parameters used in the trajectories' validation.

Parameter	Values					Units
Search Radius	3	5	7	10		Cells
Angle of Vision	45	60	90	120	180	Degrees
Road Weight	0.10	0.25	0.50	0.75	1.00	-

6.1.3 Route Similarity Assessment Method

The real trajectories of Rio de Janeiro's dataset were compared to the AxS model's artificially generated trajectories, as well as to the shortest path between origin and destination calculated using Dijkstra's algorithm (Dijkstra 1959). The shortest path was used as a baseline for assessing the model's performance. Dijkstra's algorithm was used in this study because it is a popular choice in applications where trajectories along a street network need to be calculated. It is also implemented in many GIS analysis toolkits, which is not the case of more sophisticated pathfinding methods used in more specific contexts (such as the transportation models discussed in section 3.3.3).

One problem in assessing the similarity between real and artificially generated trajectories is the different representation methods employed in each set of trajectories. Rio de Janeiro's trajectories were manually drawn as vector polylines with the road network as a visual reference, and the AxS model's trajectories are based on a grid of 200m wide cells. Hence, while Rio de Janeiro's and AxS' trajectories are different visual approximations to the street network, the shortest paths trajectories are a perfect match to the street network because they are calculated based on the actual network topology obtained from OpenStreetMap.

A buffer technique was used to overcome this problem. The accuracy of each artificial trajectory (AxS generated and shortest path) was defined by the percentage of the length of the real trajectory that is contained inside a buffer around the artificial trajectory polyline. Three buffer sizes, of 100m, 300m, and 500m, were used in this analysis, allowing the comparison to be carried out at different levels of detail.

6.1.4 Overall Route Similarity Analysis Results

An overview of the results of the 2000 model runs is presented in the dotplots in figure 6.3. In the plots, each dot represents a single model run, and the dot's position on the y axis represents the average accuracy of the artificial trajectories in that run. Each panel shows runs grouped by the value of the search radius parameter. The colour of each dot represents the road weight parameter used in that run, and the angle of vision parameter is shown on the x axis. In order to avoid dots being plotted on top of each other, a small amount of random noise was added to the x axis. This was necessary due to the small variation of accuracy between model runs of same parameters. The rows show the results for buffer widths of 100m, 300m, and 500m. The performance of the shortest paths algorithm in each buffer width is represented by the red horizontal line and percentage value on the top of each panel. The best AxS run on each panel is indicated by the black dot, alongside the run's number and accuracy value.

Results show the effect of different parameter settings on the model's outcomes. The road weight parameter has the most drastic effect on the model's performance, as the results in figure 6.3 are clearly stacked by colour, with higher road weights at the bottom of the graphs. This corroborates the sensitivity analysis results, which also identified the strong impact of the road weight parameter on the model's outcomes. The angle of vision parameter also has an important effect, as the results obtained with angle of vision set to 90º or above are significantly better than the ones obtained with angle of vision set to 60º or below. The search radius parameter mostly impacts the results when its value is low (3 cells), which generates least accurate trajectories. This parameter presents least significant effects when its value is set to 5 or more cells.

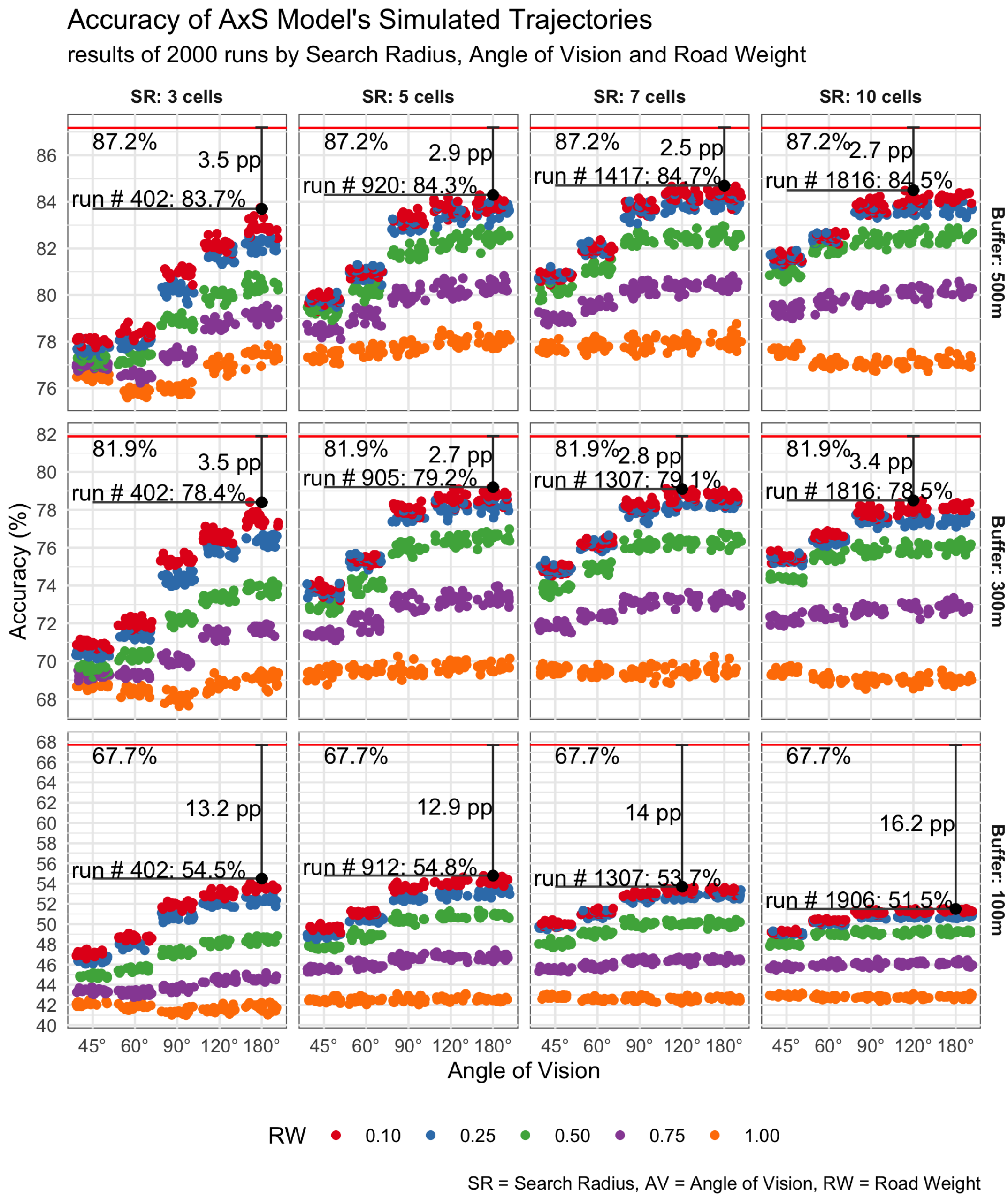

Figure 6.3: Mean accuracy of AxS generated trajectories on each of 2000 model runs in comparison to shortest path algorithm (red line). Black dots represent the best AxS run in each panel.

It is noticeable the shortest paths perform better than the AxS model at all levels of detail (100m, 300m, and 500m), as shown by the red line on each plot above the model's results. However, while the difference in accuracy is above 12pp when the buffer width is set to 100m, it is significantly smaller when the buffer width is set to 300m and 500m: 2.7pp and 2.5pp, respectively, in favour of the shortest path when compared to the best AxS runs. Considering the larger buffers, the average performance of the best AxS runs were found to be between 75% and 80% (measured with 300m buffer) and between 80% and 85% (measured with 500m buffer). These results, combined with the lower performance at the smaller buffer width (100m), suggest the model is more efficient in capturing larger scale patterns rather than fine grained details in the trajectories. This is expected because the spatial representation of the model itself, with relatively large cells of 200m side, does not allow for finer details to be captured.

A subset of the results can be seen in figure 6.4, highlighting the runs with best performing parameters presented in figure 6.3. Runs carried out with search radius of 5, 7 or 10 cells, angle of vision of 90º, 120º, or 180º, and road weight equal to 0.1 are shown in the plots. The plot shows only accuracy measured with the intermediate buffer width of 300m. It is noticeable the accuracy of the runs with same parameters present small variation, at around just 1 pp. It also can be noticed 4 sets of runs present the best overall performances, although by small margins: the ones highlighted in dashed squares in the plots, which were carried out with search radius set to 5 or 7 cells, and angles of vision set to 120º or 180º. Even the best performing model runs presented average results less accurate than the shortest paths, although by relatively small margins: between 3.9 pp and 2.7 pp in favour of the shortest paths.

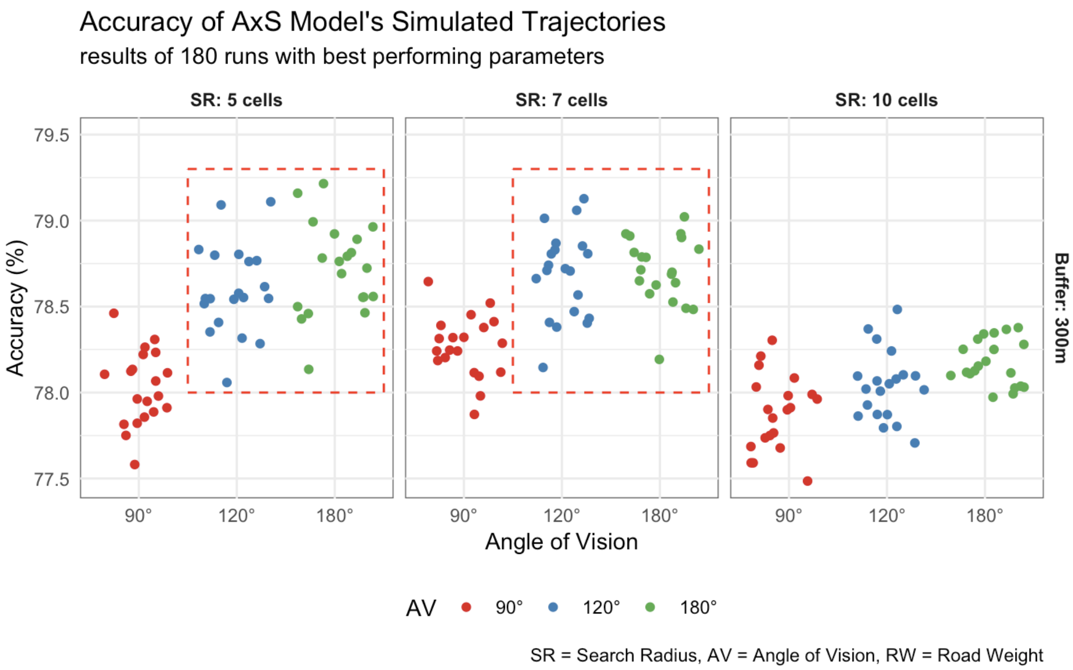

Figure 6.4: Mean accuracy of AxS generated trajectories on the 180 model runs with best performing parameters. Runs highlighted in the red dashed rectangles were carried out with search radius 5 and 7, angle of vision 120º and 180º, road weight 0.1, which tend to produce better outcomes.

One important characteristic of the AxS model is its stochasticity, which means different model runs produce different trajectories for the same origin-destination pair. This is different from Dijkstra's algorithm, which always returns the shortest path between two points. Hence, due to the model's stochasticity, it is possible the most accurate match for each trajectory might be found in separate model runs. The following analyses were designed to take this fact into consideration.

The 80 runs with best performing parameters (highlighted in figure 6.4) were combined into 4 sets of 20 runs, so that all runs in each set were carried out with the same parameter values. Hence, differences between them are due only to different stochastic choices made by agents in different model runs. Accuracy intervals were calculated for each set of aggregated runs, and are presented in figure 6.5. The plot shows minimum, maximum and mean accuracy of each combined run.

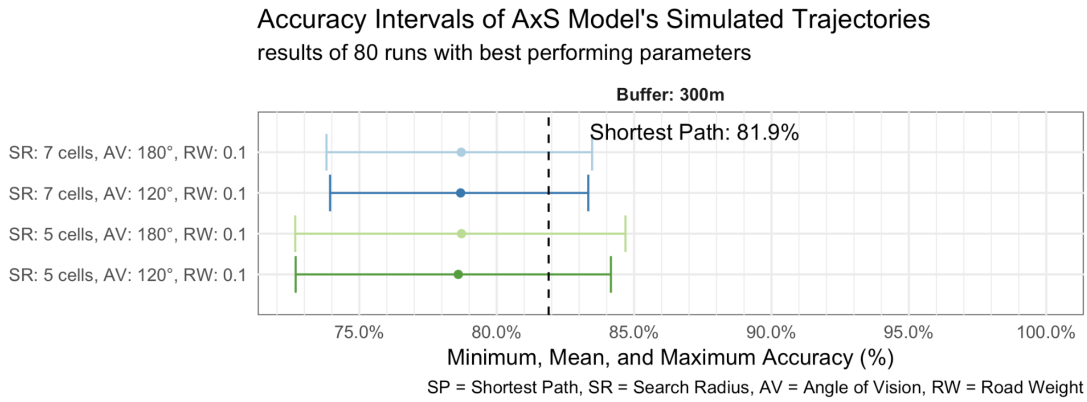

Figure 6.5: Accuracy intervals of AxS generated trajectories on the 80 model runs with best performing parameters, grouped by search radius, angle of vision and road weight.

The plot in figure 6.5 shows the average accuracy of each set of combined runs is very similar, at around 79%. The main difference between sets is in the gap between minimum and maximum accuracy limits, which is slightly wider (around 11.5 pp) when *search radius* = *5* cells in comparison to 9.5 pp when *search radius* = *7* cells. These gaps in accuracy between best and worst performing runs with same parameters demonstrate stochasticity has a significant impact in the model's results, as is expected from the model's design. For instance, at the upper limits of the accuracy interval, the model performs better than the shortest paths (around 84%, 2 pp higher than the shortest path), but the lower limits present lower accuracies (74%, around 8 pp lower than the shortest path).

These results show the model's accuracy intervals are within reasonable levels, and indicate the model is able to simulate trajectories which portray uncertainties in people's pathfinding decisions in reality. However, it is noticeable how the model's results are less accurate than the shortest paths in many cases. It can be assumed this stems mainly from the model's simple raster environmental representation, which lacks the topological structure present in the network environment used in Dijkstra's shortest paths calculation. The following sections explore these results in more detail.

6.1.5 Disaggregated Route Similarity Analysis Results

In what follows, the ability of the AxS model to accurately simulate trajectories in different situations will be further explored by disaggregating the trajectories by mode of transport and by the routes' morphological characteristics. Those tests aim to evaluate the effects of the lack of an explicit representation of the public transportation system and the simple environmental representation used in the model on the accuracy of the model's results. The disaggregation of routes based on transport mode and morphological characteristics is detailed as follows.

Although the AxS model has not been designed to generate trajectories based on any particular mode of transport, its agents' relative freedom of movement can be more closely associated to individual modes of transport, in particular walking and, to a lesser extent, driving. Comparing the trajectories generated by the model to real world trajectories that encompass different modes of transport opens up the possibility to explore the ability of the model to simulate movement made by public transport. Since trajectories by public transport have a number of constraints that were not incorporated into the model, such as fixed routes, embarking/disembarking points, and scheduled departure/arrival times, the expectation is that the model will perform better on individual than on public modes of transport. Walking trajectories are more frequent in the validation dataset (55%), followed by trajectories made by public transport (29%) and private vehicle (16%).

The simple environmental representation used in the AxS model also lacks an accurate representation of the road network's topological structure. This is particularly important in this validation exercise due to Rio de Janeiro's geography, which contains large areas of forests and hills, as well as a road network around and across those natural areas. Those features may impact on the ability of agents to navigate the city effectively, since they need to rely only on their local environmental perception. To account for the influence of those natural features in the routes of the validation dataset, those routes were analysed visually and classified into three categories according to their morphological characteristics. Those categories are: 1) *direct routes*, which connect origin and destination in relatively straight lines, without sudden changes of direction; 2) *diversion routes*, which contain large deviations from a direct path; and 3) *hybrid routes*, which contain one or more relatively straight segments mixed with large detours along the way. Most routes found in the dataset were classified in the direct category (70%), while 19% were classified as hybrid, and only 12% were categorised as diversion[3]. Examples of the three route categories can be seen in figure 6.6.

3. Note there was a small number of routes with ambiguous classifications that were arbitrarily decided, which had an insignificant effect in the aggregate results.

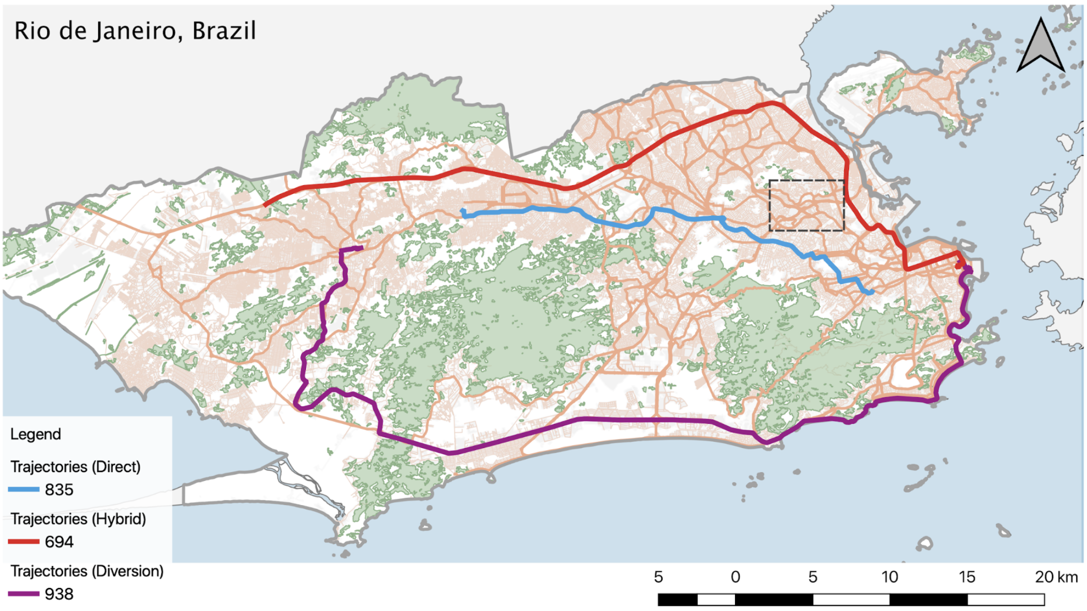

Figure 6.6: Examples of routes in the three categories: direct, diversion, and hybrid.

An overview of the results disaggregated by transport mode and morphological category can be seen in figure 6.7. The graph shows the accuracy intervals of the 20 model runs carried out with *search radius = 7 cells*, *angle of vision = 120°*, and *road weight = 0.1*. The remaining runs with best performing parameters (*search radius = 5 cells*, and *angle of vision = 180°*) present similar results, hence they are omitted from the graph for brevity.

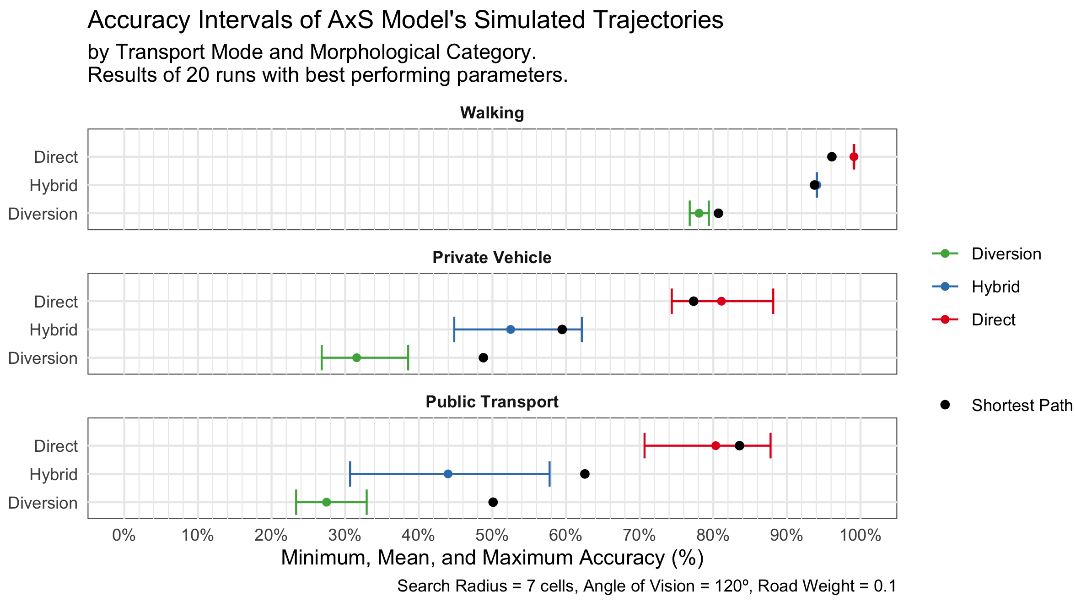

Figure 6.7: Results by transport mode and morphological category.

Results demonstrate walking trajectories are simulated with accuracy close to 100% for routes in the direct and hybrid morphological categories, and around 80% for routes in the diversion category. It is also noticeable the model outperforms the shortest path by a small margin when simulating walking routes in the direct category, while the inverse is true for routes in the diversion category. Routes in the direct category made by private vehicle and public transport are also simulated with high accuracy (80% on average, for both modes of transport). However, the gap between the lower and upper bounds is much wider in those cases when compared to walking routes, as the model's performance varies from around 70% in the worst case scenarios to almost 90% in the best cases.

The morphological category seems to have a greater impact on the artificial trajectories' accuracy than the transport mode. The model's performance on direct routes is significantly higher than on hybrid routes, which is higher than on diversion routes. The lack of a topologically accurate representation of the street network seems to be an important factor in those results, as agents in the model do not have knowledge about large diversions from the direct route that may possibly lead to a more efficient path.

The shortest path's results also present lower accuracy on diversion and hybrid routes, despite relying on the road network's topology, suggesting other factors also have an important influence on this regard. There are many reasons for an individual to deviate from the most direct route, such as knowledge of the road network and traffic conditions, public transportation options available, or even personal preferences. These factors are difficult to predict and/or account for in a simulation environment, even more so in the simplified conditions set up for this study.

Disaggregating the trajectories by their transport mode and morphological characteristics allowed the accuracy of the AxS model's pathfinding algorithm to be evaluated more precisely, as demonstrated by the results of this analysis. Specifically, the model's shortcomings when simulating trajectories with large deviations from the direct route became clear. It is important to note the Rio de Janeiro's geography, previously discussed, may play an important role in the characteristics of the routes in the validation dataset. Hence, the results are expected to be more robust in cities with fewer environmental constraints.

6.1.6 Obstacle Avoidance Analysis

This final test was carried out to verify the effect of Rio de Janeiro's geography on the agents' obstacle avoidance performance. Natural obstacles such as densely forested areas, mountains, and large bodies of water occupy a significant area of the city's territory. Agents, in the model, need to rely only on their limited spatial perception to avoid those obstacles, which may prove challenging. One specific shortcoming of the AxS model's pathfinding algorithm is that agents can get stuck behind obstacles, unable to complete their trips. In those cases, the agent is removed from the simulation and its trip remains incomplete. This particular problem does not affect graph-based pathfinding algorithms, which are guaranteed to find a path if one exists.

Not all of the 963 trajectories in the dataset contain obstacles large enough to block the agents' paths. The subset of 175 routes with obstacles was defined by routes where at least one agent has failed to arrive at the destination in the 2000 runs carried out in this exercise. Some examples of such routes are presented in figure 6.8. The map shows most of the routes with obstacles belong to the diversion category, presenting large detours around natural obstacles that agents cannot perceive due to their limited field of view.

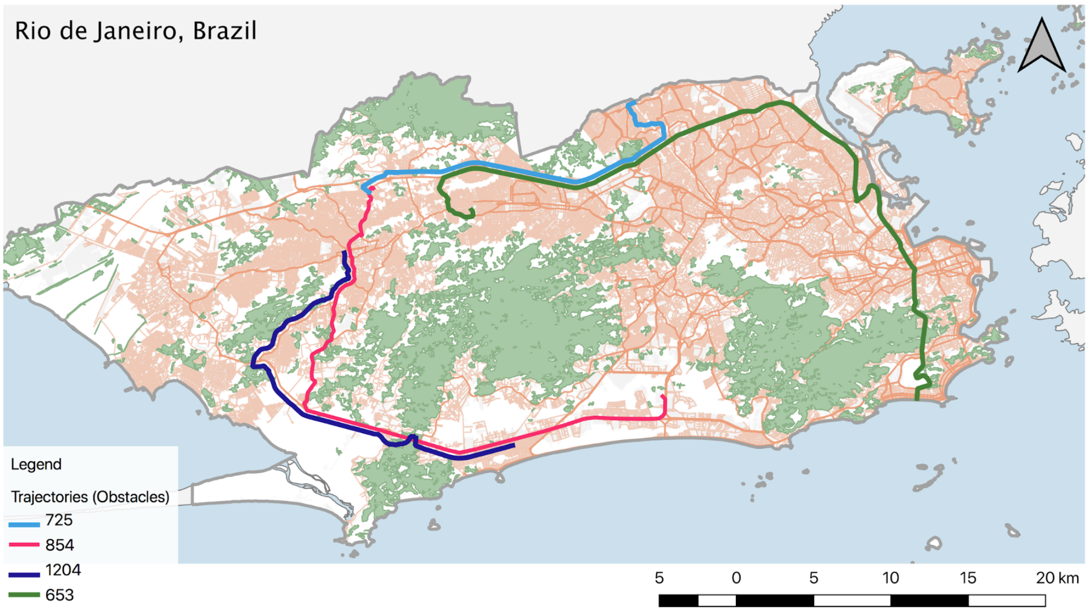

Figure 6.8: Examples of routes with obstacles, represented by the green areas in the map.

The 20 routes with the lowest success rates are shown in table 6.2. Route 725 (the light blue line in figure 6.8) proved to be the most challenging of all, as no agent was able to complete it, followed by routes 740 and 854 (the latter in pink, in figure 6.8). Eleven of the routes shown in table 6.2 were successfully completed in fewer than 10% of the 2,000 attempts.

Table 6.2: Success rates of routes with spatial constraints (out of 2000 attempts).

	Route Id	Success #	%		Route Id	Success #	%
1	725	0	0.00%	11	650	157	7.85%
2	740	1	0.05%	12	853	246	12.30%
3	854	2	0.10%	13	726	320	16.00%
4	737	3	0.15%	14	801	506	25.30%
5	1011	3	0.15%	15	1199	615	30.75%
6	1006	18	0.90%	16	670	634	31.70%
7	1204	32	1.60%	17	883	698	34.90%
8	729	63	3.15%	18	882	718	35.90%
9	653	101	5.05%	19	880	725	36.25%
10	940	116	5.80%	20	896	727	36.35%

The effect of the input parameters in the number of failed attempts can be seen in the heatmaps shown in figure 6.9. The four panels in the figure present a gradient pattern from red/orange tones in the top-left corner, indicating high failure rates, to green/blue tones in the bottom-right corner, indicating low failure rates. The heatmaps indicate a strong relationship between small agents' fields of view and higher probability of failing to avoid obstacles, corroborating the results of the sensitivity analysis (section 5.1.4). It is also noticeable a search radius of 10 cells generates slightly smaller failure rates: 8.1% in the best scenario, compared to 9.3% when search radius is set to 7 cells and 11.7% when search radius is set to 5 cells. However, artificial trajectories generated with search radius set to 10 cells are slightly less accurate than the ones generated with search radius of 5 or 7 cells, as demonstrated previously. Although the differences are not very significant, they indicate that different parameters perform better at different situations. For instance, it may be necessary to increase agents' fields of view in study areas with large obstacles to increase success rates in pathfinding, even though that may slightly decrease the overall accuracy of the results.

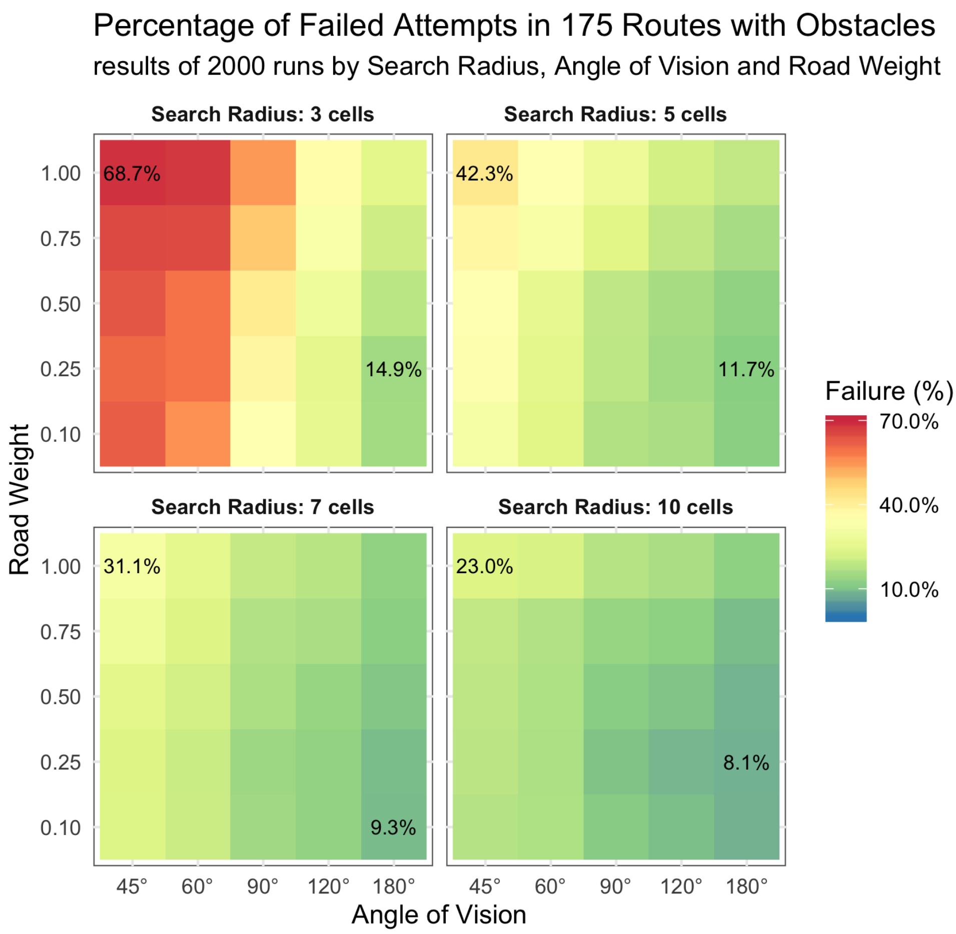

Figure 6.9: Failure rates by search radius, angle of vision and road weight.

6.1.7 Discussion on Trajectories Validation

The validation exercises reported in this section were carried out to test the accuracy of the AxS model's artificial trajectories considering modes of transport and morphological characteristics of the trajectories, as well as agents' obstacle avoidance. Overall, the tests were able to identify the model's advantages and limitations in a variety of situations. For instance, the model performed better when simulating trips that followed a relatively direct trajectory to the destination, regardless of mode of transport. This is important because the simple environmental representation used in the AxS model does not include the public transportation system nor a topologically accurate version of the road network.

The model's performance was significantly lower when simulating trajectories with large detours from the most straightforward path. This indicates people's decision process in choosing (or needing) to deviate from the most direct path is more complex than the model's simple pathfinding logic, including factors not taken into consideration in the model's design. However, only 10% of the validation dataset is comprised of routes with large detours, and there is no reason to believe this kind of route is predominant in most situations or study areas.

These exercises also demonstrated an important advantage of the agent-based pathfinding approach implemented in the AxS model over traditional graph-based approaches: in the model, agents are able to choose between alternative paths when facing more than one plausible way forward. Even though the shortest paths algorithm outperformed the AxS algorithm in many cases, the stochastic characteristic of the model's algorithm can be considered to be more alined with reality, as it allows for different individuals to follow different paths between the same origin and destination, according to their own personal preferences and spatial knowledge. The factors that influence such decisions vary at the individual level and, thus, are very difficult to predict. Allowing many agents to navigate the environment increases the odds the actual trajectory is present among the artificially generated ones.

6.2 Travel Times Validation

This section presents the exercises aimed at validating the agents' travel times in the AxS model, by measuring the difference between simulated and real world travel times by different modes of transport. As discussed previously, the AxS model does not implement transport modes explicitly, with details such as routes, timetables, stations, and capacity. Rather, the model uses movement speed as a simple proxy for transport mode. This exercise aims to verify the validity of this proxy. The following sections detail the study area and data sources used, the model's setup for the simulations, the methodology for measuring travel times accuracy, and the results.

6.2.1 Study Area and Data Sources

This validation exercise was carried in London, UK. The study area is limited to the Inner London area, which is shown in figure 6.10. Inner London provides a large enough study area for this experiment at the same time it significantly reduces the number of trips to be simulated in comparison to the entire Greater London Authority area.

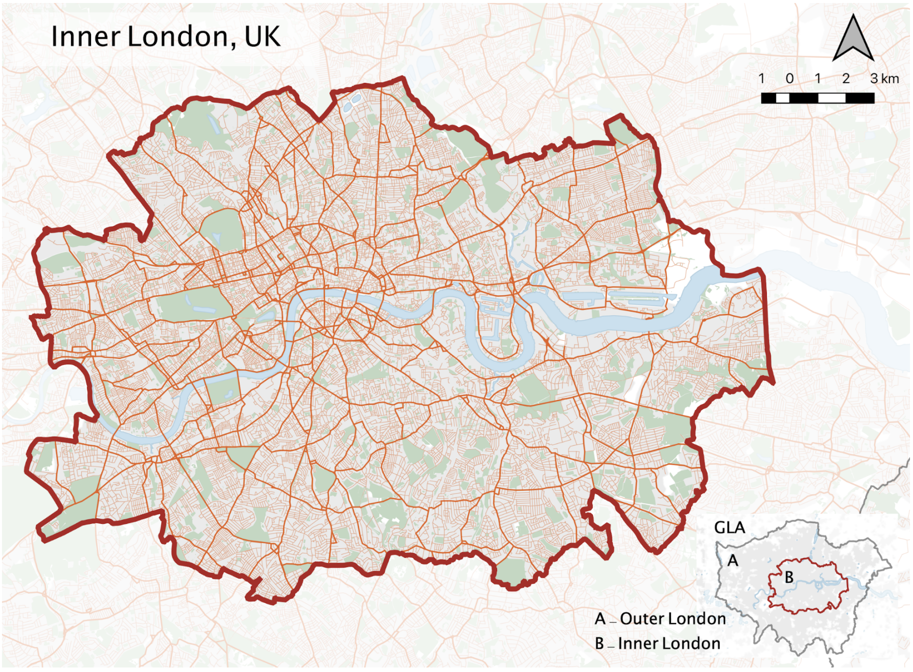

Figure 6.10: Inner London area.

Two travel times datasets were used in this analysis. The first dataset contains travel time matrices by public transport produced by the team of the RESOLUTION (REsilient Systems fOr Land Use TransportatION) research project. Travel times in this dataset were calculated by bus only and by the fastest combination of public transportation modes available, between the centroids of each pair of census areas. The second dataset used in this analysis was provided by Uber as part of the Uber Movement[4] initiative. Uber collects data from its drivers' GPS devices and aggregates them into driving time matrices, making the results available for research and planning purposes. Both datasets used MSOAs (Middle-layer Super Output Area) as spatial units, which are an intermediate-size spatial unit of the UK census, so they could be easily combined.

Inner London's spatial information, such as political boundaries and

4. Uber Movement, (c) 2019 Uber Technologies, Inc., https://movement.uber.com

subdivisions, was obtained from the UK census[5]. Information on road network was extracted from OpenStreetMap. Road segments tagged as motorway, trunk, or primary were included in the simulation, as representing the area's main road network. Those data were converted into grids of 140 by 100 square cells of 200m size to be used as inputs into the model.

6.2.2 Experiment Setup

Travel times tested in this validation exercise were based on typical travel times in Inner London, according to the RESOLUTION and Uber Movement datasets. Those typical travel times are shown in figure 6.11. The graph shows driving is faster than public transportation modes in central London by a significant margin, despite the levels of congestion in the area. The average driving speed in the study area is 21.2 km/h, while the average public transport speed is just 12.6 km/h. As expected, buses are even slower, moving at 10.2 km/h on average. The boxplots also present a large number of outliers, indicating travel speed between some zones are faster than usual, probably due to better transportation infrastructure and more direct connections. Another result obtained from the boxplots is the interquartile range (IQR), which is smaller for bus speeds than for the other transport modes. This indicates buses tend to travel at more constant speeds than other modes throughout the area, albeit slower.

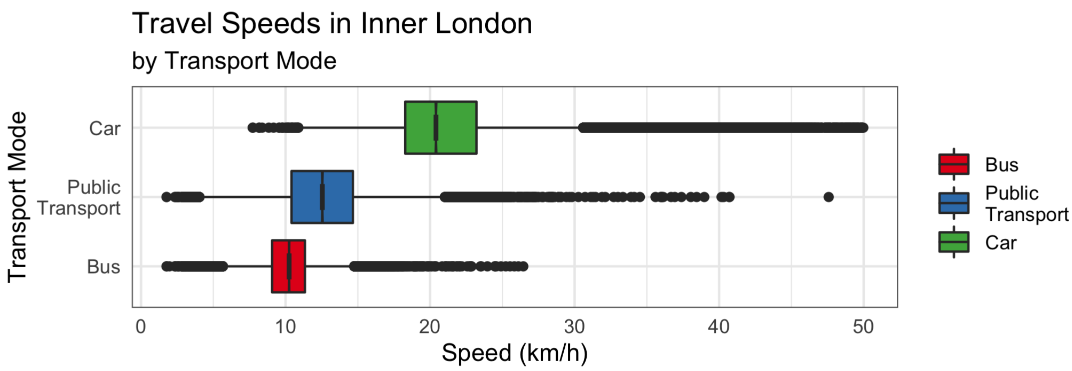

Figure 6.11: Distribution of travel speeds by transport mode in Inner London based on RESOLUTION and Uber Movement datasets.

In all simulations carried out in this validation exercise, the model's time scale factor was set to 1, so that 1 minute in real life corresponds to 1 iteration in model time. Typical travel speeds in km/h by modes of transport, from the RESOLUTION and Uber Movement datasets, were converted into model's speed units (cells per iteration), and the results can be seen in table 6.3. Travel speeds used in this test ranged from 0.3 cells per iteration (3.6 km/h) to 2.0 cells per iteration (24 km/h), at intervals of 0.1 cells per iteration (1.2 km/h), in order to cover the typical travel speeds presented in table 6.3.

5. https://census.ukdataservice.ac.uk/get-data/boundary-data.aspx

Table 6.3: Real world travel speeds converted to model travel speeds.

Mode	Speed km/h	Speed cells/iteration
Walking	5.0	0.4
Bus		
1st quartile	9.1	0.8
mean	10.2	0.9
3rd quartile	11.3	1
Public Transport		
1st quartile	10.4	0.9
mean	12.6	1.0
3rd quartile	14.7	1.2
Driving		
1st quartile	18.3	1.5
mean	21.2	1.8
3rd quartile	23.2	1.9

Centroids of MSOA's were used as origins and destinations of trips, in order to match the information from the validation datasets. All agents in a single run were set to travel at the same speed between all pairs of MSOA's in the study area, thus producing a whole travel time matrix for that speed. To account for the model's stochasticity, 5 simulations were run for each travel speed, generating a total of 90 model runs. Since there are 378 MSOAs in the Inner London area, 142,884 trips were simulated in each model run, and 12.8 million individual trips in total, for this analysis.

The simulations were conducted with *angle of vision = 120°*, *search radius = 7 cells*, and *road weight = 0.1.* Those values were chosen because they produced the best results in the trajectories validation analysis presented in the previous section.

6.2.3 Travel Times Accuracy Assessment Method

The simulated travel time matrices were compared to the real world's travel time matrices from the RESOLUTION and Uber Movement datasets to check for their accuracy. The Mean Absolute Error (MAE) measure of goodness of fit was chosen for this purpose. MAE is a popular choice of goodness of fit measure, which has the useful characteristic of producing results with the same units as the original input data, making analysis easier. MAE is calculated according to equation 6.1, where n is the number of observations, y_i and $\hat{y}_i$ are the correspondent values on each set.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_i - \hat{y}_i| \tag{6.1}$$

MAE measures how much one set of observations differs from another, on average. A MAE of zero means there is a perfect correspondence between the two sets of values. Taking the absolute differences between measured and simulated values prevents positive and negative differences from cancelling out.

6.2.4 Travel Times Analysis Results

The MAE results of the 90 model runs, by travel speed and mode of transport, are summarised in figure 6.12. The plot shows the most accurate results were found when simulating travel by car, where a mean error of 4.1 minutes was found with *travel speed = 1.6* cells per iteration. This speed is between the 1st quartile and the average driving speeds in the validation dataset (1.5 and 1.8 cells per iteration, respectively). Setting the travel speed to the average driving speed of the study area (1.8 cpi) results in a mean error of just 4.8 minutes.

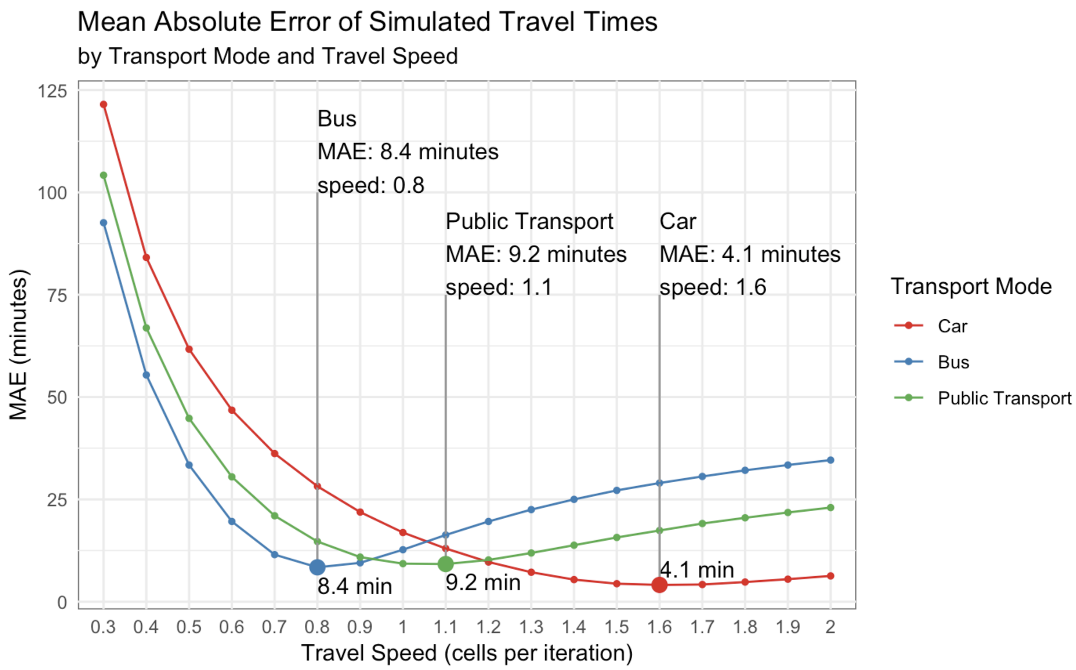

Figure 6.12: Summary of the results of 90 model runs of travel times MAE by travel speed and transport mode.

Figure 6.12 also presents results by public transportation modes. When travel by bus only is considered, higher accuracies were obtained with *travel speed = 0.8* cells per iteration (9.3 km/h), when a mean error of 8.4 minutes was found. This speed corresponds to the 1st quartile of travel speeds by bus in the validation dataset, and just below the average of 0.9 cells per iteration (10.4 km/h). That average bus speed (0.9 cpi) produced a mean error of 9.5 minutes.

Considering travel by all options of public transportation available, higher accuracies were obtained with *travel speed = 1.1* cells per iteration (equivalent to 13.2 km/h). At this speed, a mean error of 9.2 minutes was found between real and simulated travel times. This speed is between the average and 3rd quartile of travel speed by this mode of transport in the validation dataset (1.0 and 1.2 cells per iteration, respectively). The average public transportation speed of 1.0 cpi produced a slightly larger mean error of 9.3 minutes.

A second analysis was carried out to further detail the results. The cumulative percentage of trips by 1 minute MAE thresholds was calculated for the best model run of each transport mode, and the results can be seen in figure 6.13. The results reinforce that driving travel times are simulated much more accurately than travel times by public transport: 90% of driving trips are simulated with an error of up to 8 minutes, while the error of public transport trips for the same 90% figure is of up to 17 minutes. Considering a tighter margin of error of 10 minutes, the model accurately simulates 70% of bus trips and 60% of public transport trips. Very few trips were simulated with a travel time error higher than 20 minutes: 6.4% of bus trips, 7.1% of public transport trips, and only 0.4% of car trips.

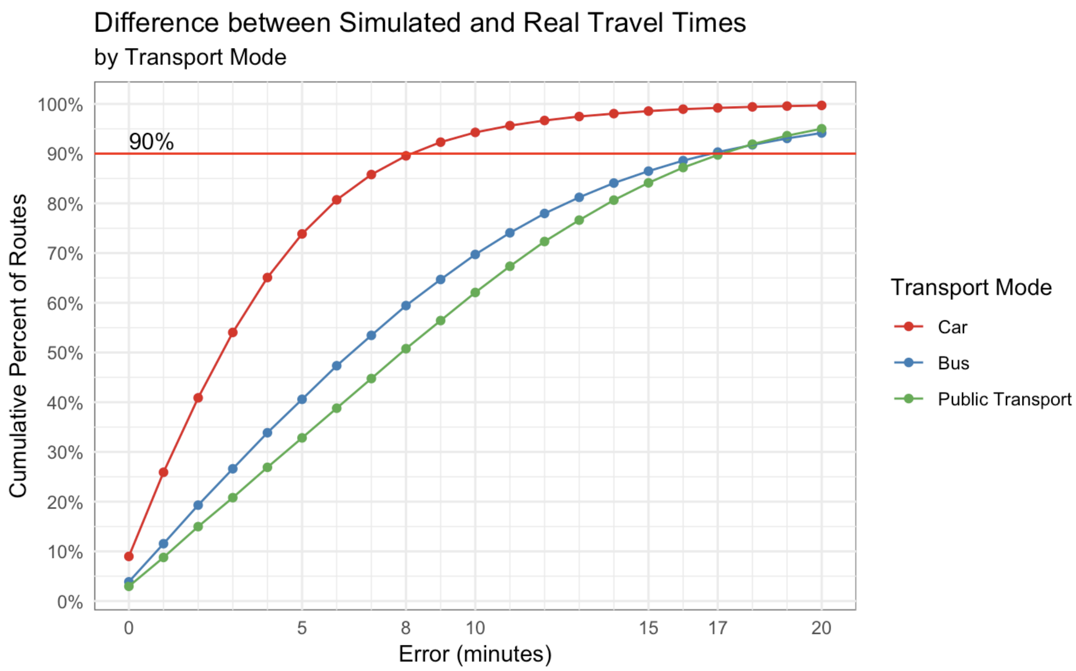

Figure 6.13: Cumulative percentage of trips by MAE threshold, by transport mode.

The higher accuracy of simulated driving travel times in comparison to other modes of transport is expected, as the road network is the main input used in the AxS model's environmental representation. Many of the specificities and uncertainties involved in travel by public transportation are missing in the model. For example, agents move at a constant speed from start to end of their trips, so time wasted in connections between modes are not accounted for. Also, every public transport trip includes slower sections made by foot, from home or work to the bus stop or train station, which are also not accounted for. Considering those limitations, the margins of error found in this analysis can be considered acceptable for the objectives of this study. This is significant and indicates the model's simple environmental representation and movement algorithm do not represent obstacles to the accurate simulation of realistic individual's travel times.

6.3 Discussion

The simulation exercises reported in this chapter were designed to provide a better understanding on the validity of the AxS model's results under different circumstances. Specifically, artificial trajectories and travel times generated by the model were compared to their real world counterparts, so the accuracy of outputs could be calculated. Overall, the results indicate the model presents good performance, producing highly accurate trajectories and travel times in the majority of the situations tested. Importantly, the limitations and cases where the model's outcomes are not as accurate were also identified and discussed throughout this analysis.

Validating trajectories and travel times separately was necessary due to the lack of a single dataset containing both information. Indeed, data availability is one of the challenges in validating models of complex social systems, as discussed in chapter 3. This exercise demonstrated how different data sources can be used in conjunction to validate agent-based models. This highlights the importance of open datasets and research collaboration, which made this analysis possible.

Both trajectories and travel times validation exercises reported here can also be described as a simplified calibration process. The results obtained provided an indicative of the range of parameters that produce the most realistic outcomes, both regarding agents' navigation and replication of travel times by transport mode. Those guidelines were used when defining the parameters for the simulation exercises presented in the following chapter.

The validation exercises also allowed for the computational performance and scalability of the AxS model to be tested. For instance, all the more than 12 million trajectories of the travel times validation analysis were simulated in a single desktop computer, an iMac with a 4 core i7 processor and 16GB of RAM, in little more than 2:30 hours. Simulating that amount of trips was only necessary for this validation experiment, though. Smaller samples of 200,000 trips were used in the simulation exercises reported in the following chapter, allowing real-time visualisation of the model's dynamics. It is not uncommon to find in the literature descriptions of analytical methods with much longer execution times and larger requirements in terms of computational resources, making this an advantage of the proposed method.

Chapter 7

Empirical Applications

This chapter presents empirical applications of the AxS model exploring accessibility and segregation in two large cities: São Paulo, Brazil, and London, United Kingdom. The objective of those applications is to demonstrate how the model can be used effectively to explore accessibility and segregation issues in real-world cities from a dynamic and individual-based perspective.

The study of São Paulo, presented in section **??**, explores inequality in access to services and opportunities between groups defined by income, level of education, and gender. The study of London, presented in section **??**, investigates ethnic segregation based on potential interaction among individuals of different ethnic groups in their usual trajectories to work. The chapter concludes with a discussion on the key findings from both empirical applications.

7.1 Accessibility Inequalities in the São Paulo Metropolitan Region

The São Paulo Metropolitan Region (SPMR), Brazil, is an official administrative area composed of the São Paulo municipality and 38 neighbouring municipalities, as shown in the map in figure 7.1. The SPMR has a total population of 19.6 million inhabitants, according to the 2010 Brazilian Census (IBGE 2010), of which 11.2 million live in the municipality of São Paulo. The southeast of the SPMR is the second most economically developed area in the region, where the three cities known as the ABC Paulista (Santo André, São Bernardo do Campo, and São Caetano do Sul) are located, with a combined population of 1.7 million inhabitants. Other important cities in the area are Guarulhos (1.3 million inhabitants), to the northeast of the region, and Osasco (700 thousand inhabitants), to the west. Most of the population is located into a continuously urbanised area with the city of São Paulo at its centre. The urban area is very extensive, which means some cities are located significantly far away from the region's centre. For example, Salesópolis is located 106 km to the east of São Paulo's city centre, and Juquitiba is located 87 km to the west, and both are 193 km apart from each other.

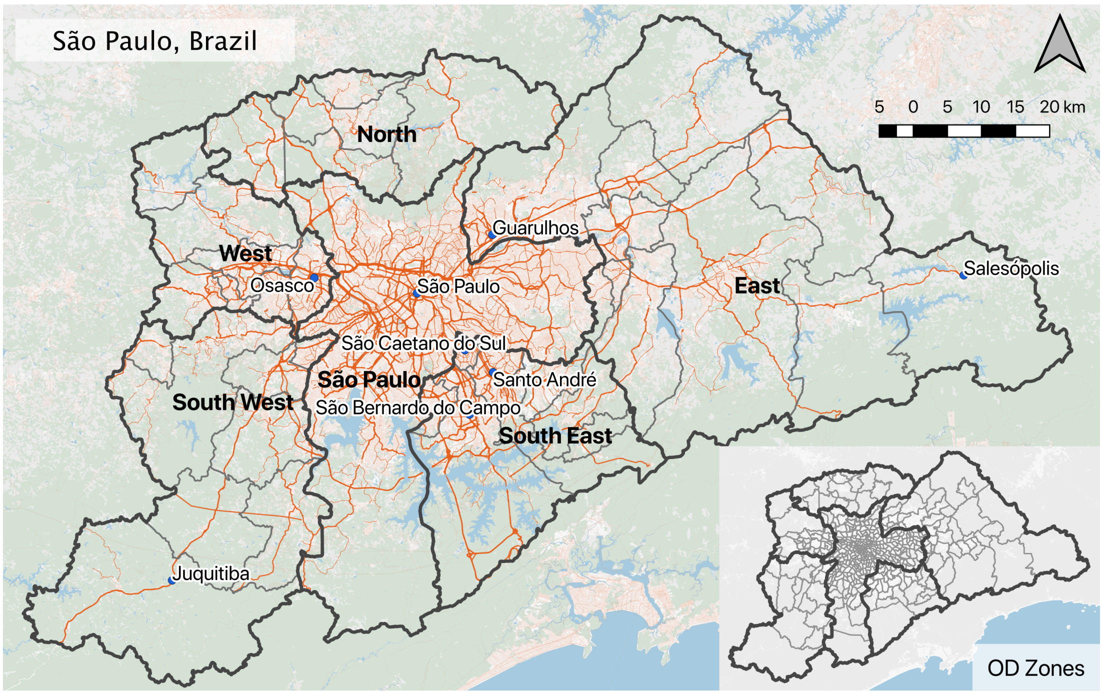

Figure 7.1: São Paulo Metropolitan Region (SPMR) subdivisions and primary road network.

In what follows, section 7.1.1 details the data sources used in this simulation, section 7.1.2 details the population distribution and modal split in the study area, and section 7.1.3 discusses the simulation process and results.

7.1.1 Data Sources

The OD dataset used in this experiment is the São Paulo Origin and Destination survey[1]. The dataset contains information on 38 million individual trips of all purposes (São Paulo 2007). Only the commuting to work trips were used in this study, in a total of around 8 million, and trajectories were simulated from home to work. The data were collected in 2007[2] by Metrô de São Paulo, the public company responsible for managing São Paulo's metro railway network and for planning public transport in the SPMR. Associated to the OD data, the dataset

1. http://www.metro.sp.gov.br/pesquisa-od/
2. A new OD survey was conducted in 2017, but the results from this survey were yet to be released at the time this experiment was conducted.

contains rich demographic information about the study area, of which income, education and gender were used in this study.

The OD survey divides the SPMR into 460 zones for data collection and processing, which can be seen in the inset map of figure 7.1. Those zones were defined to be compatible with other existing zoning systems, such as municipal boundaries, the census, and previous OD surveys. Homogeneity of land use and land cover, public transport options available, as well as physical barriers also were considered in their delimitation (São Paulo 2007).

The spatial distribution of opportunities in the study area was obtained from the CNEFE dataset (National Cadastre of Addresses for Statistical Purposes[3]), provided as part of the 2010 Brazilian Census (IBGE 2010). In this study, opportunities represent activities, services and amenities individuals can access in the city, such as shops, restaurants, healthcare facilities, schools, parks, and so on. Access to those opportunities is quantified by the AxS model's cardinal accessibility measure. The number of non-residential addresses in each street segment, available in the CNEFE dataset, was used as an indicator of the number of opportunities available at that location. Opportunities were not differentiated in terms of type, size, opening hours, or any other details. Hence, in this experiment, it is simply assumed that the larger the number of opportunities an individual can access, the more likely it is they will find suitable places to carry out their necessary activities.

Finally, spatial information on road network and urbanised areas was extracted from OpenStreetMap. Only segments tagged as motorway, trunk, or primary were included in the model, providing a level of detail compatible with the chosen grid resolution, which was set to 200m for this simulation (resulting in a grid of 825 x 550 cells). The region and spatial units' boundaries were obtained from São Paulo Metro as part of the OD survey dataset.

7.1.2 Characterisation of the Study Area

This section presents the characteristics of the study area according to the OD dataset. This means all information and percentages refer to the 8 million commuting people in the SPMR, and not to the entire population of 19.6 million people according to the census.

3. From the Portuguese Cadastro Nacional de Endereços para Fins Estatísticos.

Population Composition and Spatial Distribution

The composition of the working population in the study area is presented in tables 7.1 (by income class) and 7.2 (by education level). It is noticeable from table 7.1 that only 12% of the SPMR's working population belongs to the upper classes (A and B), while the remaining 88% of people are divided between classes C, D and E. Differences are less striking when looking at levels of education (table 7.2), as more than 60% of workers in SPMR have high school or university degrees. However, 6.7% of the population still had no formal education in the study area at the time of the survey.

Table 7.1: Proportion of individuals by income class in the SPMR.

(A) Income Class	Male	Female	Total
A	2.3%	2.4%	2.3%
B	9.7%	9.7%	9.7%
C	22.5%	23.3%	22.8%
D	39.3%	38.7%	39.1%
E	26.2%	25.9%	26.1%

Table 7.2: Proportion of individuals by education level in the SPMR.

Education Level	Male	Female	Total
University (Uni)	14.8%	21.2%	17.5%
High School (HiS)	42.3%	45.5%	43.7%
Middle School (MiS)	19.1%	15.6%	17.6%
Elementary School (ElS)	16.2%	12.1%	14.5%
No Education (NoE)	7.6%	5.5%	6.7%

The spatial distributions of places of residence and work of people in the study area are shown in figure 7.2, disaggregated by income class (maps 7.2A and 7.2C) and education level (maps 7.2B and 7.2D). The residential distribution maps by income class show people from the upper classes (7.2A1 and 7.2A2) tend to live in central areas, while the lower classes (7.2A4 and 7.2A5) live mainly in the peripheral areas of the SPMR. The residential maps by level of education show people with university degrees are concentrated in the core of the study area (7.2B1), people with high school degrees are more spread out throughout the SPMR (7.2B1), and people with lower levels of education tend to live farther away from central areas (7.2B3, 7.2B4, and 7.2B5).

The workplace maps show a polycentric distribution of jobs in the study area. The central area of the São Paulo municipality concentrates the largest number of jobs in the SPMR, for all income classes (7.2C) and educational groups (7.2D). Smaller employment centres can be found in other municipalities as well, mainly the larger ones such as the cities in the ABC region (Santo André, São Bernardo do Campo, and São Caetano do Sul), as well as Osasco and Guarulhos (the location of those cities can be seen in figure 7.1).

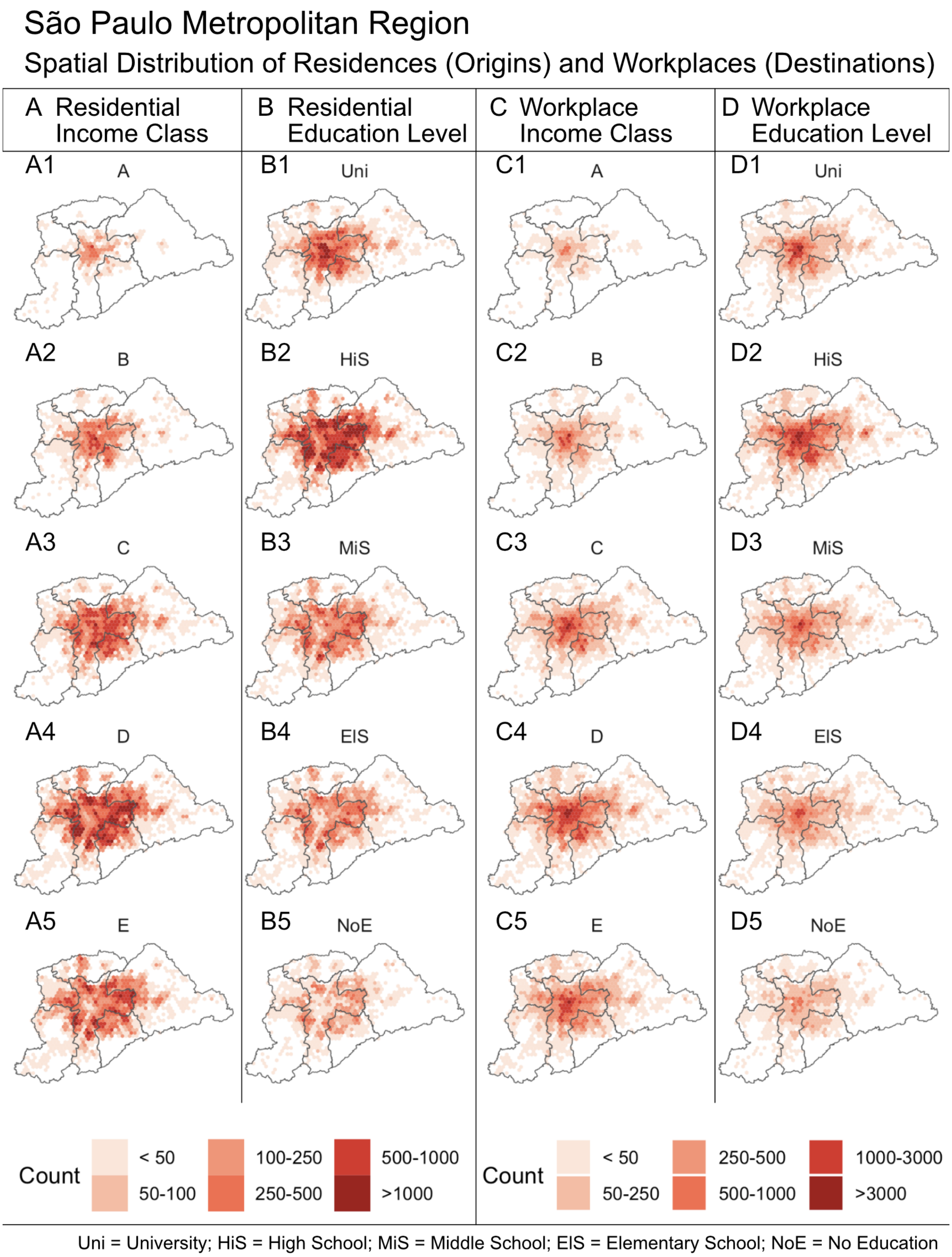

Figure 7.2: Residential and workplace distributions by income class and educational level in the SPMR.

Modal Split and Travel Behaviour

Buses are the predominant means of transport in the SPMR, used by 43.3% of the workers in the area. Private cars are used by 28.7%, while 19.5% of people walk to work. Only 3.9% of people use the metro system or suburban trains to commute, which is a very low percentage compared to similarly sized world cities. This can be explained by the limited reach and capacity of the SPMR's rail network. Gender differences are also evident on the modal split, with more females riding buses and walking, while more males drive to work. The least used modes of transport, namely bicycle and motorcycle, are used almost exclusively by males.

Table 7.3: Modal split in the SPMR.

Mode	Male	Female	Total
Bicycle	2.1%	0.2%	1.3%
Bus	38.8%	49.5%	43.3%
Car	33.5%	22.2%	28.7%
Metro / Train	3.8%	3.9%	3.9%
Motorcycle	5.3%	0.6%	3.3%
Walking	16.5%	23.6%	19.5%

There are stark differences regarding socio-economic class and transport modes, as shown in table 7.4. The higher-income individuals (income classes A and B) mostly drive to work (80% and 64.1%, respectively). Individuals of class C use cars and buses at similar rates (40.4% and 37.3%, respectively), while about half of individuals in classes D and E ride buses. More than 20% of individuals of classes D and E walk to work (20.7% and 27%, respectively), in comparison to less than 9% of individuals of classes A and B. Lower-income individuals (classes D and E) are also the ones who cycle to work more frequently (1.4% and 2.3%, respectively), although the use of this transport mode is much lower than any of the other modes. Riding motorcycles seems to be an alternative for individuals of the lower classes to increase their mobility, although their use is fairly limited, ranging from 3.2% to 3.8% for classes C, D and E. The metro and trains are used by similar rates of individuals of all classes, at rates between 3.4% and 4%.

Table 7.4: Modal split by income class in the SPMR.

	Income Class				
Mode	A	B	C	D	E
Bicycle	0.1%	0.5%	0.5%	1.4%	2.3%
Bus	7.5%	20.2%	37.3%	48.9%	51.9%
Car	80.0%	64.1%	40.4%	21.5%	11.5%
Metro / Train	3.4%	3.9%	3.8%	3.70%	4.0%
Motorcycle	0.9%	2.3%	3.2%	3.8%	3.2%
Walking	8.1%	9.0%	14.7%	20.7%	27.0%

When level of education is considered, people with university degrees present the larger differences to the other groups regarding their choice of transport mode, as shown in table 7.5. They tend to use private cars much more frequently than people from the other groups: 57.4%, compared to 18.2% to 25.3% of the other groups. The bus is the predominant means of transport for all groups, apart from the college educated, while a significant number of individuals with lower levels of education walk to work. A very small parcel of the population cycle to work. The ones who do cycle to work tend to be the less educated and lower classes, suggesting bicycles are used more due to economic reasons than by choice.

Table 7.5: Modal split by education level in the SPMR.

	Education Level				
Mode	University	High School	Middle School	Elementary School	No Education
Bicycle	0.2%	1.0%	1.7%	2.3%	2.9%
Bus	25.4%	48.9%	45.6%	43.4%	46.9%
Car	57.4%	25.3%	19.3%	20.7%	18.2%
Metro / Train	4.7%	3.8%	4.0%	3.0%	3.4%
Motorcycle	1.8%	4.0%	3.8%	2.9%	2.2%
Walking	10.3%	17.0%	25.6%	27.7%	26.4%

7.1.3 Simulation Experiment, Results and Discussion

In this experiment, a sample population of 200,000 agents was generated from the 8 million trips in the OD dataset. The sample is representative of the study area, which means the proportion of trips between origin and destination zones, the demographic characteristics of the population, and the modal split of the study area of the original OD data are the same in the sample.

The movement speed of each mode of transport was calculated based on information available in the OD dataset. Travel times and mode of transport, in the OD dataset, were informed by the survey's respondents. The movement speeds input in the simulation were calculated based on those stated travel times and the shortest road network distance between zone centroids. The resulting average movement speeds in real world (*kilometres per hour*) and model units (*cells per iteration*), per transport mode, is shown in table 7.6.

Table 7.6: Movement speeds in the SPMR (cells per iteration - cpi).

	Speed	
Mode	km/h	cpi
Bicycle	12.0	1.0
Bus	9.6	0.8
Car	19.2	1.6
Metro / Train	14.4	1.2
Motorcycle	24.0	2.0
Walking	4.8	0.4

Parameter values were chosen based on the sensitivity analysis and validation tests shown in chapters 5 and 6, respectively, and are: *search radius = 7 cells*, *angle of vision = 120°*, and *road weight = 0.1*. The simulation was run with 500 simultaneously active agents (*number of agents* parameter). Accessibility was calculated considering *time budget = 90 minutes* for all agents, regardless of social group. This value was deemed as a reasonable representation of the free time an individual may have in-between main activities, enough to carry out activities such as grocery shopping, attending to a doctor's appointment, or socialising with friends. The minimum time for participating in an activity was set to 15 minutes (*minimum activity time* parameter), which acts as a low threshold value that represents quick everyday errands. Although the *time budget* and *minimum activity time* parameters are somewhat arbitrarily selected, their setting does not introduce significant bias to the results, since they affect all agents in the simulation equally. An ideal simulation scenario would include a different time budget for each individual, and a different amount of time required for each activity, but that would require detailed data that is not available for this case study.

The resulting aggregated flow pattern produced by the simulation is shown in figure 7.3. The pattern shows how dominant the municipality of São Paulo is in the metropolitan region, attracting trips from all neighbouring municipalities. The flow pattern also helps to identify sub-centres of activity in the region, which are visible to the south, east and west of central São Paulo. An animation of the simulation process can be seen online at www.mvpsaraiva.com/thesis.

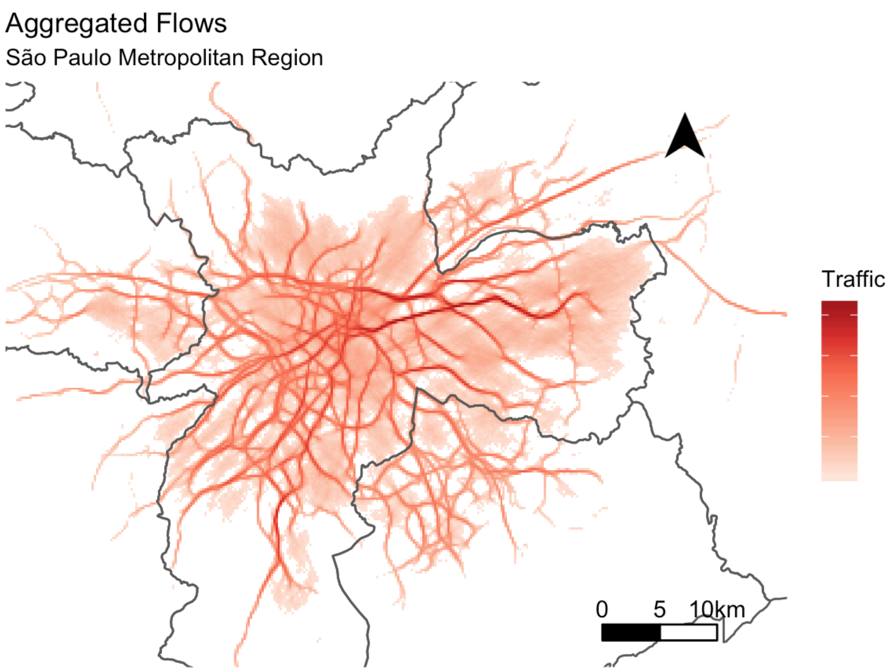

Figure 7.3: Aggregated flows in the SPMR.

The following sections report on the outputs produced by the simulation. Travel time and distance statistics by population group are presented, followed by a discussion on the accessibility levels of each group.

AxS' Outputs on Travel Distance and Time

Commuting time and distance statistics are useful to understand patterns of travel behaviour of individuals and groups. The results of those statistics generated by the AxS model are shown in the graphs in figure 7.4. The graphs present the average value of each metric for all individuals in each group, as well as the lower and upper quartiles. It is noticeable how the difference in average travel distances between groups is no higher than 1.6 km, which is the difference between income classes A and C. The difference in the upper quartile of travel distances is slightly higher, at 3.3 km between income groups A and E. In general, average travel distances do not present large differences between groups, which is significant considering the geographical extent of the study area.

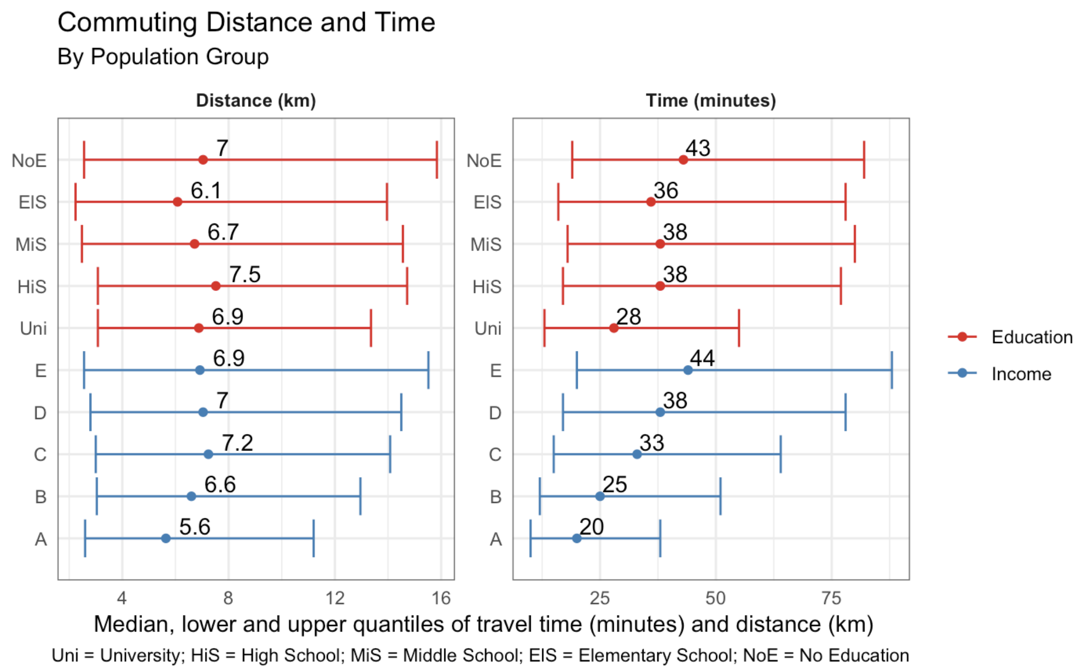

Figure 7.4: Simulated travel distances and times by educational group and income class based on AxS outputs.

More significant differences were found in travel times between groups, as well as within each group. For instance, people with university degrees tend to spend less time commuting than all other educational groups. Regarding income class, the plot shows the lower classes spend consistently longer times commuting than the upper classes. The two plots of figure 7.4 show the upper classes are able to compensate for longer travel distances by using faster means of transport, thus being able to live further away from work without being penalised for that. In contrast, the lower classes have no means of counteracting the poor public transportation infrastructure of São Paulo. This is important because many accessibility studies focus on travel distances rather than time, while the model's results indicate travel time is a more significant factor in identifying inequalities in mobility between social groups.

Even more significant differences can be found when comparing travel times and distances by mode of transport. The plot of figure 7.5 synthesises mean, lower and upper quartiles of travel time and distance by transport mode. It is noticeable that longer commuting distances in the SPMR are usually made by metro and train (17.2 km on average), which explains the long travel times of trips made by those modes of transport (72 minutes on average). Travel times by bus are fairly similar to travel times by metro and train, even though travel distances for the former are significantly shorter than for the latter. This is a consequence of the low speed of the buses in São Paulo, which affect disproportionally the lower classes who use the bus more often (as shown in table 7.4). Individual and motorised modes of transport (cars and motorcycles) allow relatively short travel times, similar to active modes of transport (walking and cycling), besides the longer distances of trips made by the former modes. Although travel times presented here are simulated and do not consider road congestion, which are a frequent phenomenon in São Paulo, the results indicate there is a clear incentive for people to use individual rather than public transport in their daily commute. This trend both worsens road congestion and harm the lower-income people who cannot afford a motor vehicle.

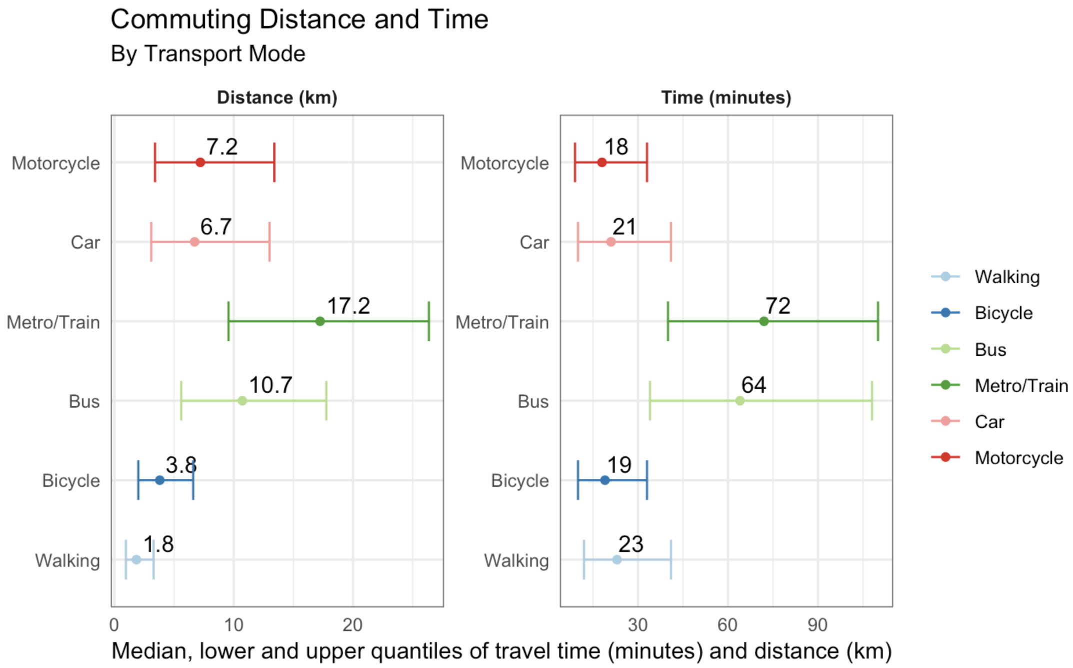

Figure 7.5: Travel distances and times by mode of transport.

Finally, interesting insights can be gained by analysing the interquartile ranges (IQR - the difference between the upper and lower quartiles) of travel distances and times by groups and transport modes, shown in the plots of figures 7.4 and 7.5. Although the groups' average values are indicative of how groups fare in comparison to each other, the overall high IQRs indicate there are large differences between individuals who belong to the same group (figure 7.4). For example, the IQRs of travel times by income class are much higher for the lower classes than for the upper classes. This means that commuting times of some individuals in class E are very short, while other individuals of the same class have extremely long commutes. The same is true for bus, metro, and train users, as shown in figure 7.5, who present much higher within-group variation in commuting time than car and motorcycle users. These results present an insight on the high inequality levels of the SPMR, which will be further explored in the remainder of this section.

AxS Accessibility Outputs

Although travel time and distance statistics presented so far are useful, accessibility metrics present a more comprehensive view on each individual's access to opportunities in a city. Such metrics combine residential and workplace locations, travel time, and travel distance into a single indicator. Since land use information for the study area is available through the aforementioned CNEFE dataset, this section focuses on cardinal accessibility, that measures the number of opportunities accessible to individuals in their free time. The geometric accessibility, that measures the size of individuals' accessible areas in their free time, is discussed when relevant.

Cardinal accessibility by mode of transport is shown in the boxplots of figure 7.6. The low accessibility values for bus, metro, and train users are striking: the median accessibility of users of those modes is zero, which means a 90 minutes time budget is not enough for many of those individuals to participate in activities other than work. This result is related to the longer commuting times of the users of those modes, and also hints at the ineffectiveness of São Paulo's public transportation system. Accessibility levels of bus, metro, and train users also present high inequality, as indicated by the high IQR of those modes in the graph. Motorcycles represent great accessibility gains for people, specifically the lower classes who are more likely to use this mode (see table 7.4). However, only 3.3% of commuters used this mode at the time of the OD survey, in comparison to the 43.3% of people who use buses and have very low levels of accessibility (as shown in table 7.3).

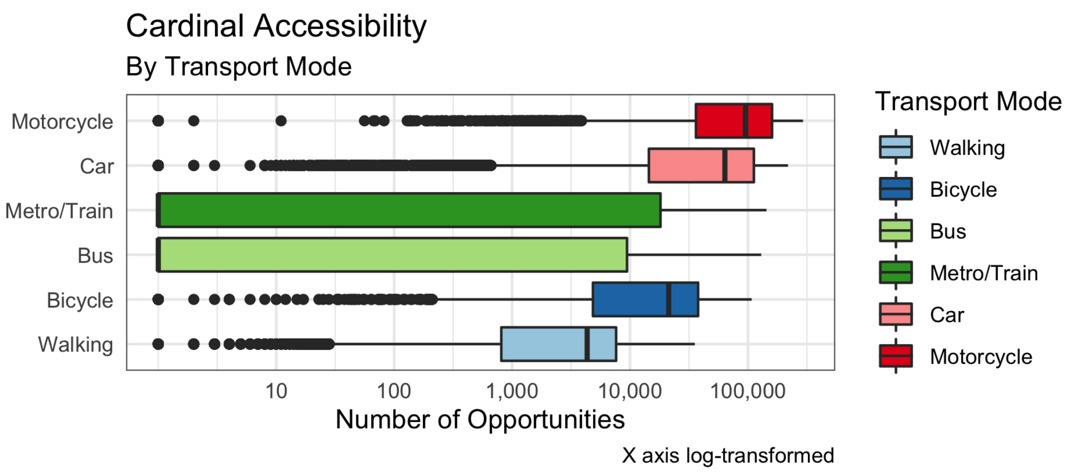

Figure 7.6: Boxplots of cardinal accessibility by mode of transport.

The cardinal accessibility plots by income class and education level (figure 7.7), present patterns in line with what was discussed so far: accessibility tends to increase with the individual's level of income or education. The plots show individuals with university degrees have the highest levels of accessibility, people with no formal education have the lowest levels, while individuals who attended from elementary to high school have similar levels of accessibility in between those two former categories. The social divide regarding income is also very clear, as the accessibility of each income group is higher than that of the groups below. The plots also show the high disparity in accessibility between individuals of the same group, indicated by the length of each group's box in the plot. An interesting pattern is that the lower quartile of almost all groups is zero, with the exception of the group of people with university degrees and income classes A and B. This reinforces the idea that a large number of people in the study area have very long commutes, which consume most of the 90 minutes time budget set for this experiment. Only the upper classes seem to be able to more easily fit extra activities in their daily schedules.

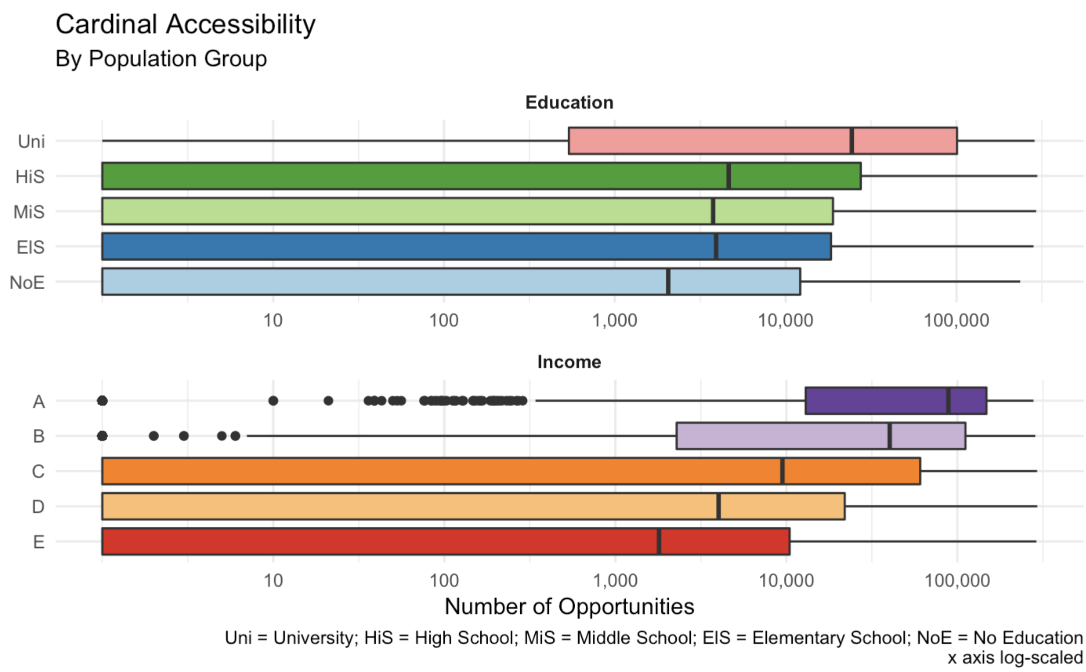

Figure 7.7: Boxplots of cardinal accessibility by education level and income class.

Disparities in accessibility level can be more clearly seen in the plots of figure 7.8. The barcharts show the percentage of individuals in each group with accessibility above and below the average accessibility of the study area. It is noticeable that only income classes A and B have more than 50% of their individuals with accessibility above average. On the opposite side, 87.4% of individuals in class E have accessibility below average, highlighting the deep inequalities in the study area. When analysing groups by education level, the divide between college educated individuals and all others is clear: the lower the level of education, the higher the chance of having accessibility below average. However, 51.9% of people with college degrees also have accessibility below average, indicating there are still many highly educated individuals with low accessibility in the SPMR.

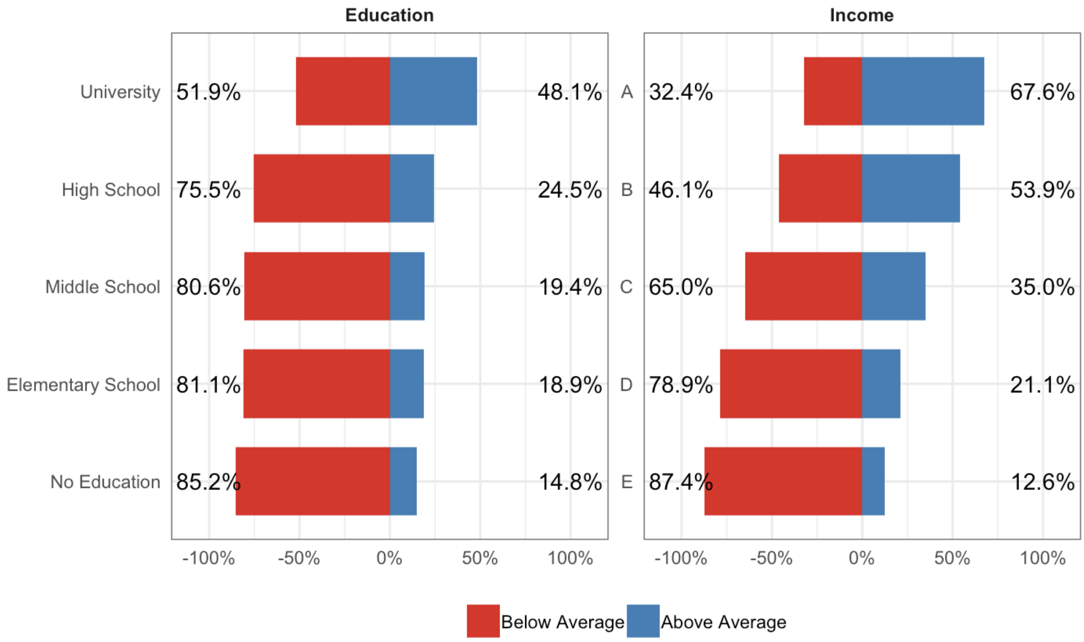

Figure 7.8: Percentage of each group's population with cardinal accessibility levels above and below the average.

The residential locations of individuals with accessibility below the lower quartile of each group were mapped and the results can be seen in figure 7.9. The red hotspots in the maps highlight the places where most people with low access to opportunities live. Those hotspots are mainly located in the east and south of the São Paulo municipality. The hotspot to the south is located at the edge of the urbanised area, although far away from the municipality's limit, while the hotspot to the east is located at the border of São Paulo with neighbour municipalities. The hotspots indicate that most people with low accessibility belong to income classes D and E, and have a high school degree.

The red hotspots in the maps indicate where accessibility-increasing policies and investments would be more effective. For instance, better transportation links between the eastern and southern edges of the São Paulo municipality to the city centre would benefit the largest number of individuals with low accessibility. Similarly, policies that create incentives for business to invest and/or relocate to those areas could also be effective in increasing the accessibility of local residents without creating extra pressures in the existing transportation system nor requiring costly and time-consuming investments in the public transport infrastructure. One such example of the effectiveness of the match between residential and workplace locations in creating accessibility can be seen in the city centre, as will be demonstrated in the following analysis.

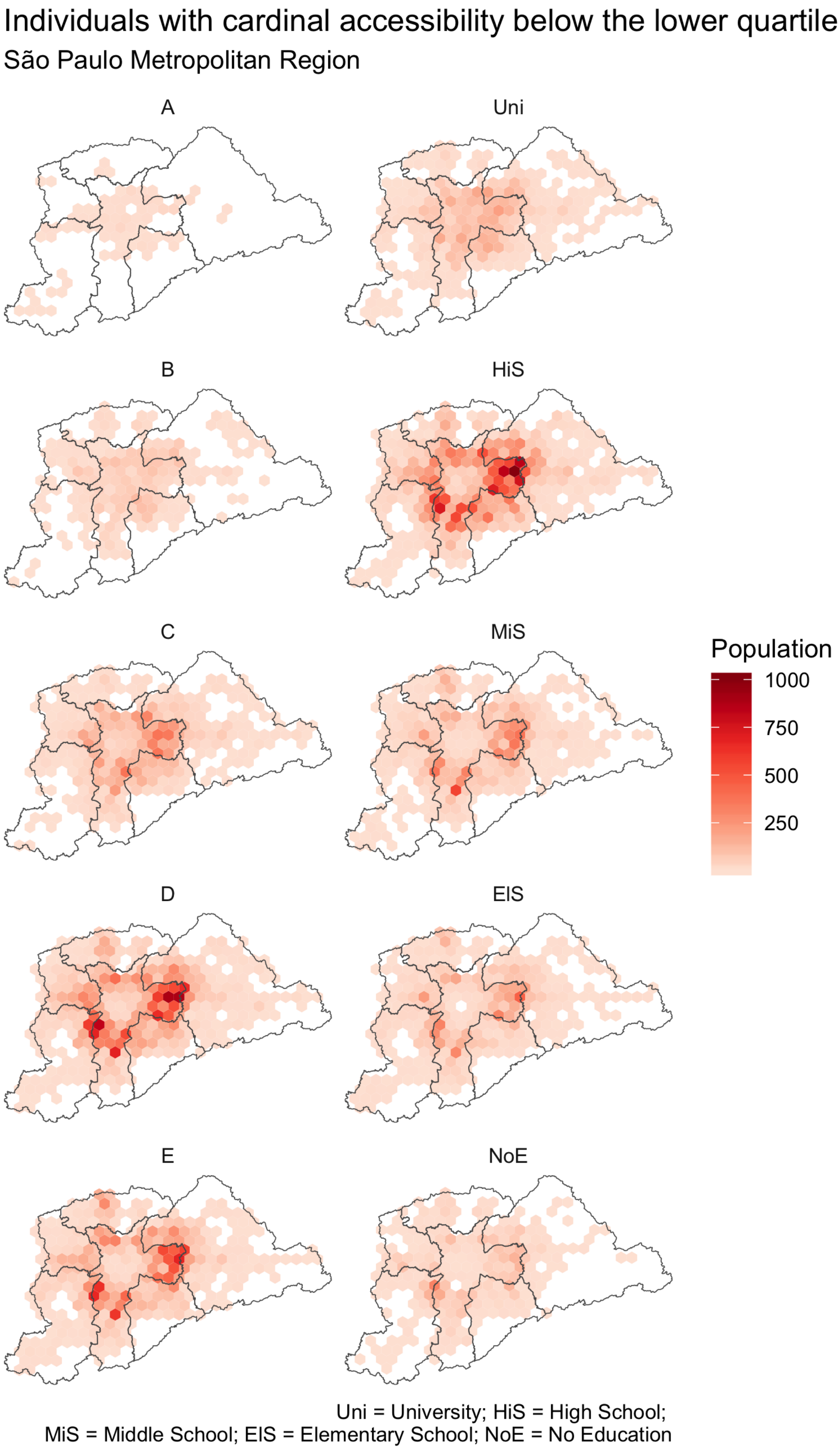

Figure 7.9: Individuals with accessibility below the lower quartile, by income class and education level.

Cardinal and geometric accessibility levels can be compared to analyse the trade-off between residential location and mobility. The difference between the two metrics can be used to highlight two groups of individuals: a) people who have access to many opportunities in a small geographic area, and b) people with access to a large area due to their high mobility, but few opportunities are located in that area. The resulting maps can be seen in figure 7.10. In those maps, high and low accessibility mean, respectively, above and below the average accessibility of each group. It is noticeable how the individuals with high cardinal accessibility and low geometric accessibility concentrate in the central areas, indicating those individuals chose to live in dense areas with many opportunities where owning a motor vehicle is deemed unnecessary. On the other hand, individuals with high geometric accessibility and low cardinal accessibility live mostly in the peripheral areas of the SPMR. These are individuals that, despite their high mobility due to car or motorcycle ownership, live and/or work in areas with few opportunities. The existence of this group of individuals demonstrates that mobility alone is not enough to improve people's access to opportunities.

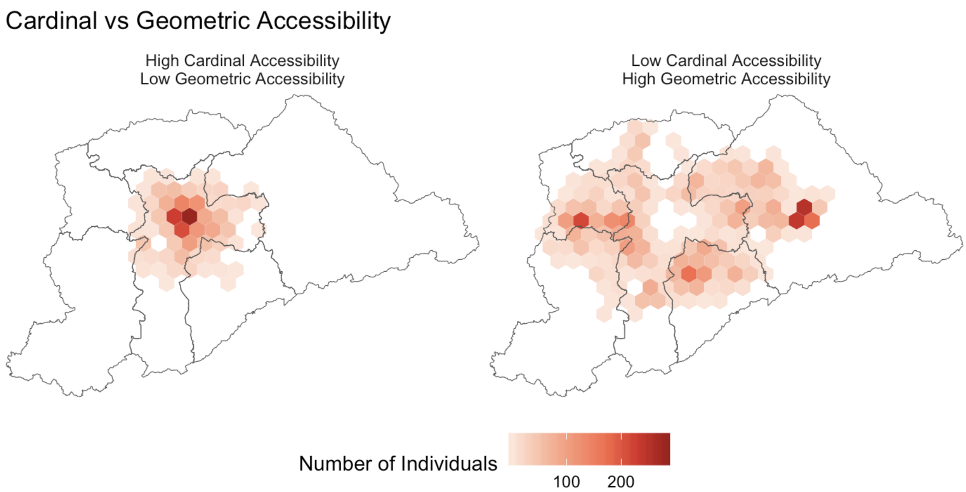

Figure 7.10: Cardinal versus geometric accessibility.

It is important to note that, even though the map is insightful, not many individuals are included in those categories shown. Specifically, only 1.2% of the population in the sample have simultaneously high cardinal accessibility and low geometric accessibility, and 2.8% of the population are in the opposite category. Hence, the individuals mapped in figure 7.10 are the exception to the rule, but they allow to demonstrate the level of detailed analyses that are possible when studying accessibility at the individual level.

Gender Gap

The last analysis to be presented in this case study is about gender differences in accessibility. This kind of disaggregation in accessibility is notably difficult to achieve with place-based measures and using traditional aggregated. The contrast in accessibility of male and female individuals, both combined and disaggregated by income class and education level, is shown in figure 7.11. The plot shows that males' median accessibility is 44.9% higher than females', in general. Considering the literature on the subject, these results are both expected and surprising, as follows. It is usually acknowledged that females have lower accessibility than males, mostly due to assuming extra responsibilities in taking care of the household (Kwan 2000a; Schwanen, Kwan, and Ren 2008). However, the results obtained here do not include any of those constraints that affect females more frequently. This means that even when conditions are kept artificially balanced, females in the SPMR have a disadvantage in access to opportunities when compared to males.

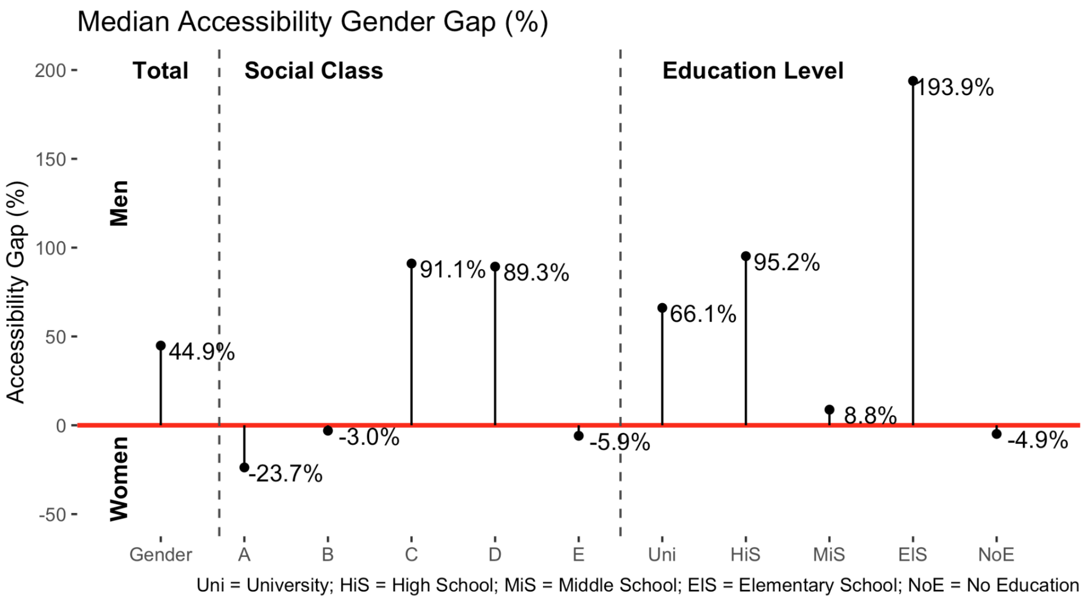

Figure 7.11: Gender gap in accessibility, by education level and income class.

The gap in accessibility can be further explored when the results are disaggregated by income class and education level. First, it is noticeable that upper class females (class A) have accessibility 23.7% higher than males in the same class, while the accessibility of females in class B is similar to the accessibility of males of the same class (3% in favour of females). However, the gap manifests more strongly in classes C and D. In those classes, males' accessibility is around 90% higher than females'. In class E, the gap tends slightly in favour of females, at 5.9%. However, accessibility of individuals of class E is already very low, as shown previously, so females in this class lack access to many opportunities even though they are slightly better off than males of the same class. These results indicate that the accessibility gap affects more strongly females in the classes C and D, who make the bulk of the workforce in the study area.

Figure 7.11 also shows gender gaps in accessibility by education level, more strikingly among people who only attended elementary school, whose gap is 193.9% in favour of males. The gap among high school graduates, which is the majority of the working population in the study area (43.7%), is also very significant at 95.2%. Only at the lowest level of education the gap tends slightly in favour of females, at 4.9%, but accessibility levels of this group are already very low compared to the other groups.

The gender differences were spatialised and can be seen in the map of figure 7.12. Overall, there is some parity between genders among residents from the central area of the SPMR. However, a large ring around the region's centre, shown in dark blue in the map, indicates areas where males have significantly more accessibility than females. Small red clusters located mostly in the peripheral areas of the metropolitan region indicate where females have larger advantages. In general, the map reflects the overall advantage males have in relation to females in terms of access to opportunities in the SPMR.

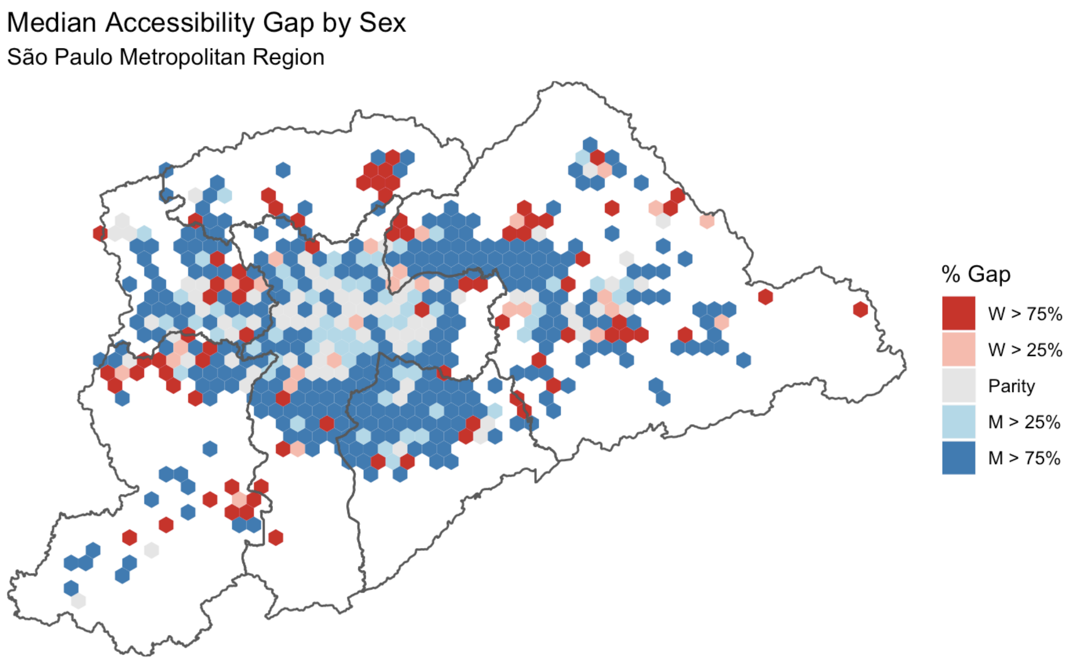

Figure 7.12: Median accessibility gap by gender in the SPMR, by place of residence.

7.1.4 Discussion on the São Paulo Case Study

This case study highlighted and quantified the high level of inequality in access to services and urban opportunities in the São Paulo Metropolitan Region. The simulations allowed for analyses that were not possible using available datasets and traditional techniques. The results demonstrated inequality in accessibility is tied to social and economic inequality as people with lower income and less education have the lowest accessibility levels. The low efficiency of the public transportation system in the SPMR, on which the lower classes particularly depend on, is a factor that tends to perpetuate those inequalities. This is more evident when travel distances are similar between higher- and lower-income individuals. In those cases, the results show higher-income individuals are able to afford faster means of transport and, consequently, have access to a significantly higher number of opportunities than lower-income individuals.

Within-group inequalities in accessibility were also identified in this case study, in a finer level of detail made possible by the individual-based approach of the methodology proposed in this thesis. For instance, it was found that many individuals in the lower-classes are able to improve their accessibility by taking individual measures, such as acquiring a car or, more frequently in the lower-income classes, a motorcycle. Living close to work and in areas of high density of opportunities is an option available to few individuals in lower-income classes.

Finally, the methodology proposed also allowed us to quantify gender differences in accessibility. The results demonstrated that, in general, males of income classes C and D have access to almost twice as many opportunities and services as females from the same groups. Taking into consideration that those females belong to groups that already have lower accessibility than the upper classes, and that many of them are heads of households, it is noticeable how low income females are specifically affected by low access to opportunities and services. High income females, on the other hand, have higher (class A) or similar (class B) accessibility to males of the same groups, partly due to the fact they do not depend on public transport. These results are an important contribution of this thesis, as measuring accessibility by gender is something that is difficult to achieve at this scale and level of detail.

7.2 Dynamic Segregation in Greater London

This section details the application of the AxS model to London, United Kingdom, as delimited by the Greater London Authority (GLA) area. The study area can be seen below, in figure 7.13. The map shows the boundary of the GLA and its subdivision in 33 boroughs, as well as the city's main road network and environmental features. The population of Greater London is estimated to be of approximately 8.8 million inhabitants as of 2018 (Office For National Statistics 2018). London is well known for the ethnic diversity of its population, hence its choice as a case study on segregation in this thesis.

In what follows, section 7.2.1 details the data sources used in this simulation, section 7.2.2 details the ethnic composition and spatial distribution of the population in the study area, and section 7.3.3 discusses the simulation process and results.

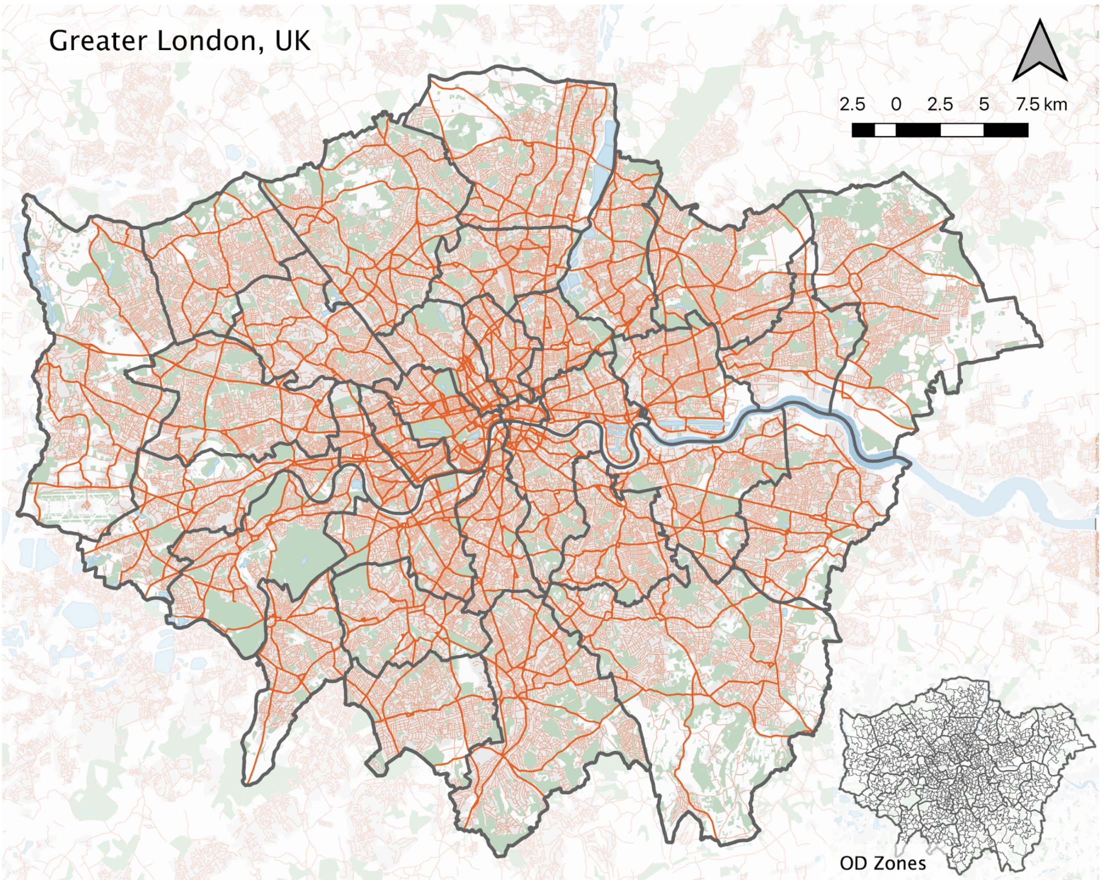

Figure 7.13: Greater London Authority area and primary road network.

7.2.1 Data Sources

OD data for this experiment were obtained from the UK census, as part of its flow datasets. Specifically, data were obtained from census table WU08AEW, which contains information about people's places of residence and work for England and Wales, disaggregated by ethnicity. Only trips starting and ending within the GLA's limits were extracted from the OD table and used in the simulation.

The inset map on the bottom-right corner of figure 7.13 shows the subdivision of the GLA into smaller census tracts at MSOA level (Middle-layer Super Output Area), which were the spatial units used in this study. MSOAs are formed by aggregating smaller Output Areas (OAs, the smallest areal units of the census) and Lower-layer Super Output Areas (LSOAs), containing on average 7,200 residents, with a minimum of 5,000. MSOAs were used in the simulations presented in this chapter because flow data disaggregated by ethnicity are made available by the UK Office of National Statistics (ONS) mainly at this scale[4].

Spatial information on geographical and political boundaries, such as the GLA limits and its subdivision in MSOAs, as well as urbanised areas, was

4. Data at smaller scales can be accessed in a secure environment provided by the ONS, due to privacy concerns. However, this option was not explored in this study.

obtained from the census[5]. Information on road network was extracted from OpenStreetMap, which included road segments tagged as motorway, trunk, or primary. Those road segments represent London's main road network at a scale compatible with the chosen grid resolution of 200m used in this study. All GIS data for the study area were converted to regular grids of 325 x 275 cells to be used as inputs into the model.

7.2.2 Characterisation of the Study Area

This section presents the characteristics of the study area obtained from the OD dataset. It is important to note the numbers and percentages presented here do not refer to the total number of jobs available in the GLA, nor its entire resident population. Rather, those numbers include only workers who have both their places of residence and employment within the GLA, and exclude those who commute from or to areas outside London.

Population Composition

UK residents are classified, according to the census, into 18 groups based on their ethnic origin. Those groups are usually aggregated into 5 supergroups based on race: White, Mixed, Asian, Black and Other. However, a different classification was used in this study, following Barros and Feitosa (2018). Instead of aggregating the ethnic groups by race, the authors proposed a grouping system based on the similarities between groups' spatial patterns of residential location, which were identified using Pearson correlation and spatial autocorrelation (Moran's index) analyses. Following that methodology, the population was classified into 4 supergroups: White British, Black, South Asian and Other. The composition of each group, as well as the number of commuting trips made by individuals of each group in the GLA, is shown in table 7.7.

Table 7.7: Flows distribution by ethnic group in the GLA.

Group	# Trips	%
G1 – White British	1,389,951	47.5%
G2 – Black (Black African, Black Caribbean, Black Other, Mixed White Black Caribbean, Mixed White Black African)	372,522	12.7%
G3 – South Asian (Asian Indian, Asian Pakistani, Asian Bangladeshi, Asian Other)	469,248	16.1%
G4 – Other (White Irish, White Other, Mixed Other, Asian Chinese, Other Arab, Other)	693,322	23.7%
Total	2,925,043	100%

5. https://census.ukdataservice.ac.uk/get-data/boundary-data.aspx

Spatial Distribution of the Population

The residential and workplace distributions in the GLA are shown in figure 7.14. It is noticeable places of residence are evenly distributed throughout the city, while workplaces are highly concentrated in the city centre. In fact, the City of London alone concentrates 8.4% of the jobs in the entire GLA, and 20% of the jobs are located in just 5 MSOAs in central London and at the Canary Wharf area (London's second CBD), as shown in table 7.8. A few other areas contain significant number of job opportunities, which can be seen in red in the map of figure 7.14, including the western boundary of the GLA where Heathrow airport is located.

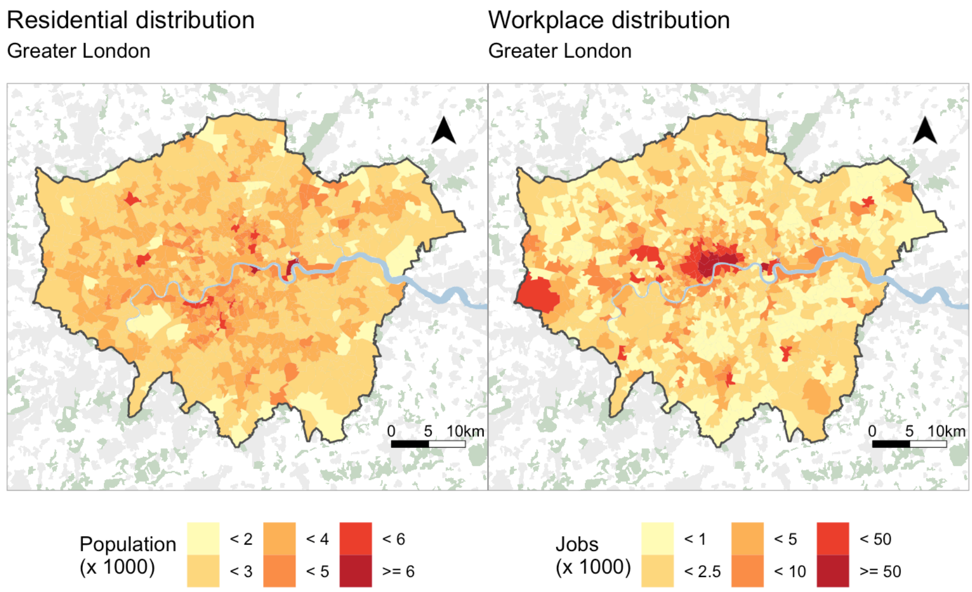

Figure 7.14: Residential (left-hand side) and workplace (right-hand side) distribution in the study area.

Table 7.8: MSOAs with larger number of jobs in the GLA.

	MSOA	Jobs	%
E02000001	City of London 001	244,231	8.4%
E02000977	Westminster 018	116,646	4.0%
E02000972	Westminster 013	104,328	3.6%
E02006854	Tower Hamlets 033 (Canary Wharf)	76,240	2.6%
E02000979	Westminster 020	60,291	2.1%
Total		601,736	20.60%

The residential and workplace location patterns of each ethnic group were mapped and can be seen in figures 7.15 and 7.16. Those maps show the residential distribution pattern presents larger variation between the groups in comparison to the workplace distribution. The similar workplace distribution maps can be explained by the aforementioned high concentration of jobs in London's central areas, which is similar for all groups.

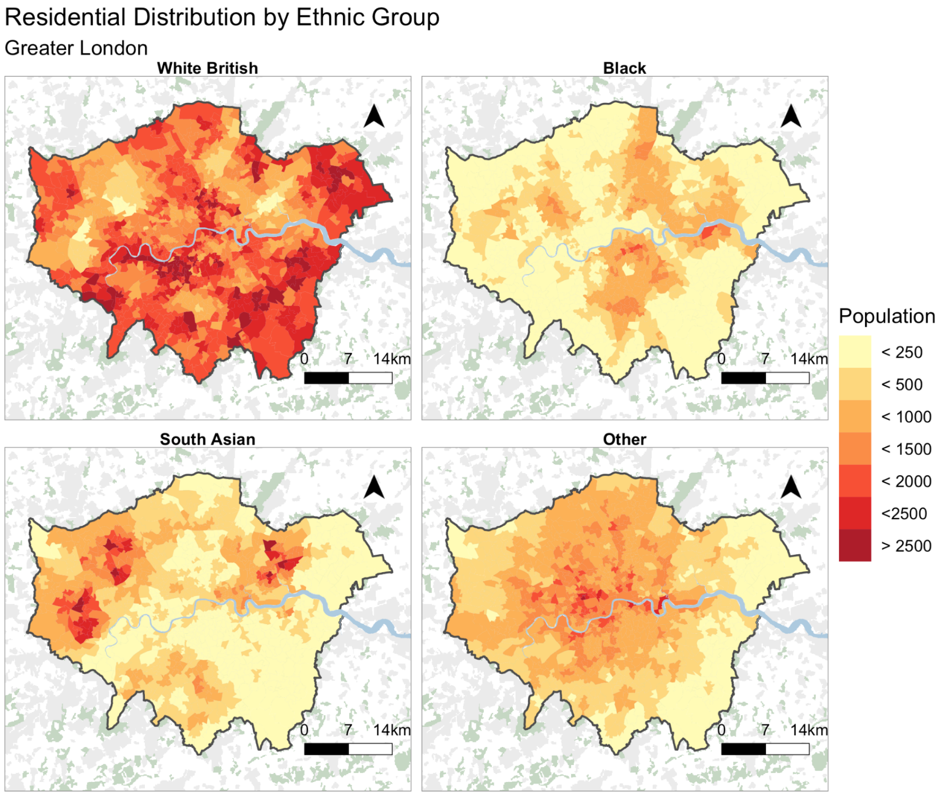

Figure 7.15: Residential distribution by ethnic group.

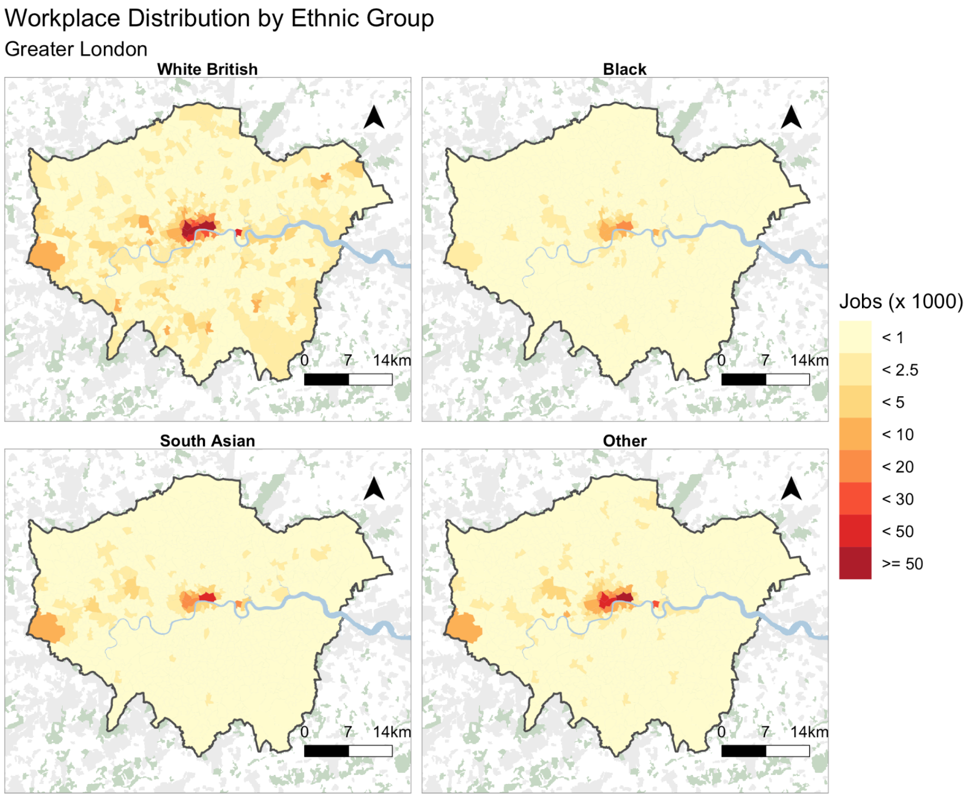

Figure 7.16: Workplace distribution by ethnic group.

The visual analysis of the groups' spatial patterns is corroborated by a Pearson's correlation analysis between the groups' places of residence and work, shown in table 7.9. On the one hand, the analysis shows there is either no correlation or negative correlations between the group's places of residence, with the lowest correlation value found between the White British and South Asian groups ($r = -0.53$). On the other hand, the correlations between places of work are very high for all groups, with values of r ranging from 0.92 (between the South Asian and Black groups) to 0.99 (between the White British and Other groups).

Table 7.9: Pearson correlation coefficients (r) between groups' places of residence and work.

	White British	Black	South Asian	Other	
White British	-	0.95	0.94	0.99	Workplace
Black	-0.31	-	0.92	0.94	
South Asian	-0.53	-0.03	-	0.95	
Other	0.12	0.09	-0.09	-	
	Residence				

7.2.3 Simulation Experiment, Results, and Discussion

In this experiment, a sample population of 200,000 agents was generated from the 2.9 million trips in the OD dataset that have their start and end points within the GLA. The proportion of trips between origin and destination zones, and the ethnic characteristics of the population obtained from original OD data were applied to the sample, so it is representative of the characteristics of the study area discussed in the previous section.

Parameter values used here are the same as in the previous experiment, chosen based on the sensitivity analysis and validation tests. Those values are: *search radius = 7 cells*, *angle of vision = 120°*, and *road weight = 0.1*. The number of simultaneously active agents was set to 500 (*number of agents* parameter). Accessibility was not calculated in this exercise, so the accessibility-related parameters (*time budget* and *minimum activity time*) were not set. Since the OD dataset does not contain information regarding transport mode, all agents were assigned *movement speed = 1 cell per iteration*[6].

The resulting aggregated flow pattern produced by the simulation is shown in figure 7.17. The map shows the pattern of flows in the GLA is highly concentrated towards the central area, which is expected due to the jobs distribution in the GLA previously shown in figures 7.14 and 7.16. An animation of the simulation process can be seen online at www.mvpsaraiva.com/thesis.

6. Information on transport mode can be found in a different census table (WU03EW) which, for its turn, does not contain information on ethnicity.

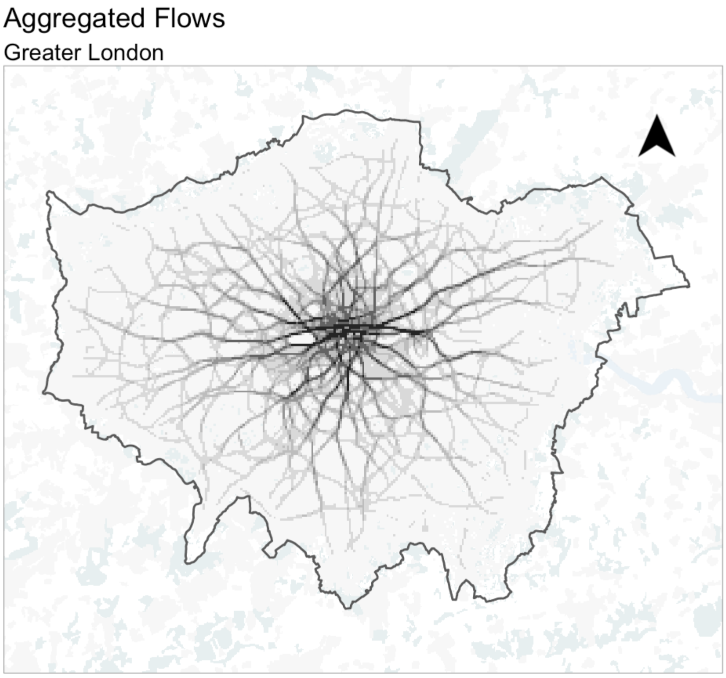

Figure 7.17: Aggregated flows in Greater London.

The following sections report on the AxS segregation outputs produced by the simulation. Visual analyses of the groups' flow patterns are presented first, followed by quantitative indicators of segregation based on diversity and copresence.

Flow Patterns / Collective Activity Spaces

The resulting flow patterns of each ethnic group is shown in figure 7.18. Each group's flow pattern reflects simultaneously their differences in residential location and similarities in workplace location. The majoritarian White British group presents the most uniform pattern, covering most of the study area. The other 3 groups, however, are more restricted to specific areas of the city and present significantly different flow patterns. For instance, the pattern of the Other group is highly concentrated in the city centre, where most individuals work, spreading towards the northwest of the city. Furthermore, this group is mostly absent in the southeast area of the city. The South Asian group's pattern presents a horizontal axis connecting population centres of this group located to the east and west of the city centre. The Black group's movement pattern, differently from the other groups' patterns, is not so concentrated towards the city centre. The Black group's pattern is also weaker than the others, reflecting both the fact this group is smaller than the others and its individuals are more spread out in the study area.

The flow patterns of the 4 groups correlate highly to each other, as shown below in table 7.10, but not as high as the correlations between the groups'

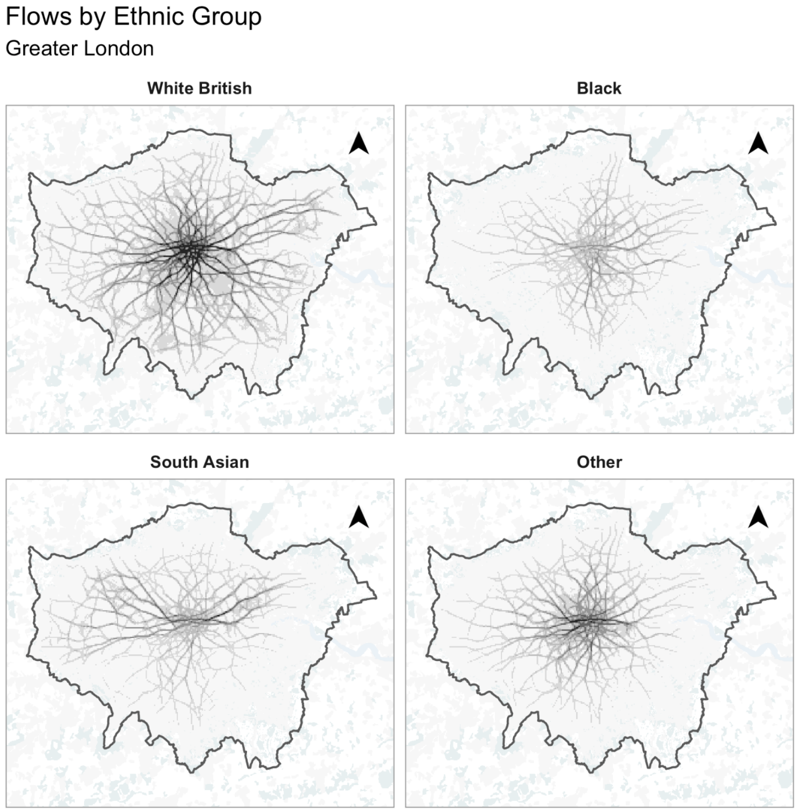

Figure 7.18: Aggregated flows by ethnic group.

workplace locations seen in table 7.9. The distinct flow pattern of the South Asian group is reflected in the slightly lower correlation coefficients between this group's pattern and the patterns of the other 3 groups, even though the correlation values (r between 0.79 and 0.84) can still be considered high.

Table 7.10: Pearson correlation coefficients (r) between groups' flow patterns.

	White British	Black	South Asian	Other
White British	-			
Black	0.90	-		
South Asian	0.79	0.81	-	
Other	0.93	0.86	0.84	-

Predominance

Predominance maps were produced by identifying the group with the highest flow count on each cell. Those maps are shown in figure 7.19, where each group is represented by a base hue and the colour brightness indicates flow counts. The map of figure 7.19a shows individuals from the White British group are predominant in most of the urban area. Their predominance is even more significant in the city centre, where a large percentage of the population commutes to. This pattern is expected, as almost 50% of the population in the study area is comprised of

White British individuals. Despite that large majority, it is noticeable some areas stand out for the predominance of South Asian individuals, mainly to the west of the city centre. Additionally, a small area to the east is mainly occupied by South Asians while Black individuals are the majority in a very small section to the north.

More interesting patterns can be seen by removing the White British group from the predominance map, as shown in figure 7.19b. From this map it can be understood that individuals from the Other group form the majority of flows in the city centre (with exception of the White British group). Those flows are completely hidden in figure 7.19a by the large number of White British individuals moving through the same areas. The patterns of the South Asian and Black groups are also clearly visible in figure 7.19b, with Black individuals moving mainly through the southeast of the urban area and South Asian individuals occupying areas to the west and east of the city centre.

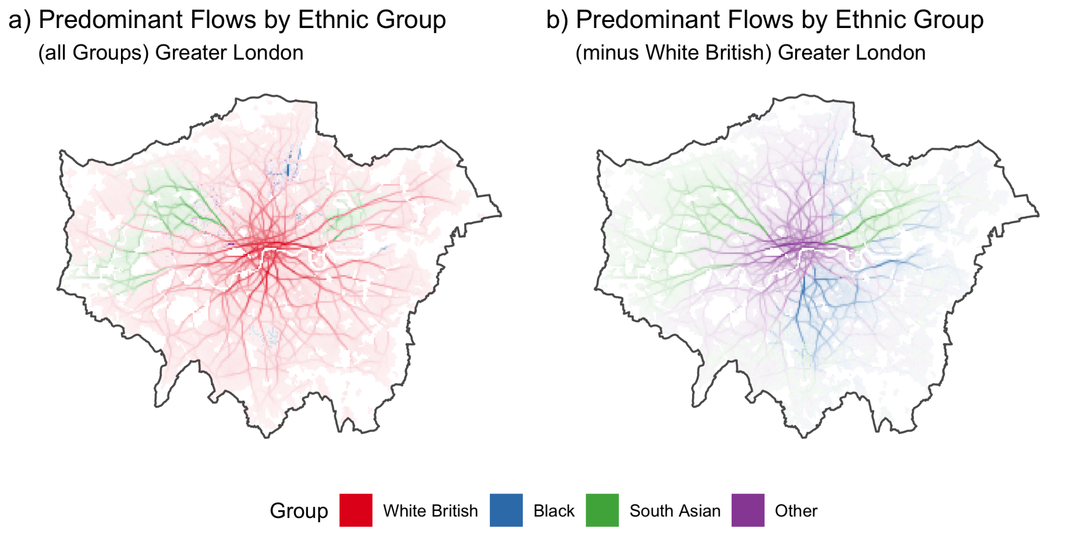

Figure 7.19: Map showing the predominant group in each area: a) all four groups, b) three minoritarian groups, excluding White British.

Diversity

The entropy-based information theory index H (Theil and Finizza 1971) was used to quantify segregation and diversity levels throughout the study area. Hence, areas used mainly by one group can be differentiated from areas where different groups are present, beyond the visual analyses previously presented. Such areas of high diversity of flow counts indicate higher potential for interaction among individuals of different groups and backgrounds.

The map of the local index H calculated for the study area is presented in figure 7.20. The map shows areas of positive values of index H, which means lower diversity, in shades of green, and areas of negative index H (high diversity) in shades of blue. It is noticeable that the city centre, that concentrates most of people's commuting destinations, is not the most diverse area of the city. This result is mainly due to the high number of White British individuals moving through this area pushing diversity down, in line with the predominance maps shown previously in figure 7.19.

The large blue areas in the map indicate the highest diversity can be found just outside the city centre, to the west, east and south. Five clusters of high diversity were pinpointed in the map by applying the k-means clustering algorithm to the 1000 cells of lowest H index value in the grid. Those clusters are identified by the five red dots in figure 7.20. These dots are located near areas of London often recognised as particularly diverse, such as the boroughs of Brent and Ealing to the west, and the borough of Haringey to the north. There is also the area known as the East End, between the City of London and Stratford. Finally, another high diversity area is located in the intersection of the boroughs of Lambeth, Croydon and Merton, in the south of London.

The map also shows the majority of the study area in shades of green, indicating low diversity overall. Those green areas cover the city centre and most of the outer boroughs of London, as well as many main roads connecting both areas. The results indicate that, although London is recognised as a very diverse city, this diversity is concentrated in specific areas of the city.

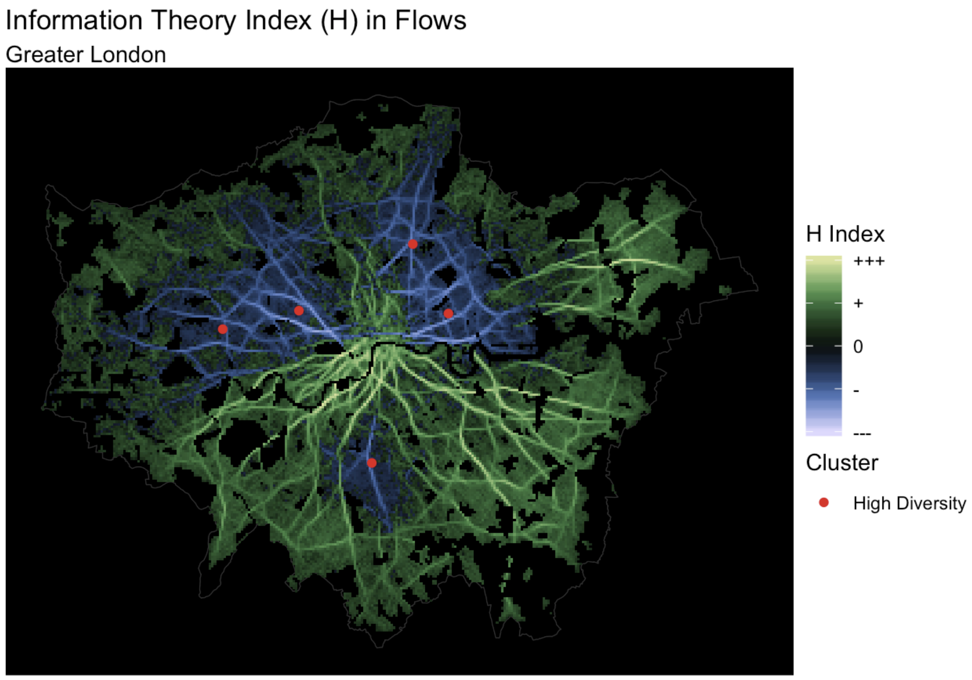

Figure 7.20: Diversity in flows in the GLA, measured using the Information Theory Index (H).

Global Copresence

The AxS model's copresence measures were used to quantify potential encounters and interactions between individuals in the study area. Copresence is an indicator of each group's isolation or exposure to other groups, and is measured based on the number of encounters between agents that take place during the course of a simulation. Specifically, the relative copresence metric was used in this study, which measures the number of encounters between individuals in comparison to the number of encounters that would be expected by random chance (as detailed in section 4.3).

The AxS model's copresence results were compared to place-based metrics of relative exposure and relative isolation[7], and the results are shown in figures 7.21 and 7.22. Negative values are represented by red bars in the plots and indicate the probability of encounter between individuals of that pair of groups is lower than random chance. Blue bars represent positive values of copresence and isolation/exposure, indicating higher probabilities of encounter.

The results of intragroup copresence, which is the probability of encounter between individuals of the same group, are shown in figure 7.21. Alongside the copresence results are the place-based relative isolation indices. A prominent feature of the plot is the intragroup copresence value of the Other group, which is the highest of the four groups. This result is even more significant considering that the place-based isolation index of the Other group is the lowest of the four groups, indicating the dynamic movement pattern of this group enables much more encounters than the residential location of its individuals alone. Another interesting result is the negative intragroup copresence of the White British group, even though its place-based isolation index is the same of the Black group (both 0.09). Furthermore, the Black group has the second highest intragroup copresence (0.15), so this stark difference between the White British and Black groups' copresence values can be attributed mainly to their distinct patterns of movement in the city. Finally, South Asian individuals have 6% higher probability of encountering other South Asian individuals than by random chance, which is much smaller than the 16% of the place-based isolation index.

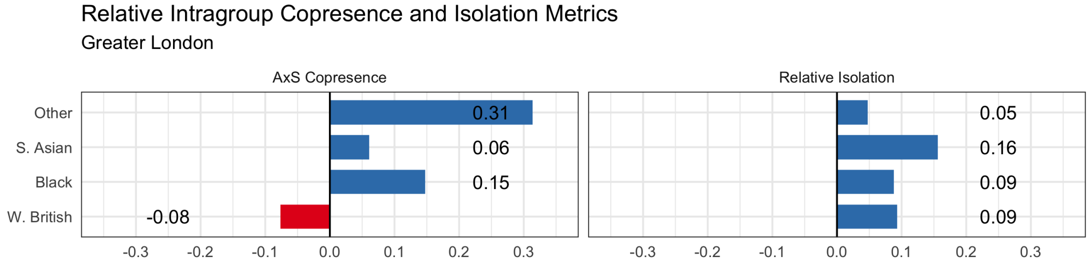

Figure 7.21: Relative intragroup copresence and relative isolation metrics.

The results of intergroup copresence, which is the probability of en-

7. Further details on the exposure and isolation indices used in this analysis can be found in appendix A.

counter between individuals of different groups, are shown in figure 7.22. These results are shown alongside the exposure indices of each pair of groups. It is noticeable the Other group is the only group with positive copresence to the other three groups, indicating this is the more integrated group in the study area. The most segregated groups in relation to each other are the South Asian and White British groups: the likelihood of encounter between individuals of the two groups is 14% lower than by random chance. One important distinction between the AxS model's copresence metric and the place-based exposure index is that the copresence metric is symmetric, while the exposure is not. That is why exposure of White British to South Asian is just -0.05, while exposure from South Asian to White British is -0.13. It is noticeable the latter value is similar to the copresence index of both groups, in one of the few occasions when the results of both metrics match.

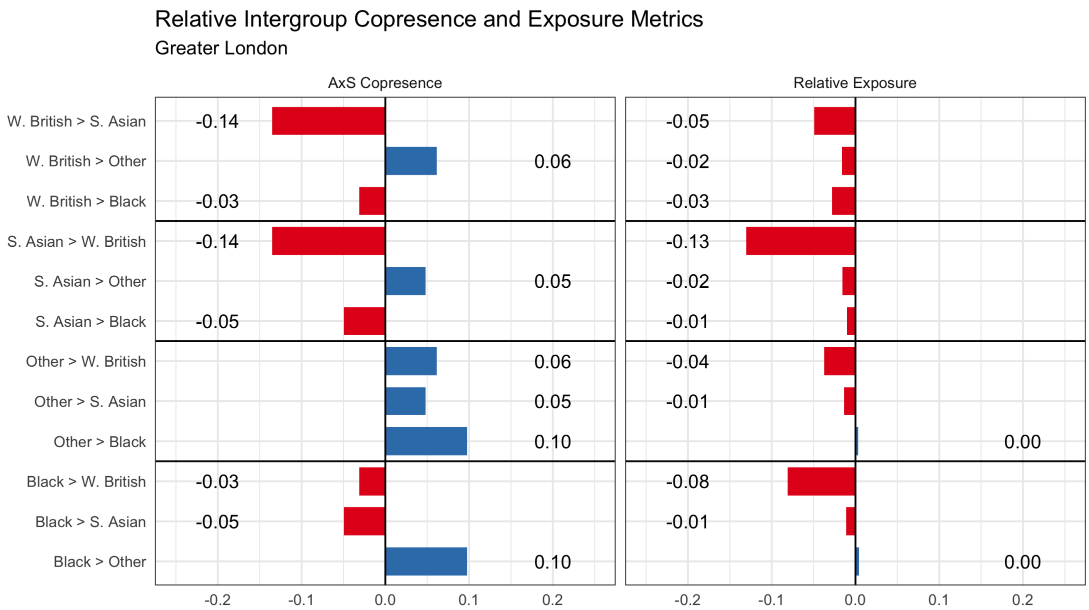

Figure 7.22: Relative intergroup copresence and relative exposure metrics.

The fact the copresence results obtained from the AxS model do not always follow the same trend of the place-based exposure and isolation indices indicate both methods are complementary to each other, and not interchangeable. It can be argued, though, the copresence index is a more comprehensive indicator as it is based not only on residential location but on individual movement patterns.

Local Copresence

The AxS model's local copresence metrics allow the study of how potential encounters between individuals are distributed in space. The local relative copresence metric, used in this analysis, compares how many encounters between individuals of two groups happened in each cell to how many encounters would be expected given the groups' size in the study area and the cell's flow count. The results are shown in the maps in figures 7.23 and 7.24. In those maps, shades of red indicate places where encounters happened more often than expected by random chance, and shades of blue indicate places where fewer encounters than expected where measured.

Intragroup copresence results (figure 7.23) show a high number of encounters among individuals of the Other group in the city centre. The White British group also has high copresence in the city centre, but there is a horizontal axis of low copresence crossing the centre of this group's map. This axis, together with the other blue areas of low copresence in the White British map resemble both the South Asian movement pattern shown previously in figure 7.18 as well as the areas of high diversity shown in the information theory index H map of figure 7.20. Black and South Asian individuals have fewer encounters in the city centre, which presents blue shades in both groups' maps, and more encounters outside but in different locations for each group. Finally, high intragroup copresence among White British individuals can be seen near the edges of the study area, in a pattern that is not present in the other groups' maps. This pattern also is similar to the large green areas of low diversity shown in figure 7.20. This indicates the movement patterns of the White British individuals, as well as their high number, have the effect of pushing diversity down in areas of the city where this group is more present.

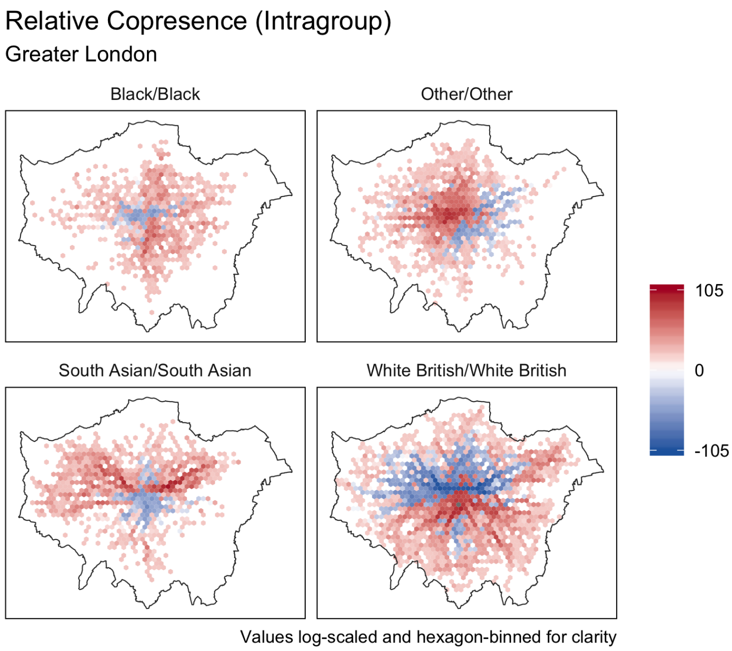

Figure 7.23: Local relative intragroup copresence.

The intergroup copresence maps (figure 7.24) show high number of encounters of individuals of the Other group with White British and South Asian individuals in the city centre, reinforcing the results that show a high presence of the Other group in central London. Conversely, the local copresence of the White British group with the Black and South Asian groups is very low in the city centre. Encounters between White British and South Asians happen more often in the northeast of the city, even though there is also a large concentration of South Asians in the west side of London, as shown earlier in the population distribution maps (figure 7.15). Furthermore, encounters between White British and Black individuals happen more often in the southeast of the study area, next to the city centre. As shown earlier in figure 7.22, intergroup copresence between the Black and Other groups is the highest in the study area (0.10), which is reflected in the local copresence map of those groups as well: the map shows red areas of high copresence between Black and Other throughout the study area, with only a few cells marked in light shades of blue indicating low copresence.

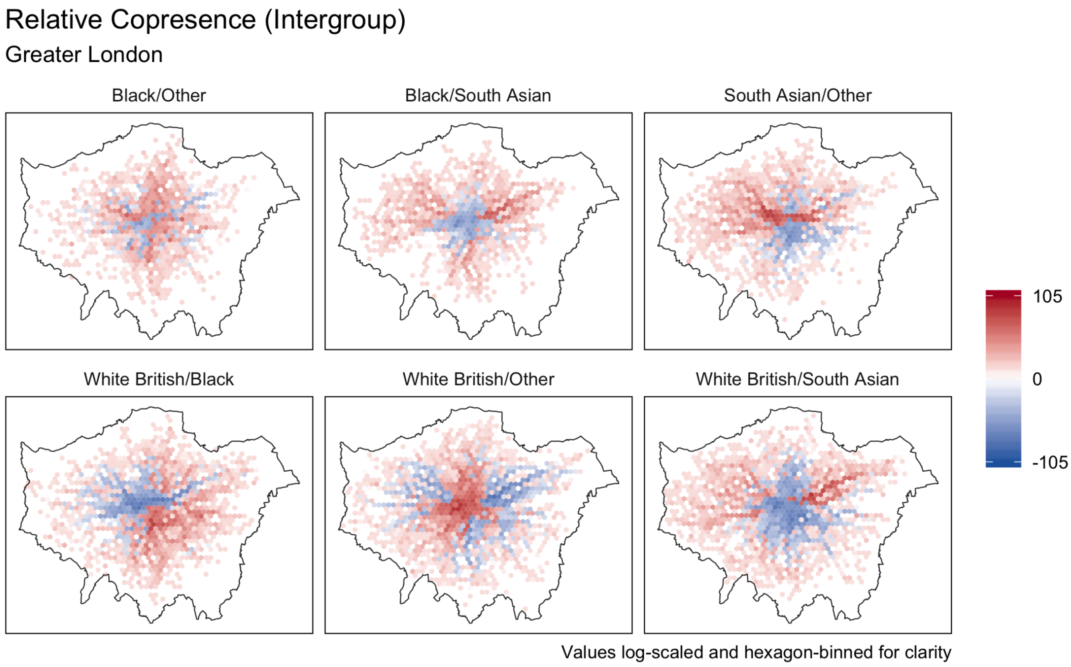

Figure 7.24: Local relative intergroup copresence.

7.2.4 Discussion on the London Case Study

London presents an interesting case study for segregation, due to its large and diverse population. This experiment explored the issue of segregation from a dynamic perspective, considering the groups' movement patterns in the urban area. Through this approach, interesting patterns of potential interaction among ethnic groups in the urban space, beyond their residential and workplace locations, were unveiled. The AxS model allows the study of such patterns both qualitatively, through the visual analysis of the model's output maps, and quantitatively, through numerical indicators of diversity and copresence.

Results demonstrated, quite surprisingly, that London's city centre is not the most diverse area of the city. Diversity, rather, is higher in areas where the number of White British individuals is lower and, thus, similar to the number of individuals of other groups. Results also indicated the Other group is the more integrated, with higher probability of interaction with the other three groups, while White British and South Asian individuals are less likely to interact with each other. Finally, significant differences were found between place-based measures of isolation and exposure and the AxS model's individual based measure of copresence, indicating people's movement patterns alter significantly their probabilities of interaction with other individuals.

Studying segregation based on individual trajectories highlighted an important limitation of this case study. It is well known that cities are not closed systems, and delimiting a city's actual (rather than administrative) limits can be a very difficult task. Yet, this case study was clearly delimited to the GLA area, which means daily commuting trips between the GLA and neighbouring towns, cities and suburban areas were not considered. The number of commuting trips in and out of the GLA can potentially affect the findings obtained in this study, as the GLA is much more ethnically diverse than the urban and rural areas around it. Hence, further investigation is needed to explore segregation patterns in London's extended metropolitan region.

7.3 Summary and Discussion

The simulation exercises presented in this chapter aimed to demonstrate the application of the AxS model to large world cities and discuss the issues of accessibility and segregation through an individual-based perspective. The case study of São Paulo explored inequalities in access to services and opportunities between population groups defined by income, education, and gender. The case study of London discussed segregation as potential interaction between individuals of different ethnic groups based on their daily commuting patterns. Even though each study focused on different issues, both were conducted using the same theoretical and methodological framework developed in this thesis and implemented in the AxS Model.

The simulation exercises presented in this chapter helped change the perspective through which accessibility and segregation are studied from a static, aggregate, and place-based perspective to a dynamic and individual-based one. Thus, the methodology proposed in this thesis has the potential to further develop an understanding of how the issues of social segregation and uneven access to opportunities and services in cities are shaped by individuals' dynamic movement patterns.

Chapter 8

Conclusions

This thesis has sought to develop a novel analytical and exploratory framework for studying accessibility and segregation in cities from an individual-based perspective. This study integrates a growing body of research on individual-based studies that can be traced back to the 1970s in the case of accessibility (Lenntorp 1976; Pirie 1979), and the 2000s in the case of segregation (Schönfelder and Axhausen 2003; Kwan 2013; Farber et al. 2015).

Based on Hägerstrand's (1970) time geographic theoretical and methodological framework, this line of individual-based studies conceptualises accessibility and segregation as the potential for interaction of individuals with places of activity (in the case of accessibility) and with individuals of different population groups (in the case of segregation). Taking advantage of the similarity between accessibility and segregation when viewed from this perspective, this thesis proposed a single theoretical and methodological framework through which both problems can be studied.

Individual-based studies face challenges such as limited data on individual activities and trajectories, as well as lack of methods for efficiently handling the available data on the scale required. Hence, despite the efforts of the research community, most individual-based studies have so far been limited in scope, covering small areas and using small population samples. To overcome those issues, this study took advantage of existing large-scale datasets as well as agent-based modelling techniques to build tools that allow the study of entire metropolitan areas and large populations, as presented in this document.

This is the concluding chapter of this thesis, in which the observations made during the course of this work are discussed and summarised. This chapter is divided in four sections. The first section discusses the main contributions of this thesis. The second section presents a critique of the methodology proposed here. The third section explores the possibilities of future work opened by this study, which is followed by the concluding remarks of this thesis.

8.1 Thesis Contributions

The methodological framework proposed in this thesis can be considered its main contribution. The framework includes the AxS model and the individual-based accessibility and segregation metrics calculated by the model during the simulation. This section also discusses the contributions of this thesis for empirical studies, which include the methods used to evaluate the AxS model's logic and outputs, and the model's application to real world cities. Those contributions are detailed in the following subsections.

8.1.1 The AxS Model

The AxS model is an agent-based model that simulates individuals who move through an urban environment between activity locations. Agents can be grouped according to their characteristics, such as ethnicity, income, and education. During movement, agents' trajectories are generated and can be visualised dynamically. Individual trajectories are used to compute output accessibility and segregation metrics. The aggregation of many individual trajectories produces movement patterns of groups of agents or of the entire population, which provide insights on city dynamics and patterns of interaction among groups. This section discusses the theoretical and methodological contributions of this thesis related to the AxS model.

Model's simplicity, scalability, and reproducibility

The AxS model was designed and developed to be an operational model, which can be applied to different study areas without undergoing significant changes. This goal has guided several design decisions made during the model's development. One of such decisions was making the model as simple to setup and use as possible. This was achieved by keeping the model's data and computational requirements low, and the results visually rich and easy to interpret.

The AxS model's data requirements are significantly lower than most agent-based models of similar purpose, such as the transportation models discussed in chapter 3. The model also significantly lowers data requirements of time geographic studies, by artificially simulating individuals' trajectories in space and time. Those trajectories are, arguably, the hardest data to obtain when large populations are considered. Most of the data required for simulations can be obtained from online sources, such as OpenStreetMap and census agencies' portals.

The model's computational power requirements are also significantly lower than most large-scale urban models. Large scale experiments, such as the metropolitan areas of São Paulo and London, were simulated dynamically onscreen in a standard desktop computer. The AxS model's rich visualisation capabilities allowed for easy communication of results to audiences of different levels of expertise, which is important to increase the model's potential reach. Although in a purely scientific context the computational power required by a model and its operational complexity are not usually important constraints, in an urban planning context, low requirements can improve the model's chances of being used as an actual decision support tool. Similarly, in educational settings, a model such as AxS can be more easily presented to and used by students, unlike larger and more complex models.

All those characteristics make the model easily reproducible, and its results easily shared. The model also can be applied to different study areas with relatively low effort, which is another advantage of the approach followed in this thesis. It is not uncommon to find, in the literature, models that are tied to a specific study area and that require significant amount of extra work to be adapted to other areas. The AxS model's characteristics make it easier for the model's results to be independently replicated and verified.

The reproducibility of an agent-based modelling study also depends on how it is disseminated. The publication of agent-based models presents specific challenges regarding the kind of content required for fully understanding a model, as discussed in chapter 3. Such content includes source code, raw data, and dynamic/interactive outputs produced by the model. In the case of this thesis, a companion website was produced to accommodate the information that cannot be included in a text document, helping the sharing of this study to wider audiences. The model's source code, input data, and video outputs of the simulations presented in this thesis are available online at www.mvpsaraiva.com/thesis. A similar approach will be adopted to articles derived from the thesis whenever such requirements are not catered by the publishers.

Time geography and agent-based modelling

Combining time geography and agent-based modelling can be considered a theoretical contribution of this thesis, as there are few examples of such combination in the literature. One of those examples is the class of activity-based models, reviewed in chapter 3, which use time geographic concepts to predict people's activity patterns and travel behaviour for purposes of traffic and transport simulation. Conversely, the AxS model has an analytical role, using time geographic concepts to explore accessibility and segregation patterns in cities.

This thesis demonstrated time geography and agent-based models are complementary. Time geography deals with individual's trajectories in space and time through a series of innovative techniques, such as using one spatial dimension to represent the time dimension in three-dimensional graphs. Two aspects make agent-based modelling a good fit for time geographic studies: first, it allows individual agents to be simulated and, second, the time dimension is explicitly represented. Those characteristics of agent-based models were leveraged in this study, allowing the development of individual-based metrics of accessibility and segregation from a time geographic perspective.

Analytical versus predictive models

This thesis contributes to the discussion of the role of agent-based models in research. Models have been traditionally designed as predictive tools, although they can also be classified as exploratory whey their objective is to simulate alternative future scenarios. Although the AxS model is an exploratory tool, its main purpose is analytical and it is not designed to explore future scenarios. As such, the model's purpose is to aid the understanding of current issues in cities, specifically accessibility and segregation, so those issues can be addressed and mitigated.

Agent-based modelling techniques present a few advantages over traditional analytical tools, such as aggregated place-based measures of accessibility and segregation, in fulfilling that analytical purpose. This includes representing individual differences and dynamic movement patterns, which is a straightforward process in agent-based models, but cannot be easily achieved with traditional aggregated and static methods.

One common belief in the field of urban analysis is that models are more complex than standard analytical methods. Indeed, it can be argued most models have a significant level of complexity, particularly the transportation models discussed in chapter 3. However, that is not necessarily true, and the AxS model is an example of a relatively simple model, conceived with modest data and computational power requirements. Hence, an important contribution of this thesis is to demonstrate models can be simple and effective analytical tools.

The role of agent-based modelling in a world of (big) data

Another contribution of this thesis is the innovative use of datasources. The new sources of data (and big data) that continuously become available for research bring many opportunities as well as challenges. As discussed in chapter 2, studying accessibility and segregation at the individual level requires significant amounts of data and at a level of detail that is seldom available. Attempts have been made by researchers to use new data sources such as social media, mobile phone call records, and smartphone apps with varying degree of success (see table 2.2).

One of the main shortcomings of new (big) data sources is they rarely contain demographic information, which are essential for studies of segregation, as well as studies focused on inequalities of accessibility. Demographic data are usually available from censuses in aggregated form, or collected in more detail at small scales. Joining all required information for a specific study from different data sources (in particular big data) can be very challenging, if at all possible. This study used traditional data sources containing demographic information, and used simulation techniques to artificially generate the missing pieces of information (specifically, trajectories and travel times). Hence, this thesis demonstrated that combining real world data with artificially generated data can be a way forward in cases where the required data are not available, adding value to existing datasets.

8.1.2 Individual-Based Measures of Accessibility and Segregation

Individual-based measures of accessibility and segregation were proposed in this thesis, based on existing time geographic and place-based metrics. One of the challenges of measuring accessibility and segregation using time geographic space-time prisms is the complexity of the three-dimensional geometric operations involved in the process. Chapter 2 of this thesis presented a review on the variety of techniques developed by researchers to overcome that challenge. Most of the solutions proposed consisted in developing alternative two-dimensional representations of three-dimensional space-time prisms. Measures calculated on two-dimensional areas, however, do not take the temporal dimension into account, which is an essential aspect of time geography. Although the AxS model also uses a two-dimensional spatial representation (a grid) to calculate the output metrics, the dynamic nature of agent-based models allows the temporal dimension to be explicitly represented in the model.

Accessibility measures

Accessibility measures, in this thesis, were calculated on individual's potential path areas, as simulated by the AxS model. Those metrics have an important conceptual difference from traditional place-based accessibility metrics, in addition to their individual-based nature. While accessibility usually measures the amount of opportunities people can access, in this study it measures the number of opportunities an individual can access *in their free time*, after commuting time is discounted. Essentially, it measures how easily an individual can access those opportunities without changing their schedules or compromising their other responsibilities. The measures proposed here also capture the combined effect of an individual's place of residence, place of work, and travel time in their accessibility. As demonstrated in chapter 7, those measures can capture accessibility inequalities at a significantly finer level of detail than traditional place-based measures of accessibility are able to.

Segregation measures

This thesis proposed copresence indices as measures of exposure/isolation between population groups, and adapted the information theory index (H) to be used as a measure of the evenness/clustering dimension of segregation.

Copresence quantifies actual encounters between agents in the model during the simulation, which represent potential encounters in the real world. This kind of metric can only be obtained in a dynamic simulation environment, such as an agent-based model. The information theory index (H), which is a place-based measure of segregation based on diversity, was adapted to raster environment to be used in the AxS model. That adaptation allowed the measurement of diversity at street level, based on people's dynamic commuting patterns rather than static places of residence. Thus, both measures are innovative relatively to traditional static and aggregated measures of segregation and, thus, contributions of this thesis.

8.1.3 Evaluation Method

Evaluating agent-based models is a challenge and a topic of much academic debate. The data necessary for fully validating an agent-based model is rarely available, due to the usually complex nature of the phenomena being modelled. In this thesis, a combination of techniques was used to evaluate the AxS model, both using abstract scenarios and real world data.

The verification and sensitivity analysis tests presented in chapter 5 were carried out in abstract environments. This decision was made to test the model's behaviour in a series of controlled situations. This allowed the effects of different parameters, and environmental conditions to be verified, and the effects of the model's rules to be better understood. The NetLogo agent-based modelling environment was very useful in this process, as it allows the outputs to be dynamically visualised onscreen so any abnormal behaviour can be more easily detected during the testing phase.

The model was further evaluated through validation tests, reported in chapter 6. Such tests were made by comparing the model's outputs to real world data, in order to make sure the model's results replicate reality up to the necessary level of accuracy. Those tests also served as a simplified calibration process, indicating the parameter values that produced the most accurate results.

A decision was made to validate agent's behaviour rather than the accessibility and segregation outputs. Validating the output metrics by comparing them to existing place-based or individual-based metrics would not make sense, mainly due to the innovative approach used in this thesis. Instead, agents' trajectories and travel times in different transport modes were compared to real world trajectories and travel times. Three different datasets were used in the validation process. That combination of multiple datasets for validation purposes is another example of the possibilities opened up by new sources of openly available data.

8.1.4 Empirical Applications

Models are often described as virtual laboratories where hypotheses can be tested and refined. While simulations in abstract scenarios test those hypotheses in simplified conditions, those intentionally do not replicate the complexity of real cities. Empirical applications can shed light on important aspects of reality that would be otherwise unnoticed. Such applications also help to better understand the model's advantages and limitations, and highlight areas for improvement. In this sense, models that can be applied to real world cities, such as the AxS model, make an important contribution to better understanding those cities and the complex phenomena that characterise them.

Chapter 7 presented two empirical applications of the AxS model. São Paulo and London were chosen as study areas, so the model could be tested in two large cities vastly different from each other. Those are not case studies in the traditional sense, as this thesis did not aim to deeply study the cities of London and São Paulo. Instead, the main aim of those studies was to demonstrate how the model can be used to explore accessibility and segregation in the real world.

The study of São Paulo revealed the well known inequality of the Brazilian society also manifests through accessibility. The proposed individual-based approach allowed for fine-grained inequalities to be identified. The study found that the lower income and education classes have significantly lower access to opportunities than the upper classes. The study also uncovered a striking gender gap in accessibility, clearly showing females from lower classes have much lower accessibility than males of the same classes. The study also revealed within-group inequalities, showing some individuals have significantly larger accessibility than other individuals in the same group.

The study of London explored ethnic segregation, unveiling patterns of potential interaction among ethnic groups in the urban space beyond people's residential and workplace locations. The analysis of the information theory index (H) results identified clusters of high diversity around the city centre but, somewhat surprisingly, not in the city centre itself. Copresence results indicated exposure dynamically measured in individuals' trajectories can be significantly different than when measured considering people's places of residence only. Interestingly, some results of copresence were found to be not only different, but also opposite to place-based exposure/isolation results. This indicates individual and place-based measures of segregation are fundamentally different from each other and can be used complementarily in order to obtain a more complete picture of urban segregation patterns.

8.2 Limitations of this Study

Many of the aforementioned contributions of this thesis can be attributed to the simplicity of the methodology proposed here. However, accessibility and segregation are complex phenomena that involve many factors that were not included in the AxS model, or were included in an oversimplified manner. This section discusses the limitations of this study, resulting from the model's design and implementation simplifications.

Representation of space and agents' movement

The choice of environmental representation was made with simplicity and scalability in mind. However, a raster environment is inherently more limited than a network-based environment for simulating movement. The main shortcoming of raster environments is the lack of an accurate representation of the street-network's topological structure. This means factors such as connectivity, turn restrictions, and directionality of streets are absent in the model's environment. Those limitations negatively affect the model's results both at larger and smaller scales, as discussed in the following paragraphs.

The lack of a topological structure makes it more difficult to identify routes where the most efficient trajectory contains large deviations from a relatively direct path. Such deviations can be found in many circumstances in the real world. For example, a motorway may circle around a densely urbanised area to avoid heavy urban traffic. In those situations, the longer motorway route can be more efficient in terms of travel time than the most direct route through heavily congested areas. The model's pathfinding algorithm is notably less accurate on those situations, as demonstrated by the validation exercises reported in chapter 6.

The aforementioned route deviations can also be due to the existence of natural obstacles (or spatial constraints) impossible to traverse, such as bodies of water, mountains, or forests. Finding a path around obstacles is relatively simple in network environments, but a significantly harder task in a raster environment. Consequently, the implementation of spatial constraints in the model is suboptimal. Movement, in the model, is just facilitated by the road network, but not constrained to it. Hence, agents need to actively look for those spatial constraints and find a way around them. In some situations, agents can get blocked behind those obstacles and be prevented from reaching their destinations.

The model's relatively coarse spatial resolution (200 m, in the simulation exercises reported in this thesis) does not allow for capturing fine grained details in the simulated trajectories. At that resolution, only the main road network can be represented. Consequently, the local road network is assumed to be embedded in regular urban cells, where agents can freely walk through. In reality, those local road networks may have many features that restrict movement within them, such as cul-de-sacs and dead ends, that cannot be represented in the model.

Representation of local environmental knowledge

As previously discussed, in reality, individuals make decisions based on their partial spatial knowledge. Specifically regarding wayfinding, spatial elements such as Kevin Lynch's (1960) landmarks and boundaries, as well as the road network's hierarchic structure, are important in individuals' decision-making process. However, those elements are absent from the model. Partial knowledge of the environment was indirectly simulated through agents' field of view, determined by *angle of vision* and *search radius* parameters. This is a very simplified representation of the complex cognition processes that take place in people's brains.

Furthermore, spatial knowledge varies in an individual basis. It is expected to be more comprehensive near places each individual is more familiar with, such as residential and workplace locations, as well as other places the individual spends time in. Conversely, spatial knowledge is non-existent or very limited in places the individual has never been to, or has not visited in a long time. Thus, spatial knowledge is thought to be gradually built over time as the individual interacts with their environment. All those complex dynamics are absent from the model. Even though differences in spatial knowledge between individuals could be represented, in the model, by different values of *angle of vision* and *search radius*, that possibility was not explored in this study.

Traffic dynamics

The AxS model simulates trajectories and city wide movement patterns from individual behaviour, but it is not a traffic model. Complex on-road dynamics, such as interactions between drivers, lane-changing behaviour, and traffic congestion, are not considered. Those features are absent from the model because they are out of the scope of this thesis, and it would be very complex and time consuming to build.

Traffic congestion, however, is a recurrent problem in large cities and can potentially affect the accessibility and segregation metrics calculated by the model. Travel speeds and travel times are strongly dependent on congestion levels, which vary between hours of the day and areas of the city. Those travel times are used, in the model, to calculate individuals' time budgets and accessibility metrics. Consequently, the resulting accessibility metrics may be overestimated when traffic congestion is not taken into consideration. Similarly, the model's segregation metrics are calculated on individuals' trajectories, which shape can be affected by congestion levels as individuals find alternative routes in order to avoid congested areas. Hence, the lack of a representation of congested areas is a limitation of this study.

Public transportation system

The AxS model does not include an explicit implementation of the public transportation system. Transport modes are represented by different movement speeds as a simple proxy. Agents using public transport move at a uniform speed during their entire trajectories, following the same pathfinding rules as pedestrians and drivers, which is not ideal. This implementation does not account for the complexity and importance of the urban transportation system, however. Important elements such as departure/arrival timetables, fixed routes, and designated embarking/disembarking points are absent from the model's implementation.

The simplified representation of the public transportation system has an impact on the accuracy of the model's results. Simulated trajectories do not follow actual routes of any specific transport mode. Simulated travel times are also affected, as time wasted while waiting for a bus or train, and during connections between different lines and modes, is not accounted for. Consequently, the validation tests reported in chapter 6 demonstrated that simulated trajectories and travel times of trips made by public transport modes are less accurate than those of trips made by car, which are not affected by those simplifications.

Activity system and temporal patterns

The current implementation of individuals' activity schedules in the model is very simple, supporting a single activity per agent, which generates a single trip. In the empirical applications of chapter 7, that trip was the commute to work. Implementing full travel and activity diaries as input and adding functionality to simulate trips with multiple stops would be relatively straightforward. However, those improvements depend on data availability, which is very restricted at that level of detail. As a matter of fact, the lack of data was one of the motivations of this study.

Temporal patterns, in the model, are also very simplified in comparison to those patterns in the real world. For example, cities present morning and afternoon peak hours, and individuals work during different hours of the day or night. In the model, all trips are simulated as if happening during the same time window, which can be associated to the morning peak hour. All individuals are also assumed to work during the same hours, which is not true in real life. Hence, the simulations presented in this thesis are to be interpreted as indicators of urban dynamics, and not as accurate reproductions of real-time traffic patterns.

Accessibility measures

The accessibility metrics proposed in this thesis are calculated on individuals' potential path areas (PPA) generated by the model. Those PPAs, however, are very simplified representations of the areas individuals can actually access. They are simple circular areas around people's places of residence and work, delimited based on individuals' movement speeds and time budget. Those circular areas assume uniform travel velocities in all directions, and no restrictions to movement imposed by the road network or transportation system, which are unrealistic assumptions. Also, opportunities located in the trajectories between people's places of residence and work, that could potentially be accessed in a short stop during commute, are not counted as accessible. Those simplifications were made, mainly, for scalability reasons as PPAs are calculated for each individual agent in the simulation.

Other limitations of the accessibility results presented in this study are related to the availability of data on activities and opportunities accessible to people. Data on opening hours of activities and activity durations (the minimum time required to participate on an activity) are seldom available in large scales. Data on quality and variety of activities and on individuals' needs are also very difficult to obtain. Individual-based accessibility metrics could, potentially, sort available opportunities based on characteristics such as affordability and desirability, for example. This would allow analyses based on the variety of opportunities inside the individual's PPA. Additional questions could be answered, such as: Does the mix of opportunities inside the individual's PPA satisfies all of the individual's needs? Is there an overabundance of one type of activity in detriment of another (for example, many fast food restaurants, and few healthier alternatives)? Evidently, those kinds of questions were not explored in this study.

Solving those data limitations is out of the reach of this thesis, as this would require extremely detailed and costly data collection. A simpler assumption was made in this study: the more activity locations are accessible, the more likely an individual is of finding a place that satisfies their needs at different situations. This assumption, however, does not solve the problem completely but leaves open opportunities for future experimentation as more detailed land-use data becomes available.

Segregation measures

The segregation measures proposed in this thesis are calculated based on the artificial trajectories simulated by the model. Hence, the main shortcoming of those metrics is related to the accuracy of the simulated trajectories. As established by the validation tests reported in chapter 6, the accuracy of the trajectories can vary significantly due to factors such as length, transport mode, presence of obstacles along the way, and existence of large detours in the route. As such, the segregation metrics results need to be analysed in light of the limitations of the model's pathfinding algorithm.

The proposed copresence metric is based on the number of encounters between individual agents in the model's environment during simulation. This means copresence is a measure of interaction in both space and time, in line with the time geographic framework. However, copresence measures *actual* encounters in the model, that represent only *potential* encounters in reality. Directly measuring actual encounters of people in reality is a significantly more challenging task and is beyond the scope of this thesis. The model's low spatial resolution also represents a significant simplification to the idea of an encounter between agents: in the study of London (section 7.2), an encounter means two agents shared a 200m x 200m cell at a given time. This is obviously a simplification and by no means indicate a meaningful interaction has occurred. In that sense, copresence indicates those individuals were potentially exposed to each other at a given time, so the minimum conditions for interaction were met.

The absence of the public transportation representation, in the model, also affects the segregation metrics outputs. When any two agents encounter each other during the simulation, that encounter counts towards the copresence results. However, that encounter is actually not possible if those two agents are using different modes of transport. Furthermore, probabilities of interaction vary widely between transport modes. While pedestrians can interact with other pedestrians, and bus/train riders may interact with other bus/train riders, car drivers are very unlikely to interact with each other in any socially meaningful way.

The simple temporal patterns simulated by the model also negatively affects the segregation results. As previously discussed, in the simulations conducted in this thesis, all individuals were assumed to work and commute during the same hours of the day. However, as that activity pattern is very simplified, it is possible for people who travel through the same areas at entirely different times, thus never encountering each other in real life, to be simulated as if they shared the same space at the same time. Those situations can cause an overestimation of interaction potential. However, this is more a limitation of the available datasources than of the model's design.

8.3 Future Work

The theoretical and methodological framework presented in this thesis represents a first step towards integrating agent-based models and individual-based accessibility and segregation metrics. Several options for future work emerged from the research developed in this thesis, mainly in two areas: model improvement, and further empirical applications. Those are discussed in this section.

8.3.1 Model Improvement

The AxS model embodies the methodological framework proposed in this thesis into an operational agent-based model. Once the model is implemented, it is relatively straightforward to develop new features and improve existing ones. The limitations previously discussed are a good indication of the areas the model needs to be improved. The following extensions of the model should be considered in the continuation of this research.

Representation of space

The AxS model's raster environment, despite its advantages, can be considered the main limitation of its design. As previously discussed, this kind of environment is simple and lightweight, making the model easier to implement and use even for very large study areas. However, it has severe limitations regarding agents' pathfinding and navigation, as it lacks the capacity to represent the topological structure of urban street networks. This is probably the main reason most agent-based models of transport and traffic simulation, such as the ones reviewed in chapter 3, use network-based environments.

Upgrading the model's environment would bring a number of advantages. For example, a network environment allows much more precise representations of travel distances and movement speeds, which are currently limited due to the model's relatively large cell sizes. Navigation problems such as agents getting blocked by spatial constraints would also be solved, as the street network can easily guide agents around those obstacles. Furthermore, the connectivity between different areas of the road network could be accurately modelled. Currently, adjacent urban areas in the model are always considered connected, which may not be true in reality due to the characteristics of the local road network. Features such as crosses on non-planar space (overpasses and underpasses), dead-ends, cul-de-sacs, turn restrictions, and one-way streets can significantly alter the connectivity between urban areas. Those characteristics can only be accurately modelled in a network environment.

Those changes can negatively affect the model's scalability, which is one of its current strengths. Significantly more computational power is expected to be required by a network-based model in order to run large-scale simulations similar to the ones discussed in chapter 7, in comparison to the current version of the model. Data requirements, however, can be more easily overcome. Datasources such as OpenStreetMap contain topologically accurate street network information of most world cities at the level of detail required by network-based simulations. OpenStreetMap data contains detailed information about the street network, including the directionality of streets, pedestrian- or vehicle-only routes, and cycling infrastructure, freely available to download and use.

Adapting the AxS model to a network-based environment is a critical update towards adding realism to the simulations. However, this is not a trivial task, as most of the model must be rebuilt for that purpose. Consequently, all the processes of verification, sensitivity analysis, and validation would have to be carried out again for the new version of the model, as its behaviours and outcomes would significantly change due to such a large intervention. However, significant increases in the level of detail of the model's environment and the accuracy of its results are expected to be obtained from this change. Thus, this update constitutes an important avenue to be explored in the future.

Public transportation system

The public transportation system is, possibly, the most important element missing in the AxS model. The reach, capacity, and efficiency of the public transportation system greatly affect people's trajectories and travel times, which impact the resulting accessibility and segregation metrics. Including an explicit representation of the pubic transportation system in the model will greatly increase its usefulness. Currently, using movement speed as a proxy for transport mode allows only an overall representation of the efficiency of each mode. However, this proxy effectively considers the public transportation network is available throughout the study area, closely following the main road network. It disregards differences in frequency of service and density of lines available in different areas of the city. Adding those features to the model would allow accessibility and segregation to be studied at an even finer level of detail than it is possible in the current version of the AxS model.

Representation of local environmental knowledge

Local environmental knowledge, in the AxS model, is simulated through the agents' limited environmental perception: agents only know and react to what is in their current field of view, defined by *angle of vision* and *search radius* parameters. It is possible, however, to implement more sophisticated forms of local knowledge. One option is to give agents a hierarchical knowledge of the city. In this option, agents would know in detail the road network near the origin and destination points of their trips, and only know the primary road network in the remaining areas of the city. Important urban landmarks and main junctions in the street network could also be added, which agents could use as guides when navigating through areas they are unfamiliar with.

Traffic dynamics

As previously discussed, the AxS model is not, nor does it aim to be, a traffic model. Implementing complex traffic dynamics would be time consuming and counterproductive. However, simplified proxies of traffic dynamics could be relatively easily implemented. A simple proxy for traffic congestion, for example, can be implemented both statically and dynamically in relatively easy ways. One option to statically represent traffic congestion is by adding an input layer indicating the more congested areas of the city being studied, as this kind of data is available for many cities. Movement through those areas would then be penalised, so agents would have to decide between slowing down or deviating from those areas. Congestion could also be implemented dynamically, through the density of agents on the model's environment: once the number of agents in any given cell reaches certain threshold, the cell would be marked as congested, and the aforementioned movement penalties would apply.

Pathfinding algorithm

The pathfinding algorithm used in the AxS model was developed with two main objectives. The first was to work in the raster environment used by the model, because most existing algorithms work on network environments. The second objective was to work in line with agent-based modelling concepts, which means the algorithm had to: a) be non-deterministic, generating different paths on different model runs when more than one alternative path exists; and b) rely on the agents' limited environmental perception only. Within those constraints, there are numerous options to improve the algorithm. In fact, the algorithm would need to be adapted for each of the model improvements discussed above.

Changing the model's environment to a network will require the development of a new pathfinding algorithm. Simply using an existing algorithm such as Dijkstra's or A* is not a suitable option because those algorithms do not comply with objective 'b' above, since they are deterministic and rely on the complete knowledge of the road network. Adapting one of those algorithms to work step-by-step, in a similar way the current AxS algorithm works, is possibly a good starting point. Adding public transportation to the model would also require significant adaptations to the pathfinding algorithm, as it would need to work on multiple networks and consider multiple transport modes with specific timetables. Adapting the algorithm to include simple forms of individual environmental knowledge and traffic dynamics can be done relatively straightforwardly, as previously discussed in the relevant subsections. Obviously, those features can become too complex very quickly, so it is important to keep the model's main objective in mind when implementing improvements. In the case of this thesis, the objective was to study individual-based accessibility and segregation in a simple manner and not building a fully functional multi-modal traffic model.

8.3.2 Further Empirical Studies

Applying the model to different situations and further exploring the cities already studied in this thesis is relatively easy, since the environmental setup is done and most of data collection has been carried out. A few possibilities of additional empirical applications are discussed in this section.

São Paulo

The empirical study of São Paulo (section 7.1) used data from the São Paulo Metro OD Survey, collected in 2007. A new OD survey was carried out in 2017, which will provide an updated picture of mobility in the São Paulo Metropolitan region once its results are released. At the time this research was being conducted, data collected in 2017's OD survey was being processed but not yet available. Once the new dataset is released, a study can be made comparing the results presented in this thesis to results obtained using the updated data. Such study could give interesting insights on how the social, economic, and political changes that happened in Brazil during that decade affected accessibility inequalities in São Paulo.

London

The study of London (section 7.2) was based on OD data disaggregated by ethnicity, but aggregated at the census tract level. Although those data are easily obtainable, they only provide the total number of people of each ethnicity travelling between pairs of zones. Additional information such as level of education and socio-economic classification, are available separately, and combining those accurately is not possible. The UK census also collects more detailed microdata, with richer demographic and transportation information that could be used to enrich the results presented here. However, due to privacy concerns, microdata is only available in a secure environment provided by the ONS (UK's Office for National Statistics), but this would require running the model on ONS machines.

Areas with no data available

The possibility of using artificially generated OD matrices as input for the AxS model can be explored in the future. Artificial OD matrices could be obtained from external models, such as the spatial interaction and activity-based models discussed in chapter 3, based on census demographic data. This option can be useful in areas where no other relevant datasets are available. Such combination of artificial ODs (generated by an external model) and artificial trajectories (generated by the AxS model), however, needs to be carefully considered when carrying out this kind of experiment, as the results may be affected by unknown interactions between models.

A simple pilot study was carried out for the city of Bagé, in Southern Brazil, using randomly generated ODs (Saraiva and Barros 2017). In that study, AxS generated trajectories were found to be correlated to network-based measures of closeness and betweenes centrality, as well as to the city's land use distribution. The pilot study of Bagé demonstrated the model can also be useful in studies of urban morphology, beyond the individual-based accessibility and segregation studies which are the focus of this thesis.

Further integrating accessibility and segregation

This thesis focused on developing the AxS model and the individual-based accessibility and segregation metrics, demonstrating their application on different contexts. The model was used to study accessibility in São Paulo and Segregation in London, but no combined exercise simultaneously exploring accessibility and segregation in the same study area was carried out. Conducting such combined exercise exploring the theoretical and empirical relationships between accessibility and segregation in the level of detail those complex phenomena require was out of the scope of this thesis, but is a possible and interesting avenue for further research.

This thesis advanced the existing body of research towards the goal of integrating accessibility and segregation into a single theoretical and methodological framework. Now that the necessary tools are built and tested, other avenues are open to further explore the relationships between accessibility and segregation from an individual-based perspective through empirical studies in real world cities.

8.4 Closing Remarks

This study is part of a growing body of research stemming from Hägerstrand's (1970) original question "what about people in regional sciences?". This thesis has contributed to answering that question by introducing a novel theoretical and methodological framework for studying accessibility and segregation from an individual-based perspective. That framework combines time geographic concepts and agent-based modelling techniques in innovative ways, making definite contributions to advancing both fields. The methods developed here allowed the study of individual-based accessibility and segregation at metropolitan scales and at a finer level of detail than previously possible with the limited data commonly available. As the issues of social segregation and associated unequal access to urban opportunities and services in cities deepen and receive more attention, understanding how they unfold via the complex dynamics of interaction of individuals becomes even more relevant. The work developed in this thesis contributes towards this goal, thus opening up new avenues for understanding how accessibility and segregation patterns are shaped by those complex dynamics and interactions.

References

Abar, Sameera, Georgios K. Theodoropoulos, Pierre Lemarinier, and Gregory M. P. O'Hare. 2017. "Agent Based Modelling and Simulation Tools: A Review of the State-of-Art Software." *Computer Science Review* 24 (May): 13–33. ISSN: 1574-0137. doi:10.1016/j.cosrev.2017.03.001.

Adnan, Muhammad. 2015. "SimMobility: A Multi-Scale Integrated Agent-Based Simulation Platform." In *Transportation Research Board 95th Annual Meeting,* 18.

Aksyonov, K. E. 2011. "Social Segregation of Personal Activity Spaces in a Posttransformation Metropolis (Case Study of St. Petersburg)." *Regional Research of Russia* 1, no. 1 (January): 52–61. ISSN: 2079-9705, 2079-9713. doi:10.1134/S2079970511010023.

Allen, Peter M. 2012. *Cities and Regions as Self-Organizing Systems : Models of Complexity.* Routledge, June. ISBN: 978-1-135-30172-9. doi:10.4324/9780203990018.

Antonini, Gianluca, Michel Bierlaire, and Mats Weber. 2006. "Discrete Choice Models of Pedestrian Walking Behavior." *Transportation Research Part B: Methodological* 40, no. 8 (September): 667–687. ISSN: 01912615. doi:10.1016/j.trb.2005.09.006.

Arentze, Theo, and Harry Timmermans. 2004. "A Learning-Based Transportation Oriented Simulation System." *Transportation Research Part B: Methodological* 38, no. 7 (August): 613–633. ISSN: 01912615. doi:10.1016/j.trb.2002.10.001.

Axhausen, Kay. 2000. "Activity-Based Modelling: Research Directions and Possibilities."

Azevedo, Carlos Lima, Neeraj Milind Deshmukh, Balakumar Marimuthu, Simon Oh, Katarzyna Marczuk, Harold Soh, Kakali Basak, Tomer Toledo, Li-Shiuan Peh, and Moshe E. Ben-Akiva. 2017. "SimMobility Short-Term: An Integrated Microscopic Mobility Simulator." *Transportation Research Record: Journal of the Transportation Research Board* 2622 (January): 13–23. ISSN: 0361-1981. doi:10.3141/2622-02.

Babakan, Ali Shirzadi, and Abbas Alimohammadi. 2016. "An Agent-Based Simulation of Residential Location Choice of Tenants in Tehran, Iran." *Transactions in GIS* 20, no. 1 (February): 101–125. ISSN: 1467-9671. doi:10.1111/tgis.12144.

Barros, Joana. 2012. "Exploring Urban Dynamics in Latin American Cities Using an Agent-Based Simulation Approach." In *Agent-Based Models of Geographical Systems,* edited by Alison J. Heppenstall, Andrew T. Crooks, Linda M. See, and Michael Batty, 571–589. Dordrecht: Springer Netherlands. ISBN: 978-90-481-8926-7 978-90-481-8927-4. doi:10.1007/978-90-481-8927-4_28.

Barros, Joana, and Flávia Feitosa. 2018. "Uneven Geographies: Exploring the Sensitivity of Spatial Indices of Residential Segregation." *Environment and Planning B: Urban Analytics and City Science* (March): 17. ISSN: 2399-8083, 2399-8091. doi:10.1177/2399808318760572.

Basak, Biswanath, and Sumana Gupta. 2017. "Developing an Agent-Based Model for Pilgrim Evacuation Using Visual Intelligence: A Case Study of Ratha Yatra at Puri." *Computers, Environment and Urban Systems* 64 (July): 118–131. ISSN: 01989715. doi:10.1016/j.compenvurbsys.2017.01.006.

Batty, Michael. 2000. "Less Is More, More Is Different: Complexity, Morphology, Cities, and Emergence." *Environment and Planning B: Planning and Design* 27, no. 2 (April): 167–168. ISSN: 0265-8135. doi:10.1068/b2702ed.

———. 2008. "Fifty Years of Urban Modelling: Macro Statics to Micro Dynamics." In *The Dynamics of Complex Urban Systems,* edited by P. Giordano, A. Vancheri, S. Albeverio, and D. Andrey, 1–20. Heidelberg, DE: Physica-Verlag.

Batty, Michael. 2009a. "Accessibility: In Search of a Unified Theory." *Environment and Planning B: Planning and Design* 36, no. 2 (April): 191–194. ISSN: 0265-8135, 1472-3417. doi:10.1068/b3602ed.

———. 2009b. "Urban Modeling." In *International Encyclopaedia of Human Geography,* edited by N. Thrift and R. Kitchin, 51–58. Oxford, UK: Elsevier.

———. 2012. "A Generic Framework for Computational Spatial Modelling." In *Agent-Based Models of Geographical Systems,* edited by Alison J. Heppenstall, Andrew T. Crooks, Linda M. See, and Michael Batty, 19–50. Dordrecht: Springer Netherlands. ISBN: 978-90-481-8926-7 978-90-481-8927-4. doi:10.1007/978-90-481-8927-4_2.

———. 2013. *The New Science of Cities.* Reprint edition. MIT Press. ISBN: 978-0-262-53456-7.

Batty, Michael, Jake Desyllas, and Elspeth Duxbury. 2003. "The Discrete Dynamics of Small-Scale Spatial Events: Agent-Based Models of Mobility in Carnivals and Street Parades." *International Journal of Geographical Information Science* 17, no. 7 (October): 673–697. ISSN: 1365-8816. doi:10.1080/1365881031000135474.

Batty, Michael, and Andrew Hudson-Smith. 2014. "Visual Analytics for Urban Design." *Urban Design* 132:38–41.

Batty, Michael, and Paul M. Torrens. 2005. "Modelling and Prediction in a Complex World." *Futures* 37, no. 7 (September): 745–766. ISSN: 0016-3287. doi:10.1016/j.futures.2004.11.003.

Bell, W. 1954. "A Probability Model for the Measurement of Ecological Segrega-Tion." *Social Forces,* no. 43: 357–364.

Benard, Stephen, and Robb Willer. 2007. "A Wealth and Status-Based Model of Residential Segregation." *The Journal of Mathematical Sociology* 31, no. 2 (March): 149–174. ISSN: 0022-250X. doi:10.1080/00222500601188486.

Benenson, Itzhak, Karel Martens, and Slava Birfir. 2008. "PARKAGENT: An Agent-Based Model of Parking in the City." *Computers, Environment and Urban Systems,* GeoComputation: Modeling with Spatial Agents, 32, no. 6 (November): 431–439. ISSN: 0198-9715. doi:10.1016/j.compenvurbsys.2008.09.011.

Benenson, Itzhak, Itzhak Omer, and Erez Hatna. 2002. "Entity-Based Modeling of Urban Residential Dynamics: The Case of Yaffo, Tel Aviv." *Environment and Planning B: Planning and Design* 29, no. 4 (August): 491–512. ISSN: 0265-8135. doi:10.1068/b1287.

Bruch, Elizabeth E., and Robert D. Mare. 2006. "Neighborhood Choice and Neighborhood Change." *American Journal of Sociology* 112, no. 3 (November): 667–709. ISSN: 0002-9602. doi:10.1086/507856.

Buliung, Ronald N., and Pavlos S. Kanaroglou. 2006. "A GIS Toolkit for Exploring Geographies of Household Activity/Travel Behavior." *Journal of Transport Geography* 14, no. 1 (January): 35–51. ISSN: 0966-6923. doi:10.1016/j.jtrangeo.2004.10.008.

Burns, Lawrence D. 1979. *Transportation, Temporal, and Spatial Components of Accessibility.* Lexington Books. ISBN: 978-0-669-02916-1.

Casas, Irene, Mark W. Horner, and Joe Weber. 2009. "A Comparison of Three Methods for Identifying Transport-Based Exclusion: A Case Study of Children's Access to Urban Opportunities in Erie and Niagara Counties, New York." *International Journal of Sustainable Transportation* 3, no. 4 (June): 227–245. ISSN: 1556-8318. doi:10.1080/15568310802158761.

Cavill, Nick, Sonja Kahlmeier, Harry Rutter, Francesca Racioppi, and Pekka Oja. 2008. "Economic Analyses of Transport Infrastructure and Policies Including Health Effects Related to Cycling and Walking: A Systematic Review." *Transport Policy* 15, no. 5 (September): 291–304. ISSN: 0967-070X. doi:10.1016/j.tranpol.2008.11.001.

Chaix, Basile, Yan Kestens, Camille Perchoux, Noëlla Karusisi, Juan Merlo, and Karima Labadi. 2012. "An Interactive Mapping Tool to Assess Individual Mobility Patterns in Neighborhood Studies." *American Journal of Preventive Medicine* 43, no. 4 (October): 440–450. ISSN: 0749-3797, 1873-2607. doi:10.1016/j.amepre.2012.06.026.

Chang, Justin S. 2006. "Models of the Relationship between Transport and Landuse: A Review." *Transport Reviews* 26, no. 3 (May): 325–350. ISSN: 0144-1647. doi:10.1080/01441640500468432.

Clark, William. 1991. “Residential Preferences and Neighborhood Racial Segregation: A Test of the Schelling Segregation Model.” *Demography* 28, no. 1 (February): 1–19. ISSN: 0070-3370, 1533-7790. doi:10.2307/2061333.

Couclelis, H., R. G. Golledge, N. Gale, and W. Tobler. 1987. “Exploring the Anchor-Point Hypothesis of Spatial Cognition.” *Journal of Environmental Psychology* 7, no. 2 (June): 99–122. ISSN: 0272-4944. doi:10.1016/S0272-4944(87)80020-8.

Crooks, Andrew. 2010. “Constructing and Implementing an Agent-Based Model of Residential Segregation through Vector GIS.” *International Journal of Geographical Information Science* 24, no. 5 (April): 661–675. ISSN: 1365-8816. doi:10.1080/13658810903569572.

Crooks, Andrew, Christian Castle, and Michael Batty. 2008. “Key Challenges in Agent-Based Modelling for Geo-Spatial Simulation.” *Computers, Environment and Urban Systems,* GeoComputation: Modeling with Spatial Agents, 32, no. 6 (November): 417–430. ISSN: 0198-9715. doi:10.1016/j.compenvurbsys.2008.09.004.

Crooks, Andrew, and Alison Heppenstall. 2012. “Introduction to Agent-Based Modelling.” In *Agent-Based Models of Geographical Systems,* edited by Alison J. Heppenstall, Andrew T. Crooks, Linda M. See, and Michael Batty, 85–105. Springer.

Dai, Jicai, Xia Li, and Lin Liu. 2013. “Simulation of Pedestrian Counter Flow through Bottlenecks by Using an Agent-Based Model.” *Physica A: Statistical Mechanics and its Applications* 392, no. 9 (May): 2202–2211. ISSN: 0378-4371. doi:10.1016/j.physa.2013.01.012.

Dailisan, Damian N., and May T. Lim. 2016. “Agent-Based Modeling of Lane Discipline in Heterogeneous Traffic.” *Physica A: Statistical Mechanics and its Applications* 457 (September): 138–147. ISSN: 0378-4371. doi:10.1016/j.physa.2016.03.104.

Dijkstra, E W. 1959. “A Note on Two Problems in Connexion with Graphs.” *Numerische Mathematik* 1 (1): 269–271.

Dijst, Martin. 1999. “Action Space as Planning Concept in Spatial Planning.” *Netherlands journal of housing and the built environment* 14, no. 2 (June): 163–182. ISSN: 0920-1580, 1573-7772. doi:10.1007/BF02496820.

Doniec, Arnaud, René Mandiau, Sylvain Piechowiak, and Stéphane Espié. 2008. "A Behavioral Multi-Agent Model for Road Traffic Simulation." *Engineering Applications of Artificial Intelligence* 21, no. 8 (December): 1443–1454. ISSN: 09521976. doi:10.1016/j.engappai.2008.04.002.

Downs, Roger M, and David Stea. 1973. *Cognitive maps and spatial behavior: Process and products.* na.

Duncan, Otis Dudley, and Beverly Duncan. 1955. "A Methodological Analysis of Segregation Indexes." *American Sociological Review* 20 (2): 210–217. ISSN: 0003-1224. doi:10.2307/2088328.

Ettema, D. F., T. A. Arentze, and H. J. P. Timmermans. 2007. "Towards an Integrated LUTI Model of Long Term and Short-Term Mobility Decisions of Households Using Social Learning." In *Proc., 10th Int. Conf. on Computers in Urban Planning and Urban Management.* Univ. de São Paulo Brazil.

Farber, Steven, Tijs Neutens, Harvey Miller, and Xiao Li. 2013. "The Social Interaction Potential of Metropolitan Regions: A Time-Geographic Measurement Approach Using Joint Accessibility." *Annals of the Association of American Geographers* 103 (3): 483–504. ISSN: 0004-5608. doi:10.1080/00045608.2012.689238.

Farber, Steven, Morton O'Kelly, Harvey Miller, and Tijs Neutens. 2015. "Measuring Segregation Using Patterns of Daily Travel Behavior: A Social Interaction Based Model of Exposure." *Journal of Transport Geography* 49, no. Supplement C (December): 26–38. ISSN: 0966-6923. doi:10.1016/j.jtrangeo.2015.10.009.

Feitosa, Flávia, Quang Bao Le, and Paul Vlek. 2011. "Multi-Agent Simulator for Urban Segregation (MASUS): A Tool to Explore Alternatives for Promoting Inclusive Cities." *Computers, Environment and Urban Systems* 35, no. 2 (March): 104–115. ISSN: 0198-9715. doi:10.1016/j.compenvurbsys.2010.06.001.

Fontaine, Corentin M., and Mark D. A. Rounsevell. 2009. "An Agent-Based Approach to Model Future Residential Pressure on a Regional Landscape." *Landscape Ecology* 24, no. 9 (November): 1237. ISSN: 0921-2973, 1572-9761. doi:10.1007/s10980-009-9378-0.

Fossett, Mark, and Warren Waren. 2005. "Overlooked Implications of Ethnic Preferences for Residential Segregation in Agent-Based Models." *Urban Studies* 42, no. 11 (October): 1893–1917. ISSN: 0042-0980. doi:10.1080/00420980500280354.

Freeman, Linton. 1978. "Segregation in Social Networks." *Sociological Methods & Research* 6 (4): 411–429.

El-Geneidy, Ahmed, and David Levinson. 2006. *Access to Destinations: Development of Accessibility Measures.* Technical report.

———. 2011. "Place Rank: Valuing Spatial Interactions." *Networks and Spatial Economics* 11, no. 4 (December): 643–659. ISSN: 1566-113X, 1572-9427. doi:10.1007/s11067-011-9153-z.

Geurs, Karst, and Bert van Wee. 2004. "Accessibility Evaluation of Land-Use and Transport Strategies: Review and Research Directions." *Journal of Transport Geography* 12, no. 2 (June): 127–140. ISSN: 09666923. doi:10.1016/j.jtrangeo.2003.10.005.

Gilbert, Nigel. 2008. *Agent-Based Models.* 153. Thousand Oaks, California: Sage Publications.

Gilbert, Nigel, and Klaus Troitzsch. 2005. *Simulation for the Social Scientist.* 2nd ed. Maidenhead, England ; New York, NY: Open University Press. ISBN: 978-0-335-21600-0.

Golledge, Reginald G. 1997. *Spatial Behavior: A Geographic Perspective.* Guilford Press, January. ISBN: 978-1-57230-050-7.

Golledge, Reginald G., Mei-Po Kwan, and Tommy Gärling. 1994. "Computational Process Modeling of Household Travel Decisions Using a Geographical Information System." *Papers in Regional Science* 73, no. 2 (April): 99–117. ISSN: 1435-5957. doi:10.1111/j.1435-5597.1994.tb00605.x.

Greenberg Raanan, Malka, and Noam Shoval. 2014. "Mental Maps Compared to Actual Spatial Behavior Using GPS Data: A New Method for Investigating Segregation in Cities." *Cities* 36, no. Supplement C (February): 28–40. ISSN: 0264-2751. doi:10.1016/j.cities.2013.09.003.

Grimm, Volker, and Uta Berger. 2016. "Robustness Analysis: Deconstructing Computational Models for Ecological Theory and Applications." *Ecological Modelling* 326 (April): 162–167. ISSN: 0304-3800. doi:10.1016/j.ecolmodel.2015.07.018.

Grimm, Volker, and Steven F. Railsback. 2005. *Individual-Based Modeling and Ecology.* Princeton University Press. ISBN: 978-0-691-09666-7.

———. 2012. "Pattern-Oriented Modelling: A 'multi-Scope' for Predictive Systems Ecology." *Philos Trans R Soc Lond B Biol Sci* 367, no. 1586 (January): 298–310. ISSN: 1471-2970. doi:10.1098/rstb.2011.0180.

Grimm, Volker, Eloy Revilla, Uta Berger, Florian Jeltsch, Wolf M. Mooij, Steven F. Railsback, Hans-Hermann Thulke, Jacob Weiner, Thorsten Wiegand, and Donald L. DeAngelis. 2005. "Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology." *Science* 310, no. 5750 (November): 987–91. ISSN: 1095-9203. doi:10.1126/science.1116681.

Hägerstrand, Torsten. 1970. "What about People in Regional Science?" 24 (1): 7–21.

Handy, Susan, and Debbie Niemeier. 1997. "Measuring Accessibility: An Exploration of Issues and Alternatives." *Environment and planning A* 29 (7): 1175–1194.

Hansen, Walter. 1959. "How Accessibility Shapes Land Use." *Journal of the American Institute of Planners* 25, no. 2 (May): 73–76. ISSN: 0002-8991. doi:10.1080/01944365908978307.

Hao, Haiming, Wanjing Ma, and Hongfeng Xu. 2016. "A Fuzzy Logic-Based Multi-Agent Car-Following Model." *Transportation Research Part C: Emerging Technologies* 69 (August): 477–496. ISSN: 0968-090X. doi:10.1016/j.trc.2015.09.014.

Hart, Peter E., Nils J. Nilsson, and Bertram Raphael. 1968. "A Formal Basis for the Heuristic Determination of Minimum Cost Paths." Conference Name: IEEE Transactions on Systems Science and Cybernetics, *IEEE Transactions on Systems Science and Cybernetics* 4, no. 2 (July): 100–107. ISSN: 2168-2887. doi:10.1109/TSSC.1968.300136.

Heppenstall, Alison, Nick Malleson, and Andrew Crooks. 2016. ““Space, the Final Frontier”: How Good Are Agent-Based Models at Simulating Individuals and Space in Cities?” *Systems* 4, no. 1 (January): 9. ISSN: 2079-8954. doi:10.3390/systems4010009.

Horni, Andreas, Kai Nagel, and Kay W. Axhausen, eds. 2016. *The Multi-Agent Transport Simulation MATSim.* London: Ubiquity Press, August. ISBN: 978-1-909188-75-4. doi:https://doi.org/10.5334/baw.

Horni, Andreas, Darren M. Scott, Michael Balmer, and Kay W. Axhausen. 2009. “Location Choice Modeling for Shopping and Leisure Activities with MATSim: Combining Microsimulation and Time Geography.” *Transportation Research Record* 2135, no. 1 (January): 87–95. ISSN: 0361-1981. doi:10.3141/2135-11.

Horton, Frank E., and David R. Reynolds. 1970. “Action Space Formation: A Behavioral Approach to Predicting Urban Travel Behavior.” *Highway Research Record,* no. 322.

Huck, Jonny J., J. Duncan Whyatt, John Dixon, Brendan Sturgeon, Bree Hocking, Gemma Davies, Neil Jarman, and Dominic Bryan. 2019. “Exploring Segregation and Sharing in Belfast: A PGIS Approach.” *Annals of the American Association of Geographers* 109, no. 1 (January): 223–241. ISSN: 2469-4452. doi:10.1080/24694452.2018.1480930.

IBGE. 2010. *Brazilian Demographic Census 2010.* Brasília, Brazil.

Ingram, David. 1971. “The Concept of Accessibility: A Search for an Operational Form.” *Regional Studies* 5, no. 2 (July): 101–107. ISSN: 0034-3404, 1360-0591. doi:10.1080/09595237100185131.

Jakle, John, Stanley Brunn, and Curtis Roseman. 1976. *Human Spatial Behavior: A Social Geography.* Duxbury Press.

Jang, Woo, and Xiaobai Yao. 2014. “Tracking Ethnically Divided Commuting Patterns Over Time: A Case Study of Atlanta.” *The Professional Geographer* 66, no. 2 (April): 274–283. ISSN: 0033-0124. doi:10.1080/00330124.2013.784952.

Jiang, Bin, and Tao Jia. 2011. "Agent-Based Simulation of Human Movement Shaped by the Underlying Street Structure." *International Journal of Geographical Information Science* 25, no. 1 (February): 51–64. ISSN: 1365-8816, 1362-3087. doi:10.1080/13658811003712864.

Johnston, Ron, Michael Poulsen, and James Forrest. 2004. "The Comparative Study of Ethnic Residential Segregation in the USA, 1980-2000." *Tijdschrift voor Economische en Sociale Geografie* 95, no. 5 (December): 550–569. ISSN: 0040-747X, 1467-9663. doi:10.1111/j.0040-747X.2004.00339.x.

Jordan, René, Mark Birkin, and Andrew Evans. 2014. "An Agent-Based Model of Residential Mobility: Assessing the Impacts of Urban Regeneration Policy in the EASEL District." *Computers, Environment and Urban Systems* 48 (November): 49–63. ISSN: 0198-9715. doi:10.1016/j.compenvurbsys.2014.06.006.

Kamruzzaman, Md., and Julian Hine. 2012. "Analysis of Rural Activity Spaces and Transport Disadvantage Using a Multi-Method Approach." *Transport Policy* 19, no. 1 (January): 105–120. ISSN: 0967-070X. doi:10.1016/j.tranpol.2011.09.007.

Kamruzzaman, Md., Julian Hine, Banihan Gunay, and Neale Blair. 2011. "Using GIS to Visualise and Evaluate Student Travel Behaviour." *Journal of Transport Geography* 19, no. 1 (January): 13–32. ISSN: 0966-6923. doi:10.1016/j.jtrangeo.2009.09.004.

Kazagli, Evanthia, Michel Bierlaire, and Gunnar Flötteröd. 2016. "Revisiting the Route Choice Problem: A Modeling Framework Based on Mental Representations." *Journal of Choice Modelling* 19 (June): 1–23. ISSN: 17555345. doi:10.1016/j.jocm.2016.06.001.

Kim, Hyun-Mi, and Mei-Po Kwan. 2003. "Space-Time Accessibility Measures: A Geocomputational Algorithm with a Focus on the Feasible Opportunity Set and Possible Activity Duration." *Journal of Geographical Systems* 5, no. 1 (May): 71–91. ISSN: 1435-5930. doi:10.1007/s101090300104.

Korhonen, TIMO, SIMO Hostikka, Simo Heliövaara, Harri Ehtamo, and KATRI Matikainen. 2007. "Integration of an Agent Based Evacuation Simulation and the State-of-the-Art Fire Simulation." In *Proceedings of the 7th Asia-Oceania Symposium on Fire Science & Technology,* 20–22.

Kravari, Kalliopi, and Nick Bassiliades. 2015. “A Survey of Agent Platforms.” *Journal of Artificial Societies and Social Simulation* 18 (1): 11.

Kuijpers, Bart, Harvey Miller, Tijs Neutens, and Walied Othman. 2010. “Anchor Uncertainty and Space-Time Prisms on Road Networks.” *International Journal of Geographical Information Science* 24, no. 8 (June): 1223–1248. ISSN: 1365-8816. doi:10.1080/13658810903321339.

Kuijpers, Bart, and Walied Othman. 2009. “Modeling Uncertainty of Moving Objects on Road Networks via Space–Time Prisms.” *International Journal of Geographical Information Science* 23, no. 9 (September): 1095–1117. ISSN: 1365-8816. doi:10.1080/13658810802097485.

Kuipers, Benjamin, Dan G. Tecuci, and Brian J. Stankiewicz. 2003. “The Skeleton In The Cognitive Map: A Computational and Empirical Exploration.” *Environment and Behavior* 35, no. 1 (January): 81–106. ISSN: 0013-9165, 1552-390X. doi:10.1177/0013916502238866.

Kurani, K. S., and Meh Lee-Gosselin. 1997. “Synthesis of Past Activity Analysis Applications.” In *Activity-Based Travel Forecasting ConferenceDepartment of TransportationFederal Transit AdministrationFederal Highway AdministrationOffice of the Secretary of TransportationEnvironmental Protection Agency.* February.

Kwan, Mei-Po. 1998. “Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-Based Framework.” *Geographical Analysis* 30, no. 3 (July): 191–216. ISSN: 1538-4632. doi:10.1111/j.1538-4632.1998.tb00396.x.

———. 2000a. “Gender Differences in Space-Time Constraints.” *Area* 32 (2): 145–156. ISSN: 1475-4762. doi:10.1111/j.1475-4762.2000.tb00125.x.

———. 2000b. “Interactive Geovisualization of Activity-Travel Patterns Using Three-Dimensional Geographical Information Systems: A Methodological Exploration with a Large Data Set.” *Transportation Research Part C: Emerging Technologies* 8, no. 1 (February): 185–203. ISSN: 0968-090X. doi:10.1016/S0968-090X(00)00017-6.

———. 2012. “The Uncertain Geographic Context Problem.” *Annals of the Association of American Geographers* 102, no. 5 (September): 958–968. ISSN: 0004-5608. doi:10.1080/00045608.2012.687349.

Kwan, Mei-Po. 2013. "Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility." *Annals of the Association of American Geographers* 103, no. 5 (September): 1078–1086. ISSN: 0004-5608. doi:10.1080/00045608.2013.792177.

Kwan, Mei-Po, and Xiao-Dong Hong. 1998. "Network-Based Constraints-Oriented Choice Set Formation Using GIS." *Geographical Systems* 5:139–162.

Kwan, Mei-Po, and Joe Weber. 2003. "Individual Accessibility Revisited: Implications for Geographical Analysis in the Twenty-First Century." *Geographical Analysis* 35, no. 4 (October): 341–353. ISSN: 1538-4632. doi:10.1111/j.1538-4632.2003.tb01119.x.

Laurie, Alexander J., and Narendra K. Jaggi. 2003. "Role of 'Vision' in Neighbourhood Racial Segregation: A Variant of the Schelling Segregation Model." *Urban Studies* 40, no. 13 (December): 2687–2704. ISSN: 0042-0980. doi:10.1080/0042098032000146849.

Lee, Brian H Y, Paul Waddell, Liming Wang, and Ram M Pendyala. 2010. "Reexamining the Influence of Work and Nonwork Accessibility on Residential Location Choices with a Microanalytic Framework." *Environment and Planning A: Economy and Space* 42, no. 4 (April): 913–930. ISSN: 0308-518X. doi:10.1068/a4291.

Lee, Jae Yong, and Mei-Po Kwan. 2011. "Visualisation of Socio-Spatial Isolation Based on Human Activity Patterns and Social Networks in Space-Time." *Tijdschrift voor economische en sociale geografie* 102, no. 4 (September): 468–485. ISSN: 1467-9663. doi:10.1111/j.1467-9663.2010.00649.x.

Lenntorp, Bo. 1976. *Paths in Space-Time Environments: A Time-Geographic Study of Movement Possibilities of Individuals.* Royal University of Lund, Department of Geography. ISBN: 978-91-40-04376-4.

Li, Zhibin, Wei Wang, Pan Liu, and David R. Ragland. 2012. "Physical Environments Influencing Bicyclists' Perception of Comfort on Separated and on-Street Bicycle Facilities." *Transportation Research Part D: Transport and Environment* 17, no. 3 (May): 256–261. ISSN: 1361-9209. doi:10.1016/j.trd.2011.12.001.

Lieberson, Stanley. 1981. "An Asymmetrical Approach to Segregation." In *Ethnic Segregation in Cities,* edited by Ceri Peach, Vaughan Robinson, and Susan Smith, 61–82. London: Croom Helm.

Liu, S., S. Lo, J. Ma, and W. Wang. 2014. "An Agent-Based Microscopic Pedestrian Flow Simulation Model for Pedestrian Traffic Problems." *IEEE Transactions on Intelligent Transportation Systems* 15, no. 3 (June): 992–1001. ISSN: 1524-9050. doi:10.1109/TITS.2013.2292526.

Lu, Yang, Muhammad Adnan, Kakali Basak, Francisco Câmara Pereira, Carlos Carrion, Vahid Hamishagi Saber, Harish Loganathan, and Moshe E. Ben-Akiva. 2015. "Simmobility Mid-Term Simulator: A State of the Art Integrated Agent Based Demand and Supply Model." In *94th Annual Meeting of the Transportation Research Board, Washington, DC.*

Lynch, Kevin. 1960. *The Image of the City.* MIT Press. ISBN: 978-0-262-62001-7.

Manley, Ed, J.D. Addison, and Tao Cheng. 2015. "Shortest Path or Anchor-Based Route Choice: A Large-Scale Empirical Analysis of Minicab Routing in London." *Journal of Transport Geography* 43 (February): 123–139. ISSN: 09666923. doi:10.1016/j.jtrangeo.2015.01.006.

Manley, Ed, and Tao Cheng. 2018. "Exploring the Role of Spatial Cognition in Predicting Urban Traffic Flow through Agent-Based Modelling." *Transportation Research Part A: Policy and Practice* 109 (March): 14–23. ISSN: 09658564. doi:10.1016/j.tra.2018.01.020.

Manson, Steven M., Shipeng Sun, and Dudley Bonsal. 2012. "Agent-Based Modeling and Complexity." In *Agent-Based Models of Geographical Systems,* edited by Alison J. Heppenstall, Andrew T. Crooks, Linda M. See, and Michael Batty, 85–105. Springer.

Massey, Douglas S., and Nancy A. Denton. 1988. "The Dimensions of Residential Segregation." *Social Forces* 67, no. 2 (December): 281–315. ISSN: 0037-7732. doi:10.1093/sf/67.2.281.

McNally, Michael G. 2007. "The Four-Step Model." In *Handbook of Transport Modelling,* 35–53. eprint: http://www.emeraldinsight.com/doi/pdf/10.1108/9780857245670-003.

McNally, Michael G, and Craig Rindt. 2007. "The Activity-Based Approach." In *Handbook of Transport Modelling,* 17.

Menghini, G., N. Carrasco, N. Schüssler, and K. W. Axhausen. 2010. "Route Choice of Cyclists in Zurich." *Transportation Research Part A: Policy and Practice* 44, no. 9 (November): 754–765. ISSN: 0965-8564. doi:10.1016/j.tra.2010.07.008.

Milakis, Dimitris, and Konstantinos Athanasopoulos. 2014. "What about People in Cycle Network Planning? Applying Participative Multicriteria GIS Analysis in the Case of the Athens Metropolitan Cycle Network." *Journal of Transport Geography* 35 (February): 120–129. ISSN: 0966-6923. doi:10.1016/j.jtrangeo.2014.01.009.

Miller, Harvey. 1991. "Modelling Accessibility Using Space-Time Prism Concepts within Geographical Information Systems." *International Journal of Geographical Information Systems* 5, no. 3 (January): 287–301. ISSN: 0269-3798. doi:10.1080/02693799108927856.

———. 1999. "Measuring Space-Time Accessibility Benefits within Transportation Networks: Basic Theory and Computational Procedures." *Geographical analysis* 31 (1): 1–26.

———. 2005. "A Measurement Theory for Time Geography." *Geographical Analysis* 37, no. 1 (January): 17–45. ISSN: 1538-4632. doi:10.1111/j.1538-4632.2005.00575.x.

———. 2007. "Place-Based versus People-Based Geographic Information Science." *Geography Compass* 1, no. 3 (May): 503–535. ISSN: 1749-8198. doi:10.1111/j.1749-8198.2007.00025.x.

Miller, John, and Scott Page. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life.* Princeton, N.J.: Princeton University Press.

Mitchell, Melanie. 2009. *Complexity: A Guided Tour.* Oxford University Press.

Netto, Vinicius M. 2017. "'The Social Fabric of Cities': A Tripartite Approach to Cities as Systems of Interaction." *Area Development and Policy* 2, no. 2 (May): 130–153. ISSN: 2379-2949, 2379-2957. doi:10.1080/23792949.2017.1315540.

Netto, Vinicius M., João Vitor Meirelles, Maíra Pinheiro, and Henrique Lorea. 2018. "A Temporal Geography of Encounters." *Cybergeo : European Journal of Geography* (February). ISSN: 1278-3366. doi:10.4000/cybergeo.28985.

Neutens, Tijs, Tim Schwanen, and Frank Witlox. 2011. "The Prism of Everyday Life: Towards a New Research Agenda for Time Geography." *Transport Reviews* 31, no. 1 (January): 25–47. ISSN: 0144-1647. doi:10.1080/01441647.2010.484153.

Neutens, Tijs, Tim Schwanen, Frank Witlox, and Philippe De Maeyer. 2010. "Equity of Urban Service Delivery: A Comparison of Different Accessibility Measures." *Environment and Planning A* 42, no. 7 (July): 1613–1635. ISSN: 0308-518X, 1472-3409. doi:10.1068/a4230.

Neutens, Tijs, Mathias Versichele, and Tim Schwanen. 2010. "Arranging Place and Time: A GIS Toolkit to Assess Person-Based Accessibility of Urban Opportunities." *Applied Geography* 30, no. 4 (December): 561–575. ISSN: 01436228. doi:10.1016/j.apgeog.2010.05.006.

Neutens, Tijs, Nico Van de Weghe, Frank Witlox, and Philippe De Maeyer. 2008. "A Three-Dimensional Network-Based Space–Time Prism." *Journal of Geographical Systems* 10, no. 1 (March): 89–107. ISSN: 1435-5930, 1435-5949. doi:10.1007/s10109-007-0057-x.

Newsome, Tracy H., Wayne A. Walcott, and Paul D. Smith. 1998. "Urban Activity Spaces: Illustrations and Application of a Conceptual Model for Integrating the Time and Space Dimensions." *Transportation* 25, no. 4 (November): 357–377. ISSN: 0049-4488, 1572-9435. doi:10.1023/A:1005082827030.

Ngo, The An, and Linda See. 2012. "Calibration and Validation of Agent-Based Models of Land Cover Change." In *Agent-Based Models of Geographical Systems,* edited by Alison J. Heppenstall, Andrew T. Crooks, Linda M. See, and Michael Batty, 181–197. Springer.

North, Michael J., and Charles M. Macal. 2007. *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation.* Oxford University Press, USA, March. ISBN: 978-0-19-517211-9.

O'Sullivan, David, J MacGill, and Chaoqing Yu. 2003. "Agent-Based Residential Segregation: A Hierarchically Structured Spatial Model." In *Proceedings of Agent 2003 Conference on Challenges in Social Simulation,* 493–507. The University of Chicago Chicago, IL.

O'Sullivan, David, Alastair Morrison, and John Shearer. 2000. "Using Desktop GIS for the Investigation of Accessibility by Public Transport: An Isochrone Approach." *International Journal of Geographical Information Science* 14, no. 1 (January): 85–104. ISSN: 1365-8816, 1362-3087. doi:10.1080/136588100240976.

O'Sullivan, David, and George L. W. Perry. 2013. "Spatial Simulation Models: What? Why? How?" In *Spatial Simulation: Exploring Pattern and Process,* 1–28. John Wiley & Sons.

Office For National Statistics. 2018. *Estimates of the Population for the UK, England and Wales, Scotland and Northern Ireland.*

Okabe, Atsuyuki, and Masayuki Kitamura. 1996. "A Computational Method for Market Area Analysis on a Network." *Geographical Analysis* 28, no. 4 (October): 330–349. ISSN: 1538-4632. doi:10.1111/j.1538-4632.1996.tb00939.x.

Okazaki, Shigeyuki, and Satoshi Matsushita. 1993. "A Study of Simulation Model for Pedestrian Movement with Evacuation and Queuing." In *International Conference on Engineering for Crowd Safety,* vol. 271.

Openshaw, Stan. 1984. "The Modifiable Areal Unit Problem." Geo Abstracts University of East Anglia. ISBN: 0-86094-134-5.

Owen, Andrew. 2014. *This Map Wants to Change How You Think About Your Commute.*

Palmer, John R. B., Thomas J. Espenshade, Frederic Bartumeus, Chang Y. Chung, Necati Ercan Ozgencil, and Kathleen Li. 2013. "New Approaches to Human Mobility: Using Mobile Phones for Demographic Research." *Demography* 50, no. 3 (June): 1105–1128. ISSN: 0070-3370, 1533-7790. doi:10.1007/s13524-012-0175-z.

Pancs, Romans, and Nicolaas J. Vriend. 2007. "Schelling's Spatial Proximity Model of Segregation Revisited." *Journal of Public Economics* 91, no. 1 (February): 1–24. ISSN: 0047-2727. doi:10.1016/j.jpubeco.2006.03.008.

Park, Yoo Min, and Mei-Po Kwan. 2017. "Multi-Contextual Segregation and Environmental Justice Research: Toward Fine-Scale Spatiotemporal Approaches." *International Journal of Environmental Research and Public Health* 14, no. 10 (October): 1205. doi:10.3390/ijerph14101205.

Patel, Amit, Andrew Crooks, and Naoru Koizumi. 2012. "Simulating Spatio-Temporal Dyanmics of Slum Formation in Ahmedabad, India." *Sixth Urban Research and Knowledge Symposium 2012:* 27.

Patterson, Zachary, and Steven Farber. 2015. "Potential Path Areas and Activity Spaces in Application: A Review." *Transport Reviews* 35, no. 6 (November): 679–700. ISSN: 0144-1647, 1464-5327. doi:10.1080/01441647.2015.1042944.

Pelechano, N., and N. I. Badler. 2006. "Modeling Crowd and Trained Leader Behavior during Building Evacuation." *IEEE Computer Graphics and Applications* 26, no. 6 (November): 80–86. ISSN: 0272-1716. doi:10.1109/MCG.2006.133.

Pirie, Gordon H. 1979. "Measuring Accessibility: A Review and Proposal." *Environment and Planning A* 11 (3): 299–312.

Portugali, J, I Benenson, and I Omer. 1997. "Spatial Cognitive Dissonance and Sociospatial Emergence in a Self-Organizing City." *Environment and Planning B: Planning and Design* 24, no. 2 (April): 263–285. ISSN: 0265-8135. doi:10.1068/b240263.

Portugali, Juval. 2011. *Complexity, Cognition and the City.* Springer Science & Business Media, July. ISBN: 978-3-642-19450-4.

Pred, Allan. 1977. "The Choreography of Existence: Comments on Hägerstrand's Time-Geography and Its Usefulness." *Economic Geography* 53, no. 2 (April): 207–221. ISSN: 0013-0095. doi:10.2307/142726.

Prelipcean, Adrian C., Győző Gidófalvi, and Yusak O. Susilo. 2018. "MEILI: A Travel Diary Collection, Annotation and Automation System." *Computers, Environment and Urban Systems* 70 (July): 24–34. ISSN: 0198-9715. doi:10.1016/j.compenvurbsys.2018.01.011.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria.

Rai, R., Michael Balmer, Marcel Rieser, V. Vaze, Stefan Schönfelder, and Kay Axhausen. 2007. "Capturing Human Activity Spaces: New Geometries." *Transportation Research Record: Journal of the Transportation Research Board* 2021 (December): 70–80. ISSN: 0361-1981. doi:10.3141/2021-09.

Railsback, S. F., S. L. Lytinen, and S. K. Jackson. 2006. "Agent-Based Simulation Platforms: Review and Development Recommendations." *SIMULATION* 82, no. 9 (September): 609–623. ISSN: 0037-5497. doi:10.1177/0037549706073695.

Rasouli, Soora, and Harry Timmermans. 2014. "Activity-Based Models of Travel Demand: Promises, Progress and Prospects." *International Journal of Urban Sciences* 18, no. 1 (January): 31–60. ISSN: 1226-5934, 2161-6779. doi:10.1080/12265934.2013.835118.

Reardon, Sean F. 2006. "A Conceptual Framework for Measuring Segregation and Its Association with Population Outcomes." In *Methods in Social Epidemiology,* edited by J. Michael Oakes and Jay S. Kaufmann, 169–192. San Francisco, CA: Jossey-Bass.

Reardon, Sean F., and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation." *Sociological Methodology* 32, no. 1 (August): 33–67. ISSN: 0081-1750, 1467-9531. doi:10.1111/1467-9531.00110.

Reardon, Sean F., and David O'Sullivan. 2004. "Measures of Spatial Segregation." *Sociological methodology* 34 (1): 121–162.

Rindsfüser, Guido, and Dr Franziska Klügl. 2007. "Agent-Based Pedestrian Simulation." *disP - The Planning Review* 43, no. 170 (January): 9–18. ISSN: 0251-3625. doi:10.1080/02513625.2007.10556985.

Rissel, Chris E. 2009. "Active Travel: A Climate Change Mitigation Strategy with Co-Benefits for Health." *New South Wales Public Health Bulletin* 20, no. 2 (March): 10–13. ISSN: 1834-8610. doi:10.1071/NB08043.

Rissel, Chris, Stephen Greaves, Li Ming Wen, Anthony Capon, Melanie Crane, and Chris Standen. 2013. "Evaluating the Transport, Health and Economic Impacts of New Urban Cycling Infrastructure in Sydney, Australia – Protocol Paper." *BMC Public Health* 13, no. 1 (October): 963. ISSN: 1471-2458. doi:10.1186/1471-2458-13-963.

Rybarczyk, Greg. 2014. "Simulating Bicycle Wayfinding Mechanisms in an Urban Environment." *Urban, Planning and Transport Research* 2, no. 1 (January): 89–104. ISSN: null. doi:10.1080/21650020.2014.906909.

Safi, Hamid, Behrang Assemi, Mahmoud Mesbah, and Luis Ferreira. 2017. "An Empirical Comparison of Four Technology-Mediated Travel Survey Methods." *Journal of Traffic and Transportation Engineering (English Edition)* 4, no. 1 (February): 80–87. ISSN: 20957564. doi:10.1016/j.jtte.2015.12.003.

Salvini, Paul, and Eric Miller. 2005. "ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems." *Networks and Spatial Economics* 5, no. 2 (June): 217–234. ISSN: 1566-113X, 1572-9427. doi:10.1007/s11067-005-2630-5.

Samuelson, Douglas A, Matthew Parker, Austin Zimmerman, Loren Miller, Stephen Guerin, Joshua Thorp, and Owen Densmore. 2008. "Agent-Based Simulations of Mass Egress After Improvised Explosive Device Attacks": 12.

Santos, Adella, Nancy McGuckin, Hikari Nakamoto, Danielle Gray, and Susan Liss. 2011. "Summary of Travel Trends: 2009 National Household Travel Survey."

São Paulo, Metrô de. 2007. *Pesquisa Origem e Destino 2007: Síntese das Informações da Pesquisa Domiciliar.* Technical report.

Saraiva, Marcus, and Joana Barros. 2017. "The AXS Model: an agent-based simulation model for urban flows." In *GeoComputation 2017, Leeds, UK.*

Schelling, Thomas. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1 (2): 143–186.

———. 1978. *Micromotives and Macrobehavior.*

Schönfelder, Stefan, and Kay Axhausen. 2003. “Activity Spaces: Measures of Social Exclusion?” *Transport Policy,* Transport and Social Exclusion, 10, no. 4 (October): 273–286. ISSN: 0967-070X. doi:10.1016/j.tranpol.2003.07.002.

Schwanen, Tim, Mei-Po Kwan, and Fang Ren. 2008. “How Fixed Is Fixed? Gendered Rigidity of Space–Time Constraints and Geographies of Everyday Activities.” *Geoforum* 39, no. 6 (November): 2109–2121. ISSN: 00167185. doi:10.1016/j.geoforum.2008.09.002.

Sener, Ipek N., Naveen Eluru, and Chandra R. Bhat. 2009. “An Analysis of Bicycle Route Choice Preferences in Texas, US.” *Transportation* 36, no. 5 (September): 511–539. ISSN: 1572-9435. doi:10.1007/s11116-009-9201-4.

Serras, Joan. 2005. “Extending TRANSIMS Technology to an Integrated Multilevel Representation.”

Sherman, Jill E., John Spencer, John S. Preisser, Wilbert M. Gesler, and Thomas A. Arcury. 2005. “A Suite of Methods for Representing Activity Space in a Healthcare Accessibility Study.” *International Journal of Health Geographics* 4 (October): 24. ISSN: 1476-072X. doi:10.1186/1476-072X-4-24.

Shi, Jianyong, Aizhu Ren, and Chi Chen. 2009. “Agent-Based Evacuation Model of Large Public Buildings under Fire Conditions.” *Automation in Construction* 18, no. 3 (May): 338–347. ISSN: 0926-5805. doi:10.1016/j.autcon.2008.09.009.

Shoko, M., and J. Smit. 2013. “Use of Agent Based Modelling to Investigate the Dynamics of Slum Growth.” *South African Journal of Geomatics* 2, no. 1 (January): 54–67. ISSN: 2225-8531.

Silm, Siiri, and Rein Ahas. 2014a. “Ethnic Differences in Activity Spaces: A Study of Out-of-Home Nonemployment Activities with Mobile Phone Data.” *Annals of the Association of American Geographers* 104, no. 3 (May): 542–559. ISSN: 0004-5608. doi:10.1080/00045608.2014.892362.

———. 2014b. “The Temporal Variation of Ethnic Segregation in a City: Evidence from a Mobile Phone Use Dataset.” *Social Science Research* 47, no. Supplement C (September): 30–43. ISSN: 0049-089X. doi:10.1016/j.ssresearch.2014.03.011.

Smith, L., R. Beckman, D. Anson, K. Nagel, and M. Williams. 1995. *Transims: Transportation Analysis and Simulation System.* Technical report LA-UR–95-1664; CONF-9504197–1. Los Alamos National Lab., NM (United States), August. doi:10.2172/88648.

Stopher, Peter R., David T. Hartgen, and Yuanjun Li. 1996. "SMART: Simulation Model for Activities, Resources and Travel." *Transportation* 23, no. 3 (August): 293–312. ISSN: 0049-4488, 1572-9435. doi:10.1007/BF00165706.

Stouffer, Samuel A. 1940. "Intervening Opportunities: A Theory Relating Mobility and Distance." *American Sociological Review* 5 (6): 845–867. ISSN: 0003-1224. doi:10.2307/2084520.

Stubenschrott, Martin, Christian Kogler, Thomas Matyus, and Stefan Seer. 2014. "A Dynamic Pedestrian Route Choice Model Validated in a High Density Subway Station." *Transportation Research Procedia* 2:376–384. ISSN: 23521465. doi:10.1016/j.trpro.2014.09.036.

Sycara, Katia P. 1998. "Multiagent Systems." *AI magazine* 19 (2): 79.

Theil, Henri, and Anthony Finizza. 1971. "A Note on the Measurement of Racial Integration of Schools by Means of Informational Concepts†." *The Journal of Mathematical Sociology* 1, no. 2 (July): 187–193. ISSN: 0022-250X, 1545-5874. doi:10.1080/0022250X.1971.9989795.

Thompson, Jason, Giovanni Savino, and Mark Stevenson. 2015. "Reconsidering the Safety in Numbers Effect for Vulnerable Road Users: An Application of Agent-Based Modeling." *Traffic Injury Prevention* 16, no. 2 (February): 147–153. ISSN: 1538-9588. doi:10.1080/15389588.2014.914626.

Timmermans, Harry, and Theo A. Arentze. 2011. "Transport Models and Urban Planning Practice: Experiences with Albatross." *Transport Reviews* 31, no. 2 (March): 199–207. ISSN: 0144-1647. doi:10.1080/01441647.2010.518292.

Torrens, Paul M. 2010. "Agent-Based Models and the Spatial Sciences." *Geography Compass* 4, no. 5 (May): 428–448. ISSN: 17498198, 17498198. doi:10.1111/j.1749-8198.2009.00311.x.

———. 2012. "Moving Agent Pedestrians Through Space and Time." *Annals of the Association of American Geographers* 102, no. 1 (January): 35–66. ISSN: 0004-5608. doi:10.1080/00045608.2011.595658.

Tsai, Jason, Natalie Fridman, Emma Bowring, Matthew Brown, Shira Epstein, Gal Kaminka, Stacy Marsella, et al. 2011. “ESCAPES: Evacuation Simulation with Children, Authorities, Parents, Emotions, and Social Comparison.” In *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 2,* 457–464. AAMAS ’11. Taipei, Taiwan: International Foundation for Autonomous Agents and Multiagent Systems. ISBN: 0-9826571-6-1 978-0-9826571-6-4.

Turner, Alasdair, and Alan Penn. 2002. “Encoding Natural Movement as an Agent-Based System: An Investigation into Human Pedestrian Behaviour in the Built Environment.” *Environment and Planning B: Planning and Design* 29, no. 4 (August): 473–490. ISSN: 0265-8135, 1472-3417. doi:10.1068/b12850.

Urry, John. 2007. *Mobilities.* Polity. ISBN: 978-0-7456-3419-7.

Vahidi, Hossein, and Wanglin Yan. 2016. “How Is an Informal Transport Infrastructure System Formed? Towards a Spatially Explicit Conceptual Model.” *Open Geospatial Data, Software and Standards* 1, no. 1 (December): 8. ISSN: 2363-7501. doi:10.1186/s40965-016-0009-9.

Vallée, Julie, and Pierre Chauvin. 2012. “Investigating the Effects of Medical Density on Health-Seeking Behaviours Using a Multiscale Approach to Residential and Activity Spaces: Results from a Prospective Cohort Study in the Paris Metropolitan Area, France.” *International Journal of Health Geographics* 11 (December): 54. ISSN: 1476-072X. doi:10.1186/1476-072X-11-54.

Van der Leeuw, S. E. 2004. “Why Model?” *Cybernetics and Systems* 35, nos. 2-3 (March): 117–128. ISSN: 0196-9722. doi:10.1080/01969720490426803.

Wagner, Neal, and Vikas Agrawal. 2014. “An Agent-Based Simulation System for Concert Venue Crowd Evacuation Modeling in the Presence of a Fire Disaster.” *Expert Systems with Applications* 41, no. 6 (May): 2807–2815. ISSN: 0957-4174. doi:10.1016/j.eswa.2013.10.013.

Waldrop, M. Mitchell. 2018. “Free Agents.” *Science* 360, no. 6385 (April): 144–147. ISSN: 0036-8075, 1095-9203. doi:10.1126/science.360.6385.144.

Wang, Donggen, and Fei Li. 2016. "Daily Activity Space and Exposure: A Comparative Study of Hong Kong's Public and Private Housing Residents' Segregation in Daily Life." *Cities* 59, no. Supplement C (November): 148–155. ISSN: 0264-2751. doi:10.1016/j.cities.2015.09.010.

Wang, Donggen, Fei Li, and Yanwei Chai. 2012. "Activity Spaces and Sociospatial Segregation in Beijing." *Urban Geography* 33, no. 2 (February): 256–277. ISSN: 0272-3638. doi:10.2747/0272-3638.33.2.256.

Weber, Joe, and Mei-Po Kwan. 2002. "Bringing Time Back In: A Study on the Influence of Travel Time Variations and Facility Opening Hours on Individual Accessibility." *The Professional Geographer* 54, no. 2 (May): 226–240. ISSN: 0033-0124. doi:10.1111/0033-0124.00328.

Wegener, Michael. 2004. "Overview of Land Use Transport Models." In *Handbook of Transport Geography and Spatial Systems,* 127–146. Emerald Group Publishing Limited.

Wilensky, Uri. 1999. *NetLogo.* Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.

Wong, David W. S., and Shih-Lung Shaw. 2011. "Measuring Segregation: An Activity Space Approach." *Journal of Geographical Systems* 13, no. 2 (June): 127–145. ISSN: 1435-5930, 1435-5949. doi:10.1007/s10109-010-0112-x.

Wooldridge, Michael. 2009. *An Introduction to Multiagent Systems.* Chichester, UK: John Wiley & Sons.

———. 2013. "Intelligent Agents." In *Multiagent Systems - 2nd Edition,* edited by Gerhard Weiss. Cambridge, MA: MIT press.

Wu, Yi-Hwa, and Harvey Miller. 2001. "Computational Tools for Measuring Space-Time Accessibility within Transportation Networks with Dynamic Flow." *Journal of Transportation and Statistics* 4 (2/3): 1–14.

Yip, Ngai Ming, Ray Forrest, and Shi Xian. 2016. "Exploring Segregation and Mobilities: Application of an Activity Tracking App on Mobile Phone." *Cities* 59, no. Supplement C (November): 156–163. ISSN: 0264-2751. doi:10.1016/j.cities.2016.02.003.

Young, Iris Marion. 2002. *Inclusion and Democracy.* New Ed edition. Oxford: Oxford University Press, June. ISBN: 978-0-19-829755-0.

Zamenopoulos, Theodore, and Katerina Alexiou. 2012. "A Complexity Theoretic View of Cities as Artefacts of Design Intentionality." In *Complexity Theories of Cities Have Come of Age: An Overview with Implications to Urban Planning and Design,* edited by Juval Portugali, Han Meyer, Egbert Stolk, and Ekim Tan, 327–346. Berlin, Heidelberg: Springer Berlin Heidelberg.

Zenk, Shannon N., Amy J. Schulz, Stephen A. Matthews, Angela Odoms-Young, JoEllen Wilbur, Lani Wegrzyn, Kevin Gibbs, Carol Braunschweig, and Carmen Stokes. 2011. "Activity Space Environment and Dietary and Physical Activity Behaviors: A Pilot Study." *Health & Place* 17, no. 5 (September): 1150–1161. ISSN: 1353-8292. doi:10.1016/j.healthplace.2011.05.001.

Zhuge, Chengxiang, Chunfu Shao, Jian Gao, Chunjiao Dong, and Hui Zhang. 2016. "Agent-Based Joint Model of Residential Location Choice and Real Estate Price for Land Use and Transport Model." *Computers, Environment and Urban Systems* 57 (May): 93–105. ISSN: 0198-9715. doi:10.1016/j.compenvurbsys.2016.02.001.

Ziemke, Dominik, Simon Metzler, and Kai Nagel. 2017. "Modeling Bicycle Traffic in an Agent-Based Transport Simulation." *Procedia Computer Science,* 8th International Conference on Ambient Systems, Networks and Technologies, ANT-2017 and the 7th International Conference on Sustainable Energy Information Technology, SEIT 2017, 16-19 May 2017, Madeira, Portugal, 109 (January): 923–928. ISSN: 1877-0509. doi:10.1016/j.procs.2017.05.424.

Appendix A

Exposure and Isolation Indices

This appendix details the place-based exposure and isolation metrics used in this study, following approach developed by Bell (1954). The *exposure index (P^*)* measures the probability of members of one group meeting members of other group, while the *isolation index (Q)* measures the probability of encounter between members of the same group.

In the following equations, N denotes population size; subscripts i and j index spatial units; and subscripts m and n index population groups. Hence:

N = total population in the study area;

N_j = total population of spatial unit j;

N_m = total population of group m in the study area;

N_{jm} = total population of group m in spatial unit j.

The exposure of group m to group n, denoted as $P^*_{(m,n)}$, is the proportion of members of group n in the spatial units of all members of m. It is calculated as per equation A.1.

$$P^*_{(m,n)} = \sum_{j=1}^{J} \left[\left(\frac{N_{jm}}{N_m} \right) \left(\frac{N_{jn}}{N_j} \right) \right] \tag{A.1}$$

The isolation of group m, denoted as $Q_{(m)}$, is a particular case of exposure, referring to the exposure of group m to itself. It is calculated as per equation A.2.

$$Q_{(m)} = \sum_{j=1}^{J} \left[\left(\frac{N_{jm}}{N_m} \right) \left(\frac{N_{jm}}{N_j} \right) \right] \tag{A.2}$$

Both exposure and isolation indices range from 0 (indicating no exposure) to 1 (indicating maximum exposure). It is important to note that exposure and isolation indices are highly sensitive to the population distribution in the study area and in each spatial unit. More specifically, the higher the proportion of a population group in the total population, higher the likelihood any individual will encounter a member of that group in the city, by random chance. To account for that distortion, relative exposure $rP^*_{(m,n)}$ and relative isolation $rQ_{(m)}$ indices can be calculated by discounting the relative size of the target population group from the original exposure and isolation indices.

Following approach described by Johnston, Poulsen, and Forrest (2004), equations A.3 and A.4 denote relative exposure and relative isolation indices, respectively.

$$rP^*_{(m,n)} = P^*_{(m,n)} - (N_n/N) \tag{A.3}$$

$$rQ_{(m)} = Q_{(m)} - (N_m/N) \tag{A.4}$$

Relative exposure and relative isolation indices present advantages over their counterparts, mainly in terms of interpretability. For example, as discussed in Johnston, Poulsen, and Forrest (2004), very low (close to 0) or very high (close to 1) values of isolation can be easy to understand: or the group is widespread in the city (a low index), or it is very concentrated in a few areas (a high index). For exposure, the interpretation is the opposite, yet still straightforward: the higher the exposure to a group, the higher the group's presence across the study area. Intermediate values, however, are not as meaningful and can lead to ambiguous interpretation. Relative indices are easier to interpret. Values of relative exposure close to 0 mean the exposure to a given population group is close to the expected given that group's proportion in the study area. Positive values mean the likelihood of encountering a member of such group is higher than it would be by random change, due to that group's distribution in the city, and the opposite is true for negative values. Finally, the magnitude of the relative values allow comparisons across study areas: a relative exposure of 0.4 is higher than a relative exposure of 0.2, regardless of city size and population distribution.