Urban segregation as a complex system: an agent-based simulation approach

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

FLÁVIA DA FONSECA FEITOSA

aus

JUIZ DE FORA, BRASILIEN

Bonn 2010

1. Referent: Prof. Dr. Paul L.G. Vlek

2. Referent: Prof. Dr. Eckart Ehlers

Tag der Promotion: 04.03.2010

Erscheinungsjahr: 2010

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn
http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert

ABSTRACT

Urban segregation represents a significant barrier for achieving social inclusion in cities. To overcome this, it is necessary to implement policies founded upon a better understanding of segregation dynamics. However, a crucial challenge for achieving such understanding lies in the fact that segregation is a complex system. It emerges from local interactions able to produce unexpected and counterintuitive outcomes that cannot be defined a priori.

This study adopts an agent-based simulation approach that addresses the complex nature of segregation. It proposes a model named MASUS, Multi-Agent Simulator for Urban Segregation, which provides a virtual laboratory for exploring theoretical issues and policy approaches concerning segregation. The MASUS model was first implemented for São José dos Campos, a medium-sized Brazilian city. Based on the data of this city, the model was parameterized and calibrated.

The potential of MASUS is demonstrated through three different sets of simulation experiments. The first compares simulated data with real data, the second tests theories about segregation, and the third explores the impact of anti-segregation policies. The first set of experiments provides a retrospective validation of the model by simulating the segregation dynamics of São José dos Campos during the period 1991-2000. In general, simulated and real data reveal the same trends, a result that demonstrates that the model is able to accurately represent the segregation dynamics of the study area.

The second set of experiments aims at demonstrating the potential of the model to explore and test theoretical issues about urban segregation. These experiments explore the impact of two mechanisms on segregation: income inequality and personal preferences. To test the impact of income inequality, scenarios considering different income distributions were simulated and compared. The results show how decreasing levels of income inequality promote the spatial integration of different social groups in the city. Additional tests were conducted to explore how the preferences of high-income families regarding the presence of other income groups could affect segregation patterns. The results reveal that the high levels of segregation were maintained even in a scenario where affluent households did not take into account the income composition of neighborhoods when selecting their residential location.

Finally, the third set of experiments provides new insights about the impact of different urban policies on segregation. One experiment tests whether the regularization of clandestine settlements and equitable distribution of infrastructure would affect the segregation trends in the city. The simulated outputs indicate that they had no significant impact on the segregation patterns. Besides this test focusing on a general urban policy, two specific social-mix policy approaches were explored: poverty dispersion and wealth dispersion. The results suggest that policies based on poverty dispersion, which have been adopted in cities in Europe and the United States, are less effective in developing countries, where poor families represent a large share of the population. On the other hand, the policy based on wealth dispersion was able to produce substantial and long-term improvements in the segregation patterns of the city.

Städtische Segregation als komplexes System: Ein agentenbasierter Simulationsansatz

KURZFASSUNG

Die städtische Segregation stellt eine bedeutende Barriere für die Erreichung der sozialen Inclusion in den Städten dar. Um diese zu überwinden, ist es notwendig, eine Politik zu betreiben, die die Dynamiken der Segregation besser versteht und berücksichtigt. Eine besondere Herausforderung für ein besseres Verständnis dieser Dynamik ist die Tatsache, dass Segregation ein komplexes System ist. Dieses System entsteht aus lokalen Interaktionen, die zu unerwarteten und nicht eingängigen Ergebnissen führt, die nicht von vornherein bestimmt werden können.

Diese Studie wendet einen multi-agenten Simulationsmodel an, das die komplexe Natur der Segregation berücksichtigt. Es schlägt ein Modell mit dem Namen MASUS (Multi-Agent Simulator for Urban Segregation) vor. Dieses bietet ein virtuelles Labor für die Untersuchung der theoretischen Aspekte und Politikansätze der Segregation. Das Modell wurde für São José dos Campos, eine mittelgroße brasilianische Stadt, eingesetzt. Das Modell wurde auf der Grundlage der Daten dieser Stadt parametisiert und kallibriert.

Das Potenzial von MASUS wird durch drei verschiedene Arten von Simulationsexperimente dargestellt. Die erste vergleicht simulierte Daten mit realen Daten, die zweite prüft Segregationstheorien, und die dritte untersucht die Auswirkungen von Antisegregationspolitik. Die erste Gruppe von Experimenten liefert eine rückblickende Validierung des Modells durch die Simulation der Segregationsdynamiken von São José dos Campos im Zeitraum 1991-2000. Die simulierten und realen Daten zeigen im Allgemeinen die gleichen Trends. Dies zeigt, dass das Modell in der Lage ist, die Segregationsdynamik im Untersuchungsgebiet korrekt darzustellen.

Die zweite Gruppe von Experimenten hat zum Ziel, das Potenzial des Modells hinsichtlich der Untersuchung und Prüfung der theoretischen Aspekte städtischer Segregation darzustellen. Diese Experimente untersuchen die Auswirkung von zwei Mechanismen auf Segregation: Einkommensungleichheit und persönliche Präferenzen. Um die Auswirkungen von Einkommensungleichheit zu prüfen, wurden Szenarien mit unterschiedlichen Einkommensverteilungen simuliert und verglichen. Die Ergebnisse zeigen wie abnehmende Einkommenshöhen die räumliche Integration von verschiedenen sozialen Gruppen in der Stadt fördern. Zusätzliche Tests wurden durchgeführt, um zu untersuchen wie die Präferenzen von Haushalten mit hohen Einkommen im Bezug auf das Vorhandensein anderer Einkommensgruppen die Segregationsmuster beeinflussen könnten. Die Ergebnisse zeigen, dass die Segregation auf hohem Niveau blieb sogar in einem Szenario wo wohlhabende Haushalte das Einkommensgefüge der Nachbarschaft bei der Wahl ihrer Wohngegend nicht berücksichtigten.

Die dritte Gruppe von Experimenten führt zu neuen Einsichten über die Auswirkungen von verschiedenen städtischen politischen Maßnahmen auf die Segregation. Ein Experiment prüft ob die Regulierung von illegalen Siedlungen und die gleichmäßige Verteilung der Infrastruktur die Segregationstrends in der Stadt

beeinflussen. Die Ergebnisse der Simulation zeigen, dass diese keine signifikante Auswirkung auf die Segregationsmuster haben. Neben diesem Test, der die allgemeine städtische Politik zum Inhalt hat, wurden zwei Ansätze der spezifischen Sozialen-Mix-Politik untersucht: Armutsverteilung und Wohlstandsverteilung. Die Ergebnisse deuten daraufhin, dass eine Politik der Armutsverteilung, die aus europäischen und nordamerikanischen Städten bekannt ist, weniger wirkungsvoll in Entwicklungsländern ist, wo arme Familien einen Großteil der Bevölkerung darstellen. Auf der anderen Seite führte eine Politik der Wohlstandsverteilung zu erheblichen und langfristigen Verbesserungen der Segregationsmuster der Stadt.

TABLE OF CONTENTS

1 INTRODUCTION

In 2008, for the first time, the majority of the population on Earth lived in urban areas. By the year 2030, the urban population will reach 4.9 billion, which is equivalent to 60% of the global population. Nearly all of this population growth will take place in the cities of developing nations (UNFPA 2007). In this urbanized global context, the need to fulfill the potential of cities as engines of economic and social development has never been greater.

While cities are often associated with poverty concentration, slum proliferation and social disorders, they have also traditionally been the centers of economic growth and innovation. Cities provide the cost-reducing advantages of agglomeration economies as well as many economic and social externalities, including social and cultural amenities, infrastructure, and skilled workers (Todaro and Smith 2008). Urban areas, in particular the large ones, can account for substantial income and wealth creation. The metropolitan region of São Paulo, for example, has 10% of Brazil's population and accounts for almost 25% of the gross domestic product (IBGE 2007,2008). The capital created by cities represents an opportunity for poverty prevention and alleviation. Nevertheless, more than enhancing progress or development, the rapid spread of urbanization in developing countries associated with misguided urban policies has created an exclusionary urban order that reflects and reproduces the injustices and inequalities of society (Rolnik and Saule Jr. 2001).

To realize the potential role of cities in fostering development, it is essential to remove the barriers that inhibit the formation of inclusive cities, i.e., cities capable of promoting growth with equity (UN-Habitat 2001a). Urban segregation represents one of these barriers, with impacts that have been reinforcing social exclusion[1] in cities of the developing world (UN-Habitat 2001b). Different types of urban segregation exist depending on the context within a city, including income, racial or ethnical segregation. By concentrating on the reality of Brazilian cities, well known for its remarkable levels of social inequality and exclusion, this study focuses on income segregation, which is

[1] Here, the idea of social exclusion extends the concept of poverty. While poverty is related to the purchasing power of individuals, social exclusion also regards ethical and cultural elements, such as discrimination and stigmatization (Sposati 1999).

defined as the separation among the residential location of families belonging to different income groups.

In Brazil and other Latin America countries, the dynamic relation between income segregation and social exclusion has often created a continuous downward spiral: exclusion promoting segregation, and segregation promoting exclusion. On the one hand, the legal market for affordable, accessible and habitable housing in these countries has proven incapable of meeting the needs of socially excluded families (UN-Habitat 2001b). For these families, informal and clandestine means of accessing and occupying urban land are often the only available alternative. Such exclusionary reality promotes the consolidation of highly segregated settlements, characterized by deprivation and non-realization of housing rights (UN-Habitat 2001b). On the other hand, segregation imposes difficulties in the daily life of disadvantaged families that perpetuate or worsen their condition of exclusion. For example, the lack of positive relations among different social groups increases prejudice and territorial stigmatization, keeps disadvantaged people away from participation at the societal level, and reduces their access to jobs and high-quality education (Bichir et al. 2004; Katzman and Retamoso 2006; Naiff and Naiff 2005; Torres 2004; Torres et al. 2005). In addition, poor segregated areas have been consistently associated with higher exposure to violence and diseases, bad accessibility that imposes time-consuming trips to work or school, and low quality of the built and natural environment (Hughes 2004; Katzman and Retamoso 2006; Sabatini et al. 2001; Torres et al. 2003).

In some developed countries, attempts to promote integration among different social groups are not new, being first recognized at the end of the nineteenth century. At this time, idealistic projects like the Bournville Village and the Garden Cities were proposed in the United Kingdom as solutions to the urban degradation observed in industrial cities. These projects aimed to accommodate all social classes in a more balanced manner, although still keeping segregation at the micro scale (Sarkissian 1976).

The claims of social mix emerged again during the post war period and beginning of the cold war in the 1940's, this time embedded in a discourse of national reconstruction and the development of universal state provision (Cole and Goodchild 2001; Sarkissian 1976). The response to that was the development of "new towns", especially in the United Kingdom and United States, which were planned in the context of the welfare state, when capitalist democracies needed to evaluate the social justice of their systems (Cole and Goodchild 2001; Sarkissian 1976). Such egalitarian vision has also influenced the creation of new towns in countries like Brazil. For example, the pilot plan of the capital Brasília, developed by Lúcio Costa in 1957, explicitly proposes residential blocks that "favor a certain degree of social coexistence, avoiding undue and undesirable class distinctions" (Costa 1991: 6). Later, however, most of Costa's original plan was modified, largely because of the growth of Brasília. Currently, the original area of the plan is merged with 20 satellite cities, which constitutes a metropolitan region with more than 3.5 million inhabitants (IBGE 2008). This region is well known for its high levels of income segregation and by the fact that only wealthy families can afford to live in the area of the pilot plan (Gouvêa 1995; Paviani 1996; Valladares 1999).

Under a different context, the contemporary interest in minimizing segregation has arisen as a response to many factors, including: (a) the development of new concepts such as underclass, social exclusion and social capital, which were often associated with studies describing the negative neighborhood effects of concentrating disadvantages (Cole and Goodchild 2001), (b) management difficulties and residualization in social housing, which was left for those who for reason of poverty, age or infirmity could not find suitable accommodation in the private sector (Cole and Goodchild 2001; Prike 1998), and (c) the emergence of protests from activists and journalists (Cole and Goodchild 2001). In several European and North American countries, traditional public housing strategies that had resulted in segregated and problematic areas were recognized as a mistake, and since then, housing and planning legislation have consistently emphasized the social mix at the neighborhood level (Allen et al. 2005; Cole and Goodchild 2001; Smith 2002). Different strategies have been followed to address this objective, including the regeneration of distressed areas, distribution of housing vouchers to move poor families out of neighborhoods with a high concentration of disadvantages, and regulations that required mixed occupancy as a condition for approving or funding new residential developments (Clampet-Lundquist 2004; Claydon and Smith 1997; Kleinhans 2004; Smith 2002).

In Brazil, the issue of segregation started to receive attention during the 1970's, in a period characterized by many critical discussions about the capitalist development in the country (Marques and Torres 2004). At this time, segregation was understood as the spatial materialization of inequalities produced by the labor market, which was driven by a peripheral and dependent type of capitalism (Bonduki and Rolnik 1979; Kowarick 1979; Maricato 1979b; Santos 1980). Within this framework, the studies were more focused on understanding processes that were considered as causal factors of segregation, and less on the phenomenon itself and its consequences (Bichir 2006).

The situation changed during the 1990's, when Brazilian debates started to address segregation as an issue of its own importance (Bichir 2006). Since then, an increasing number of studies has emphasized the negative consequences of segregation and the need for well-informed policies able to promote the spatial integration among different income groups (Torres 2004; Torres et al. 2006). Some progress in this direction can be observed and is worth mentioning. For example, the Brazilian Statute of the City, issued in 2001, recognizes a set of legal instruments that enable municipalities to promote a comprehensive regulation of clandestine settlements in public and private areas, and to restrain speculative retention of land that promotes excessive urban sprawl and forces poor families to live in distant peripheral areas (Rolnik and Saule Jr. 2001). The increasing presence of the state in poor outskirts of the city, improving access to infrastructure and other facilities (Torres et al. 2003), as well as some punctual investments focusing on the legalization and integration of slums into the legal urban fabric, like the Favela-Bairro project in Rio de Janeiro (Soares and Soares 2005), are also initiatives that can contribute to a decrease in segregation levels.

Despite these advances, there is still a wide gap between the scientific debates that advocate spatial integration of social groups and the policy practice. For example, Brazilian housing policies still rely on strategies that have been long condemned and avoided in developed countries, like the creation of large and homogeneous social housing settlements for the poor, located in cheap land at the outskirts of the city. By focusing exclusively on minimizing the housing deficit of urban areas, this type of policy displaces poor families to isolated areas, distant from the supply of equipments, services and opportunities, which very often turn into distressed neighborhoods (Luco and Rodríguez 2003; Preteceille and Ribeiro 1999; Rolnik 1997; Sabatini 2006; Torres 2004).

Designing and implementing policies that effectively minimize segregation and its negative effects is not an easy task. While Brazilian attempts in this direction are still very incipient, studies evaluating the experience of developed countries present several divergences concerning the impacts of social mix policies, even when they evaluate outcomes of the same policy strategy. Some studies identify many accomplishments of social mix policies (Feins and Shroder 2005; Popkin et al. 2004; Rosenbaum 1995; Rosenbaum and DeLuca 2000; Turbov and Piper 2005), while others focus on their failures and the need for restructuring them (Musterd and Andersson 2005; Musterd et al. 2003; Smets and den Uyl 2008; Uitermark 2003). These divergences indicate that there is no single formula for success: expected achievements are unlike to be met without well-informed policies that address the local particularities of mechanisms able to influence segregation dynamics.

Contextual mechanisms that contribute to urban segregation are many and vary from place to place (UN-Habitat 2001b). Brazilian literature has focused on at least four different and complementary mechanisms that can influence the behavior of social groups while selecting their residential location within the city: labor market, personal preferences, land and real estate markets, and state policies and investments. The first mechanism refers to the inequalities of the labor market and its socio-economic impacts, such as social exclusion, which have been considered by many as responsible for segregation and the precarious life conditions of poor families (Bonduki and Rolnik 1979; Kowarick 1979; Lago 2000; Maricato 1979b; Santos 1980).

Personal preferences are considered as a second mechanism, which is closely related to voluntary segregation. This is particularly relevant among affluent families, who are often seeking for status or want to protect themselves from problems associated with poverty. The fear of violence, in particular, is commonly used to justify the creation of gated communities, where safety is guaranteed by private security companies (Caldeira 2000; Pessoa de Souza e Silva 2007; UN-Habitat 2001b).

Land and real estate markets represent a third mechanism, and studies focusing on it stress how developers and their agents stimulate a competition for housing that reinforces the self-segregation of affluent groups and excludes poor families (Abramo

2001). Finally, the state is considered as a fourth mechanism, which permits segregation through its lack of action and promotes it through the unequal distribution of capital improvements, massive public housing projects, or regulatory devices such as exclusionary zoning (Rolnik 1997).

All these contextual mechanisms are clearly interdependent. Personal preferences, for example, are commonly affected by the real estate market, especially by entrepreneurs, who are constantly advertising new ideals of living and well-being (Caldeira 2000; Pessoa de Souza e Silva 2007). On the other hand, the real estate market is always adapting and reinventing itself in order to address the preferences of consumers (Pessoa de Souza e Silva 2007). The labor market, another mechanism influencing segregation, is directly related to the purchasing power of individuals and, therefore, is also continuously affecting personal preferences in general and the real estate market.

Improving the understanding about the relation between the aforementioned mechanisms and segregation is an essential step towards the development of social mix policies that are able to address clear goals. However, a crucial challenge for studies that seek a better comprehension of such relations relies on the fact that segregation displays many hallmark features of so-called *complex systems*. A complex system is "an entity, coherent in some recognizable way but whose elements, interactions, and dynamics generate structures and admit surprise and novelty that cannot be defined a priori" (Batty and Torrens 2005: 745). As a complex system, the dynamics of segregation are characterized by emergence, scale dependencies, interdependencies, and feedback loops. Urban segregation is a macro-scale phenomenon, but emerges from the residential choices of many individuals at the micro level (Schelling 1971). This emergent process results in a coherent form, with recognizable patterns, that adapts and organizes itself over time without any singular entity deliberately managing or controlling (Holland 1998).

The individual choices driving urban segregation dynamics are influenced by many contextual mechanisms which, as previously mentioned, are highly interdependent and constantly affecting each other. On the other hand, urban segregation is not only shaped and reshaped by the individual choices and the mechanisms influencing these choices, but is also able to influence them. In other words, there is a feedback loop between the emergent properties of segregation and the individual choices at the micro level. For example, not only do households often cluster in segregated neighborhoods, but they also recognize and react to emergent patterns of segregation: neighborhoods are named and can acquire reputations that further affect the residential choices of those living or considering living there (Gilbert 2004). The feedback loops between the different components involved in segregation dynamics introduce non-linearity into the system. As result, small differences in context or local behavior are able to produce large, unexpected, and sometimes counterintuitive outcomes that are not equivalent to the simple sum of the constituent parts (Holland 1998).

By facing the challenge of improving our understanding about segregation through the lens of complexity theory, it is likely that we will obtain a much more solid background for the development of well-informed policy strategies, which are able to properly address the phenomenon. Considering that, this study is motivated by the need for a scientific tool that is able to represent segregation as a complex system and to provide alternative scenarios that:

1. Improve the understanding about urban segregation and its relations with different contextual mechanisms, and
2. Support planning actions by offering insights about the adequacy of policy strategies.

The complex nature of segregation imposes difficulties regarding the use of traditional tools that are based on an aggregate static modeling approach, such as statistical modeling or classical optimization. Instead of focusing on the correlation between elements or relying on the idea of equilibrium, it is necessary to grasp segregation from the bottom-up, prioritizing the process rather than the product (Batty et al. 2006). By addressing the shortcomings of traditional techniques, agent-based modeling (ABM) has proven to be a promising approach for dealing with complex systems.

ABM focuses on individual decision-making units, called agents, which interact with each other and their environment (Gilbert 2008). These agents, which are autonomous and heterogeneous, are constantly acting according to a specific set of rules that can be changed through adaptation and learning (Gilbert 2008). By explicitly simulating interaction processes that occur at a micro level, ABM enables researchers to explore the emergence of macro structures from bottom-up in a very natural way (Gilbert and Troitzsch 1999; Miller and Page 2007).

Contrasting with traditional models and reflecting a movement towards relativism and post-modernism, agent-based models do not focus on making exact predictions (Batty 2009). Instead, they are mainly exploratory, more likely to be frameworks for assembling relevant information, more oriented towards understanding and structuring debates in processes of decision support that are much more consensual and participative (Batty 2009; Batty and Torrens 2005).

Thomas Schelling's model of racial segregation has been recognized as the first attempt at agent-based modeling in social sciences (Schelling 1971,1978). The model is based on a regular lattice representing the urban space on which agents, representing households, are placed at random. The agents belong to two different groups (e.g., white and black) and have a certain degree of tolerance in relation to the other group: they are satisfied with a mixed neighborhood, as long as the number of neighbors belonging to the same group is sufficiently high. What is revealing about this abstract model, and demonstrates its ability in representing emergent properties of segregation, is the counter-intuitive fact that extreme segregation patterns take place under a very mild preferential bias.

Schelling's work inspired many others, who developed variations of his model by using alternative utility functions (Bruch and Mare 2006; Clark 1991; Pancs and Vriend 2003), including individual preferences for housing or neighborhood quality (Fossett and Senft 2004), adopting different notions of neighborhoods (Fossett and Waren 2005; Laurie and Jaggi 2003; O'Sullivan et al. 2003), considering an additional hierarchical level (Omer 2005), adding game theory principles (Zhang 2004), and using vector-based representations (Crooks 2008).

Despite the existence of many agent-based models for segregation, only a few examples of models that rely on empirical data and methods can be found. Benenson and his colleagues, for example, developed an ethnical segregation model for the Yaffo area of Tel Aviv, which is occupied by Arab and Jewish residents (Benenson et al. 2002). Another example is the model of Bruch (2006), which explores the relationship between race and economic factors, and how both govern residential mobility to produce and maintain segregated neighborhoods in Los Angeles. No empirically-based model, however, has been developed to address the particularities of segregation in Brazilian cities. The research objectives of this study address this gap.

1.1 Research objectives

The goal of this study is to develop an operational agent-based simulation model of urban segregation in a spatially and temporally explicit manner, which is able to provide alternative scenarios that explore the impacts of different contextual mechanisms on the emergence of segregation patterns and support planning actions.

The specific objectives are:

1. To develop a conceptual and theoretical agent-based framework for modeling urban segregation dynamics;
2. To specify and estimate statistical models that depict the residential choice behavior of urban households (agents) and dynamics of the urban environment based on empirical data collected at São José dos Campos, a medium-sized city located in the State of São Paulo, Brazil;
3. To build an operational agent-based model for urban segregation by converting the specifications and parameters resulting from objectives (1) and (2) into a executable computer program;
4. To execute simulation experiments for testing the operational model's ability to accurately represent the real target system (validation) and to provide new insights about theories and policies on segregation.

1.2 Outline of the thesis

This thesis consists of seven chapters. After the introduction to the general problem and research objectives (Chapter 1), Chapter 2 defines the concept of urban segregation adopted in this work and describes its recent trends in Brazilian cities, impacts on the urban space and population, and different mechanisms that are able to promote and counter the phenomenon. Also included are segregation indices, which are useful tools for monitoring segregation patterns through time.

Chapter 3 redefines urban segregation under the mindset of complex systems science and introduces methods that are more appropriate to account for its complex nature. It presents conceptual and technical aspects of agent-based models (ABM), including a methodological protocol for developing ABM simulations.

Chapter 4 addresses the first specific objective. It introduces the conceptual principles and architecture of an agent-based framework named Multi-Agent Simulator of Urban Segregation (MASUS). Regarding the implementation level, the simulation protocol developed for the operational MASUS model is also presented.

Chapter 5 addresses the second specific objective. It provides empirical parameters that are used as inputs for the first operational MASUS model. The chapter begins with a brief description of the study site, which comprises the urban area of São José dos Campos, a medium-sized municipality located in the State of São Paulo, Brazil. Further, the chapter presents the empirical parameterization of the MASUS sub-model responsible for simulating the residential choice behavior of households (agents), and the empirical parameterization of MASUS sub-models that simulate dynamics of the urban environment, including urban sprawl, land value, and housing stock.

Chapter 6 addresses the third and forth specific objectives. It presents the operational MASUS model built from the specifications given in Chapter 4 and the empirical parameters provided in Chapter 5. In addition, simulation experiments that aim to validate the MASUS model and illustrate its potential for testing theories and policies on urban segregation are described.

Finally, Chapter 7 provides an evaluation of the study regarding the achievements of the objectives and recommendations about possible applications and further development of the MASUS model.

2 URBAN SEGREGATION: DEFINITIONS, TRENDS, AND MEASURES

2.1 Defining urban segregation

In general terms, the concept of urban segregation is related to the idea of distance or isolation among different social groups in an urban environment. The perception that such 'distance or isolation' can assume different meanings led White (1983) to distinguish two types of segregation: sociological and geographical. Sociological segregation regards the lack of interaction among population groups, while geographical segregation focuses on the spatial separation among the groups. These two types of segregation often present a high correlation: physical separation can promote social distance, and vice versa. However, this relationship is far from being universal. The caste system in India and the hacienda system in Latin America, for instance, are extreme cases that show the prevalence of strong social distances despite the spatial proximity of the different social groups (Rodríguez 2001; Sabatini et al. 2001).

Urban segregation has different meanings and effects depending on the specific form and structure of the cities, as well as their cultural and historical context. Its categories depend on the criteria adopted for classifying social groups, such as income, class, race, migratory origin, or ethnicity. In the United States, where segregation has received increasing attention since the beginning of the Civil Rights Movement in the 1950's, most studies focus on racial issues (Clark 1991; Duncan and Duncan 1955; Massey and Denton 1987,1993; Morgan 1983a; Schelling 1972). In Latin America, however, most studies concentrate on socioeconomic segregation (Feitosa et al. 2007; Lago 2000; Marques and Torres 2004; Ribeiro 2001; Rodríguez 2001; Sabatini and Salcedo 2007; Torres 2004; Torres et al. 2002; Villaça 1998). This interest emerges because social inequality, of income or social classes, is considered one the most outstanding features of Latin American countries, even more than poverty (Sabatini 2006).

Following the Latin American studies, this study adopts a concept of urban segregation that is explicitly spatial and regards the distances among the residences of families belonging to different income groups: the income residential segregation. An important advantage of this approach is the possibility of developing and using analytical indicators that measure segregation (see section 2.5) and allow comparisons between different periods and regions (Torres 2004).

2.1.1 Dimensions of segregation

There is a consensus among researchers that urban segregation is a multidimensional phenomenon, whose depiction demands measuring each dimension (Massey and Denton 1988; Reardon and O´Sullivan 2004; Sabatini 2006). Different dimensions of segregation produce distinct impacts on the development of urban communities and landscapes and, therefore, have different implications for public policies (Sabatini 2006). The classical paper The Dimensions of Residential Segregation, written by Massey and Denton and published in 1988, was the first to present a compound definition for segregation. Massey and Denton pointed out five dimensions of segregation: evenness, exposure, clustering, centralization, and concentration (Table 2.1). According to them, evenness and exposure are non-spatial dimensions of segregation. On the other hand, clustering, centralization, and concentration are spatial dimensions, since they need information about location, shape, or size of areal units.

Table 2.1 Dimensions of segregation according to Massey and Denton (1988).

Dimension	*Definition*
Evenness	Differential distribution of social groups in an urban environment.
Exposure	Potential contact among different social groups in an urban environment.
Clustering	Degree to which members of a certain group live disproportionately in contiguous areas.
Centralization	Degree to which a social group is near the center of an urban area.
Concentration	Relative amount of physical space occupied by a social group in an urban environment.

By arguing that segregation has no non-spatial dimension, Reardon and O'Sullivan (2004) reviewed Massey and Denton's work. According to these authors, the difference between the non-spatial dimension evenness and the spatial dimension clustering is simply an effect of data aggregation at different scales. The evenness degree at a certain scale of aggregation (e.g., census tracts) is related to the clustering degree at a lower level of aggregation (e.g., blocks). Reardon and O'Sullivan combined both concepts into the spatial evenness/clustering dimension, which refers to the balance of the population groups' distribution. Centralization and concentration were considered subcategories of the spatial evenness/clustering dimension. The authors also conceptualized the dimension exposure as explicitly spatial. They proposed the spatial exposure/isolation dimension, which refers to the chance of having members from different groups (or the same group, if we consider isolation) living side by side (Figure 2.1).

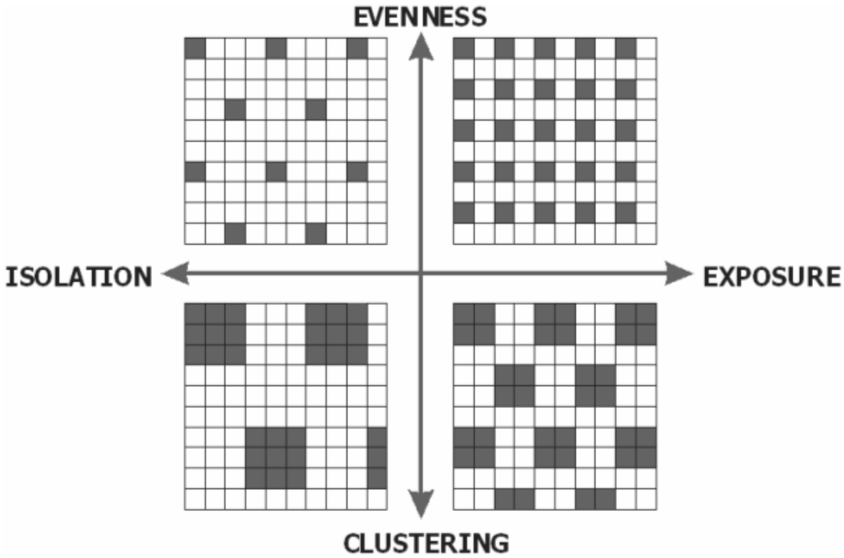


Figure 2.1 Spatial dimensions of segregation according to Reardon and O'Sullivan (2004).

This research adopts the segregation dimensions proposed by Reardon and O'Sullivan and monitors segregation by computing measures that are able to depict each spatial dimension (see section 2.5). These spatial dimensions are similar to the objective dimensions of segregation advocated by Sabatini (2006). The first objective dimension of segregation defined by Sabatini, named 'spatial concentration', is similar to the dimension spatial evenness/clustering, while the second objective dimension, called 'social homogeneity', is analogous to the dimension spatial exposure/isolation. Sabatini asserts that spatial concentration represents the first stage of segregation, and its impacts are usually less harmful than those resulting from social homogeneity, which is the second stage of segregation.

2.2 Patterns of urban segregation

The most influential approach for explaining patterns of segregation relies on the human ecology tradition associated with the Chicago School (Burgess 1924; Harris and Ullman 1945; Hoyt 1939). The Chicago School refers to a set of urban studies that emerged in Chicago during the first half of the 20th century. They became famous for their systematic and formal approach, focused on the city as a social laboratory. The efforts to understand the spatial organization of human activities yielded classical urban models that translate distinct patterns of residential segregation. Following these classical models, cities are developed through a competition for space that produces concentric zones (Burgess 1924), specific sectors (Hoyt 1939), or multiple nuclei (Harris and Ullman 1945) that accommodate households with different resources.

The concentric model, proposed by Burgess (1924), states that a city grows outward from a central point in a series of rings. Burgess observed that there was a correlation between the distance from this central point (the CBD) and the wealth of residential areas. Since Burgess's studies relied on the study of Chicago, he observed that wealthier families tended to live further away from the CBD. Later interpretations of the Burgess model pointed out an inverse correlation between the CBD and the wealth of neighborhoods. This 'center-periphery' pattern can be observed in some Latin-American cities (section 2.2.1), where wealthier families tend to concentrate in central areas, while poorer families occupy the outskirts of the city.

Hoyt (1939) proposed a second model, known as sector model, which advocates the idea that a city develops in sectors instead of rings. According to him, if a district is set up for high-income residences, any new development in that district will expand from the outer edge and, therefore, the sector shape emerges. A third model, known as multiple nuclei model, was proposed by the geographers Chauncy Harris and Edward Ullman (Harris and Ullman 1945), who wanted to demonstrate that not all cities fit into the concentric and sector model. They argued that the activities of many cities revolve around many nuclei rather than around a single CBD. According to them, the location of different land uses within a city, which includes the residential areas for distinct socioeconomic groups, cannot always be predicted. Historical, cultural, and socio-economic values will have differing impacts on cities, and the exact location of an economic or ethnic nucleus cannot be determined for all cities. The formation of these nuclei depends on a variety of factors - topographical, historical, cultural, racial, economic and political - that do not result in the same combination for each urban area (Harris and Ullman 1945).

Considering that this thesis develops an agent-based model for segregation and illustrate its potential through a case study in a Brazilian city, the next paragraphs present an overview of segregation patterns that have been observed in this country. This overview indicates some similarities between the Brazilian patterns and the classical models of the Chicago School.

2.2.1 Segregation in Brazilian cities

Brazilian studies have analyzed urban segregation since the 1970's. The studies developed during the 1970's and 1980's qualified the Brazilian segregation pattern under a 'dual' perspective (Lago 2000), i.e., characterized by a strong contrast between the wealthy center and the poor outskirts (Bonduki and Rolnik 1979; Caldeira 2000; Kowarick 1979; Lago 1998). This pattern, known as 'center-periphery', resulted from an urban growth stimulated by the influx of migrants, mostly from rural areas, seeking for employment opportunities. Its spatial arrangement, which is analogous to the concentric model proposed by Burgess (see section 2.2), keeps families belonging to different social classes far from one another. While affluent families occupy central and well-equipped neighborhoods (Figure 2.2), they are also able to influence public investments and regulations that displace the poorest families to further areas known as *periferias* (peripheries) and make the city's underdevelopment less visible (Caldeira 2000).

Periferias are socially homogeneous settlements located in the outer fringes of the city (Figure 2.2). These settlements are typically clandestine, created and sold by private developers who conducted land subdivisions without any formal review or approval by the appropriate county agencies. Due to the lack of affordable housing offers in the 'legal city', the land ownership in these settlements and the self-construction of houses became the only alternative for many poor families (Bonduki and Rolnik 1979; Maricato 1979a; Santos 1980). These families are excluded from the advantage of living in neighborhoods with basic infrastructure, facilities and urban services (Kowarick 1979; Torres et al. 2002). In particular, their accessibility to jobs is limited, since workers usually have to face long commuting trips to and from work (Caldeira 2000; Santos 1980).

(a)

(b)

Figure 2.2 (a) Wealthy center: São Paulo’s downtown (Fernandes 2005), and (b) Poor *periferia*: Jardim Ângela, a district of São Paulo (McHugh 2008).

This center-periphery pattern, however, has been overlapped by a new pattern of segregation, which arose due to political and socioeconomic changes that occurred in the 1980's (Caldeira 2000; Lago 2000; Torres et al. 2002). During the period 1981-1989, the Brazilian gross domestic product (GDP) increased at an average annual rate of only 1.6%, and the per capita income declined by 8.3% (Bresser Pereira 1992). The 1980's are known as the 'lost decade' because of the stagnation, hyperinflation, and the increase in the external debt during the period. While the economic crisis led to an impoverishment of the population and an increase in social inequalities, the growth of peripheral irregular settlements occurred at a slower pace. This fact is partially explained by the establishment of the Federal Law for Urban Land Parceling (6766/79). This law regulates the minimal requirements for approval and development of urban settlements and introduced penalties for land developers who ignore these. It also propitiated a more active presence of the state in the outskirts, improving access to infrastructure and public facilities. This expansion of the legalized city promoted a larger social diversity in areas that were only occupied by the underclass (Caldeira 2000; Lago 2000).

The impoverished population that remained unable to afford a dwelling in the 'legal city' or even to build their own house at an irregular settlement also contributed to the attenuation of the spatial duality of the urban space. This population group promoted the proliferation of *favelas*, the Brazilian equivalent of shantytowns. Unlike informal settlements, *favelas* are the product of some form of land invasion and their residents do not hold any land ownership. A particular characteristic of *favelas* is that they can emerge in different regions of the city, including those closer to wealthy neighborhoods (Torres et al. 2002). This characteristic challenges the social homogeneity of the center-periphery pattern, since it diffuses poverty through many parts of the city (Figure 2.3).

Finally, the emergence of wealthy and gated urban developments also promoted smaller geographical distances among different social classes. The spread of gated neighborhoods introduced residential alternatives for the high- and middle-income groups outside the traditional areas where these social groups concentrate (Caldeira 2000). Therefore, the separation among population groups, which had been guaranteed by absolute distances, began to be maintained by other types of obstacles, such as protection walls, which are able to reinforce exclusion and preserve homogeneous areas.

Figure 2.3 Favela Paraisópolis beside a fortified enclave in Morumbi, São Paulo (Vieira 2005).

Based on this new reality, Caldeira (2000) introduced the idea of *fortified enclaves*. Fortified enclaves are spaces for the middle and upper classes that are typically isolated from surrounding neighborhoods by physical barriers and other surveillance resources, such as guards, warning signs, and high-tech alarms (Figure 2.3). Such developments impose challenges for society through their "privatization of public space, conflict with planning norms, and interference with the integrated planning of the cities in which they are built" (Pessoa de Souza e Silva 2007: 557). Despite these negative issues related to fortified enclaves, Sabatini and his colleagues (2001; 2007) assert that these developments can also bring high-quality services and commerce to the poor areas where they are located. According to them (Sabatini et al., 2001: 9), "poor groups that end up near these projects benefit not only in objective terms (employment, services, urban facilities), but in subjective terms as well (like the sense of belonging to a place that is prospering)". These benefits relate to a decrease in the scale of segregation[2].

[2] Here, the term 'scale' refers to the level of detail in the analysis, and not to its cartographic meaning.

From another perspective, Villaça (1998) stresses a tendency related to large-scale patterns of segregation. According to him, despite the spread of gated neighborhoods and *favelas*, which establish smaller distances among different social groups, it is important to observe the macrosegregation of city. Macrosegregation is the “process in which different social classes tend to concentrate in different general regions or groups of neighborhoods of the metropolis” (Villaça 1998: 142). Villaça observed that the self-segregation of middle and high classes has increased and usually follows a certain direction of territorial expansion starting from the city’s center. This trend resembles the classical sector model proposed by Hoyt (1939), since it creates a cone-shaped wealthy axis that concentrates most high-income families. This axis, however, is not necessarily homogeneous. In fact, it is commonly characterized by a degree of social diversity, including the presence of some low-income families (Sabatini 2006). Even so, for the wealthy residents of this area, the need for circulating through other parts of the city and the possibilities of confronting other realities are reduced (Villaça 1998).

Besides the large-scale segregation patterns promoted by wealthy families, the cities keep attracting poor families that locate in large peripheral settlements. Therefore, despite the more active presence of the state in these areas, the city still decays, socially and physically, towards its outskirts, except in the ‘wealthy cone’ area (Sabatini 2006).

By comparing the traditional center-periphery pattern and the recent trends of segregation (see Figure 2.4), it can be seen that segregation in Brazilian cities has become more complex and is ruled by antagonistic forces that deal with different scales. This has operational consequences and indicates the importance of considering the issue of spatial scale when studying segregation. For example, due to the social diversity of high-income neighborhoods (wealthy axis), an analysis based on smaller scales would lead to the conclusion that these places are less segregated, when, in fact, they can be highly segregated at larger scales. On the other hand, the presence of a wealthy gated community in a poor region of the city decreases the large-scale segregation of the area, even though gated communities are very homogeneous and present a high degree of segregation at smaller scales.

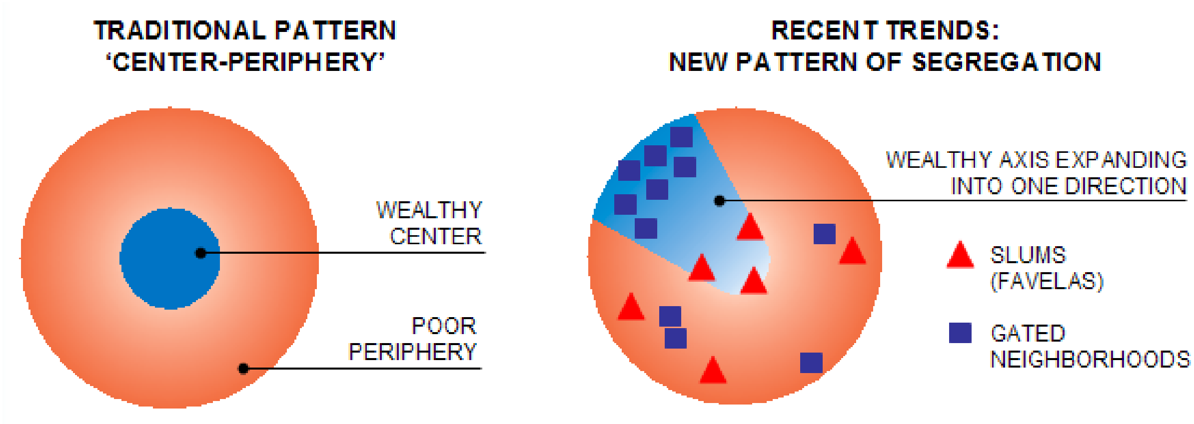


Figure 2.4 Patterns of segregation in Brazilian cities.

2.3 Impacts of segregation

While studying the impacts of segregation on community development, it is important to recognize that segregation is not a problem, but a phenomenon that can produce distinct outcomes depending on specific contexts (Sabatini 2006). Nevertheless, the acute spatial concentration of disadvantages, such as poverty, has consistently led to several negative consequences for the life of urban inhabitants and the ability of cities to contribute to social and economic development (Katzman and Retamoso 2006; Préteceille 2003; Rodríguez 2001; Sabatini et al. 2001; Torres et al. 2003). For this reason, the issue of segregation has received increasing attention in policy and academic debates of many developing countries.

Considering the reality of Brazilian cities, Torres et al. (2006) assert that segregation is not a mere 'sociological curiosity', but is associated with important repercussions for the economic and social opportunities of individuals and families living in the most segregated areas. Reinforcing this point, many authors have pointed out features of Brazilian segregation dynamics that contribute to increase and/or perpetuate poverty (Hughes 2004; Marques and Torres 2004; Ribeiro and Santos Junior 2003; Torres 2004; Torres and Marques 2001; Torres et al. 2006).

Because low-income families can only afford to live in depreciated areas of the city, a common a priori characteristic of Brazilian neighborhoods with a high concentration of poverty is the poor quality of built and natural environment and the higher exposure to natural disasters and diseases (Torres 2004). Moreover, the irregular status of dwellings located in segregated neighborhoods and/or the lack of a political voice of their inhabitants often restrain the access to public policies and investments that could contribute to the improvement of these areas (Torres 2004; Torres et al. 2006). As a result, segregation affects the access by poor families to schools, health services, and public utilities in general (Rodríguez 2001; Sabatini 2006; Torres 2004). Regarding the reality of São Paulo, Torres and Marques (2001) conducted spatial and quantitative analyses that empirically showed how extremely segregated areas, which they call *hiperperiferias*, overlap the worst socio-economic indicators with flooding events and land sliding risks, heavily polluted environment, and inefficient social services.

For the poor families living in segregated peripheries, accessibility-related problems are also a daily reality, e.g., longer commuting distances to work and school. Also, unlike in middle- and upper-class neighborhoods, the concentration of low-income consumers is not likely to sustain strong local business and services that could contribute to the creation of local employment opportunities and decrease the need of time-consuming trips within the city.

Besides impacts concerning territorial and accessibility issues, the lack of positive relations among different social groups can increase prejudice and neighborhood stigmatization, keep disadvantaged people away from participation at a societal level, and reduce their opportunities for jobs and skill upgrading (Atkinson 2005; Briggs 2005; Katzman and Retamoso 2006; Torres 2004). In Brazil, several studies have focused on the prejudice against inhabitants of segregated neighborhoods, especially *favelas*, and how segregation limits their prospects for upward mobility. Naiff and Naiff (2005) analyzed, by means of interviews, the perception of middle-class citizens towards *favela* residents in Rio de Janeiro. Their study revealed an increasing sense of denial, distrust and stigmatization against the *favela* residents, who are often seen as responsible for the high criminality rates in the city. Complementing these findings, Rocha and Araújo (2008) and Cecchetto and Monteiro (2006) present testimonies from young *favela* dwellers that describe how the location of their residences decreases their chances of getting a job, and report that providing false address information to potential employers is often a strategy adopted to avoid discrimination.

The spatial concentration of disadvantages can also promote problems that emerge from the absence of social capital (Cole and Goodchild 2001). Social capital is a set of informal values and norms that are shared among people and allow cooperation between them (Fukayama 1995). Contemporary scientific discourses commonly assert that the lack of social capital between different social groups, also known as ‘bridging social capital’ (Putnam 1995), hinders disadvantaged groups to acquire support networks that could assist their upward mobility.

Another effect attributed to the absence of bridging social capital is the lack of positive role models. For a disadvantaged family, interaction with people who are in steady employment and who give importance to education may result in the former acquiring a set of mainstream values from the latter. These values may raise new patterns of behavior, aspirations, and motivations that contribute, for instance, to better performance in school and attendance to colleges, or to improved motivation for finding work (Tunstall and Fenton 2006). Rosenbaum et al. (1998) assert that such interaction can also reduce crime rates, arguing that illegal behavior is less commonly disapproved of in areas of deep poverty concentration. Many negative impacts attributed to the absence of bridging social capital can be found in Brazilian cities, where the isolation of poverty has been consistently associated with lower performance in school, higher incidence of teenage pregnancy, as well as higher rates of unemployment and violence (Bichir et al. 2004; Hughes 2004; Torres et al. 2005). In São Paulo, for instance, the life expectation of the residents of Guainases, a highly segregated and violent neighborhood, has been reported as being 12 years lower than that of individuals living in wealthy neighborhoods (Hughes 2004).

When analyzing the impacts of segregation, it is also relevant to take the different dimensions and scales of segregation into consideration. Sabatini (2006) asserts that the spatial concentration of a social group (dimension evenness/clustering) may have a positive side. For example, it can help to preserve the cultural identities of an ethnic group, or promote social and political empowerment of the urban poor. The social homogeneity (dimension isolation/exposure), however, tends to promote problems like those mentioned above (Sabatini 2006). Such problems are accentuated when the isolation of the poor occurs in broader scales of segregation, e.g., in large and homogeneous peripheries (Rodríguez 2001; Sabatini et al. 2001; Sabatini et al. 2005).

Finally, it is important to mention that segregation concerns impacts that affect not only poor families, but also other inhabitants of the city. For example, segregation contributes to an increase in violence, which, in the case of Brazil and many other Latin American countries, promoted the development of a culture of fear and the self-segregation of wealthy families, who perceive the contact with poor individuals as increasingly threatening. This resulted in the spread of fortified enclaves for middle and upper classes, which fragment the city and promote the decline of its public spaces (Caldeira 2000; Pessoa de Souza e Silva 2007). By hindering the contact between social classes, fortified enclaves also become a key element of a spiral process where the increase in segregation fostered by these developments lead to higher rates of violence, which increase the culture of fear and, consequently, stimulate the further proliferation of fortified enclaves.

In summary, the severe segregation in Brazilian cities imposes innumerous negative impacts to the daily life of the urban population, contributes to the perpetuation of poverty, and impairs the cities' capacity to promote economic and social development. Therefore, reducing the current levels of urban segregation is critically important for the Brazilian society as a whole.

2.4 Promoting and countering urban segregation

The negative impacts ascribed to the concentration of deprivation are unlikely to be resolved without policies that effectively address the causes of segregation. It is impossible to assign the emergence of segregation to a single cause. Researchers have identified different and complementary mechanisms that influence how distinct social groups interact and occupy urban spaces. Nevertheless, it is important to keep in mind that this is not a unidirectional process. Instead, it is characterized by constant feedback loops, where the so-called causal mechanisms of segregation can also be affected by segregation in the long term.

Considering existing studies, it is possible to identify approaches focusing on four different sets of causal mechanisms of segregation: personal preferences, labor market, land and real estate markets, and the controlling power of the State[3]. The first approach concentrates on personal preferences: social segregation can increase because people prefer to live among neighbors similar to themselves. This voluntary segregation

[3] The last three factors (labor market, land market, and controlling power of the state) are mentioned in Torres et al. (2003).

can be considered as comprehensible instead of socially condemnable. It often results from the families' attempt to reinforce their social identities through shared values and to improve their quality of life (Marcuse 2005; Sabatini 2006). This social practice is particularly common among advantaged families, who usually prefer to live in areas of concentrated wealth and keep themselves apart from urban problems related to poverty (Caldeira 2000; Pessoa de Souza e Silva 2007; UN-Habitat 2001b). Studies on segregation modeling have a strong tradition of considering personal preferences to understand the emergence of the phenomenon (Sakoda 1971; Schelling 1971).

The second approach considers the inequalities of the labor market and its socio-economic impacts as being responsible for segregation and the precarious life conditions of part of the urban population (Katzman and Retamoso 2006; Kowarick 1979; Lago 2000; Morris 1995; Ribeiro 2001; Turok and Edge 1999; Webster 1999). Jargowsky (1997), for instance, asserts that the growth of the US economy brought positive impacts in areas of poverty concentration. Nevertheless, in a Latin America context, Sabatini (2006) advocates that the population impoverishment due to economic crises may promote a backward progression in the segregation process, and mention the case of São Paulo during the 1980's as an example.

The third approach focuses on the dynamics of land and real estate markets. It stresses how real estate agents stimulate a competition for land and housing that reinforces the self-segregation of higher income groups and the exclusion of disadvantaged families (Abramo 2001). In Brazil, the speculative nature of urban land markets tends to increase segregation, e.g., when neighborhoods begin to attract wealthy residents and owners decide to raise land prices based on the expected land use for this area. In general, land valuation seems to be an important motivation behind the voluntary segregation of affluent families. It is interesting to notice, however, that the relation between land value and segregation is self-reinforcing: the increase in land prices promoted by the voluntary segregation is a factor that limits the access of poor families to serviced land, which consequently contributes to the overall segregation of the city and to further gaps between land prices of different neighborhoods (Sabatini 2000,2006).

From another land-market perspective, private settlers who conduct illegal land subdivisions in cheap areas located in the outskirts of the city, the so-called *periferias*, also influence segregation by increasing poverty concentration (Smolka 2005). Nevertheless, the maximization of the profits of real estate agents is not always associated with an increase in segregation. An example is the case of many high-income neighborhoods that have been densified through high-rise constructions for families with lower income. These projects increase the profits of real estate investors and, at the same time, contribute to a social diversification of wealthy neighborhoods. Another example is the spread of gated neighborhoods for upper classes in areas occupied by the poor, which does not necessarily decrease segregation, but contributes to a reduction in its scale (Sabatini 2006).

The state can play an active role in mitigating segregation impacts related to the labor market and to the land and real estate market. Nevertheless, its ability to influence people's personal preferences is much more limited and unnecessary, since voluntary segregation is not essentially negative (Sabatini 2006). The approach that focuses on the labor market to explain the emergence of segregation calls for structural macroeconomic policies, such as fiscal and monetary policies, as well as investments in public education and health care. Regarding the land and real estate market, the state can settle initiatives to regulate its dynamics, like for example, policies to diversify land uses and promote developments for upper classes in areas occupied by disadvantaged families. In addition, the state can control land speculation and regularize illegal settlements.

Measures to diversify land uses and promote developments for upper classes in poor neighborhoods represent an effort to regulate the market towards a decrease in the scale of segregation. This stimulus can occur through public investments in infrastructure, changes in the norms of land use, tax exemption measures, and concessions (Sabatini 2006). Such initiatives are more effective if complemented by policies that contain land speculation by capturing capital gains and controlling urban sprawl (Sabatini 2006). The Brazilian Statute of the City (Rolnik and Saule Jr. 2001) issued in 2001 offers a set of instruments that can help local policy makers in this direction. For instance, to restrain the speculative retention of land, the statute establishes that vacant or underutilized lands located in areas with good infrastructure are subject to taxes that are progressive over time. These lands are also subject to compulsory building and subdivision, according to the local master plan (Rolnik and

Saule Jr. 2001). These instruments control excessive urban sprawl, which promotes the large-scale segregation of the poor and increases the need for investments to expand infrastructure networks (Rolnik and Saule Jr. 2001). The statute also recognizes legal instruments that enable municipalities to promote a comprehensive regularization of illegal settlements in private and public areas. These instruments include the regulation of the constitutional rights to usucaption (adverse possession) and the concession of the real right to use (a sort of leaseholding) (Fernandes 2006,2007). Combined with land speculation control measures, these initiatives can contribute to democratize the conditions of access to urban land and housing (Fernandes 2006,2007; Rolnik and Saule Jr. 2001). However, some cases of irregularity demand the removal of poor families to more adequate areas, either to protect them from natural disasters or to guarantee environmental standards (Sabatini 2006). The Brazilian Provisional, introduced in 2001, settles conditions for the municipal authorities to conduct this sort of action (Fernandes 2006).

This discussion demonstrates the importance of governmental institutions in regulating mechanisms that promote segregation. Governmental *laissez-faire* approaches that ignore such mechanisms are in fact contributing to the perpetuation of urban segregation. In addition, governmental regulations or interventions can also aggravate the problem. For these reasons, some researchers indicate the controlling power of the state as another cause of segregation. According to this approach, the state can intensify segregation through its permissiveness, urban legislation, or investments (Rolnik 1997). For example, the widespread practice of exclusionary zoning to separate different activities and groups has played a key role in excluding disadvantaged families from privileged areas of the city (Ihlanfeldt 2004). Zoning codes define standards of land occupation that often rely less on technical aspects and more on the practices and logic of market investments. An example is the requirement for minimum lot sizes, which cannot be afforded by poorer families and exclude them from certain neighborhoods (Rolnik 1997).

Other state interventions that promote segregation concern the unequal distribution of urban investment (Marques and Bichir 2002; Préteceille 2003; Smolka 1992; Sugai 2002). For instance, punctual investments that increase the land value of a neighborhood can drive low-income families away from this area. Policies aimed at controlling segregation should consider democratizing the distribution of investments, including the access to infrastructure and urban facilities (Torres 2004). The Brazilian Statute of the City recognizes several mechanisms to ensure the democratic participation of citizens and other stakeholders in planning and managing the city. These mechanisms include: participatory budget practices, public hearings, consultations, creation of councils, environmental and neighborhood impact studies, and popular initiatives for the proposal of urban laws (Fernandes 2007). These measures help to undermine the public investments biased toward wealthy areas (Sabatini 2006), and some of them have already been carried out in Brazilian cities, e.g., participatory budget practices in Porto Alegre and Belo Horizonte (Wood and Murray 2007).

Social housing projects focusing on maximizing dwelling offers are another state intervention that can promote segregation. These projects are usually homogeneous settlements located on cheap land at the outskirts of the city, far from the supply of equipments, services and opportunities (Luco and Rodríguez 2003; Sabatini 2006; Smith 2002; Torres 2004; van Kempen 1994). Such social housing projects, which are very common in Brazilian cities, reinforce the trend to displace poor families from the best locations, increase the scale of segregation, and therefore worsen its negative effects.

In the United States and some European countries, these traditional public housing strategies that had resulted in large areas of poverty concentration were recognized as a mistake. Therefore, minimizing urban segregation - or at least its scale - became a target explicitly expressed in many policy debates (Cole and Goodchild 2001; Smith 2002). To integrate different social groups, three strategies are currently the most intensely followed in these developed countries: dispersal of poverty, regeneration of troubled neighborhoods, and regulation for new developments.

Strategies for promoting integration through the spatial dispersion of poverty focus on moving low-income households out of distressed areas into middle-class neighborhoods. Some housing programs in the Unites States adopt this strategy, like the Moving to Opportunity and the HOPE VI (Housing Opportunities for People Everywhere). The program Moving to Opportunity gives housing vouchers to low-income families for renting private dwellings in neighborhoods with a poverty rate of less than 10% (Smith 2002). The program HOPE VI adopts additional strategies for dispersing poverty, such as replacing distressed and high-density public housing with fewer affordable residential units in middle-class neighborhoods (Popkin et al. 2004). The high costs of these initiatives, however, represent an obstacle to its adoption in developing countries. Besides, the strategy is more appropriate for cities where the poor are a minority (Sabatini 2006), which is not the case in developing countries. Considering a Latin American context, Sabatini (2006) asserts that dispersing wealthy families seems a more effective way to promote positive changes in segregation patterns.

The second strategy commonly adopted in developed countries focuses on regenerating problematic public housing. This implies measures to improve local services and social programs, oppose delinquencies and territorial stigmas, demolish high-density constructions, build high-quality houses, and encourage middle-class households to move into these areas. This strategy has been also adopted in developing countries: a good example is the Favela-Bairro project in Rio de Janeiro, which integrates existing *favelas* into the fabric of the city by upgrading their infrastructure and services (Soares and Soares 2005).

The third strategy involves regulating new developments by requiring mixed occupancy as a condition for approval or funding. This requirement is often expressed as percentages of affordable land or built area within the new residential developments (Sabatini 2006). The Section 106 of the UK's Town and Country Planning Act 1990, for instance, allows local authorities to negotiate with developers for some affordable units in new developments in exchange for planning permission (Claydon and Smith 1997).

There are several divergences about the impact of policies aimed at minimizing segregation. Some studies conducted in developed countries identify many accomplishments and characterize these policies as successful (Feins and Shroder 2005; Turbov and Piper 2005). On the other hand, other studies focus on the failure of these policies and the need for restructuring them (Clampet-Lundquist 2004; Silverman et al. 2005; Smets and den Uyl 2008). Such divergences reinforce the importance of constantly monitoring and adjusting policies in order to get the expected results. Most importantly, the design of these policies must consider the particularities of cities, which differ in segregation patterns, population composition, levels of deprivation, culture, structure of housing markets, and many other features that demand specific approaches.

2.5 Measuring urban segregation

Given the increasing importance of urban segregation in policy and scientific debates, several researchers have proposed indices to measure the different dimensions of the phenomenon (Bell 1954; Duncan and Duncan 1955; Feitosa et al. 2007; Jargowsky 1996; Morgan 1975; Reardon and O'Sullivan 2004; Sakoda 1981; Wong 1993). The first generation of segregation indices was proposed during the 1950's in the United States and focused on measuring segregation between two population groups (black and white). The dissimilarity index *D* (Duncan and Duncan 1955) and the exposure/isolation index (Bell 1954) are the most distinguished measures of this period.

In the 1970's, segregation studies started to focus on multigroup issues, including the segregation among social classes or among White, Blacks and Hispanics. To meet these needs, a second generation of segregation indices was developed by generalizing versions of existing two-group measures (Jargowsky 1996; Morgan 1975; Reardon and Firebaugh 2002; Sakoda 1981). However, these measures are insensitive to the spatial arrangement of population among areal units. This shortcoming leads to what White (1983) identified as the *checkerboard problem*. Given two checkerboards, the first with an alternation of black-and-white squares, and the second with all the black squares located on one side of the board, the results of non-spatial segregation indices indicate a maximum segregation degree for both arrangements. The non-spatial indices are not able to show the second arrangement as more segregated than the first (Figure 2.5).

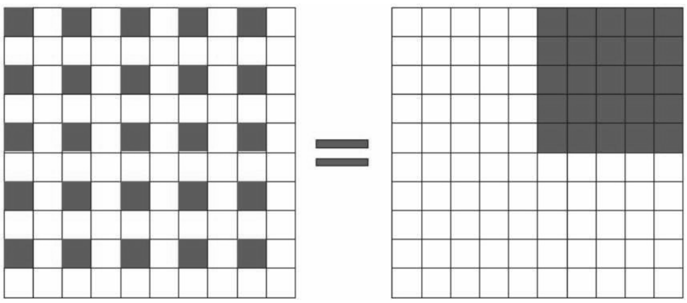


Figure 2.5 The checkerboard problem (White 1983).

To overcome the checkerboard problem, several researchers proposed spatial measures of segregation (Feitosa et al. 2007; Jakubs 1981; Morgan 1983b; Morrill

1991; Reardon and O'Sullivan 2004; White 1983; Wong 1993,1998). This study adopts two spatial indices proposed by Feitosa et al. (2007) to measure different dimensions and scales of segregation. The first, named generalized spatial dissimilarity index, captures the dimension evenness/clustering (section 2.11). The second, called spatial isolation index, captures the dimension exposure/isolation (section 2.11). Global and local versions of these measures are used in a complementary manner to depict segregation patterns. While global indices summarize the segregation degree of the entire city, local indices show segregation as a spatially variant phenomenon that can be displayed in maps.

The indices adopted in this study rely on the idea that an urban area comprises different localities, which are places where people live and exchange experiences with their neighbors. The intensity of these exchanges varies according to the distance among population groups, given a suitable definition of distance. The population characteristics of a locality are expressed by its *local population intensity*, which is calculated by using a kernel estimator. A kernel estimator is a function that estimates the intensity of an attribute in different points of the study area (Silverman 1986).

To calculate the local population intensity of a locality *j*, a kernel estimator is placed on the centroid of areal unit *j* and estimates a weighted average of population data. The weights are given by the choice of a distance decay function (e.g., Gaussian) and a bandwidth parameter (Figure 2.6). This procedure allows researchers to specify functions that formalize a hypothesis about how population groups interact across spatial features. The specification of different bandwidths, for instance, enables analyses in multiple scales: the indices are able to start from the most detailed data and generalize them for analyzing segregation in broader scales.

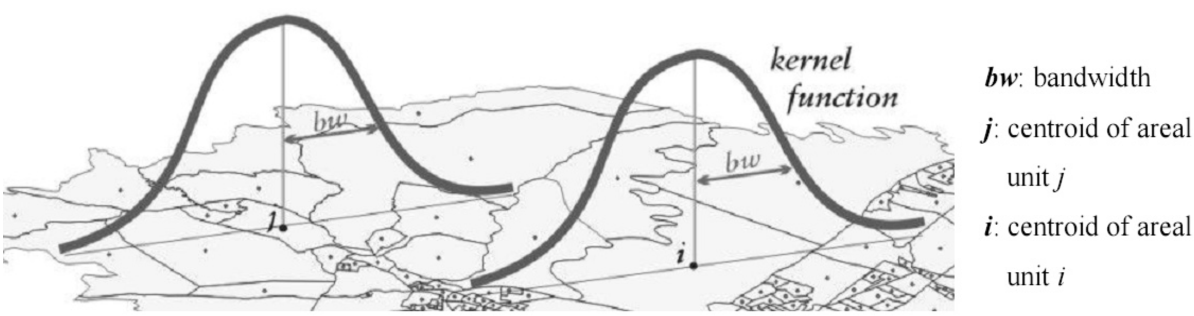


Figure 2.6 Gaussian kernel estimator (Feitosa et al. 2007).

The local population intensity is a geographically weighted population average that considers the distance between groups. Formally, the local population intensity of a locality j ($\breve{L}_j$) is calculated as (Feitosa et al. 2007):

$$\breve{L}_j = \sum_{j=1}^{J} k\left(N_j\right), \tag{2.1}$$

Where: N_j is the total population in areal unit j; J is the total number of areal units in the study area; and k is the kernel estimator which estimates the influence of each areal unit on the locality j.

The local population intensity of group m in the locality j ($\breve{L}_{jm}$) is calculated by replacing the total population in areal unit j (N_j) with the population of group m in areal unit j (N_{jm}) in equation (2.1):

$$\breve{L}_{jm} = \sum_{j=1}^{J} k\left(N_{jm}\right). \tag{2.2}$$

2.5.1 Measuring the spatial dimension evenness/clustering

The global version of the *generalized spatial dissimilarity index* ($\breve{D}(m)$) measures the average difference of the population composition of the *localities* from the population composition of the city as a whole. The formula of $\breve{D}(m)$ is:

$$\breve{D}(m) = \sum_{j=1}^{J} \sum_{m=1}^{M} \frac{N_j \left|\breve{\tau}_{jm} - \tau_m\right|}{2N\tau_m\left(1-\tau_m\right)}, \tag{2.3}$$

Where:

$$\breve{\tau}_{jm} = \frac{\breve{L}_{jm}}{\breve{L}_j}. \tag{2.4}$$

In equation (2.3), N is the total population of the city; N_j is the total population in areal unit j; τ_m is the proportion of group m in the city; $\breve{\tau}_{jm}$ is the local proportion of group *m* in locality *j*; *J* is the total number of areal units in the study area; and *M* is the total number of population groups. In equation (2.4), $\breve{L}_{jm}$ is the local population intensity of group *m* in locality *j*; and $\breve{L}_j$ is the local population intensity of locality *j*.

The index $\breve{D}(m)$ measures the proportion of people who would have to move from their localities to achieve an even population distribution. It varies from 0 to 1, where 0 stands for the minimum degree of evenness and 1 for the maximum degree. Despite these established meanings, it is still hard to interpret the values obtained within this [0,1] interval: does a $\breve{D}(m)$ value equal to 0.6 reveal a situation of severe segregation or not? This is not a trivial question, since the values of segregation measures are sensitive to the scale of the data: indices computed for smaller areal units tend to present higher values than indices computed for larger areal units (Feitosa et al. 2007). This is called the grid problem (White 1983) and it is inherent to all segregation measures.

In the case of spatial segregation measures, as the ones presented in this section, the scale variability is also related to the bandwidth used in the computation of the measures. An index computed with a small bandwidth will have higher values than another that is computed with a large bandwidth. Because of that, it is unfeasible to establish fixed thresholds that assert whether the index results indicate a severe segregation level or not. Instead, the interpretation of global indices of segregation is more useful when relational, for example, focused on the comparison of values obtained for an urban area in different points in time. Based on that, it is possible to draw conclusions about segregation trends along the years.

The local version of the *generalized spatial dissimilarity index* ($\breve{d}_{j(m)}$) is obtained by decomposing the index $\breve{D}(m)$. It shows how much each locality contributes to the global $\breve{D}(m)$ measure of the city (Feitosa et al. 2007). The local index $\breve{d}_{j(m)}$ can be displayed as a map and used to identify critical areas. The formula of $\breve{d}_{j(m)}$ is:

$$\breve{d}_j(m) = \sum_{m=1}^{M} \frac{N_j \left| \breve{\tau}_{jm} - \tau_m \right|}{2N\tau_m(1-\tau_m)} \, , \qquad (2.5)$$

Where: the equation parameters are the same as in equation (2.3).

2.5.2 Measuring the spatial dimension exposure/isolation

The global version of the *spatial isolation index* ($\breve{Q}_m$) measures the average proportion of group *m* in the *localities* of each member of the same group (Feitosa et al. 2007):

$$\breve{Q}_m = \sum_{j=1}^{J} \frac{N_{jm}}{N_m} \left(\frac{\breve{L}_{jm}}{\breve{L}_j} \right), \tag{2.6}$$

Where: N_{jm} is the population of group *m* in areal unit *j*; N_m is the population of group *m* in the study region, $\breve{L}_{jm}$ is the local population intensity of group *m* in locality *j*, and $\breve{L}_j$ is the local population intensity of locality *j*.

The isolation index varies from 0 (minimum isolation) to 1 (maximum isolation). The results of the index $\breve{Q}_m$ depend on the overall composition of the city. For example, if the proportion of the group *m* increases in the city, the index $\breve{Q}_m$ tends to become higher.

The local version of the spatial isolation index ($\breve{q}_m$) can also be obtained by decomposing $\breve{Q}_m$ (Feitosa et al. 2007):

$$\breve{q}_m = \frac{N_{im}}{N_m} \left(\frac{\breve{L}_m}{\breve{L}_j} \right). \tag{2.7}$$

Where: the equation parameters are the same as in equation (2.6).

In general, measures of segregation are useful tools for describing the phenomenon in its multiple scales and dimensions. By computing these measures to different dates, it is possible to analyze several aspects of segregation: Is the global segregation of a city increasing or decreasing? Is this trend applied to both dimensions of segregation? What is happening at smaller/larger scales? Where are the most critical areas of poverty isolation?

Measures of segregation, in particular the local ones, can be also used to explore the relationship between the segregation of social groups and other urban indicators. For example, local indices of segregation estimated at different scales and compared with violence rates can reveal whether poor families isolated at broader scales are more vulnerable to violent events than those who are segregated at smaller scales or not segregated. Such experiments can contribute to the debate about different patterns of segregation and their impacts.

Nevertheless, despite the value of these measures, they represent only static snapshots of segregation at a certain moment. They are unable to help researchers to understand the underlying dynamics of the phenomenon or how different contextual mechanisms (such as those described in section 2.4) can lead to the emergence of specific patterns of segregation. The next chapter introduces a set of concepts and methods related to the theory of complexity that contribute to overcome this limitation.

3 URBAN SEGREGATION AS A COMPLEX SYSTEM: CONCEPTS AND METHODS

3.1 The complex nature of urban segregation

Urban segregation is an explicit spatial phenomenon that emerges from the interaction between many individuals and displays markedly different global patterns depending on specific socioeconomic contexts. A better understanding of segregation is challenged by the fact that it exhibits many of the characteristic hallmarks of a so-called complex system. According to Batty and Torrens, a complex system is an “entity, coherent in some recognizable way but whose elements, interactions, and dynamics generate structures and admit surprise and novelty that cannot be defined a priori” (Batty and Torrens 2005: 745). The idea that “the whole is more than the sum of the parts” (Simon 1996: 231) is crucial to understand complex systems, and it is what differentiates them from those that are merely complicated.

A complicated system consists of many elements that are independent of each other. For this reason, with a reductionist thinking, scientists can understand complicated systems by reducing them to their atomic elements and then studying these elements in isolation (Holland 1998; Miller and Page 2007). When the dependence among the atomic elements starts to play a role, the system shifts from complicated to complex, and the same reductionist approach fails to provide insights about it (Levy 1992; Miller and Page 2007). In a complex system, many heterogeneous and autonomous elements interact at the micro-level and give rise to the global properties of the system. These properties, which are called *emergent* (Holland 1998), then feedback into the system’s micro-level (Figure 3.1).

The idea of emergence also applies to urban segregation, since its macro-structure emerges from the interaction between many individuals (households) at a micro-level, who are constantly making choices about their residential location. As a complex system, segregation cannot be simply understood through the investigation of its ‘micro-elements’ alone like, for instance, through studies about individual reasons for residential mobility (Clark and Onaka 1983; Knapp et al. 2001; South and Deane 1993). It is also difficult to understand segregation through studies situated at the other extreme, i.e., studies that focus only on the ‘macro-structure’, like those that emphasize the measurement of segregation patterns resulting from the spatial arrangement of social groups (Feitosa et al. 2007; Morrill 1991; Reardon and O'Sullivan 2004; White 1983 ; Wong 1993; 1998).

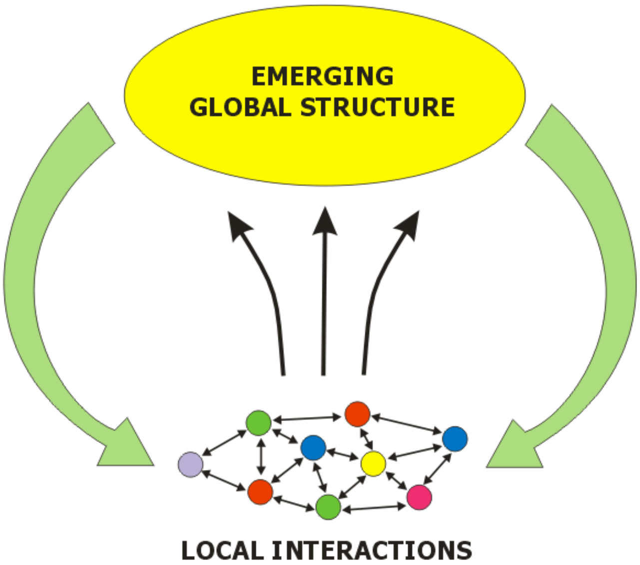


Figure 3.1 The principle of emergence, a hallmark of complexity (Adapted from Psycology Wiki 2009).

These contrasting approaches dealing with the micro- and macro-dimensions of segregation are actually complementary. It is important to recognize factors that influence the residential mobility of different types of households (micro-level), as well as to have tools that are able to describe and quantify patterns of segregation (macro-level). Nevertheless, these conventional approaches based on reduction or aggregation fail to provide insights about the 'macro-micro' relations that underlie the dynamics of segregation. In this direction, the science of complex systems and its ability to explore what is 'in-between' the usual scientific boundaries (Miller and Page 2007) becomes particularly pertinent for the challenge of conciliating processes that operate at local scales with those at larger scales (Torrens 2000).

In addition to emergent properties, many other system characteristics belonging to the lexicon of the complex systems science are useful to explain urban segregation, including non-linearity, adaptation, self-organization, and path dependence. Segregation results from *non-linear interactions* between many independent households that are able to generate unexpected and counter-intuitive global patterns (Schelling 1971). The changes in the system are induced by the residential decisions of *heterogeneous* households, which can differ regarding several attributes (e.g., knowledge, needs, income, race, etc.). These decisions are *context dependent* and constantly *adapting* to the current circumstances. They are also not necessarily brilliant (Miller and Page 2007), since households have limited knowledge and capacity to process information, i.e., their decisions result from their *bounded rationality* (Benenson and Torrens 2004).

There are many context-related mechanisms that can influence the households' decisions about moving into a specific location: land market, personal preferences, labor market, and public policies and investments (section 2.4). All these factors are *dynamic* and not only influence, but can also be influenced by the households' choices. For example, while the households' residential choices are constantly constrained by the land market, the spatial distribution of households resulting from these choices consolidates neighborhoods with certain reputations and characteristics that affect the land market dynamics (feedback mechanism). The different levels of interactions in the system are also *self-organized*, being able to produce recognizable patterns without any centralized authority deliberately managing or controlling (Holland 1998). In addition, segregation is also characterized by *path dependence*, since earlier states and choices are able to affect future possibilities. Finally, the dynamics of segregation have a strong *spatial* component: Households are constantly evaluating their local environment and their decisions are not only influenced by their location, but are also about their future location.

3.2 Social simulation as a tool for exploring the 'in-between'

Social scientists have traditionally relied on research methods that consist either of mathematical and equation-based models or verbal descriptions based on historical and ethnographic observation (Hanneman et al. 1995). Verbal descriptions offer a high flexibility regarding the type of problems that can be analyzed. Nevertheless, this method is often vague, inconsistent and difficult to verify. Apparently coherent and logical arguments may, in fact, contain critical flaws (Holland 1998; Miller and Page 2007). Studies relying on this type of approach are the most commonly found in the Brazilian literature about segregation (Caldeira 2000,2005; Schiffer 2001; Villaça 1998,2001).

At the other extreme, mathematical or equation-based models rely on precise and general statements about a system. They provide new insights through the interpretation of quantities such as equilibrium, maxima and minima, and partial derivatives of dependent variables with respect to independent variables (Hanneman et al. 1995). This rigor and generality of mathematical analytical models (often used deductively) represent attributes that are desirable in science, although they also impose constraints on the study of systems that are dynamic, spatial, non-equilibrated, heterogeneous and comprising individuals with bounded rationality. Traditional economic models often focus on situations involving very few of infinitely many rational agents that have access to all available information, and process this efficiently towards optimization (Miller and Page 2007). The work of Yizhaq, Portnov, and Meron (2004) exemplifies the application of this type of modeling to describe segregation. Another common quantitative approach relies on statistical methods to filter out the noise and extract the regular part of the agents' behavior, which are often used inductively. It is the case of discrete choice models (Ben-Akiva and Lerman 1987), which decompose the agents' utility of choosing a residential location into a random part (the noise) and a deterministic part (the regular). Nevertheless, this emphasis on the average behavior may be incomplete, or in some cases, even misleading (Miller and Page 2007).

While verbal descriptions offer flexibility, mathematical models offer the rigor. However, to deal with an 'in-between' field such as a complex system, it is necessary to rely on models that are able to bridge the gap between both these attributes. Computer-based simulation represents a promising alternative in this direction. It is more flexible than mathematical models, and more rigorous than verbal descriptions (Hanneman et al. 1995). Over the past decade, computer-based simulations have become much more widely accepted among social scientists. They are able to deal with complex system issues that cannot be 'solved' through mathematical equations. At the same time, computer-based simulations demand much higher precision than verbal descriptions, since they force modelers to specify computer programs in a complete and exact manner (Gilbert 2008). These simulations allow researchers to explore a variety of new questions through a sort of 'laboratory' on their desktop: Simulations can be started, stopped, examined, modified and restarted to test new hypotheses (Holland 1998).

According to Gilbert and Troitzsch (1999), the logic of simulation as a method consists of building an abstract representation from a target system based on certain assumptions. This model is then converted into a computer program, which can be executed and generates simulated data. The obtained results should be compared with data collected from the target system to check whether the model generates outputs which are similar to those produced in the real world (Figure 3.2). This process of developing a simulation model demands a strategy 'in-between' deductive and inductive approaches. Like deduction, one starts with a set of assumptions, but then relies on experiments to generate data that can be analyzed inductively. For this reason, simulation has been known as a *third way of doing science* (Axelrod 2003).

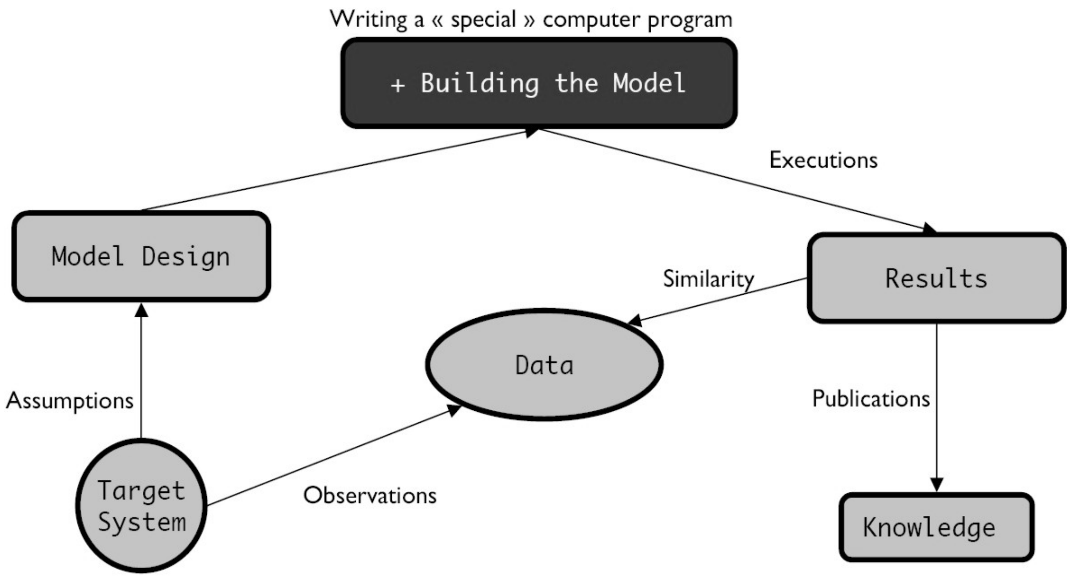


Figure 3.2 The logic of simulation as a method. Diagram designed by Drogoul et al. according to their interpretation of Gilbert's and Troitzsch's proposition (Drogoul et al. 2003).

During the last half-century, three approaches have gained special attention in the field of social simulation: system dynamics, microsimulation, and agent-based models (Gilbert and Troitzsch 1999). Originally developed in the 1950's by Jay W. Forrester, system dynamics is a macro-level approach that relies on systems of difference and differential equations (Gilbert and Troitzsch 1999). It is structured in terms of temporal cause-and-effect relationships, and focuses on feedback linkages among the components of the target system (Roberts 1983). System dynamics has been useful for exploring the non-linearity of some complex systems over time, since it relies on elements like feedbacks, stocks, and flows. However, this approach deals with aggregates rather than with atomic elements, which restricts the possibility to model heterogeneity and more complex recursive reasoning processes (Gilbert 2008).

While system dynamics simulates the behavior of an aggregate agent, microsimulation focuses on simulating the development of a population of individual agents over time. This approach, which started to be developed in the social sciences during the 1970's, uses data on large samples of individual units (e.g., households, vehicles, or firms), along with rules to simulate the evolution of the sample individuals (Gilbert and Troitzsch 1999; Gilbert 2008). Despite the focus on the individual as the unit of analysis, microsimulation kept the emphasis on higher levels of aggregation forecasting (e.g., national unemployment rate), without pretending to explain (Gilbert and Troitzsch 1999). Other disadvantages that disqualify the use of microsimulation to study segregation include its inability to model interactions between elements and the lack of spatial references (Gilbert and Troitzsch 1999; Gilbert 2008).

The third approach, known as agent-based model (ABM), addresses the shortcomings of the previous techniques by enabling the representation of individual decision-making units interacting with each other and their environment. These decision-making units, called *agents*, are autonomous and possibly heterogeneous (Gilbert 2008). Agents are presumed to be interacting according to a specific set of rules that can be changed through adaptation and learning (Gilbert 2008). The explicit simulation of such interaction process, which occurs at the micro-level, allows researchers to explore the emergence of macro-structures from bottom-up in a very natural way (Gilbert and Troitzsch 1999; Miller and Page 2007). For this reason, agent-based models represent a promising approach to understand complex social systems, including urban segregation.

Schelling's model of racial segregation (Schelling 1971) is a classical example illustrating an agent-based simulation and its ability to provide further insights about complex systems. He distributed white and black agents on a lattice and considered that these agents had a degree of tolerance in relation to the other racial group: They were satisfied with a mixed neighborhood, as long as the number of neighbors with the same color was sufficiently high. It would be reasonable to think that if agents do not insist on living with the same race, no segregation pattern will emerge. However, Schelling demonstrated the unexpected fact that patterns of intense racial segregation appear under these conditions. This is an example of how agent-based models can contribute to understanding the emergence of counter-intuitive global structures from local individual interactions.

3.2.1 Purposes of social simulation

Earlier approaches to simulation have commonly been used with the purpose of obtaining accurate and quantitative predictions of a real-world system. Such purpose relies on a positivist view of the role of simulation models. It believes in a fully observable law governing systems that can be reproduced in a model and extrapolated to predict the future (Wu 2002). Nevertheless, with the advances of complexity and the chaos theory, this view has been often criticized or considered too difficult to achieve (Batty and Torrens 2005; Macy and Willer 2002; Wu 2002).

The chaos theory is frequently mentioned to explain why the precise prediction of the future state of a system can be so difficult (Holland 1998). It advocates that small and local change can lead to major transformations in the system's evolution. Therefore, the prediction of a chaotic system depends on the perfect knowledge of the initial conditions and all the values of all the relevant variables, which is usually impossible to obtain. This idea is summarized by the well-known 'butterfly effect' example, where the flapping of a butterfly's wing can eventually cause worldwide changes in the weather (Holland 1998). Complex systems are considered to be at the edge of chaos and order: not so active, but also not static. Therefore, the key to improve the knowledge about them is to determine mechanisms that provide some *structure* to the system, as well as the minors that can be ignored (Holland 1998; Miller and Page 2007). Prediction, in this case, depends on the required level of detail.

From the recognition of challenges imposed by the complexity of real-world systems, researchers have become increasingly interested in predictions that are not necessarily quantitative. According to Troitzsch, prediction can have at least three meanings, each responding to one of these three different questions about the behavior of the target system (Troitzsch 1997, as quoted in Troitzsch, 2009: 1.1):

1. "Which kinds of behavior can be expected under arbitrarily given parameter combinations and initial conditions?"

2. “Which kind of behavior will a given target system (whose parameters and previous states may or may not have been precisely measured) display in the near future?”
3. “Which state will the target system reach in the near future, again given parameters and previous states which may or may not have been precisely measured?”

The traditional view of prediction, which focuses on quantitative results, answers the question (3), and it is the goal of models that seek to reproduce the dynamics of the target system as exactly as possible. This approach demands the use of models that Gilbert (2008) call ‘facsimile’. Their parameters are usually calibrated to precisely replicate a known situation (present or past), and the models are then used to predict the future or what could happen if something were changed (Batty 2005; Gilbert 2008). Nevertheless, when the system’s trajectory reacts in a chaotic manner, i.e., highly sensitive to initial conditions and parameters, quantitative prediction of future states are likely to be impossible to achieve (Troitzsch 1998).

Conscious of this limitation, most simulations developed in the social sciences concern qualitative predictions, which are able to answer the questions (1) and (2). This sort of simulation is mainly focused on identifying global patterns that can emerge when certain local rules are applied or some specifics laws are considered (Troitzsch 2004). Despite their use of quantitative procedures, facsimile models can be used for qualitative predictions, especially when studying phenomena like segregation. A model that includes segregation measures, for instance, is likely to face difficulties in producing outcomes that can be rigorously compared with the measurements from real-world data. It is not reasonable to expect that the local segregation index computed for a ‘cell’ (section 2.5) achieves the same value when computed for the simulated and real data. Indeed, this is also not important. The most important thing is to observe the trends of segregation in different scales and dimensions along time, and how they change once the modeler explores different parameters and conditions.

In addition to facsimile models, models that Gilbert (2008) designate as *abstract* and *middle-range* can also be used to achieve qualitative predictions. Abstract models are those without any intention to simulate a specific empirical reality, and therefore are based on ‘artificial societies’. Even so, abstract multi-agent models should be able to demonstrate the emergence of expected macro-level patterns from the interaction of agents that follow plausible rules (Gilbert 2008). The Epstein and Axtell’s Sugarscape model is a classical example of an abstract model (Epstein and Axtell 1996).

Like abstract models, middle-range models are general and not applicable to a specific observation, but they do focus on a particular empirical phenomenon (Gilbert 2008). The aim of this type of model is to extract some conclusions about a target system that can be widely applied, including specific results that we can expect under certain circumstances. For example, Schelling’s model of racial segregation reveals that we can expect to find segregation patterns even if households do not mind having others from different races in their neighborhood (Schelling 1971). Besides this model, others have been developed to simulate segregation dynamics in a general manner, including many variations of Schelling’s model (Bruch and Mare 2006; Laurie and Jaggi 2003; O'Sullivan et al. 2003; Zhang 2004), and the SimSeg model (Fosset and Senft 2004).

Finally, it is still important to consider simulations for other exploratory purposes that do not lead to any type of prediction. For example, simulations can be useful to provide a rigorous demonstration that something is possible, illustrate a certain dynamic for educational purposes, or simply suggest new ideas about a complex situation (Holland 1998).

3.3 Agent-based models: basic concepts

Agent-based models (ABM) have been increasingly recognized as a useful approach for studying complex social systems in general and urban segregation in particular (Benenson et al. 2002; Bruch and Mare 2006; Crooks 2008; Schelling 1971). Briefly, an ABM consists of multiple agents interacting within an environment. From this succinct definition, it is possible to extract the most basic components of an ABM, i.e., agents, interactions, and environment, whose concepts are introduced in the following.

3.3.1 Agents

In an ABM framework for social simulation, agents are computer programs used to represent social actors, e.g., individuals, households, or institutions (Gilbert 2008). With the increasing popularity of this sort of simulation, many discussions have arisen around the definitions of agent, and how they differentiate from computer programs in general.

After examining many of the existing definitions, Franklin and Graesser (1997: 27) proposed the following one: “An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future”.

Although the properties of agents differ according to specific applications, Franklin and Graesser’s definition emphasizes agent’s features that have been conventionally identified as important: autonomy, social ability, reactivity, and proactivity (Wooldridge and Jennings 1995). First, agents are autonomous; they are a separate *locus* of control, fully responsible for their actions and in charge of accomplishing their role (Wooldridge and Jennings 1995). Although centralized authorities may exist as environmental constraints, there is no global or external flux of control dictating the agent’s actions. This ‘self-organization’ of autonomous agents is what promotes the emergence of global patterns from the bottom-up (Macy and Willer 2002). Second, agents have social ability and are able to interact with each other. Third, agents are reactive and capable of responding to stimuli coming from their environment. In addition, agents are proactive, which means that they exhibit goal-directed behavior by taking their own initiative (Wooldridge and Jennings 1995; Zambonelli et al. 2001).

Concerning the problem of conceptualizing and designing agents, Gilbert (2008) advocates another set of properties that he considers more helpful to have in mind during this process: perception, performance, memory, and policy (Gilbert 2008: 21-22). Agents are able to perceive the characteristics and dynamics of their environment, including the presence of other agents in the surroundings. Agents are also capable of performing a set of behaviors, which often includes motion (they can move within the environment), communication (they send and receive messages), and/or action (they can change the environment). Agents should have a memory and be able to record their past perceptions and performances. Finally, agents have a policy, i.e., “a set of rules, heuristics, or strategies” that establishes what they will do next, or even how they learn and adapt (Gilbert 2008: 22).

Considering the example of a segregation model, agents represent households and should be implemented in a way that allows them to recognize the attributes of different neighborhoods, including the agents living there (perception), store and retrieve all the neighborhood’s perceptions (memory), compare them, evaluate whether it would be better to live in a neighborhood different to their current one (policy), and finally, move if they consider appropriate (performance).

3.3.2 Environment

Environment defines the space in which agents operate, serving as a support to their actions. The meaning and role of an environment depends on the system that is being modeled. In some situations, it may be neutral, with minimal or no effect on the agents or, in analogy to the real world, the environment may have an active role in providing the context for agents to perform their actions, to acquire information about the problem they have to solve, and to communicate with each other (Gilbert 2008; Weyns et al. 2005). In the latter case, the environment can be specified as an independent piece of software that encapsulates its own roles in the ABM, including particular characteristics and dynamics that directly influence the agent's behavior and the emergence of complex structures (Gilbert 2008; Weyns et al. 2005). It can be implemented as agents, but in a simplified manner: The environment has its own attributes and set of rules for changing its state, but it does not need to achieve goals or perform elaborated actions such as moving and send messages (Gilbert 2008; Weyns et al. 2005).

Focusing on the relation agent-environment, Russell and Norvig (2003) advocates that agents perceive the environment through sensors, and act upon it through effectors. The properties of environments may vary significantly, and can be classified as the following (Russel and Norvig 2003: 46):

1. Accessible versus inaccessible: reveals whether the agents can access complete and accurate information about environment's state, or not.
2. Deterministic versus non-deterministic: reveals whether the next state of the environment is entirely determined by its current state and the actions performed by agents, or not.
3. Episodic versus non-episodic: reveals whether the agent's decisions within an 'episode' do not influence its decisions in the next episode, or the opposite.
4. Static versus dynamic: reveals whether the environment only changes when agents act, or not.
5. Discrete versus continuous: reveals whether the number of states and actions in the environment are limited, or not.

Depending on the system that is being modeled, the environment can be more appropriately represented as a geographical space, an analogy to space, or a network (Gilbert and Troitzsch 1999; Gilbert 2008). Environments as geographical spaces are particularly suitable for problems where absolute distances matter, such as segregation. For other problems, however, it may be convenient to model the space as an analogy to some features other then geography, e.g., “knowledge space” (Gilbert 2008). There is still another sort of application where the most important is the relationships between agents (e.g., trades) and the environment can be represented as a network of links and nodes (Gilbert 2008; Tesfatsion 2003).

In most ABM, the environments represent a geographical space where agents are located and, in many cases, able to move around (Gilbert and Troitzsch 1999; Gilbert 2008). In a model of segregation, for instance, this type of representation can provide a geographical reference that allows agents to have a notion of proximity and identify other agents in their vicinity. In addition, it provides information about other urban features that are also relevant to the households’ decisions on residential locations, e.g., land price, dwellings availability, and quality of infra-structure.

The features of such spatially explicit environments can be abstractly simulated or directly portrayed from real landscapes. The inclusion of detailed representations of the real word is facilitated by integrating geographical information systems (GIS) into the model (Castle and Crooks 2006; Crooks 2006; Gilbert 2008; Parker 2005). A GIS is a computational system that is designed to assemble, store, update, analyze, and display geographically referenced data (Worboys and Duckham 2004). It can contain multiple layers with different features and attributes about the real world, e.g., roads, buildings, and land use.

Once the integration between GIS and ABM is established, the modeled environment may rely on detailed geographical data from a GIS, and possibly also write the output of its simulated dynamic into a format readable by GIS (Parker 2005). Although environments in ABM have been often represented as a two-dimensional grid, ABM integration with GIS has allowed the use of the so-called vector GIS, i.e., the use of polygons for representing the environment. For example, polygons could be used to represent a variety of land-parcel shapes and sizes in an urban environment (Crooks 2006). The use of polygons to represent the environment introduces new operational challenges to the model, which can be related, for instance, with the definition of neighborhoods or the agent's capability to detect environmental features. Nevertheless, these challenges can also create opportunities, e.g., the ability to include topological relations such as adjacency or intersection (Crooks 2006).

3.3.3 Interactions

Interactions represent the main feature that distinguishes ABM from other simulation approaches, like microsimulation and system dynamics. The agents' potential to locally interact with each other and their environment is the key to the simulation of the emergent properties of complex systems (Axelrod 2003; Holland 1998). For this reason, all ABM include some sort of interaction that involves transmission of knowledge or materials that can affect the behavior of the recipients (Gilbert 2004). The nature and sophistication level of these interactions may vary substantially depending on the roles assumed by the agents in a simulated system. In some cases, agents interact by simply perceiving the presence of their pairs in the surroundings, while other situations demand interactions based on the development and use of complicated communication means (Gilbert 2008; Zambonelli et al. 2001). In general, ABM can present direct agent-agent interactions, indirect agent-agent interactions, and agent-environment interactions.

Interactions between agents (agent-agent) usually have an ontological correspondence to social relations that take place in the real world (Gilbert 2004). As in the real world, these interactions can be direct or indirect. Agents can directly interact with each other by giving and receiving resources (e.g., money or food), or by exchanging information through messages. In the latter case, the communication between agents can demand the specification of a 'language' (Gilbert and Troitzsch; Gilbert 2008). In these situations, the agent's interaction can go beyond the clear and direct agent-to-agent message exchange and include some ambiguity in the communication. This ambiguity allows, for instance, the simulation of agents misunderstanding received messages and transmitting them in a different manner. Such problem has been particularly explored in studies dealing with the evolution of language (Gilbert 2008; Smith et al. 2003; Steels 1997). Agents can indirectly interact with others by observing them, copying their behavior, or even avoiding them (Gilbert 2008). In a segregation model, for instance, households can have indirect interactions by detecting the status of households that live in a specific neighborhood, or by trying to imitate the residential standards of households with similar or higher status.

In general, agent-agent interactions can also be defined as cooperative or competitive. They are cooperative when agents exchange knowledge to coordinate activities, improve their collective performance, and accomplish their goals as a team (Jennings et al. 1998; Zambonelli et al. 2001). On the other hand, competitive interactions occur when agents are self-interested and try to maximize their individual benefit, often at the expense of others (Jennings et al. 1998; Zambonelli et al. 2001). Interactions between agents in a segregation model occur on competitive bases, since the macro-patterns of segregation can be seen as the outcome of a continuous contest for the most convenient residential locations in the city (Feitosa et al. 2008; Villaça 1998).

Interactions between agents and their environment also play a vital role in multi-agent models. As Russell and Norvig (2003) state, agents are constantly being influenced by the environment through their sensors and are influencing it through their effectors. These agent-environment interactions are often used to mediate indirect interactions between agents, since agents are able to detect the impacts of another agent's action on their shared environment and act in response to it (Gilbert 2008; Le 2005). This can be illustrated with another example pertinent to the dynamics of segregation: Poor households can decide to move to another location as a response to the increase in prices promoted by the presence of many new affluent households in their neighborhood.

Agents' interactions involving the environment are typically local, with agents having only a limited sphere of influence through which they can sense and alter the environment (Jennings and Wooldridge 2000). Nevertheless, this is not necessarily true for all applications. For instance, an agent that interacts with the environment by changing its residential location can have a higher propensity to move to a closer neighborhood, but it should also be able to evaluate the possibility of living in further places. In real life, people can acquire some knowledge about many neighborhoods that are not necessarily close to their own. This knowledge can be gained through their personal contacts, media, or simply circulating in the city.

3.4 Agent architectures

An agent architecture can be defined as a "structural model of the components that constitute an agent as well as the interconnections of these components together with a computational model that implements the basic capabilities of the agent" (Lind 2001: 184). There are many different approaches of agent architectures. Each of them has its own merits depending on the tasks that agents have to perform and how they should interact with other agents and the environment. In essence, they represent distinct approaches to design how agents perceive other agents and their environment (perception), store and retrieve these perceptions (memory), evaluate the circumstances and decide what to do (policy), and finally perform the action considered the most appropriate and viable (performance).

Russell and Norvig (2003) propose the following classification of agent architectures: simple reflex, model-based reflex, goal-based, utility-based and learning agents. Simple and model-based reflex agents simply act as a response to a stimulus and do not have a reasoning model. Their functioning relies on production-rule systems, which basically consist of a set of rules about behavior, called productions (Luger 2005; Nilsson 1998; Russel and Norvig 2003). These rules present a conditional (IF) and action component (THEN), i.e., if the current situation matches the condition established by a certain rule then the agent performs the action related to the same rule. A production system also contains a working memory, which is a database that stores the agent's current state, and a rule interpreter. The rule interpreter is a program that selects the productions that should be executed (Gilbert 2008; Klahr et al. 1987). The difference between simple and model-based reflex agent types consists in how agents define and interpret the current situation. While simple reflex agents define the current situation only by their perceptions at the moment, model-based agents define it also by an internal state stored in the agent memory. This internal state contains a representation of the environment, a 'world view' model, which estimates how the environment evolves and how the agent's actions can affect it. Due to this mechanism, agents have knowledge about the part of their environment that they cannot currently perceive.

Goal-based agents have goal information describing desirable situations. Unlike reflex agents, who react immediately to stimuli, goal-based agents have a symbolic reasoning model. They are able to take future events into consideration and 'plan' a sequence of actions to reach their goal. For that, condition-action rules are replaced by a goal-seeking framework. Artificial intelligence techniques involving search and planning provide solutions to finding the action sequences of goal-based agents (Luger 2005; Nilsson 1998; Poole et al. 1998; Russel and Norvig 2003).

Utility-based agents are similar to goal-based agents, but the goals are not so clear and only differentiate 'desirable' from 'non-desirable' states. This type of agents has a utility function that weights the importance of different goals, and provides a 'performance measure' that allows the comparison of different states. It is a function that maps a state to a measure that quantifies the associated degree of the agent's happiness (Russel and Norvig 2003). Unlike goal-based agents, utility-based agents do not choose an action that achieves a goal, but an action that increases their utility or happiness. In case of conflicting goals, e.g., high quality and low price, the utility function determines the suitable trade-off (Russel and Norvig 2003).

Learning agents represent the most sophisticated type of agent. They act independently, learning and adapting to changing circumstances. Learning agents are able to analyze themselves in terms of behavior, error and success. Based on their past critical analyses, they are able to learn which perceptions of the environment are desirable, and how to behave in order to improve their future performance. An advantage of learning agents is their capacity to operate in unknown environments and become more competent with time (Russel and Norvig 2003). Different techniques can be used to design learning agents, including genetic algorithms (Gilbert and Troitzsch 1999; Holland 1975; Mitchell 1996) and neural networks (Bar-Yam 1997; Gilbert and Troitzsch 1999; Gilbert 2008).

Simple and model-based reflex agents can be categorized as reactive, i.e., they choose their actions immediately by following rules that address a specific situation (Wooldridge and Jennings 1995). This type of agent does not use complex symbolic reasoning and is only sufficient for limited environments, where the possible situations can be covered by the production rules. On the other hand, goal-based, utility-based, and learning agents are called deliberative, since they have a central reasoning system that constitutes their 'intelligence' and are able to carry out intentional plans to accomplish their goals (Ginsberg 1989; Wooldridge and Jennings 1995). It is also possible to design hybrid agents, which use reactive and deliberative approaches to obtain the best properties of each. Hybrid agents can respond quickly to simple and well-known situations, and also have the ability to make plans and evaluate unforeseen situations during their decision-making process (Wooldridge and Jennings 1995). For instance, an agent's behavior can be guided by simple rules while selecting a subset of alternatives from the all the possible ones and then rely on utility functions to choose the most appropriate from this subset.

3.5 Methodological protocol for developing ABM simulations

The increasing use of ABM for studying complex problems has established some pragmatic steps, or a sort of 'protocol', for developing an agent-based simulation (Dooley 2002; Gilbert and Troitzsch 1999; Gilbert 2004,2008; Le 2005; Richiardi et al. 2006). This protocol relies on the basic ideas of the 'logic of simulation as a method' (Gilbert and Troitzsch 1999) and commonly consists of the following steps: (1) problem analysis and objective formulation, (2) conceptual modeling, (3) theoretical specification, (4) programming, (5) verification, (6) validation, (7) analyses of simulated results, and (8) documentation of scientific findings. These methodological steps served as basis for the simulation model in this study and are presented in the following. However, it is important to emphasize that these steps are introduced in a idealized consecutive order, while in practice most of them are iterative and may take place sequentially or in parallel (Gilbert 2008).

3.5.1 Problem analysis and objective formulation

Like in any research process, modelers should start by analyzing the problem of interest and specifying a question that represents the objective of the research. This main question should be able to generate some specific objectives or research questions, where the level of detail can be associated with the main elements of the model (Gilbert 2004). Usually, this stage involves observations and a review of existent theories about the target system, which helps the modeler to articulate his/her beliefs about the actual system's behavior, identify factors that seem to be relevant to the problem, and specify the assumptions on which the model will be developed (Gilbert 2004; Troitzsch 2004).

3.5.2 Conceptual modelling and theoretical specification

After establishing the research objectives and assumptions that will guide the model-building process, the specifics of the model should start to be conceptualized. Since any model is a simplified representation of the real world, it is necessary to determine the level of detail that will be considered and, therefore, the type of model that is going to be built (abstract, middle-range, or facsimile). Based on that, this is the moment to define how this real-world system, with its actors, relations, and environment, will be translated into abstract components of a computational model.

The conceptual model provides a general view of agents, environment, interactions and, if pertinent, external factors that may influence the modeled system. Real-world actors, e.g., people and institutions, will be represented as agents with a set of attributes. These attributes can be dynamic and represent the state of agents, or can simply consist of static properties that differentiate one agent from the others (Gilbert 2004). Key aspects about the environment that will support the agents' actions also need to be defined, including what they represent (e.g., geographical space or networks), their relevance in the model, and whether they will be modeled as another type of agent with its own attributes and dynamics or not. The conceptual model includes also the linkages between different agents and the environment, and may contain external factors, e.g., public policies, which influence the modeled system although their dynamics are not being simulated.

After the definition of the conceptual model, it is time to conduct the theoretical specification of its components. The theoretical specification consists of building the model's architecture, including details about the modules that comprise the whole system, the linkages, and algorithms (Le 2005; Le et al. 2008). A very important part of this stage is the definition of an approach for designing the agents, i.e., the agent's architecture. There are many types of agent's architectures, and their suitability varies according to the goals of the simulations (section 3.4). The theoretical specification provides the guidelines for the model implementation and, in case of empirically-based models, also guides the data collection and the methods for empirical estimation of the simulation parameters (e.g., statistical models, descriptive statistics, and spatial analyses). For models based on empirical data, the theoretical specification, data collection and parameterization are conducted in a very iterative manner. For instance, the lack of certain types of data may demand changes in the parameterization and theoretical specification. Another example is the case when the empirical parameterization reveals unexpected aspects that are relevant for the behavior of the target system, and motivates a review of the theoretical specification.

3.5.3 Programming

Once the model has been specified, these specifications are converted into an executable computer program. The program can be written from scratch, with an object-oriented programming language, although it is usually easier to use one of the available ABM simulation platforms (Gilbert 2008). In general, these platforms provide simulation frameworks that benefit the development of agent-based applications in many aspects. One clear benefit provided by any ABM platform is the fact that they relieve researchers from programming the parts of simulation that are not content-specific, e.g., basic algorithms and graphic libraries (Gilbert and Bankes 2002; Tobias and Hofman 2004). In addition, they improve the reliability and efficiency of the simulations, since many parts of the program have been developed by professional developers (Tobias and Hofman 2004).

There are different types of platforms to support ABM simulations. Some are libraries of standardized routines that researchers can include in their simulation programs, while others are complete modeling environments with their own programming language. Commonly used ABM platforms include NetLogo (Wilensky 1999), Repast (North 2006), MASON (Luke et al. 2005) and Swarm (Minar et al. 1996). There is no 'ideal' platform, and the choice of one of them should consider the user's expertise in programming, the purposes of the study, and the expected complexity of the model (Gilbert 2008). Many researchers have provided reviews of and comparisons between some platforms, including Gilbert and Bankes (2002), Serenko and Detlor (2002), Tobias and Hofman (2004), Castle and Crooks (2006), Railsback et al. (2006), and Gilbert (2008).

This research adopted NetLogo, a multi-agent programming language and modeling environment developed at the Center for Connected Learning and Computer-Based Modeling (CLL). As a multi-agent programming language, NetLogo supports agents called turtles that are able to move on a grid of patches. Turtles and patches can interact with each other and perform multiple tasks concurrently (Tisue and Wilensky 2004). Thus, NetLogo is suitable for simulating complex phenomena evolving over time, allowing modelers to give instructions to many independent agents and explore the emergence of global patterns from the local interactions between these agents and their environment.

NetLogo provides an entry-level programming interface that reduces programming efforts, but it is still powerful enough for sophisticated modeling and allows experienced programmers to add their own Java extensions (Railsback et al. 2006; Sklar 2007). In addition, NetLogo provides a built-in graphical interface, an extensive and comprehensive documentation, and has an active user community that answers user's questions very efficiently (Gilbert 2008; Railsback et al. 2006). As a result, NetLogo currently stands out as the most popular agent-based modeling environment, being used across a wide range of disciplines and educational levels (Gilbert 2008; Sklar 2007).

3.5.4 Verification

The process of developing an operational simulation model includes its verification, which consists of checking if the program executes exactly what is stated in its theoretical specification. The verification or 'debugging' is particularly difficult to execute in simulations of complex systems, since the outcomes can be unexpected and it is often unclear whether they emerged from the agents' features and interactions, or from some unknown bug (Gilbert and Terna 1999). In addition, simulations are often stochastic, with a random component that simulates the effects of uncertainty, and repeated runs can generate different outcomes (Gilbert and Troitzsch 1999).

There are many techniques and good practices that assist the reduction of bugs (Gilbert 2008; Schut 2007), including: (a) writing elegant codes, (b) recording many intermediate outputs and checking the simulation step by step (e.g., comparing them with calculations done in a spreadsheet), (c) testing the model with parameter values from known scenarios, and (d) testing the model with extreme parameters to see whether outputs are reasonable. Additional verification techniques can be found in Gilbert (2008), Schut (2007), and Wooldridge (1997). The use of platforms like NetLogo is also useful during the verification process, as these offer run-time testing and debugging environment. In addition, NetLogo's simple programming language can be more easily reviewed. Nevertheless, it is always important to keep in mind that no simulation is totally free of bugs (Gilbert 2008) and, therefore, verification is a time-consuming stage that must be always carefully conducted.

3.5.5 Validation and analyses of results

After checking whether the simulation program is working according to its specification, it is important to validate the model. Validation is the process of ensuring whether the simulation output is a suitable representation of the real target system (Gilbert and Troitzsch 1999; Gilbert 2008). There are many challenges involving the validation of social systems, a fact that often encourages criticisms of using multi-agent models in social sciences. An important challenge is related to the difficulties in acquiring suitable data for systematic validation (Troitzsch 2004). In addition, different types of simulation models can be developed for many purposes (section 3.2.1), including quantitative and qualitative predictions, and this implies different criteria for validation (Gilbert 2008).

Abstract models, for instance, should be seen as part of the process of theory development and thus, their validation needs to consider whether: (a) the model produces expected and interpretable macro-patterns, (b) the agent's behavioral rules that generates these macro-patterns are plausible, and (c) the model is able to generate further theories (Gilbert 2008). For other models, which also focus on qualitative prediction but are more closely related to a specific social phenomenon, the validation should consider whether the dynamics of the model are similar to the ones observed in the real world (Gilbert 2008). This is the case of middle-range models and facsimile models for qualitative prediction, where comparisons of real and simulated outputs can rely on 'statistical signatures' (Gilbert 2008; Moss 2008) or methods of pattern recognition (Bishop 1995; Yilmaz 2006). Finally, there are facsimile models for quantitative prediction, which use validation techniques to assure that the model reproduces the state of a particular target system as exactly as possible (Gilbert 2008).

According to Gilbert (2008), models can be validated in terms of the fit between a theory and the corresponding model for that theory, and in terms of the fit between the simulated outputs and the real target system that the model aims to simulate. The first type of validation can be conducted through sensitivity analysis, while the latter type demands comparisons of the model with empirical data (Gilbert 2008).

Sensitivity analysis is a careful investigation of how the simulated outputs vary when one or more of the model's factors are modified. The term factor refers to a parameter, input variable, or module of a simulation model (Kleijnen 1995,1999). Different factor settings can correspond to different assumptions about the relationship between these factors and the target system. Considering Schelling's model of segregation (1971), for instance, a sensitivity analysis can be conducted to investigate how changes in the households' preferences affect global patterns of segregation.

Sensitivity analysis requires many simulation runs, with factors changing from run to run. Since models often have many factors, sensitivity analysis including all conceivable combinations of factors would demand an enormous number of runs (Gilbert 2008). For this reason, a central problem of sensitivity analysis is how to select a sub-set of factor combinations from all the possible ones. Possible approaches for that include the use of prior knowledge to restrict the range of factors, and different techniques for sampling the factors' space (Gilbert 2008).

For those models that are expected to match real-world states, comparisons with empirical data are necessary. These comparisons can analyze the agreement between reality and the model's outputs in qualitative or quantitative terms, depending on the type of model and purpose of the simulation. Thomsen et al. (1999) propose a trajectory of successive validation levels for simulation models based on empirical data and which purpose is to be used prescriptively. The lowest level of this trajectory uses exploratory techniques for validations, including sensitivity analysis, and techniques for checking how well the simulation system can capture and simulate important features of the target system. The highest level concerns the validation of reasoning, representation and usefulness. It verifies how the simulation works in replicating, predicting, and changing the performance of a real system (Thomsen et al. 1999). Four types of experiments comprise this part of the validation trajectory proposed by Thomsen and his colleagues: retrospective, gedanken, natural history, and prospective experiments with interventions.

Retrospective experiments aim to replicate past states of the target system based on retrospective data (Thomsen et al. 1999). They calibrate the model by fine-tuning its parameters in a way that reproduces a state of the system that is known (past or present). To increase the confidence in the model, one can compare the simulation outputs with data about a different known state of the system that was not used in the model's calibration (Fagiolo et al. 2006). In this case, the quality of the model's outputs is assessed through qualitative or quantitative comparison of simulated results with empirical data. For qualitative comparisons the experiment can rely, for instance, on distribution plots of certain attributes of the agent's population. Considering the case of segregation simulations, plots and maps of segregation indices can be prepared for both simulated and empirical data, and be visually compared. For quantitative comparisons, a mathematical statistics approach can be used for validation. During this process, it is important to keep in mind that data deriving from simulations are time series, and therefore, autocorrelated. Thus, statistical procedures for time-series data should be considered. Many authors have presented specific techniques for comparing distributions and time-series data, including Law and Kelton (2000) and Kleijnen (1999).

Based on the retrospective validation, gedanken experiments can be conducted to answer 'what-if' questions. Specific parameter combinations and conditions are set to simulate hypothetical scenarios (what-if), and the outputs are compared with the results expected from theories and/or expert opinions. As in the retrospective experiments, the comparison of results can be quantitative or qualitative, although the latter is much more likely in this case, since theories and expert opinions are often expressed in terms of trends instead of precise values.

Both retrospective and gedanken experiments demonstrate the representational validity of the model, i.e., its ability to simulate salient features of a target system. In addition, they can provide insights about the cause-and-effect relationship between different parameter combinations and the macro-behavior of the system (Thomsen et al. 1999). The remaining types of experiments suggested by Thomsen and his colleagues, i.e., natural history and prospective experiments, test the suitability of the model for supporting decision making. Therefore, while retrospective and gedanken experiments are related to past or present states of the target system, the focus of natural history and prospective experiments switches to future states. Natural-history experiments simulate the future state of a system by keeping the model's factors as they are expected to be in the next years. Prospective intervention experiments attempt to predict alternative futures, i.e., how alternative policies and interventions could influence future states of the target system.

The four experiments described above still reflect a very traditional view of modeling, which attempts to get the present right and then conduct further simulations to predict the future. Thus, this procedure should be considered with reservations when simulating complex systems. Given the challenges imposed by this type of system to reach quantitative predictions, it is important to keep in mind that simulation experiments should be less focused on accurate forecasting, and more oriented towards understanding and structuring debate (Batty and Torrens 2005). The results of simulation runs are more useful when considered in terms of how the different factors of the model (parameters, variables, and module structure) are related, and how they contribute to changes in the behavior of the target system. Such results, usually presented as graphs and statistics, should return to the initial research questions and theories considered for the work. By improving the understanding about the target system, analysis of the simulation results can often provide insights about the implications of current and alternative polices.

4 MASUS: A MULTI-AGENT SIMULATOR FOR URBAN SEGREGATION

Studies focusing on urban segregation have been challenged by the complex features of the phenomenon, which include emergence, self-organization, and non-linearity (Chapter 3). Regarding this problem, this chapter introduces the framework of a multi-agent simulator for urban segregation, called MASUS, which represents the complexity of segregation dynamics and supports further understanding and debate on issues related to the phenomenon. The chapter is subdivided in four main sections. The first presents an overview of the methodological steps for developing the operational MASUS. The second introduces one of these modeling steps, the MASUS conceptual framework. Based on this framework, the third section presents specifications of the MASUS architecture, including its modules and algorithms. Finally, the forth section focuses on the model's implementation level by introducing the MASUS simulation protocol.

4.1 Overview of methodological steps

Based on the methodological protocol for MAS simulations (section 3.5), the process for developing the operational MASUS comprises 10 steps (Figure 4.1):

1. *Problem analysis and objective formulation* (Chapter 1), which follows the theoretical background presented in Chapter 2 and 3.
2. *Conceptual model framework* (section 4.2), which relies on the theoretical and methodological background presented in Chapter 2 and 3.
3. *Theoretical specification* (section 4.3).
4. *Data collection* (Chapter 5), which was conducted in the city of São José dos Campos to serve as basis for the empirical parameterization and simulation experiments using the MASUS model. The data include information about household composition and mobility behavior, as well as characteristics of the different neighborhoods in the city. Since the data availability influences the feasibility of the theoretical specification, steps 3 and 4 took place in a very iterative manner.
5. *Empirical parameterization* (Chapter 5), which comprises statistical models that describe the residential mobility of households (agent interaction) and

dynamics of the urban landscape (environment), including urban sprawl and land market. Since the empirical parameterization often reveals new aspects about the dynamic of the target system, e.g., the relevance of certain environmental characteristics on the behavior of households, it may motivate a review on the theoretical specification of the model.

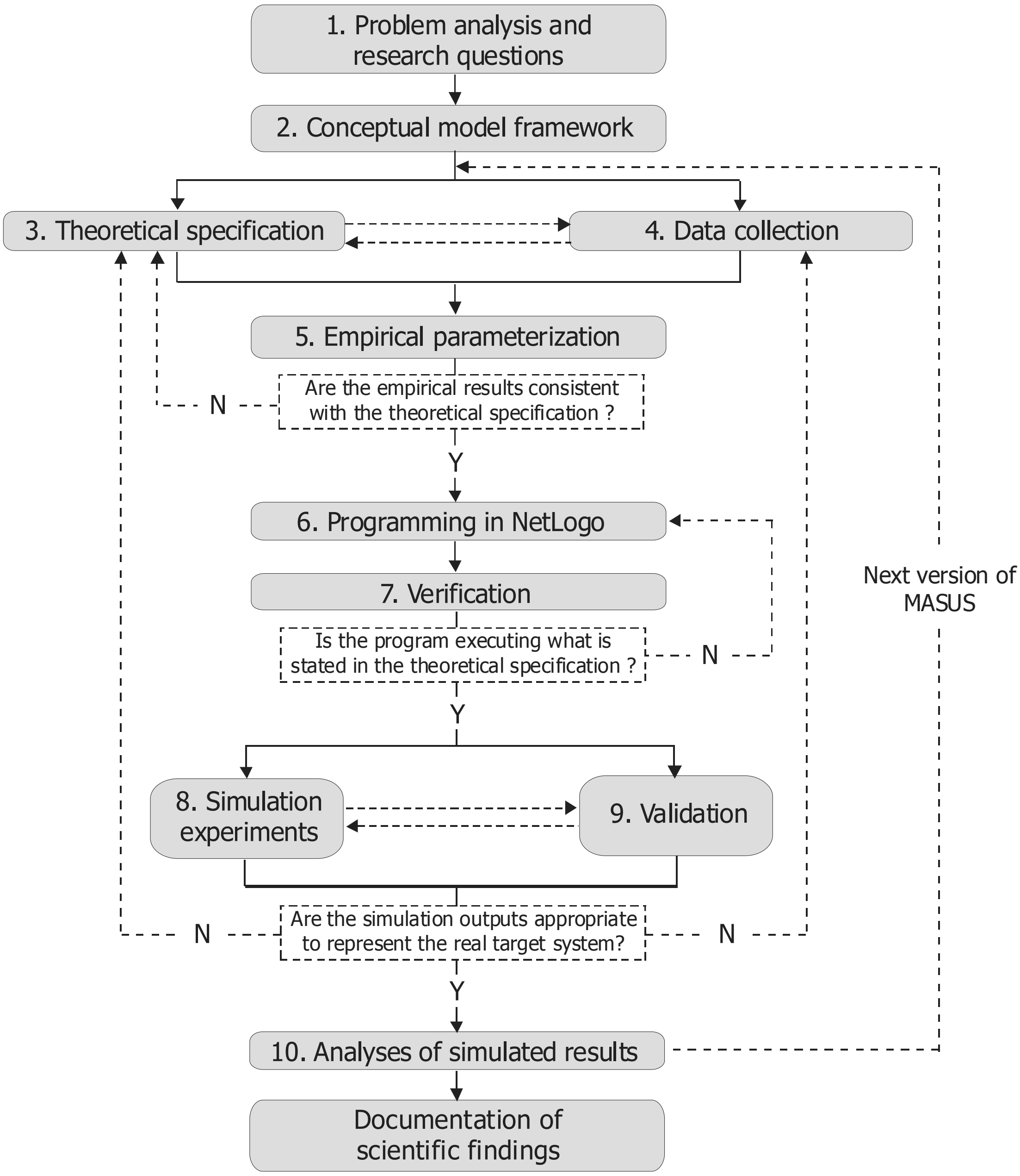


Figure 4.1 Methodological steps for developing the operational MASUS.

6. *Programming in NetLogo* (section 6.1), which converts the theoretical specification into an executable MASUS model using the platform NetLogo (Wilensky 1999).
7. *Verification*, which occurs iteratively with step 6 and consists of checking whether the program really executes what is theoretically specified.
8. *Simulation experiments* (section 6.2), which uses real data to set the initial conditions, and the results from the empirical parameterization are adopted as the basis to set the parameter values.
9. *Validation* (section 6.2), which is performed through experiments that aim at checking whether the simulated outputs are a good representation of the real target system. The validation may include comparisons with empirical data and sensitivity analysis (section 3.5.5). It is not a deterministic process, especially when dealing with complex systems, but it is an important step for defining the confidence level that one should place in the simulation outputs.
10. *Analyses of simulated results* (section 6.2).

4.2 Conceptual MASUS framework for modeling urban segregation

By nature, urban segregation is an emergent system. Its macro-structure arises from the residential choices of many households at the micro-level. For this reason, understanding the factors that influence the residential mobility of households is a key issue in any segregation model that considers the complex nature of the phenomenon. The study of residential mobility has a long tradition. Economists, sociologists, geographers, and psychologists have investigated several decision factors that potentially contribute to the locational behavior of households. These decision factors can be generally classified in four main types:

1. Household attributes, such as size, number of children, income, tenure status (renter or owner), as well as the householder's age, gender, education, and work status. These factors are related to demographic events that influence the households' mobility, like leaving the parental housing, getting married, changing jobs or income, divorce, having children, death of a partner, etc. Cohort is also a factor that influences these demographic events, since attitudes towards marriage, career, and parenthood may change within a

generation (Dieleman 2001; Ettema et al. 2005; Magnusson 2006; Mulder and Hooimeijer 1999).

2. Environment attributes, which include attributes related to different aspects and scales:

2.1. Land and housing attributes, such as price (including taxes), size, type, tenure of dwelling (Clark and Davies Withers 1999; Magnusson 2006; Mulder and Hooimeijer 1999).

2.2. Housing supply and density, which are attributes related to the demand for housing, land-use dynamics, and planning restrictions (Strassmann 2001). These attributes will adjust to reconcile consumer tastes with the existing housing stock at each point of time (McFadden 1977).

2.3. Physical accessibility, which is important to determine the travel costs. It includes the transport network, as well as the housing location with respect to workplaces, commerce, and services (Clark and Davies Withers 1999; Dieleman 2001).

2.4. Environment quality, which includes the availability of infrastructure, public services, green areas, etc. (Borgers and Timmermans 1993).

3. Neighborhood population composition. In societies with a strong stratified structure, residential mobility is influenced by attributes that go beyond physical environment characteristics (Phe and Wakely 2000), e.g., the status attached to a place with a 'desirable' social composition of neighbors.

4. External factors, which include demographic and economic changes, policies at different levels of governance, wealth level and distribution, and tenure structure (Dieleman 2001).

The conceptual framework used as the basis for specifying the MASUS model includes the aforementioned aspects of residential mobility in three distinct components (Figure 4.2):

1. Urban population system, which considers the household attributes and neighborhood population composition.
2. Urban landscape system, which includes the environment attributes.

3. Experimental factors, which considers external factors. These factors are not simulated in the MASUS model, but their features can be considered and modified by the user during simulation experiments.

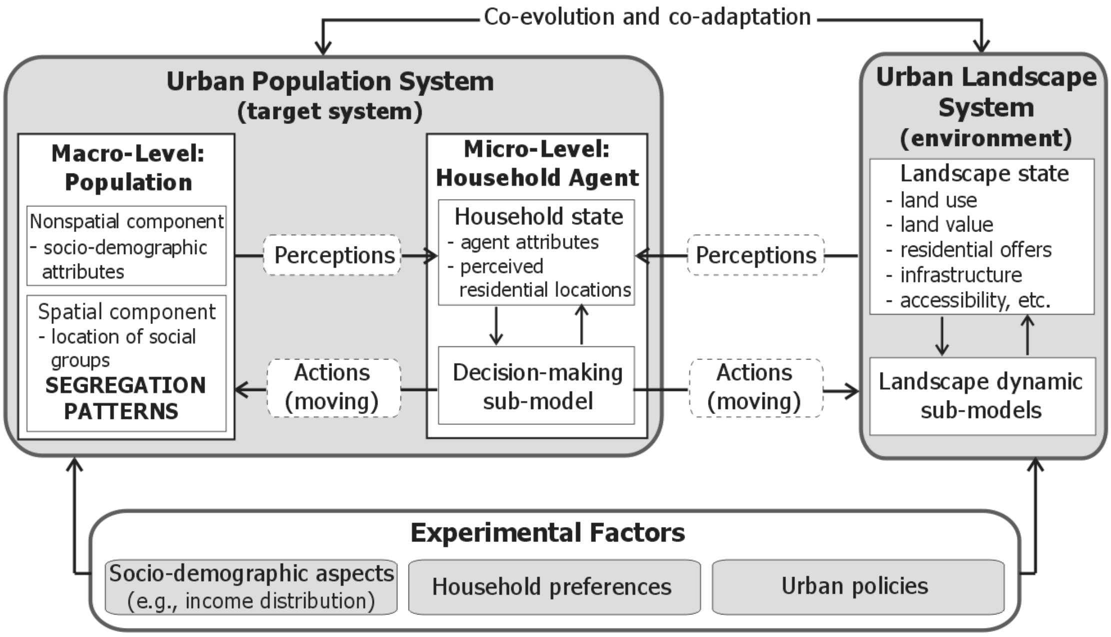


Figure 4.2 The conceptual MASUS framework.

4.2.1 Urban population system

The urban population system is the targeted system of the MASUS model. It represents self-organized processes at both micro- and macro-levels. In general, the micro-level of a system regards heterogeneous elements interacting with each other and their environment. These interactions give rise to global properties at the macro-level of the system, which then feedback to its micro-level. Given the purpose of the MASUS model, the micro- and macro-levels of the urban population system focus on aspects considered relevant for the emergence of a specific global property of this system: the segregation by income.

Based on the above, the heterogeneous elements of the micro-level of the urban population system are the residents of the city, represented by household agents. The household agents have their specific state and autonomy based on their decision-making sub-model. The macro-level of the system represents the urban population in its totality, which is self-organized and has emerged from the activities of household agents over space and time. The urban population is characterized by non-spatial and spatial components. The non-spatial component corresponds to the entire aggregation of household attributes, e.g., the income and education levels of the population as a whole. The spatial component corresponds to the residential location of households belonging to different social groups, i.e., the segregation patterns of the city. The measurement of these segregation patterns corresponds to the output of the MASUS model.

By guiding the households' residential mobility, the decision-making sub-model of household agents represents the main 'engine' of the system. Once a household agent decides to act, i.e., to move to another residential location, it is contributing to a change in the spatial arrangement of social groups in the city and, therefore, to dynamics of segregation (macro-level of urban population system). In addition, the households' decisions also influence certain features of the urban landscape system, like land value and residential offers.

The decision-making sub-model of household agents includes a mechanism that considers, directly or indirectly, the decision factors mentioned in section 4.2: household attributes, environment attributes, neighborhood population composition, and external factors. The locational behavior of a household agent depends on its state, which comprises the household attributes and perceptions about the current and alternative residential locations (Figure 4.2). These perceptions regard the environment attributes and neighborhood population composition of these locations: The first is related to the urban landscape system, while the latter is related to the urban population system itself (macro-level feedback to the micro-level). This means that households' decisions not only influence the urban landscape and the macro-level of the urban population system, but they are also influenced by these factors. The external factors are categorized in the MASUS framework as experimental factors that are able to influence the urban population or the urban landscape system and, therefore, also affect households' decisions in an indirect manner.

4.2.2 Urban landscape system

The urban landscape system represents the environment where household agents are situated and act. It also provides a spatially explicit context for their decisions about whether to move or not. Given the relevance of the urban landscape for the households' decisions, it is conceptualized as a grid of patches or cells, which are simplified agents with their own state and transitional dynamics.

The landscape-patch state is described by a list of spatial variables that are relevant for the households' residential choice, e.g., land value, quality of infrastructure, accessibility, and dwelling offers. The dynamics of landscape attributes occur in parallel with the residential mobility of households. In addition, the set of rules and sub-models driving these dynamics is diversified and must follow a multiple time-scale approach. For instance, the number of dwellings offered in a landscape patch changes according to different processes: (a) a deterministic rule that is applied continuously as households move in or out, and (b) a stochastic sub-model that is applied at the end of each simulation cycle, which simulates the increase in the total number of dwellings due to new developments or the decrease due to the expansion of non-residential land uses. There are also landscape dynamics that operate in a larger time scale, such as accessibility to roads. For simplification purposes, attributes that follow in this category are considered static in the MASUS model or have its features updated by the user during the simulation.

4.2.3 Experimental factors

The experimental factors represent exogenous parameters and input data that can be modified by the user to test theories or policy approaches regarding segregation. Studies focusing on the causes of segregation have emphasized the role of different and complementary mechanisms, including personal preferences, labor market, land and real estate markets, and the controlling power of the state (section 2.4).

The personal preferences are represented as the experimental factor household preferences (Figure 4.2). Their influence on urban segregation can be explored through changes in the parameters of the decision-making sub-model of household agents. For example, the effect of the households' preferences for living in neighborhoods with a high proportion of families belonging to the same social group can be explored by changing the weight of this variable in their decision-making mechanism.

The experimental factor socio-demographic aspects can indirectly represent the effects of the labor market by allowing experiments with different population income levels (average and distribution). This type of experiment can be conducted to test hypotheses relating income inequality and segregation patterns, an issue that causes controversy in scientific debates about segregation (Sabatini 2006). Socio-demographic aspects resulting from general population dynamics, e.g., growth (including migration and natural growth) and aging, are also considered in the model.

Experiments involving the dynamics of land and real estate markets and the controlling power of the state can be conducted through the factor urban policies. The expected results of a variety of regulations for land and real estate markets can be considered in the model to simulate their impacts on the segregation patterns of the city. It is possible to include the results of policies to stimulate the diversification of land uses, control land speculation, regularize clandestine settlements, provide equal access to basic infrastructure, and stimulate the construction of developments for middle and upper classes in poor neighborhoods. Social-mix policies that have been adopted in developed countries for mitigating segregation can also be explored, e.g., policies focusing on the spatial dispersion of poverty or the requirement of mixed occupancy as a prerequisite for approving new developments.

4.3 Theoretical specification of MASUS architecture

Based on the MASUS conceptual framework, this section provides the specification of the model's architecture by representing its elements, internal structure, and interrelations. For each component of the conceptual framework, one module is specified in detail:

1. The URBAN-POPULATION module represents the conceptual system of urban population (target system).
2. The URBAN-LANDSCAPE module represents the conceptual system of urban landscape (environment).
3. EXPERIMENTAL-FACTOR module represents external factors that may influence urban segregation.

4.3.1 URBAN-POPULATION module

The URBAN-POPULATION module is the most important module of the MASUS framework. As the name suggests, it represents the system of urban population and its dynamics. The module is organized in three interrelated levels: household agent

(*HouseholdAgent*), household social group (*HouseholdGroup*), and population (*Population*) (Figure 4.3).

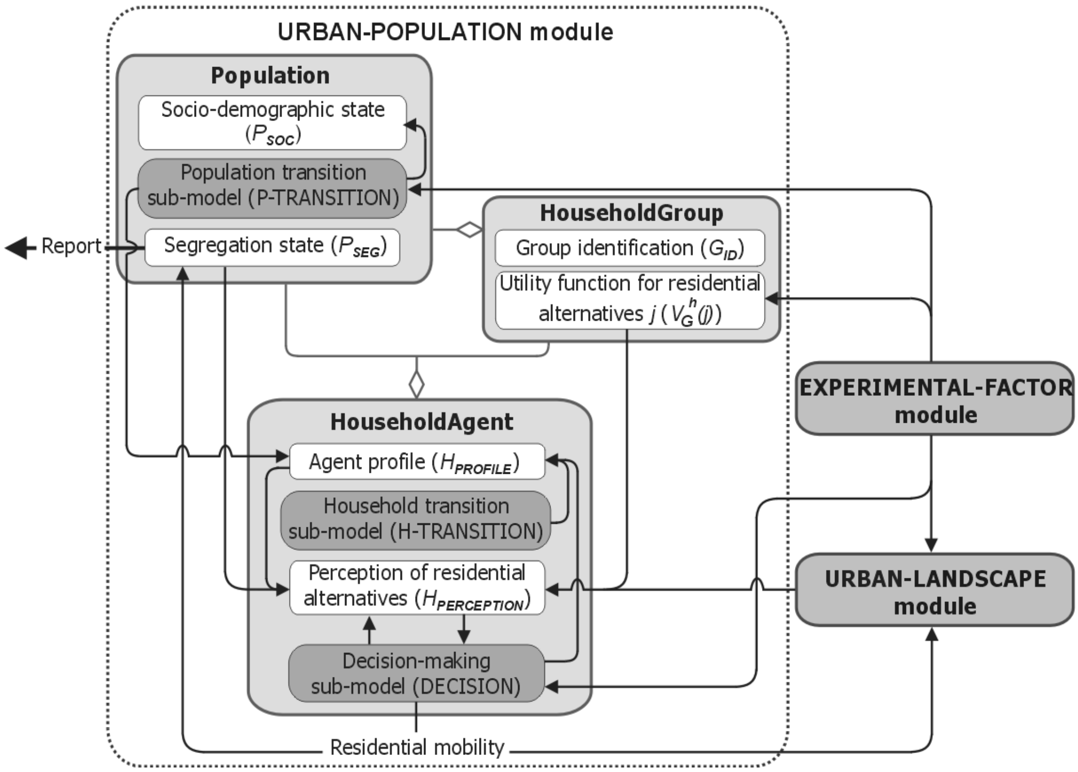


Figure 4.3 Architecture of the URBAN-POPULATION module.

Structure of the household agent (*HouseholdAgent*)

The household agent (named *HouseholdAgent*) represents one or more persons living in a residence. It is the minimal unit of the system of urban population, since household members are not represented. The *HouseholdAgent* structure can be formally expressed as:

$$HouseholdAgent = \{H_{profile}, H_{perception}, \text{H-TRANSITION}, \text{DECISION}\} \quad (4.1)$$

Where: $H_{profile}$ is the agent profile; $H_{perception}$ is the agent's perception about some residential locations in the city, including its own; H-TRANSITION is the household transition sub-model that guides some of the dynamics inherent to the agent's profile (e.g., aging of the head of household); and DECISION is the decision-making sub-model that rules the behavior of the household regarding its residential mobility. The components $H_{profile}$ and $H_{perception}$ constitute the state of the household agent, while the H-TRANSITION and DECISION are internal models of the household agent.

Agent profile ($H_{profile}$)

The agent profile ($H_{profile}$) includes household variables that are relevant to their locational behavior. The relevance of these variables varies according to the empirical context that is being considered and, therefore, their selection should take into consideration the results of statistical analysis about residential mobility in the study area. In the current specification of MASUS, the $H_{profile}$ is expressed as follows:

$$H_{profile} = \{ H_{id}, H_{inc}, H_{edu}, H_{age}, H_{group}, H_{size}, H_{kids}, H_{tenure}, H_{location} \} \quad (4.2)$$

Where: H_{id} is the identification code; H_{inc}, H_{edu}, and H_{age} are head of household's variables that indicate income, education, and age, respectively; H_{group} is the identification of the household's social group, which is defined by the income (H_{inc}); H_{size} is the household's size; and H_{kids} is a binary variable to indicate the presence of children; H_{tenure} is the tenure status (renter or owner); and the variable $H_{location}$ indicates the place where the household is located.

With the exception of H_{id}, all the variables of the agent profile are dynamic. The dynamics of the variables H_{inc}, H_{edu}, H_{age}, H_{group}, H_{size}, H_{kids}, and H_{tenure} are ruled by the household transition sub-model (H-TRANSITION) and the population transition sub-model (P-TRANSITION). The sub-model DECISION drives the household's residential mobility and, therefore, the dynamics of their residential location ($H_{location}$) and tenure status (H_{tenure}).

Household transition sub-model (H-TRANSITION)

The household transition sub-model (H-TRANSITION) is an internal model of the household agent that represents natural dynamics of its profile, such as aging of the head of household. It consists of a set of rule-based functions:

$$\text{H-TRANSITION} = \{ F_{H\text{-}age} , F_{H\text{-}inc+}, F_{H\text{-}inc-}, F_{H\text{-}kids}, F_{H\text{-}dissolve} \} \tag{4.3}$$

Where: $F_{H\text{-}age}$, $F_{H\text{-}inc+}$, $F_{H\text{-}inc-}$, and $F_{H\text{-}kids}$ are functions performing dynamics of the household's variables H_{age}, $H_{inc,}$ and H_{kids}. Indirectly, these functions can also change the values of the variables H_{group} (by changing the variable $H_{inc,}$) and H_{size} (by changing the variable H_{kids}). $F_{H\text{-}dissolve}$ is a function that 'dissolves' the household agent and allows to represent, for instance, households moving to another city.

The function $F_{H\text{-}age}$ adds 1 year to the age of the household head after each time step. It is represented as follows:

$$F_{H-age} = {}^{t+1}H_{age} = {}^{t}H_{age} + 1 \tag{4.4}$$

The functions $F_{H\text{-}inc+}$ and $F_{H\text{-}inc-}$ simulate social mobility by adding a probability of increasing or decreasing the household's income, which is measured in minimum wages. They are expressed as:

$$F_{H-inc+} = {}^{t+1}H_{inc} = \begin{cases} {}^{t}H_{inc} + 1, & \text{if } q \le \theta_{inc+} \\ {}^{t}H_{inc}, & \text{otherwise} \end{cases} \tag{4.5}$$

$$F_{H-inc-} = {}^{t+1}H_{inc} = \begin{cases} {}^{t}H_{inc} - 1, & \text{if } q \le \theta_{inc-} \\ {}^{t}H_{inc}, & \text{otherwise} \end{cases} \tag{4.6}$$

Where: q is a random number distributed evenly over [0,1]. θ_{inc+} and θ_{inc-} are values within [0,1] representing the chance for a head of household to have its income increased/decreased by 1 minimum wage. Depending on the new value of H_{inc}, the household's social group (H_{group}) may also change.

The function $F_{H\text{-}kids}$ simulates changes in the household variable 'presence of kids' (H_{kids}). Households with 2 or more people, young head of household, and no children have a chance θ_{kids+} to change their H_{kids} status. On the other hand, households with children and a senior head of household have a chance θ_{kids-} to change their H_{kids} status, as these children are probably getting older and leaving the parental housing. The function F_{H-kids} is formally expressed as:

$$F_{H-kids} = {}^{t+1}H_{kids} = \begin{cases} 1, & if({}^{t}H_{kids} = 0 \quad and \quad {}^{t}H_{size} > 1 \quad and \quad {}^{t}H_{age} < m \quad and \quad q \leq \theta_{kids+}) \\ 0, & if({}^{t}H_{kids} = 1 \quad and \quad {}^{t}H_{age} > n \quad and \quad q \leq \theta_{kids-}) \\ {}^{t}H_{kids}, & otherwise \end{cases} \tag{4.7}$$

Where: q is a random number distributed evenly over [0,1]; m is the limit age for a head of household to be considered 'young'; n is the minimal age for a head of household to be considered 'senior'; θ_{kids+} and θ_{kids-} are values within [0,1] representing the chance of a household of having its H_{kids} status modified to 1 and 0, respectively. Once this modification occurs, the household size (H_{size}) automatically changes.

The function $F_{H\text{-}dissolve}$ 'dissolves' household agents that changed their characteristics due to demographic events that are not simulated by MASUS (e.g, divorce, death of a member, moving out of the city, etc.). To represent these cases, household agents have a chance $\theta_{dissolve}$ of being dissolved, and new agents are created by the population transition sub-model (P-TRANSITION). The $F_{H\text{-}dissolve}$ is formally expressed as:

$$F_{H-dissolve} = \{Agent\} = \begin{cases} \{\ \}, & if \quad q \leq \theta_{dissolve} \\ \{Agent\}, & otherwise \end{cases} \tag{4.8}$$

Where: *{Agent}* is a set containing one household agent; q is a random number distributed evenly over [0,1]; $\theta_{dissolve}$ is a value within [0,1] representing the chance of a household agent of being dissolved. The value $\theta_{dissolve}$ can be multiplied by a factor α to increase/decrease the dissolution probability of certain groups of households. Households headed by older people, for example, have a higher chance of being subjected to structural changes, since their heads are more likely to be replaced or, in the case of small households, to move to another place like a nursing home or a relative's house.

Decision-making sub-model (DECISION)

The decision-making sub-model (DECISION) is an internal mechanism of the *HouseholdAgent* that guides the agent's decision and action regarding its residential location. In this sub-model, the household chooses between alternatives like:

1. Stay in current location;
2. Move within the same neighborhood;
3. Move to a neighborhood that is similar to the original one, e.g., from one poor irregular settlement to another (*n* locations are randomly selected);
4. Move to a different type of neighborhood, e.g., from a socially diverse neighborhood to a gated settlement with a high concentration of affluent households (*m* locations are randomly selected).

Residential locations are represented as landscape patches of 100 m × 100 m, while neighborhoods are represented as sets of landscape patches corresponding to census tracts. Neighborhoods are classified in four different types: (1) neighborhood with a high concentration of affluent households, (2) socially diverse neighborhood, (3) neighborhood with a high concentration of low-income households, and (4) a neighborhood type that is similar to (3), but covers clandestine settlements, like *favelas* and irregular *periferias* (section 2.2). More details about the classification of different types of neighborhoods are presented in section 5.2.2.

The household agent's decision-making sub-model executes the following main steps (see Figure 4.4): (1) select a set of residential alternatives, (2) compute the household's perception of residential alternatives ($H_{perception}$), (3) compute the household's probability to choose each alternative, (4) choose a residential alternative, (5) perform action (move or stay), and (6) update agent profile and urban landscape.

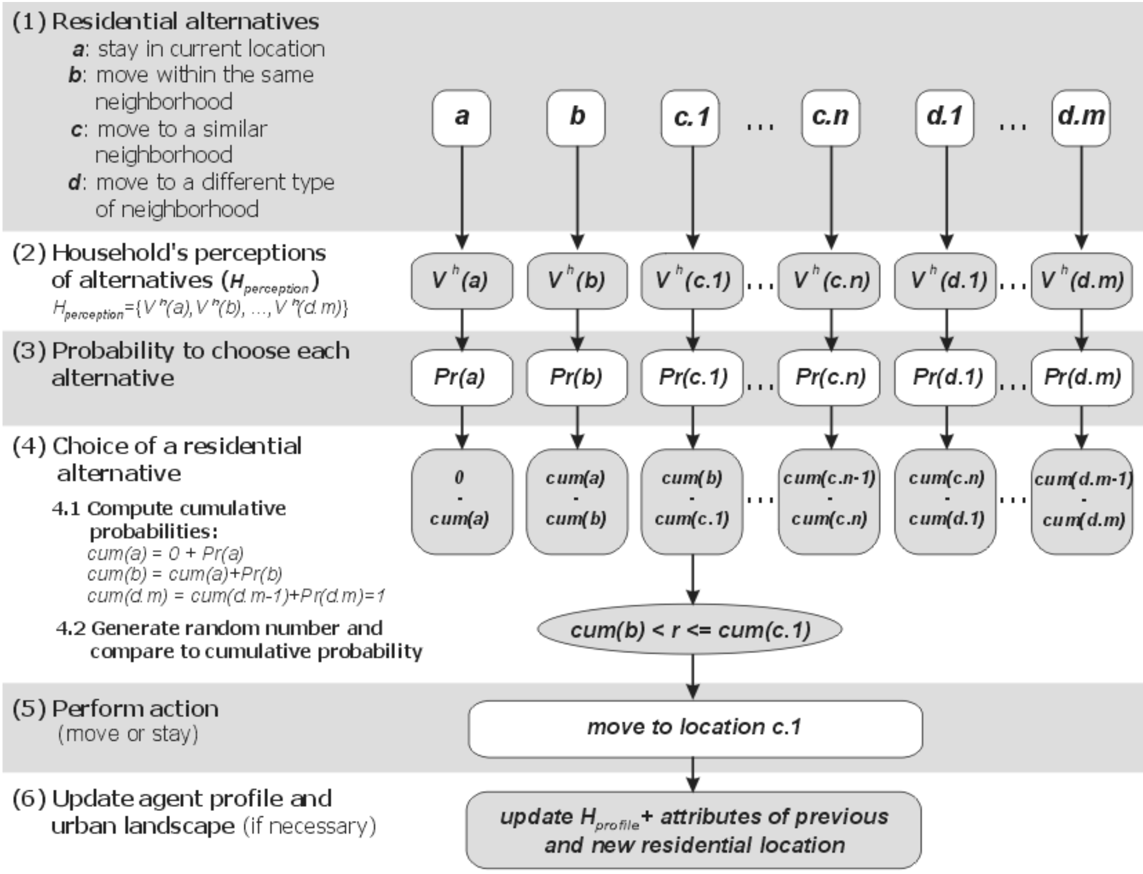


Figure 4.4 Main steps of the household agent's decision-making sub-model (DECISION).

In the first step, *select a set of residential alternatives*, the household agent chooses locations from a valid set, which excludes places without available dwellings. Because the model assumes that agents can evaluate the possibility of living in any neighborhood of the city, the selection imposes no restriction regarding the distance between the alternative and the household's current location. This modeling decision takes into consideration the fact that real households can acquire knowledge about many neighborhoods – including some in further locations - through their social contacts or other information source (e.g., newspapers). Given a set of valid locations, the household agent *h* living in the location *a* neighborhood type μ ($L_{neigh}= \mu$) selects a choice set that consists of (a) location *a* (not move), (b) location *b* within the same neighborhood, (c) *n* randomly selected locations in neighborhoods with $L_{neigh}= \mu$ (same type of neighborhood), and (d) *m* randomly selected locations in neighborhoods with $L_{neigh}\neq \mu$ (other type of neighborhood).

The second step, *compute the household's perception of residential alternatives* ($H_{perception}$), consists of obtaining the household's utility for each selected residential alternative *j* ($V^h(j)$). The function $V^h(j)$ is a nested logit utility function that considers the household's attributes ($H_{profile}$), the environment attributes of alternative *j* (L_{state}), and the neighborhood population composition of alternative *j* (P_{seg}). The utility function and its reference parameters are obtained from the estimation of a 3-level nested logit model (Greene 2000), which jointly models household's mobility choice (first level: stay or move), neighborhood type choice (second level), and residential location choice (third level). More details about this nested logit model are provided in Chapter 5. Assuming that there is a structural difference in the way that households from different social groups evaluate residential locations, the utility function and its parameters are group specific and, therefore, provided by the level *household social group (HouseholdGroup).*

After obtaining the utilities for all the selected alternatives ($H_{perception}$), it is possible to perform the third step, which consists of *computing the household's probability to choose each alternative* (see Chapter 5). The fourth step of the household agent's decision-making sub-model is to *choose a residential alternative*. For that, the sub-model generates a random number between 0 and 1, and compares it to the cumulative probabilities of each residential alternative. The chosen alternative is the one that has the cumulative probability interval which contains the random number (see Figure 4.4). After that, the household agent *performs the action* that corresponds to the chosen alternative, i.e., changes or keeps the household location ($H_{location}$).

Finally, in case the performed action involves moving to a new location, the last step of the DECISION sub-model is to *update the agent profile and urban landscape state of its previous and new residential locations (patches).* Once a household moves, it changes its location ($H_{location}$). The household may also change its tenure status (H_{tenure}), considering that this type of change is usually related to residential mobility. The household location ($H_{location}$) corresponds to the identification of the patch where the household lives, and changes automatically as the agent move to a new location. The dynamics in the household's tenure status are guided by the rule-based function $F_{H\text{-}tenure}$, which is only applied if ${}^{t+1}H_{location} \neq {}^{t}H_{location}$. According to this function, renters have a probability θ_{owner} of moving to an owned dwelling (renters becoming owners), and owners have a probability θ_{renter} of moving to a rented dwelling. The $F_{H\text{-}tenure}$ is formally expressed as:

$$F_{H-tenure} = {}^{t+1}H_{tenure} = \begin{cases} 1 \quad (renter), & if\,(H_{tenure} = 0 \quad and \quad q \le \theta_{renter}) \\ 0 \quad (owner), & if\,(H_{tenure} = 1 \quad and \quad q \le \theta_{owner}) \\ {}^{t}H_{tenure}, & otherwise \end{cases} \tag{4.9}$$

Where: q is a random number evenly distributed over [0,1], θ_{renter} and θ_{owner} are values within [0,1] representing the chance that a household becomes a renter or owner, respectively.

A household that moves modifies not only its profile and the spatial arrangement of the population, both part of the system of human population, but also the dwelling offers in its previous and new residential locations (L_{offers}), which is an attribute of landscape patches (system of urban landscape). The function $F_{L\text{-}offers}$ simulates this change by adding 1 dwelling offer to the previous location and subtracting 1 dwelling offer to the new location. Formally, the $F_{L\text{-}offers}$ is expressed as:

$$F_{L-offers} = {}^{t+1}L_{offers} = \begin{cases} {}^{t}L_{offers} - 1, & if\,(L_{id} = {}^{t}H_{location}) \\ {}^{t}L_{offers} + 1, & if\,(L_{id} = {}^{t+1}H_{location}) \\ {}^{t}L_{offers}, & otherwise \end{cases} \tag{4.10}$$

Where: L_{offers} is the number of offers in the landscape patch, L_{id} is the identification of the landscape patch, and $H_{location}$ is the identification of the landscape patch where the household is located.

Structure of the household social group (*HouseholdGroup*)

The household social group (named *HouseholdGroup*) is a collection of household agents that belong to the same social group and adopt the same criteria to evaluate their residential alternatives. In the current MASUS version, social groups are defined by the income of the head of household. The formal expression of the *HouseholdGroup* is:

$$HouseholdGroup = \{G_{id},\ V_G^h(j)\} \tag{4.11}$$

Where: G_{id} is the group identification that matches with the H_{group} stored in the household profile. $V_G^h(j)$ is the nested logit utility function for residential alternatives that represents the household group criteria for evaluating residential alternatives. The specification and estimated parameters of the nested logit utility function for each household social group are presented in Chapter 5.

Structure of the population (*Population*)

The *Population* class is the collection of all household agents and represents the macro-level of the system of urban population. It is formally expressed as:

$$Population = \{P_{soc},\ \text{P-TRANSITION},\ P_{seg}\} \tag{4.12}$$

Where: P_{soc} is the socio-demographic state of the population, including its size and the total population composition in terms of the variables that comprise the household's profile. P-TRANSITION is the population transition sub-model responsible for keeping the socio-demographic state of the population according to expected levels provided by the user. P_{seg} is the segregation state of the population, which represents the spatial arrangement of household social groups. P_{soc} and P_{seg} are the non-spatial and spatial component of the population state, respectively.

Socio-demographic state of the population (P_{soc})

The socio-demographic state of the population (P_{soc}) corresponds to the non-spatial characteristics of the population as a whole. It consists of the procedure $F_{pop\text{-}stat}$, which computes and plots basic statistics of the population evolving over time, including: (a) total number of households, (b) total number of households belonging to each social group, (c) average of monthly income, and (d) income distribution (Gini index and Lorenz curve).

Population transition sub-model (P-TRANSITION)

Since demographic prediction is not among the purposes of the MASUS model, the size and the socio-demographic composition of the population regarding the variables that comprise the household's profile (income, education, size, age, children, and tenure status) follow annual control values that are provided by the user. The population transition sub-model (P-TRANSITION) is responsible for creating households with profiles that meet the expected (user-specified) socio-demographic composition of the population as a whole. The P-TRANSITION consists of a set of procedures:

$$\text{P-TRANSITION} = \{F_{create}, F_{P\text{-}inc}, F_{P\text{-}edu}, F_{P\text{-}size}, F_{P\text{-}age}, F_{P\text{-}kids}, F_{P\text{-}tenure}\} \tag{4.13}$$

Where: F_{create}, $F_{P\text{-}inc}$, $F_{P\text{-}edu}$, $F_{P\text{-}size}$, $F_{P\text{-}age}$, $F_{P\text{-}kids}$, $F_{P\text{-}tenure}$ are functions controlling the total characteristics of the population in relation to size, income, education, household size, head of household age, presence of children, and tenure status.

The F_{create} is a procedure that creates agents in order to achieve the number of households expected in the next year (${}^{t+1}P_{total}$), which is equivalent to:

$${}^{t+1}P_{total} = {}^{t}P_{total} + ({}^{t}P_{total} * \alpha_{growth}) \tag{4.14}$$

Where: P_{total} is the total population, and α_{growth} is the user-defined annual growth rate. The procedure F_{create} creates (${}^{t+1}P_{total}$ - ${}^{t}P_{total}$) new household agents and locates them in a set M. These household agents belonging to the set M have their $H_{profile}$ defined by the functions $F_{P\text{-}inc}$, $F_{P\text{-}edu}$, $F_{P\text{-}size}$, $F_{P\text{-}age}$, $F_{P\text{-}kids}$, $F_{P\text{-}tenure}$, and choose their location according to their internal decision-making sub-model (DECISION).

Since there is no information about the previous residential location of the new households, their DECISION sub-model does not include the residential options 'stay in the current location', 'move within the same neighborhood', or 'move to a similar neighborhood'. Instead, n alternatives are randomly chosen in the city and evaluated under the same conditions, i.e., following the group-specific criteria that match the group identification (H_{group}) of the new household agent.

The $F_{P\text{-}inc}$ is a procedure that controls the income composition of the population as a whole. First, it computes the total population of each income group i (${}^{t+1}P_{inc(i)}$) according to user-defined group proportions ($\tau_{inc(i)}$), and then stipulates the number of households that should be added to the group i ($P_{new\text{-}inc(i)}$):

$$ {}^{t+1}P_{inc(i)} = {}^{t+1}P_{total} * \tau_{inc(i)} \tag{4.15} $$

$$ P_{new\text{-}inc(i)} = {}^{t+1}P_{inc(i)} - {}^{t}P_{inc(i)} \tag{4.16} $$

Afterwards, the totals $P_{new\text{-}inc(i)}$ are used to generate the H_{inc} of the household agents located in the set M. For instance, if the total $P_{new\text{-}inc(i)}$ computed for the income group '10 to 20 minimum wages' is equal to n, it means a H_{inc} between 10 and 20 minimum wages will be addressed to n households selected from the set M. In case the value $P_{new\text{-}inc(i)}$ is higher than the number of households in set M, the sub-model will meet the total expected for the income group i by changing the H_{inc} of households that are already located in the city and belong to a similar income group. In case the value $P_{new\text{-}inc(i)}$ is negative, the sub-model selects $|P_{new\text{-}inc(i)}|$ households belonging to the income group i and changes their H_{inc} to values within the income interval of other groups. Since the household's social group (H_{group}) is defined by the income, this variable is updated as soon as a new H_{inc} is attributed to the agent.

Following the same logic of the $F_{P\text{-}inc}$, the P-TRANSITION sub-model executes other procedures to control the population composition in relation to the variables education of the head of household ($F_{P\text{-}edu}$), household size ($F_{P\text{-}size}$), age of the head of household ($F_{P\text{-}age}$), presence of children ($F_{P\text{-}kids}$), and tenure status ($F_{P\text{-}tenure}$). All these procedures rely on annual distributions (or proportions) provided by the user. Because descriptive statistics based on empirical data demonstrated that the distribution of variables like education and household size varies according to income groups (e.g., low-income households usually have a lower education level and larger number of members), the user can provide differentiated proportion values for these variables per income group.

Segregation state of the population (P_{seg})

The segregation state of the population (P_{seg}) comprises segregation indices that are used to evaluate and analyze the simulation experiments. In Chapter 2, the importance of analyzing different dimensions and scales of segregation was enphasized. Considering that, the segregation state of the population is formally expressed as:

$$P_{seg} = \{ \left(\breve{D}(m), \breve{d}(m), \breve{Q}_m, \breve{q}_m\right)_{bw=\psi}, \left(\breve{D}(m), \breve{d}(m), \breve{Q}_m, \breve{q}_m\right)_{bw=\chi} \} \tag{4.17}$$

Where: $\breve{D}(m)$ and $\breve{d}(m)$ are the global and local version of the generalized spatial dissimilarity index, respectively, which measure the segregation dimension evenness/clustering (see Chapter 2); $\breve{Q}_m$ and $\breve{q}_m$ are the global and local version of the spatial isolation index, respectively, which measures the segregation dimension exposure/isolation; and $bw=\psi$ and $bw=\chi$ are the bandwidths of the moving windows used to compute the segregation indices in different scales.

The global indices $\breve{D}(m)$ and $\breve{Q}_m$ are displayed as values between 0 and 1, which summarize the segregation degree of the whole city. The local indices $\breve{d}(m)$ and $\breve{q}_m$ show how the different localities contribute to the global indices, and they can be presented as segregation maps (see section 2.5).

4.3.2 URBAN-LANDSCAPE module

The URBAN-LANDSCAPE module represents the urban landscape system (environment) and its dynamics (Figure 4.5). It plays an important role in the MASUS framework, since household agents are not only situated within the landscape, but also make decisions about it (where to live), based on the landscape's characteristics and their personal profiles.

The URBAN-LANDSCAPE module is organized in two interrelated levels: entire landscape (*EntireLandscape*), and landscape patch (*LandscapePatch*). The *EntireLandscape* represents the macro-level of the urban landscape system. It plays a limited role in the model, since the agent's decisions do not consider this landscape level. This is because agents can only access information about some landscape portions (patches). Nevertheless, the state of the entire landscape (EL_{state}) provides information for the computation of global variables (EL_{global}) that are necessary for simulating the dynamics of some attributes of the landscape patches.

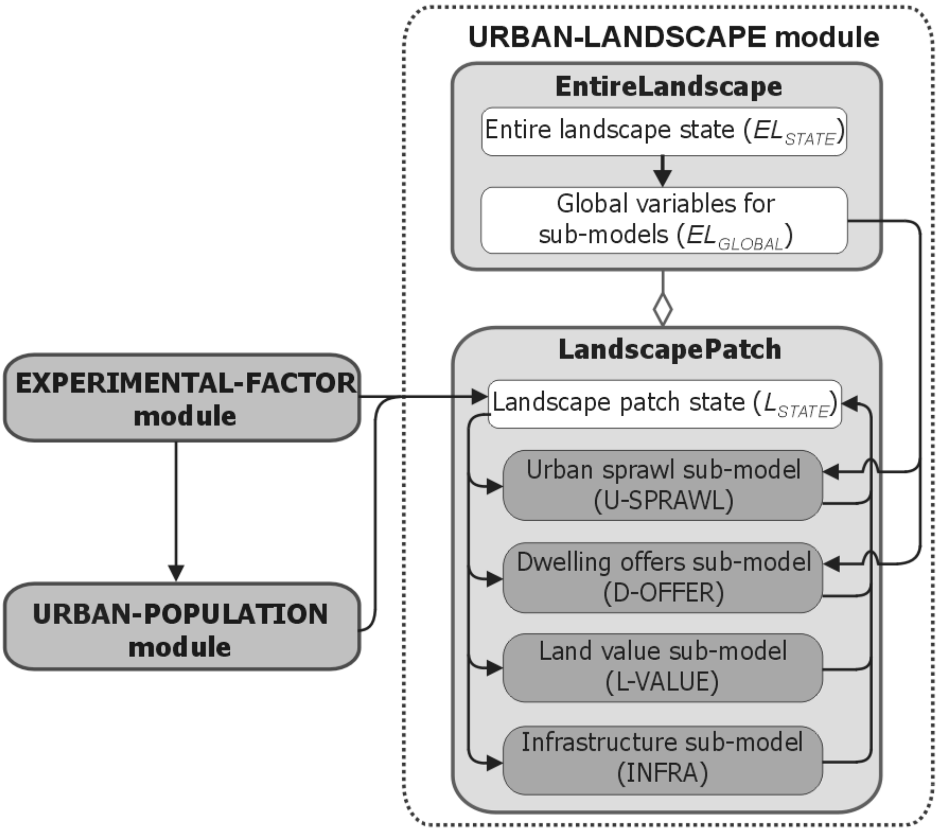


Figure 4.5 Architecture of the URBAN-LANDSCAPE module.

Structure of the landscape patch (*LandscapePatch*)

The *LandscapePatch* represents the micro-level of the urban landscape system. It is a portion of the environment measuring 100 m by 100 m, which corresponds to the minimal unit of the urban landscape system. Its structure can be formally expressed as:

$$LandscapePatch = \{L_{state}, \text{U-SPRAWL, D-OFFER, L-VALUE, INFRA}\} \quad (4.18)$$

Where: L_{state} is the landscape-patch state; U-SPRAWL, D-OFFER, L-VALUE, and INFRA are sub-models for urban sprawl, dwelling offers, land value, and infrastructure, respectively.

Landscape-patch state (L_{state})

The landscape-patch state includes environment variables that are relevant, directly or indirectly, to the locational behavior of households. For this reason, the selection of these variables depends on the results of statistical analysis about residential mobility in the study area. In the current specification of MASUS, the L_{state} is formally expressed as:

$$L_{state} = \{ L_{physical}, L_{access}, L_{zoning}, L_{market}, L_{U\text{-}SPRAWL}, L_{D\text{-}OFFER}, L_{L\text{-}VALUE}, L_{INFRA} \} \quad (4.19)$$

Where: $L_{physical}$ is a set of variables related to physical aspects of the landscape patch; L_{access} is a set of variables related to the accessibility of the patch; L_{zoning} is a set of variables related to the zoning legislation; and L_{market} is a set of variables related to the real-estate market. $L_{U\text{-}SPRAWL}$, $L_{D\text{-}OFFER}$, $L_{L\text{-}VALUE}$, and L_{INFRA} are sets of variables that are exclusively relevant to the sub-models U-SPRAWL, D-OFFER, L-VALUE, and INFRA, respectively.

The set $L_{physical}$ consists of the binary variable *urban use* (L_{urban} : 1=urban and 0=not urban), *terrain slope* (L_{slope}), *infrastructure* (L_{infra}), and *type of neighborhood* (L_{neigh}). The dynamics of the variables L_{urban} and L_{infra} are ruled by the sub-models U-SPRAWL and INFRA, respectively. The variable set $L_{physical}$ is formally expressed as:

$$L_{physical} = \{ L_{urban}, L_{slope} , L_{infra}, L_{neigh} \} \quad (4.20)$$

The set L_{access} consists of the variables distance to the Central Business District ($L_{d\text{-}CBD}$) and distance to roads ($L_{d\text{-}roads}$):

$$L_{access} = \{ L_{d\text{-}CBD}, L_{d\text{-}roads} \} \quad (4.21)$$

The set L_{zoning} consists of the variable *floor area ratio* (L_{FAR}), which is the limit imposed for the ratio between the total floor area of buildings and the size of the land, and the binary variables *zone of residential use* (L_{res}), *zone of mixed use* (L_{mixed}), *central zone* ($L_{central}$), *zone of industrial transition* ($L_{ind\text{-}tran}$), *zone of social interest* (L_{social}, e.g., social housing projects), *zone of non-residential use* ($L_{non\text{-}res}$). The variable *zone of non-*

residential use includes areas that are either protected, for industrial use only, or for aero activities. The variable set L_{zoning} is formally represented as:

$$L_{zoning} = \{ L_{FAR}, L_{res}, L_{mixed}, L_{central}, L_{ind\text{-}tran}, L_{social}, L_{non\text{-}res} \} \tag{4.22}$$

The set L_{market} consists of the variables *land value* (L_{value}), *total number of dwellings* (L_{dwe}) and *dwelling offers* ($L_{dwe\text{-}offer}$). The dynamics of these variables are guided by the sub-models L-VALUE and D-OFFER, respectively. The variable set L_{market} is formally expressed as:

$$L_{market} = \{ L_{value}, L_{dwe}, L_{dwe\text{-}offer} \} \tag{4.23}$$

Urban sprawl sub-model (U-SPRAWL)

The U-SPRAWL sub-model simulates the expansion of the city's urbanized areas. It comprises two phases: (1) the *transition phase* ('how many?'), which quantifies the sprawl that is expected to occur during the period $t \rightarrow t+1$, i.e., how many patches will convert their use from non-urban to urban, and (2) the *allocation phase* ('where?'), which indicates the location of the new urban patches.

The transition phase relies on the Markov chain to assess the total number of patches converting their use from non urban to urban during the time interval $t \rightarrow t+1$. The Markov chain is a mathematical model for describing a certain type of process that moves in a sequence of steps through a set of states (Lambin 1994). The central mechanism of a Markov chain is a probability P_{ij}, which refers to the transition from a state i to a state j in a given time interval (Brown, 1970). For the U-SPRAWL model, the state of the system is defined by the number of patches that are urbanized/not-urbanized. The Markov model can be expressed, in matrix notation, as (Baker, 1989):

$$\Pi(t+1) = P^{n} . \Pi(t) \tag{4.24}$$

Where: $\Pi(t)$ is a column vector, with k elements, representing the fraction of land area in each of the s states at time t. For the U-SPRAWL model, the $\Pi(t)$ has two elements, one is the number of urban patches, and the other is the number of non-urban patches at time t. $\Pi(t+1)$ is a column vector showing the fraction of occupation of s states at time $t+1$, P is a matrix whose elements are global transition probabilities P_{ij}, accounting for the probability of a certain patch to change from state *i* to *j* during the time interval $t \rightarrow t+1$, and *n* is the number of time steps between *t* and *t*+1. For example, if *n* corresponds to one year, then *n* would be 10 if the addition in time corresponds to 10 years.

The global transition probabilities P_{ij} can be statistically estimated from a sample of transitions occurring during a certain time interval. Given a_{ij} indicating transitions between pairs of states over a time interval, the transition probabilities P_{ij} are estimated as:

$$P_{ij} = a_{ij} / \sum a_{ij} \tag{4.25}$$

In the particular case of the U-SPRAWL sub-model, it is necessary to compute the global transition probability $P_{NU \rightarrow U}$, which accounts for the probability of a patch to change from 'non-urban' (NU) to 'urban' (U). The global transition probability $P_{NU \rightarrow U}$ is stored as a global variable of the *EntireLandscape* (EL_{global}). Following the Markov model equation (4.24), $P_{NU \rightarrow U}$ is used to compute the total number of new urban patches during the period $t \rightarrow t+1$ ($T_{urban-new}$). $T_{urban-new}$ is also stored as a global variable of the *EntireLandscape* and retrieved later during the allocation phase.

Once the number of new urban patches is known ($T_{urban-new}$ = 'how many?'), the *allocation phase* is responsible for indicating which non-urban patches convert their use to urban during the period $t \rightarrow t+1$ ('where?') and, therefore, for updating the landscape patch variable 'urban use' (L_{urban}). The allocation phase relies on binary logistic regression to compute the local transition probability of a non-urban patch becoming urban ($p_{NU \rightarrow U}$). Binary logistic regression is a type of regression analysis where the dependent variable is a dummy variable (e.g., 1=urban and 0=not urban). The statistical model for logistic regression is (Moore and McCabe 2003):

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_n X_n \tag{4.26}$$

Where: p is the probability that the event Y occurs (e.g., convert the land use to urban), X is the explanatory variable, and β is the logistic model parameter.

The local transition probability of a non-urban patch becoming urban ($p_{NU \to U}$) can be estimated as follows:

$$p_{NU \to U} = 1 \Big/ 1 + \exp(-\beta_0 - \beta_1 X_1 - \beta_n X_n) \tag{4.27}$$

Where: the explanatory variables X represent one of the variable sets of the landscape-patch state ($L_{U\text{-}SPRAWL}$), and β represents parameters estimated from empirical data (see Chapter 5).

The U-SPRAWL sub-model computes $p_{NU \to U}$ for all non-urban patches (L_{urban}=0). Afterwards, it ranks these patches according to their $p_{NU \to U}$ in a decreasing order. Following this rank, the rule-based procedure $F_{NU \to U}$ is executed for each non-urban patch until the total number of new urban patches reaches the value $T_{urban-new}$ (computed in the transitional phase). The $F_{NU \to U}$ is formally expressed as:

$$F_{NU \to U} = {}^{t+1}L_{urban} = \begin{cases} 1 \ \ (urban), & if \quad q \leq p_{NU \to U} \\ {}^{t}L_{urban}\,(non-urban), & otherwise \end{cases} \tag{4.28}$$

Where: q is a random number evenly distributed over [0,1].

Dwelling offers sub-model (D-OFFER)

The D-OFFER sub-model simulates the dynamics of the landscape-patch variables *total number of dwellings* (L_{dwe}) and *dwelling offers* ($L_{dwe\text{-}offer}$). It simulates the gain of dwellings in some areas of the city promoted by new residential developments, and the loss of dwellings in other parts of the city due to the progression of non-residential uses in certain neighborhoods. Like the U-SPRAWL model, it also comprises a *transition phase* and an *allocation phase.* The transition phase ('how many?') quantifies the overall gain and loss of dwelling offers during the period $t \rightarrow t+1$, while the allocation phase ('where?') indicates which patches gained and lost dwellings during the same period.

The transition phase determines the total number of dwellings in the city (T_{dwe}), the total dwelling loss due to the progress of non-residential use ($T_{dwe\text{-}loss}$), and the total number of new dwellings ($T_{dwe\text{-}gain}$). These values are stored as global control variables of the *EntireLandscape* (EL_{global}), and retrieved later during the allocation phase. The total number of dwellings in the city (T_{dwe}) is formally expressed as:

$$T_{dwe} = T_{dwe\text{-}occup} + (T_{dwe\text{-}occup} * \tau_{stock}), \tag{4.29}$$

Where: $T_{dwe\text{-}occup}$ is the total number of occupied dwellings in the city ($T_{dwe\text{-}occup}$), which is also equal to the total number of households; ($T_{dwe\text{-}occup} * \tau_{stock}$) represents the housing stock of the city, which is equivalent to a proportion τ_{stock} of the occupied dwellings.

The total dwelling loss due to the progress of non-residential use ($T_{dwe\text{-}loss}$) during the period $t \rightarrow t+1$ is equal to a proportion τ_{loss} of the total number of dwellings (T_{dwe}):

$$T_{dwe\text{-}loss} = {}^{t+1}T_{dwe} * \tau_{loss}, \tag{4.30}$$

The total number of new dwellings ($T_{dwe\text{-}gain}$) during the period $t \rightarrow t+1$ is equal to the difference between the number of dwellings in t+1 and t (${}^{t+1}T_{dwe}$ - ${}^{t}T_{dwe}$), plus the dwelling loss in the period $t \rightarrow t+1$ ($T_{dwe\text{-}loss}$):

$$T_{dwe\text{-}gain} =({}^{t+1}T_{dwe} - {}^{t}T_{dwe}) + T_{dwe\text{-}loss} \tag{4.31}$$

The allocation phase is responsible for indicating where the gain and loss of dwellings will take place and for updating the landscape-patch variables *total number of dwellings* (L_{dwe}) and *dwelling offers* ($L_{dwe\text{-}offer}$). It allocates the total number of dwelling offers based on two linear regression models (Neter et al. 1996): one that estimates the patches' loss of dwellings due to the expansion of non-residential uses (e.g., expansion of commercial use in residential areas), and another that estimates the patches' gain of dwellings due to new investments in residential developments. For each urban patch, the D-OFFER sub-model computes local transition loss (Y_{loss}) and local transition gain (Y_{gain}) of dwellings:

$$Y_{loss} = \beta_0 + \beta_1 X_1 + \beta_n X_n \tag{4.32}$$

$$Y_{gain} = \beta_0 + \beta_1 X_1 + \beta_n X_n \tag{4.33}$$

Where: the explanatory variables X represent one of the variable sets of the landscape-patch state ($L_{D\text{-}OFFERS}$), and β represents parameters estimated from empirical data (see section 5.3.2).

The total dwelling gain and loss occurring in the urban patches must meet the value of the global control variables $T_{dwe\text{-}loss}$ and $T_{dwe\text{-}gain}$. Considering this, the local transitions Y_{loss} and Y_{gain} are normalized by the factors θ_{loss} and θ_{gain}, respectively. These factors are computed as follows:

$$\theta_{loss} = \frac{T_{dwe-loss}}{\sum_{j=1}^{J} Y_{loss(j)}} \tag{4.34}$$

$$\theta_{gain} = \frac{T_{dwe-gain}}{\sum_{j=1}^{J} Y_{gain(j)}} \tag{4.35}$$

Where: $T_{dwe\text{-}loss}$ is the total dwelling loss and $T_{dwe\text{-}gain}$ is the total number of new dwellings in the city during the period $t \rightarrow t+1$. $Y_{loss(j)}$ and $Y_{gain(j)}$ represent the local transition loss and gain of the landscape patch j, respectively.

The new values for the landscape-patch variables *total number of dwellings* (L_{dwe}) and *dwelling offers* ($L_{dwe\text{-}offer}$) are computed as follows:

$$^{t+1}L_{dwe} = {}^{t}L_{dwe} + (Y_{gain} * \theta_{gain}) - (Y_{loss} * \theta_{loss}) \tag{4.36}$$

$$L_{dwe-offer} = L_{dwe} - L_{pop} \tag{4.37}$$

Where: L_{pop} is the total number of households living on the landscape patch.

Infrastructure sub-model (INFRA)

The INFRA sub-model simulates the dynamics of the landscape-patch variable *infrastructure* (L_{infra}). In the current version of MASUS, the variable L_{infra} is a composed index that ranges from 0 to 1, and represents the provision of water, sewage, and garbage collection. The sub-model assumes that these services always improve and will eventually be provided to all the inhabitants of the city. Based on that, the sub-model relies on a linear regression model where the dependent variable represents the improvement in infrastructure (Y_{infra}) during the period $t \rightarrow t+1$. For each urban patch, the L-VALUE sub-model computes the Y_{infra} as follows:

$$Y_{infra} = \beta_0 + \beta_1 X_1 + \beta_n X_n \tag{4.38}$$

Where: the explanatory variables X belong to the variable set $L_{L\text{-}VALUE}$ of the landscape-patch state, and β represents parameters estimated from empirical data (see section 5.3.3).

The landscape variable L_{infra} is updated after each annual cycle as follows:

$$^{t+1}L_{infra} = {}^{t}L_{infra} + Y_{infra} \tag{4.39}$$

Land value sub-model (L-VALUE)

The L-VALUE sub-model simulates the dynamics of the landscape-patch variable *land value* (L_{value}). For each urban patch, the L-VALUE sub-model computes the land value based on the results of the linear regression model previously estimated from empirical data:

$$L_{value} = \beta_0 + \beta_1 X_1 + \beta_n X_n \qquad (4.40)$$

Where: the explanatory variables X represent those contained in the variable set $L_{L\text{-}VALUE}$ of the landscape-patch state, and β represents parameters estimated from empirical data (see section 5.3.4).

4.3.3 EXPERIMENTAL-FACTOR module

The EXPERIMENTAL-FACTOR module consists of specification templates that can be set to test theories and policy approaches on segregation. These specifications can affect the system behavior through four pathways:

1. Changing global variables of the population transition sub-model (P-TRANSITION) that affect the social composition of the population and, therefore, the profile of households;
2. Changing parameters that drive the behavior of household agents in the decision-making sub-model (DECISION);
3. Changing the structure of the decision-making sub-model (DECISION);
4. Changing the state of the urban landscape.

The specification templates regard three different experimental factors: *socio-demographic aspects*, *household preferences*, and *urban policies*. In the current MASUS version, the specification templates for the experimental factor *socio-demographic aspects* allow exploring the relation between income inequality, seen as a product of the labor market, and segregation. The user can choose to execute alternative scenarios with low, original, and high inequality levels. These different inequality scenarios were developed by changing the global variables that control the income composition of the population in the P-TRANSITION sub-model.

The templates for the experimental factor *household preferences* focus on exploring how the preferences of affluent households for having neighbors similar to themselves can influence segregation dynamics. Different scenarios can be simulated by changing the parameter β_{neigh} that establishes the relevance of the neighborhood income composition to the DECISION sub-model of affluent households. To implement this change, the parameter β_{neigh} is multiplied by $\theta_{neigh.}$, a factor that ranges from 0 to 3. The MASUS interface allows users to select a value for θ_{neigh}, which implies the following:

1. If $\theta_{neigh} = 0$, affluent households do not consider the income composition of neighborhoods when selecting their residential locations;
2. If $\theta_{neigh} = 1$, the preference of affluent households for having neighbors similar to themselves is equal to the original one (estimated from empirical data);
3. If $\theta_{neigh} = 3$, the preference of affluent households for having neighbors similar to themselves is three times higher than the original one.

The experimental factor *urban policies* provides specification templates for the following policies:

1. Regularization of irregular settlements: To test the effect of this policy on segregation patterns, the EXPERIMENTAL-FACTOR module provides a template where clandestine settlements are converted to regular. This is possible by changing the landscape-patch variable *type of neighborhood* (L_{neigh}).
2. Universalization of infrastructure: To test this policy, the EXPERIMENTAL-FACTOR module provides a template where the value of the landscape-patch variable *infrastructure* (L_{infra}) is maximal for all urbanized patches.
3. Poverty dispersion: This template tests the effect of policies that provide housing vouchers to move low-income households from distressed areas to middle-class neighborhoods. For that, the EXPERIMENTAL-FACTOR module changes the decision-making sub-model of the *n* households who received the benefit. These households can only move to neighborhoods with a low concentration of poverty.
4. Wealth dispersion: This template tests the effect of policies that intervene on the real estate market by providing incentives for constructing developments for middle and upper classes in poor neighborhoods. This template selects non-occupied areas close to poor neighborhoods and changes the variable *type of neighborhood* (L_{neigh}) of these patches to a type that is attractive to affluent households.

4.4 MASUS simulation protocol

At the implementation level, the simulation protocol performed by MASUS consists of the following steps (Figure 4.6):

1. Set up the initial state of the system.
2. Start the main time loop (annual cycle):

2.1. Execute the decision-making sub-model (DECISION) for all households.

2.2. Calculate segregation indices and other population statistics.

2.3. Report simulated outputs (statistics, maps and graphs).

2.4. Update population and landscape state for the next cycle.

2.5. Update year (${}^{t+1}year = {}^{t}year + 1$) and repeat the annual cycle.

The step *set up the initial state of the system* imports GIS data that represents the population and landscape state of the study area in the beginning of the simulation (t_0). In addition, it sets up parameters according to the user-defined scenario, and executes sub-models that simulate the changes occurring during the period $t_0 \rightarrow t_0 + 1$. These sub-models are the household transition (H-TRANSITION), population transition (P-TRANSITION), urban-sprawl (U-SPRAWL), and dwelling-offers (D-OFFER).

After setting up the initial state of the system, it is possible to start the *annual cycle*, which is the main time loop of the simulation. The first procedure of the annual cycle is to *execute the decision-making sub-model (DECISION),* which is responsible for the household's decision about moving to another residential location. The DECISION sub-model is executed for all households, including the new ones created by the sub-model P-TRANSITION for the period $t \rightarrow t + 1$.

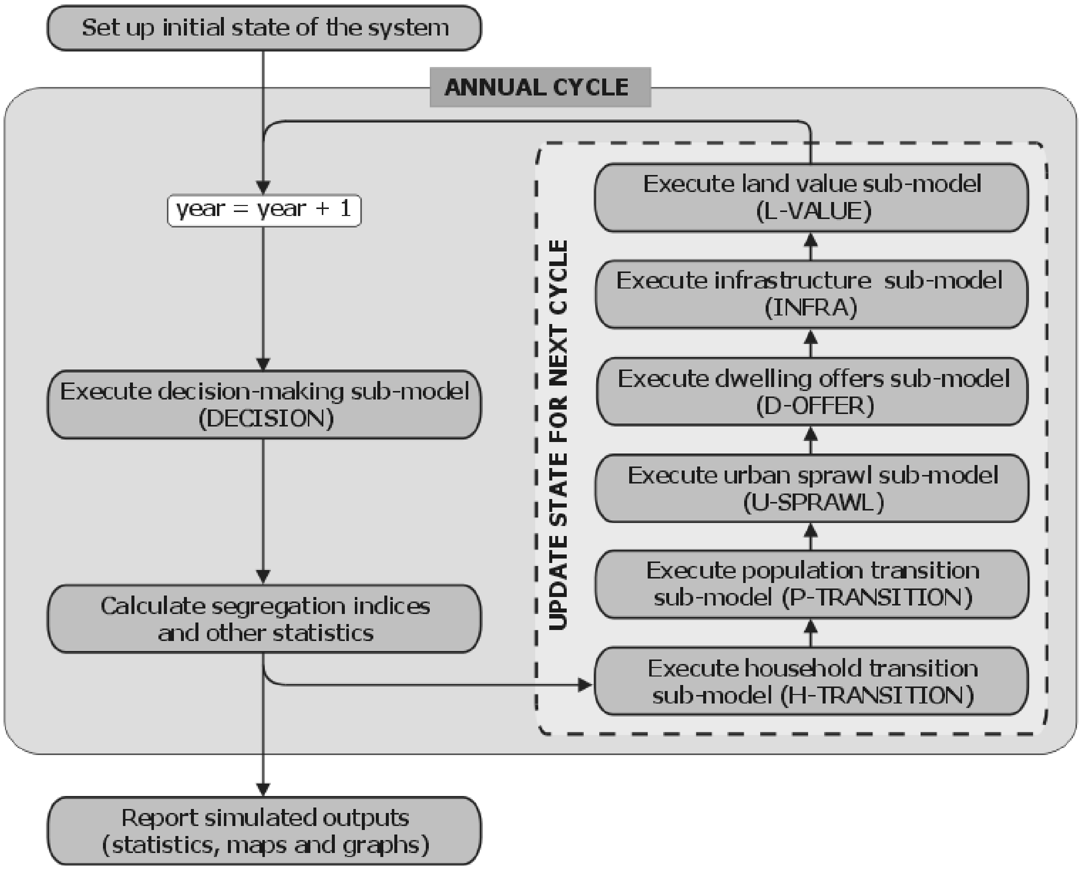


Figure 4.6 Flow chart showing the main steps of the MASUS simulation process.

The next procedures of the annual cycle are to *calculate* and *report segregation indices and other population statistics*. The MASUS program computes and reports the global and local segregation indices (section 2.5). The global indices of segregation are reported and plotted in graphs, while the local indices are shown as maps. Other reported and plotted population data include the total number of households, the number of households belonging to each social group, the Gini index, and the Lorenz curve.

After reporting the simulation outputs, the simulation program *updates the population and landscape state for the next cycle*. This step includes executing sub-models belonging to the URBAN-POPULATION module (H-TRANSITION and P-TRANSITION) and to the URBAN-LANDSCAPE module (U-SPRAWL, D-OFFER, INFRA, L-VALUE). Finally, the program updates the year and repeats the annual cycle.

5 EMPIRICAL PARAMETERIZATION OF THE MASUS MODEL: URBAN DYNAMICS IN SÃO JOSÉ DOS CAMPOS, BRAZIL

Recent developments in agent-based modeling (ABM) have demonstrated an increasing interest in combining social simulation models with empirical methods (Janssen and Ostrom 2006). According to Janssen and Ostrom (2006), three main reasons underlie this interest for empirically based ABM. The first is related to the existence of a large number of theoretical models, which makes it feasible to use ABM as a method for gaining new scientific insights. The second is the larger availability of relevant data. Finally, the third reason is that the increasing use of laboratory experiments in social sciences has challenged some of the simple models of human interactions in social-dilemma situations and emphasized the relevance of empirically based models (Janssen and Ostrom 2006).

The dissemination of empirically based ABM motivates the debate about one of the main challenges of contemporary social sciences, which is the development of models that are generalizable but still applicable in specific cases (Janssen and Ostrom 2006). As an empirically based model, MASUS provides different degrees of generalization in each of its specification levels. The MASUS conceptual framework (section 4.2) is the most generalizable level. It provides a general view of agents, environment, and interactions that give rise to different patterns of segregation. This overview is not context specific and also not exclusive for the simulation of income segregation. It can serve as basis for simulating other types of segregation, like racial or ethnic.

The MASUS theoretical specification (section 4.3) describes the model's architecture, including its modules, linkages and algorithms. For the most part, this theoretical specification is still general enough to be applicable to different types of segregation in different contexts. Nevertheless, it cannot achieve the same level of generality as the conceptual framework, since some specifics that are necessary for the MASUS implementation (e.g., variables and parameters) depend on data collection and empirical parameterization.

The MASUS model was first implemented for São José dos Campos, a medium-sized city located in the State of São Paulo, Brazil. Based on the data of this city, the aim of this point of the study is to provide empirical parameters for the sub-models of MASUS. This chapter comprises three main sections. The first section provides a brief description of the study area. Due to the nature of this work, this description emphasizes aspects related to the socioeconomic development of São José dos Campos and its implications for the segregation patterns of the city in the past years. The next section focuses on the residential choice behavior of the households in São José dos Campos. It provides the parameterization of the most important MASUS sub-model, i.e., the decision-making (DECISION) sub-model. The parameters of this sub-model indicate the effect of household and neighborhood characteristics on the residential choice of households belonging to different social groups. The final section focuses on the empirical parameterization of the urban landscape system. It presents the estimation of parameters for the sub-models that simulate dynamics of urban sprawl (U-SPRAWL), dwelling offers (D-OFFERS), infrastructure quality (INFRA), and land value (L-VALUE).

5.1 Study area: São José dos Campos, Brazil

São José dos Campos is a Brazilian municipality located in the State of São Paulo, between the metropolitan areas of São Paulo (91 km away) and Rio de Janeiro (334 km). The municipality has an estimated population of 609,229 (IBGE 2008) and a total area of 1,100 km^2. The site selected for the first implementation of the MASUS model corresponds, according to the macro zoning plan for São José dos Campos (PMSJC 2003), to the municipality's urbanized areas and areas for urban expansion (Figure 5.1).

São José dos Campos has a strong industrial sector, serving as host to most of the Brazilian aerospace sector and many other industries, such as automotive, defense, pharmaceutical, telecommunications, and petrochemical. In 2006, the city had the 22nd highest GDP in Brazil, and a per-capita GDP of R$ 25,419, while the country's average was R$ 12,688 (IBGE 2007). Despite these positive economic indicators, São José dos Campos is far from becoming a city that promotes the social inclusion of its inhabitants. Instead, the city has presented increasing rates of inequality. In 1991, the poorest 20% of households earned 3.4% of the total income, while the wealthiest 20% held 58.3%. In 2000, this disparity increased: The poorest 20% of the households earned 2.5% of the total income, while the wealthiest 20% earned 61.6%. The Gini index, which measures income inequality, also reflects this increasing disparity: Its value was equal to 0.55 in 1991 and became 0.59 in 2000.

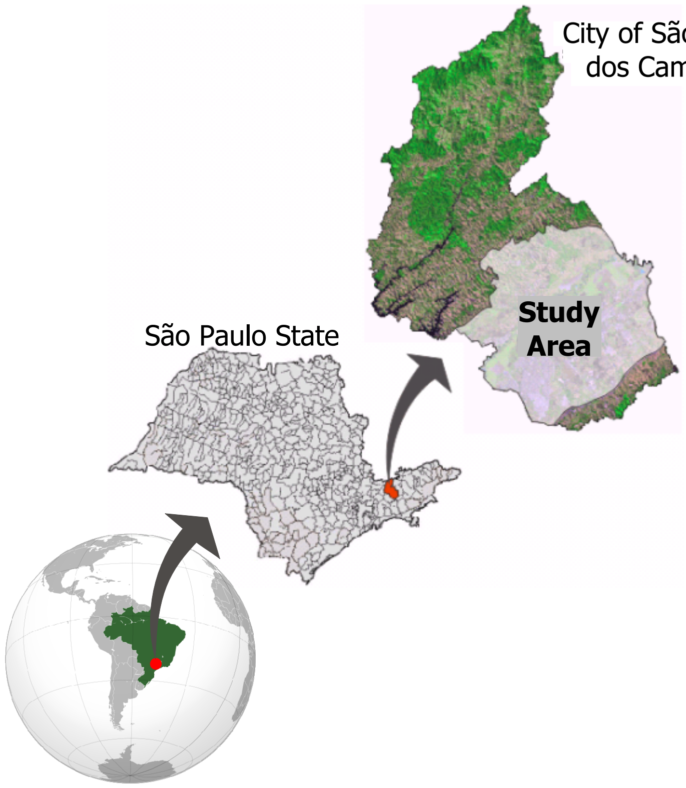


Figure 5.1 Location of the study area. Adapted from Feitosa (2005).

The dynamics of segregation patterns in São José dos Campos follow the trends that have been generally described in the literature about urban segregation in Brazilian cities (see section 2.2.1). During the 1950´s and 1960´s, there was a strong industrialization process in São José dos Campos, which attracted a large number of qualified and non-qualified workers from other Brazilian regions. The city presented high annual population growth rates in these decades: 5.6% during the 1950´s and 6.7% during the 1960´s, while the average annual rate in Brazil was 3.2% and 2.8%, respectively. This population growth generated an accentuated expansion of the city, which was characterized by the expansion of the central and traditional nucleus, and the configuration of peripheral and distant settlements, the so-called *periferias*. The segregation pattern known as 'center-periphery' prevailed at this time: Families belonging to higher strata occupied the center of the city and the adjacent areas in the west, while lower strata families occupied peripheral areas. Such pattern was reinforced by public investments, which were mainly concentrated on the central areas, while the poor *periferias* were characterized by precarious or inexistent infrastructure and services.

The population growth of São José dos Campos continued to intensify during the 1970´s and beginning of the 1980´s, when the city developed a strong aerospace and military sector to address the demands of the dictatorship established in the country. The city attracted a great number of non-qualified workers, who occupied illegal settlements in the *periferias* and slums (*favelas*) in central areas. Encouraged by exclusionary zoning policies and laws regulating the establishment of gated communities, medium and higher classes continued their expansion to the western region of the city.

After 1985, the decline of the military industry and the commercial openness of Brazil promoted a serious economic crisis in São José dos Campos. For this reason, population growth rates began to seriously decline. The economic recovery of the city only started to occur in the middle of the 1990´s. The economic changes during the period of crisis modified the segregation pattern of income groups in the city. This pattern became more complex and ruled by forces that deal with different scales of segregation. Considering a broader scale of segregation, medium- and high-income groups expanded from the center towards the western part of the city, while low-income families continued to establish large homogeneous settlements in the periphery. On the other hand, the proliferation of *favelas* and gated neighborhoods in wealthy and poor regions of the city is related to a decrease in the scale of segregation. Quantitative studies about São José dos Campos segregation patterns during the period 1991-2000 have demonstrated an increase in income segregation considering local and broader scales. This increase had been strongly promoted by the isolation of high-income families (Feitosa 2005; Feitosa et al. 2007).

5.2 Residential choice behavior of households

Urban segregation is a macro structure that emerges from the residential choice behavior of many households at the micro level. Representing this behavior is, therefore, a first condition for the simulation of segregation. In the MASUS framework, this task is performed by the DECISION sub-model (see section 4.3.1): Based on a discrete choice approach, the sub-model guides the households' decisions about whether to move to another residential location or not. This section presents the specification and estimation of discrete choice models used as input for the DECISION sub-model of the first operational MASUS model. The purpose of these discrete choice models is to assess how household and neighborhood characteristics influence the residential choices of households belonging to different income groups.

5.2.1 Analytical framework

The discrete choice approach adopted in this study is able to jointly model a household's mobility choice, neighborhood type choice, and specific neighborhood location choice. This is done by way of a nested multinomial logit model (Ben-Akiva 1973; Greene 2000; Train 2003). Like the multinomial logit model (MNL), the nested multinomial logit model (NMNL) is based on the micro-economic random utility theory, which states that individuals make their choices among options to maximize their utility, subject to constraints such as lack of knowledge and information (McFadden 1973). The NMNL, however, arose from an attempt to overcome constraints imposed by the MNL. The latter approach requires the independence of irrelevant alternatives (IIA) assumption, i.e., the unobserved utility of alternatives must be uncorrelated. The NMNL is a generalization of MNL that allows for a particular pattern of correlation in unobserved utility (Greene 2000; Train 2003). By clustering the related alternatives into subgroups, the IIA assumption is preserved within the subgroup but relaxed between them. The DECISION sub-model adopts a nested logit approach because it is likely that the expected utilities associated with the unobserved effect of its choice set are correlated.

The nested logit framework for the DECISION sub-model is organized in three levels (Figure 5.2). The first level (*i*) concerns the household decision about *moving* or *staying* and focuses on how personal attributes such as age and tenure status can influence the mobility rate of different income groups. The second level (*j*) focuses particularly on the *neighborhood type choice*. Having decided to move, the household can choose between:

1. *moving within its current neighborhood*, which is a decision that represents only an adjustment of household needs and does not promote change in the segregation patterns;

2. *moving to the same type of neighborhood*, e.g., from a poor irregular settlement to another one, which is a sort of residential choice that can be related to new trends of segregation patterns, but does not contribute towards the change of neighborhood profiles;
3. *moving to a different type of neighborhood*, which concerns residential choices that are able to promote a significant change in the spatial distribution of different income groups in the city (for details about the neighborhood types see section 5.2.2).

The third level concerns the *neighborhood location choice* (*k*), and complements the second level by including particular neighborhood characteristics that may influence the household choice for a certain location. Neighborhoods are randomly sampled, since estimation of discrete choice models has been shown to yield consistent estimates of the parameters, though with some loss of efficiency (Ben-Akiva and Lerman 1987).

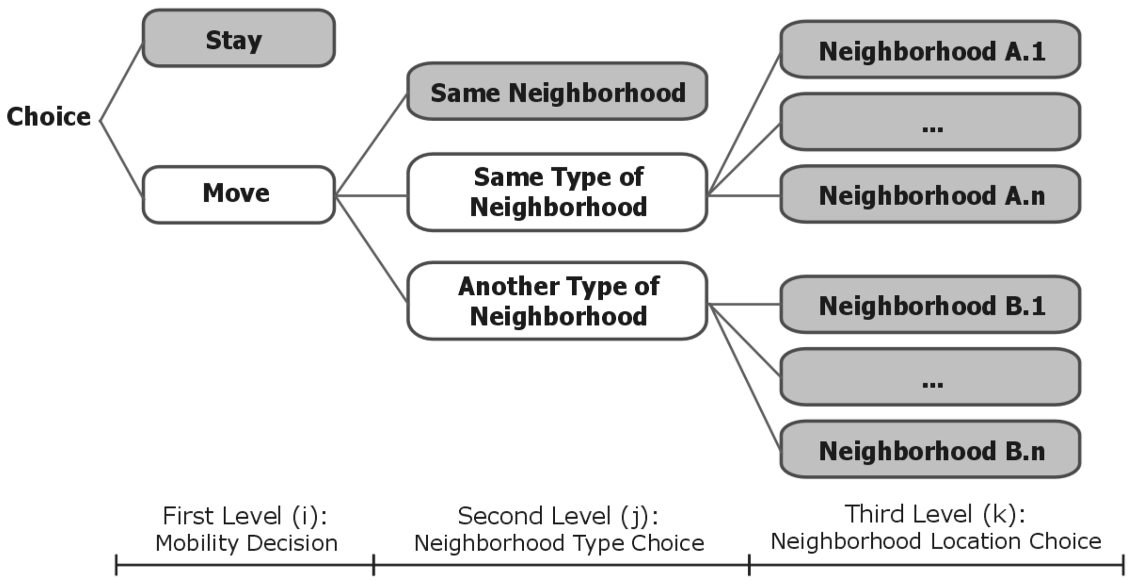


Figure 5.2 Nesting structure of the NMNL.

Considering that $X_{k|i,j}$, $Y_{j|i}$ and Z_i refer to the vectors of explanatory variables specific to the categories (*k*|*i,j*), (*j*|*i*) and (*i*), respectively, the probability of choosing a particular branch *k* in limb *j*, trunk *i* is (Greene 2000):

$$\Pr(k) = \Pr(k \mid i, j) \cdot \Pr(j \mid i) \cdot \Pr(i) \tag{5.1}$$

The conditional probability Pr(k|i,j) and Pr(j|i) in equation (5.1) are the functions of the forms:

$$\Pr(k \mid i, j) = \frac{\exp\left(\frac{1}{\tau_{j|i}}\left(\beta' X_{k|i,j}\right)\right)}{\sum_n \exp\left(\frac{1}{\tau_{m|i}}\left(\beta' X_{n|ij}\right)\right)} \tag{5.2}$$

and

$$\Pr(j \mid i) = \frac{\exp\left(\frac{1}{\tau_i}\left(\alpha' Y_{j|i} + \tau_{j|i} I_{j|i}\right)\right)}{\sum_m \exp\left(\frac{1}{\tau_l}\left(\alpha' Y_{m|i} + \tau_{m|i} I_{m|i}\right)\right)}, \tag{5.3}$$

Where: $I_{j|i}$ is the inclusive value for category (j|i) and $\tau_{j|i}$ is the dissimilarity parameter (or inclusive value parameter).

The $I_{j|i}$ transfers information from the neighborhood location choice model (third level) to the neighborhood *type* choice model (second level). Formally, $I_{j|i}$ is the log of the denominator of the conditional probability Pr(k|i,j):

$$I_{j|i} = \ln\left(\sum_n \exp\left(\frac{1}{\tau_{m|i}}(\beta' X_{n|ij})\right)\right). \tag{5.4}$$

The dissimilarity parameter $\tau_{j|i}$ provides a summary measure of the degree of similarity of the alternatives within the nest j, while the term $\tau_{j|i} I_{j|i}$ represents the expected utility that the decision maker receives from the choice among the alternatives in nest j.

The probability of choosing i, Pr(i) is:

$$\Pr(i) = \frac{\exp(\gamma' Z_i + \tau_i I_i)}{\sum_l \exp(\gamma' Z_l + \tau_l I_l)}, \tag{5.5}$$

Where:

$$I_i = \ln\left(\sum_m \exp\left(\frac{1}{\tau_l} \left(\alpha' Y_{m|i} + \tau_{m|i} I_{m|i} \right) \right) \right). \tag{5.6}$$

Generally, the dissimilarity parameter τ can differ over nests, reflecting different correlations among unobserved factors within each nest, but its value must lie within a particular range for the model to be consistent with utility-maximizing behavior (McFadden 1977). If $\tau = 1$, there is no correlation among the unobserved components of utility for alternatives within a nest, and the choice probabilities become standard logit probabilities; if $0 < \tau < 1$, the model is consistent with utility maximization for all possible values of the explanatory variables; if $\tau > 1$, the model is only consistent for some range of explanatory variables (Börsch-Supan 1990; Herriges and Kling 1996); and if $\tau < 0$, the model is inconsistent with utility maximization by implying that improving the attribute of an alternative can decrease its probability of being chosen. In case of degenerate nests, i.e., nests with only one alternative, the dissimilarity parameter can be constrained to 1. In our model, it is the case of the first-level nest 'stay', and the second-level nest 'move within the same neighborhood'.

We use the full information maximum likelihood (FIML) method to maximize the following log likelihood function:

$$\ln L = \sum_n \ln P(k \mid i, j) + \ln P(j \mid i) + \ln P(i) \tag{5.7}$$

5.2.2 Neighborhood types in São José dos Campos

To define and characterize the neighborhood types considered in the second level of the nested logit framework (Figure 5.2) for São José dos Campos, three maps of segregation were considered:

1. *Local dissimilarity index map*. This map shows how the proportions of income groups in each locality differ, on average, from the households' income composition of the whole city. For example, if the proportion of income groups in a tract and its surroundings is equal to the proportion of these groups in the city, the local dissimilarity of this tract is equal to zero.
2. *Local isolation index map for high-income households*. This map depicts the potential contact between the members of this social group. A very populated tract with a high proportion of affluent households living in it and its surroundings will present a very high index. The high-income group includes the households whose family heads have an income higher than 20 minimum wages[4].
3. *Local isolation index map for low-income households*, which is similar to the previous one, but computed for poor households (family head income inferior than 2 minimum wages).

These maps were developed by Feitosa et al. (2007) based on local segregation measures computed for the Census 2000 data on household income (see section 2.5).

The segregation indices presented in these three maps cover different dimensions of segregation (see section 2.1.1). While the local dissimilarity index refers to the balance of the distribution of social groups (dimension evenness/clustering), the local isolation index refers to the chance of having members from the same group living side by side (dimension exposure/isolation). Based on these maps and on information about irregular settlements, four main types of neighborhoods were defined (Table 5.1 and Figure 5.3):

1. *Type A:* Areas with high indices of dissimilarity and isolation of affluent households. It corresponds to neighborhoods with high land values and housing quality, good infrastructure and services, as well as many gated and guarded settlements and apartment complexes.
2. *Type B*: Areas with low indices of dissimilarity and isolation with groups. It is the most socially diverse type of neighborhood. These neighborhoods are well

[4] Minimum wage is the lowest level of work compensation secured by law. The Brazilian minimum wage was R$ 151 per month (U$ 85) in 2000, and R$ 465 per month (U$ 232) in 2009.

served with infrastructure, and often concentrate many services and commercial activities.

3. *Type C*: Areas with high indices of dissimilarity and isolation of poor households. Despite the poverty concentration of these neighborhoods, they are regular and have basic infrastructure. They include social housing projects.

4. *Type D*: Irregular settlements like *periferias* and *favelas*, with high indices of dissimilarity and isolation of poor households. Since these settlements are not part of the 'legal city', their residents do not pay taxes and the areas do not receive public investments.

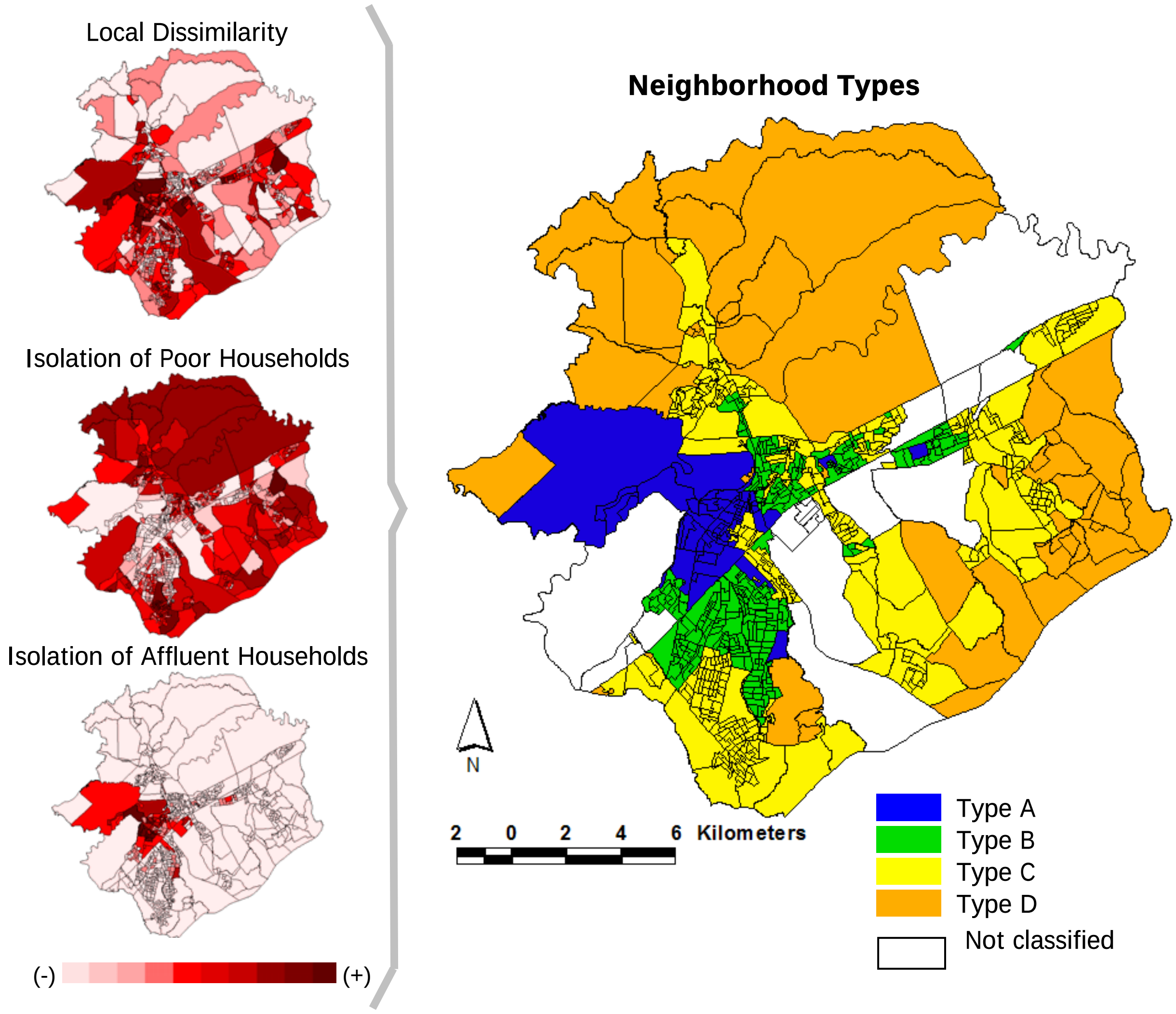


Figure 5.3 Neighborhood types in São José dos Campos.

Table 5.1 Average values of variables considered for the classification of neighborhood types

Variables	Neighborhood Types			
	A	*B*	*C*	*D*
Dissimilarity index (average)[1]	8.4	2.3	4.3	4.2
Isolation index for affluent households (average)[1]	42.9	2.2	0.3	10.3
Isolation index for poor households (average)[1]	0.3	1.8	6.1	4.2
Settlement legal condition	regular	regular	regular	irregular

[1] Because the magnitude of local segregation indices is very small, the values presented in this table were multiplied by 10^4.

5.2.3 Selection of explanatory variables and hypothesis

The selection of neighborhood and household variables relied on the hypothesis about determinants of household mobility and neighborhood choice. Because in this study it is assumed that households with different income levels have distinct residential choice behaviors, the variable 'income of household head' was used to stratify the model estimation based on three income intervals: up to 4 minimum wages, from 4 to 10 minimum wages, and more than 10 minimum wages.

In order to conduct the empirical tests, the variables originated from the working hypotheses were matched to the appropriate level of the nested structure. These variables include alternative-specific constants (C), household-specific variables (H), neighborhood-specific variables (N), and interactions between household- and neighborhood-specific variables (HN) (Tables 5.2, 5.3, and 5.4).

For the first level of the NMNL, which concerns the choice of moving or not, the hypotheses focus on household characteristics that may influence mobility behavior. Three hypotheses were developed and tested (Table 5.2):

Hypothesis 1

Mobility decreases as the age of the household head increases. The life cycle of families has been one of the most extensively used concepts to explain residential mobility (Davies and Pickles 1985; Ford and Smith 1990; Graham and Isaac 2002; Rossi 1955; Speare 1970). Households go through different stages of a life cycle, and while in these stages, they show different tendencies to change their residential location. Many demographic events that motivate a change of residence occur when people are younger, such as leaving their parents house, marriage, childbirth and job change (Barbon 2004; Chang et al. 2003; Clark and Onaka 1983; Huang and Clark 2002; Kan 1999; Sabagh et al. 1969). To test this hypothesis, we included the variable *age of household head* (H_{age}) in the NMNL model, and its estimated coefficient is expected to have a negative sign (decreased mobility).

Hypothesis 2

Renters have higher mobility rates than owner-occupiers. This assumption is also consistent with the literature on residential mobility (e.g., Speare, 1970) and was tested in the NMNL model through a dummy variable representing the households that live in a rented dwelling (H_{renter}). The estimated coefficient of this variable is expected to have a positive sign (increased mobility).

Hypothesis 3

Renters with limited financial resources are more vulnerable to housing insecurity and more likely to present higher mobility rates due to their inability to pay rents and bills, or due to the irregular status of their residences. We interacted the variable H_{renter} with the income of the household head ($H_{renter*income}$) in order to empirically test if the mobility of renters decreases when their income increases. The estimated coefficient of the resulting variable is expected to have a negative sign (decreased mobility).

Table 5.2 Explanatory variables for the first level of the NMNL: Mobility decision

Variable	*Description*	*Hypothesis/ Expected effect*	*Source*
H_{age}	Age of household head H_{age}	1 / (-)	UPHD survey
H_{renter}	Household tenure status (1 if household lives in a rented dwelling, 0 if otherwise)	2 / (+)	UPHD survey
$H_{renter*income}$	H_{renter} interacted with the income of the household head H_{income}	3 / (-)	UPHD survey

While the first level of the NMNL concerns household attributes, the second and third levels focus on how households assess the characteristics of potential residential locations. However, the second level considers the impact of these characteristics in terms of the household's neighborhood type choice, and the third level concerns their impact on the neighborhood choice in general, regardless of the second-level alternatives (move within the same or to the same type or to another type of neighborhood). The coefficients of the residential location variables were first estimated for the second-level alternatives. In case the coefficients of a variable were not significantly distinguishable among these alternatives, the variable was then considered in the third-level of the model as generic, i.e., with a common coefficient for all choice alternatives (Table 5.3 and 5.4).

This study considers general and income-group-specific hypotheses regarding aspects that contribute to the attractiveness of neighborhoods. The neighborhood-related hypotheses tested for all income groups are:

Hypothesis 4

Families face costs for moving and are, therefore, after controlling for other characteristics, more likely to stay in their current residence than to move. Thus, the estimated coefficients for the alternative-specific constants *move within the same neighborhood*, *move to the same type of neighborhood* and *move to another type of neighborhood* (C_{move1}, C_{move2}, and C_{move3}) are expected to have a negative effect on the household's utility.

Hypothesis 5

Once households are considering moving to a different neighborhood, they prefer to choose a place that is closer to the original one in order to keep their social bonds. As people gain familiarity with an area, they are more likely to develop friendships and to appreciate the local facilities and services (Abramo 2002; Speare 1974). Therefore, familiarity usually increases the attractiveness of a place. This assumption was tested through the inclusion of the variable *distance to the original place of residence* (N_{dist}) in the model. The estimated coefficient for this variable is expected to have a negative sign.

Hypothesis 6

Households generally choose to spend a smaller portion of their income on housing (Waddell et al. 2007). Studies conducted in São Paulo support this hypothesis by showing that most families (42%) spend up to 25% of their income in housing, while 37% spend between 25% and 40% of their income (Barbon 2004). To test this hypothesis, the *average land price of the neighborhood divided by the income of the household head* ($HN_{price\ /\ inc}$) was included in the model. This variable represents the housing affordability for the household, and its estimated coefficient is expected to have a negative sign.

Table 5.3 Explanatory variables for the second and third level of the NMNL: Neighborhood type choice and specific neighborhood choice for all income groups

Variable	*Description*	*Hypothesis/ Expected effect*	*Source*
C_{move1} C_{move2} C_{move3}	Alternative-specific constant for 'moving within the same neighborhood' (C_{move1}), 'moving to the same type of neighborhood' (C_{move2}), and 'moving to another type of neighborhood' (C_{move3}): 1 if alternative is true, 0 if otherwise	4 / (-)	UPHD survey + neighborhood type map
N_{dist}	Distance between the original place of residence and the neighborhood alternative.	5 / (-)	GIS-based calculation
$HN_{price\ /\ inc}$	Average land price of the neighborhood (N_{price}) divided by the income of the household head (H_{inc}).	6 / (-)	UPHD survey + property advertisements
N_{offers}	Total number of real estate offers in the neighborhood.	7 / (+)	property advertisements
N_{CBD}	Distance between the neighborhood alternative and the Central Business District (CBD)	8 / (-)	GIS-based calculation

Hypothesis 7

New investments in housing and land development provide a higher availability of dwellings in certain area of the city, which attract new residents and consolidate residential expansion vectors. To consider the impact of housing availability, the variable *number of real estate offers* (N_{offers}) was included in the NMNL model.

Hypothesis 8

Accessibility increases the attractiveness of neighborhoods, since areas with higher accessibility tend to concentrate services, commercial activities and jobs (Waddell 1996). We tested the impact of accessibility on residential location choice by including the variable *distance to the Central Business District (CBD)* (N_{CBD}) in the model. Using this variable as a proxy of accessibility implies a monocentric city assumption, which has been overcame in most metropolises and medium-sized cities, including São José dos Campos, where alternative employment centers have emerged in the last years. However, the CBD remains the most accessible area of the city, positively correlated with the density of businesses and availability of public transportation. An exploratory analysis based on São José dos Campos' empirical data about the number of commercial establishments and bus frequency revealed that these variables have a negative and nonlinear correlation with the distance to the CBD (R^2 equivalent to 0.41 and 0.54, respectively).

While the hypotheses 4 to 8 were generalized for all income groups, the following set of hypotheses focuses on differentiating the residential choice behavior of each group:

Hypothesis 9

Households tend to choose places with a higher proportion of neighbors belonging to their own income group. To test this hypothesis, variables representing the proportion of income groups in the neighborhoods (N_{lower}, N_{medium}, and N_{higher}) were included in the model of the respective group. The estimated coefficients for these variables are expected to have a positive sign.

Hypothesis 10

Once moving to another neighborhood, poorer families have a higher probability of moving to neighborhoods type C or D. Nevertheless, household characteristics such as education and family size can impact the probability of moving to more or less segregated areas. For instance, poor families with a better educated household head have a higher chance to move to less segregated neighborhoods (type B), and a smaller chance to move to irregular neighborhoods (type D). Regarding the family size, we assume that poor families with a large number of members have higher chances to move to irregular neighborhoods (type D). To test this hypothesis, dummies of neighborhood types were interacted with household attributes and included in the model ($H_{edu} * N_B$, $H_{edu} * N_D$, and $H_{size} * N_D$).

Hypothesis 11

Middle-class families are more likely to choose type B neighborhoods, although the probability of these families moving to a poorer neighborhood increases if the area provides good infrastructure and services. The interaction between the variable representing infrastructure and the dummy variable for neighborhood type C ($N_{infra} * N_C$) allows testing this hypothesis.

Hypothesis 12

Families with higher income are more likely to choose type A and B neighborhoods. The probability of choosing type A neighborhood, however, increases if the family has children. Studies on residential location choice have shown that households with children are more attracted to peaceful and safe neighborhoods that provide spaces for children to play (Gayda 1998). These are the main appeals of gated neighborhoods, which is the dominant kind of development in type A neighborhoods. We include the variable $HN_{kids,A}$ ($N_A * H_{kids}$) to test if type A neighborhoods significantly increase the utility of families with children.

Table 5.4 Group-specific explanatory variables for the second- and third-level of the NMNL: Neighborhood type choice and specific neighborhood choice

Variable	*Description*	*Hypothesis/ Expected effect*	*Source*
N_{lower}	Proportion of families in the neighborhood with lower income (head of household income up to 4 minimum wages)	9 / low-income group: (+)	Census 2000
N_{medium}	Proportion of families in the neighborhood with medium income (head of household income between 4 and 10 minimum wages)	9 / middle-income group.: (+)	Census 2000
N_{higher}	Proportion of families in the neighborhood with higher income (head of household income superior to 10 minimum wages)	9 / high-income group.: (+)	Census 2000
N_A, N_B, N_C, N_D	Type A, B, C, and D neighborhoods, respectively (1 if true, 0 if otherwise)	10-12 / (various)	neighborhood type map
$H_{edu} * N_B$	Years of education of the household head H_{edu} interacted with the dummy variables N_B	10 / low-income group:(+)	UPHD survey
$H_{edu} * N_D$	Years of education of the household head H_{edu} interacted with the dummy variable N_D	10 / low-income group: (-)	UPHD survey
$H_{size} * N_D$	Household size H_{size} interacted with the dummy variable N_D	10 / low-income group: (+)	UPHD survey
$N_{infra} * N_C$	Neighborhood infrastructure index N_{infra} (multiplication of the proportion of water supply, sewage, and waste disposal) interacted with the dummy variable N_C	11 / middle-income group: (+)	Census 2000
$HN_{kids,A}$	Presence of kids in the household interacted with the dummy variable N_A	12 / high-income group: (+)	UPHD survey

5.2.4 Data sources

To investigate the determinants of household mobility and neighborhood choice, this work relies on household-level data from the 'Survey for Urban Planning Instrumentation and Evaluation of the Housing Deficit in São José dos Campos' (UPHD). This survey was conducted in 2003 by the Population Studies Center of the University of Campinas (NEPO/UNICAMP) and the Municipal Government of São José dos Campos (PMSJC). The 7,910 respondents of the UPHD survey were selected from a total of 141,814 households distributed in 24 socio-economic regions, which corresponds to approximately 5% of the population. For each respondent, the survey provides retrospective residential mobility history as well as detailed information about demographic, socio-economic, and housing characteristics.

The UPHD survey was used in this study to analyze the residential mobility history of households during the 12 months that preceded the interviews. The survey provides information about the households' residential choice, including the name of their previous and current neighborhood. Based on this data, it was possible to define the dependent variable of the NMNL model. The UPHD data also provide information about household characteristics that may affect residential mobility behavior (household-specific variables), such as age (H_{age}), income (H_{income}), education (H_{edu}), tenure status (H_{renter}), family size (H_{size}), and presence of children (H_{kids}).

The neighborhood-specific variables were extracted from different sources:

1. The Brazilian Census 2000, which provided the proportion of income groups (N_{lower}, N_{medium}, and N_{higher}) and infrastructure data (N_{infra}).
2. The map of neighborhood types (Figure 5.3), which provided the dummy variables for the neighborhood types A, B, C, and D (N_A, N_B, N_C, and N_D).
3. GIS-based calculations, which provided the variables *distance to CBD* (N_{CBD}) and *distance between the original and alternative neighborhood* (N_{dist}). All distances are provided in meters.
4. Property advertisements collected from local newspapers printed from March 2002 to February 2003, which provided the variables land price (N_{price}) and real estate offers (N_{offers}). The data collection was conducted at the Municipal Archive of the City of São José dos Campos.

5.2.5 Results and discussion

Nested logit models were estimated using the *nlogit* command of STATA, Version 10, which uses a parameterization that is consistent with random utility maximization (RUM). To estimate the NMNL models proposed in this study, it is necessary to include the attributes of neighborhoods that are within a household's choice set. This is straightforward for the alternatives that involve neighborhoods that are known, like the one chosen by the household or the one where the household was living before (alternatives 'not move' or 'move within the same neighborhood'). However, it is also necessary to include the attributes of those alternative neighborhoods that were not chosen but, given the large number of neighborhoods in a city, it is not possible to consider all of them. The problem of estimating individual choice models when the number of alternatives is impractically large has been discussed in the literature for household mobility choice (Ben-Akiva and Lerman 1987), and it has been proved that it is possible to estimate a model on a subset of alternatives without inducing inconsistency. Hence, for representing the third-level neighborhood alternatives within the second-level nests 'move to the same type of neighborhood' and 'move to another type of neighborhood', 10 neighborhoods addressing each nest condition were randomly selected.

The NMNL were estimated for households with lower, middle, and high income (Table 5.5, 5.6, and 5.7, respectively). The coefficients of all NMNL levels were estimated with respect to the choice 'stay', and the model was fitted with the constraint that the inclusive value parameter for degenerated branches is equal to 1.

Table 5.5 NMNL estimations for lower-income households (N observations = 63228, N cases = 2874, choice 'stay' as the base case)

Level	*Choice*	*Variable*	*Coef.*	*Std. err.*
1st	Move	Age of the household head (H_{age})	-0.043 ***	0.005
		Renter (H_{renter})	3.080 ***	0.269
		Renter * household income ($H_{renter*income}$)	-1.2(10^-3) ***	4.2(10^-4)
2nd	Move within the same neighborhood	Constant (C_{move1})	-1.592 **	0.799
		Real estate offers (N_{offers})	-2.8(10^-3)	1.7(10^-3)
		Distance to CBD (N_{CBD})	1.1(10^-5)	3.7(10^-5)
		Prop. of lower-income families (N_{lower})	-0.647	0.908
	Move to the same type of neighborhood	Constant (C_{move2})	-3.810 ***	1.377
		Real estate offers (N_{offers})	1.9(10^-3) ***	6.9(10^-4)
		Distance to CBD (N_{CBD})	6.7(10^-5) **	3.1(10^-5)
		Prop. of lower-income families (N_{lower})	0.953 *	0.570
		Type C neighborhood (N_C)	0.991	0.686
		Type D neighborhood (N_D)	0.582	1.37
		Education status * Type B ($H_{edu} * N_B$)	0.073	0.056
		Education status * Type D ($H_{edu} * N_D$)	-0.051	0.109
		Household size * Type D ($H_{size} * N_D$)	0.040	0.175
	Move to another type of neighborhood	Constant (C_{move3})	- 6.163 ***	2.202
		Real estate offers (N_{offers})	3.0(10^-3) ***	1.1(10^-4)
		Distance to CBD (N_{CBD})	10.3(10^-5) **	4.6(10^-5)
		Prop. of lower-income families (N_{lower})	1.520 *	0.907
		Type C neighborhood (N_C)	2.379 **	1.085
		Type D neighborhood (N_D)	2.254 *	1.278
		Education status * Type B ($H_{edu} * N_B$)	0.195 **	0.087
		Education status * Type D ($H_{edu} * N_D$)	-0.057	0.049
		Household size * Type D ($H_{size} * N_D$)	0.065	0.075
3rd	Generic variables	Land price/ income ($HN_{price/inc}$)	-1.9(10^-3)	2.4(10^-3)
		Distance from original residence (N_{dist})	-1.3(10^-4) ***	5.1(10^-5)
Dissimilarity Parameters		τ_{move} (first level)	0.658 *	0.28
		$\tau_{move\ 2}$ (second level)	0.449 **	0.179
		$\tau_{move\ 3}$ (second level)	0.791 *	0.319

Likelihood-ratio test for IIA ($\tau = 1$): χ^2 = 13.93 ***
Wald test: χ^2 = 440.94 ***
Likelihood-ratio index (McFadden's R^2) = 0.231

***, **, and * indicate statistical significance at the 99%, 95%, and 90% levels.

Likelihood-ratio tests for the independence of irrelevant alternatives (IIA), recommended by Greene (2000), suggest that the nesting is appropriated for the models. To perform this test, the model is fitted with and without the restriction that the inclusive parameters of the non-degenerated branches are equal to one. A chi-squared test statistic is computed by taking twice the difference in the log likelihood functions with the degree of freedom equal to number of restrictions imposed. The test statistic of the models estimated for lower-, middle-, and higher-income households are 51.19, 30.90, and 12.03, respectively, and the critical values for a one-tailed 1% test is 1.35. Hence, we reject the null hypothesis that the inclusive parameters of the non-degenerated branches are equal to one.

The values of the inclusive parameters need to lie within the interval (0,1) in order to be considered consistent with utility-maximizing behavior for all possible values of the explanatory variables (Börsch-Supan 1990; Herriges and Kling 1996; McFadden 1977). A two-tailed test at a 95% confidence level suggests that this parameter estimate is significantly different from 1 and 0, which indicates a degree of similarity among unobserved factors within each non-degenerated nest. Both tests indicate that, for this dataset, the specified nested logit models are adequate to characterize household mobility and neighborhood choice.

The goodness of fit for the models was assessed through the Wald's chi-squared statistic and the likelihood-ratio index (or McFadden's pseudo R^2). The Wald tests show that the empirical NMNL is highly significant ($p<0.001$) in explaining household mobility and neighborhood choice of all income groups. The likelihood-ratio indices of the nested logit models estimated for households with lower, middle, and higher income are, respectively, 0.231, 0.225, and 0.206. These likelihood-ratio indices indicate the gain in the likelihood function due to the independent variables, i.e., how well the estimated model performs compared with a model in which all the parameters are zero. The 'percentage of correctly predicted choices' was not considered to evaluate the goodness-of-fit of the models. This statistic is based on the idea that the best prediction for each case is the alternative with the highest probability, which, according to Train (2003 : 73), is a notion 'opposed to the meaning of probabilities and the purpose of specifying choice probabilities'. In models like the ones presented in this section, where the predicted probability of 'staying' is the highest for all cases, we would assume that the alternatives associated with 'moving' would never be chosen when, in fact, there is a probability that it may occur.

Table 5.6 NMNL estimations for middle-income households (N observations = 47432, N cases = 2156, choice 'stay' as the base case)

Level	*Choice*	*Variable*	*Coef.*	*Std. err.*
1st	Move	Age of the household head (H_{age})	-0.046 ***	0.007
		Renter (H_{renter})	2.243 ***	0.466
		Renter * household income ($H_{renter*income}$)	4.5(10^-5)	2.9(10^-4)
2nd	Move within The same neighborhood	Constant ($C_{move\ 1}$)	-2.123 ***	0.524
		Real estate offers (N_{offers})	4.5(10^-4)	4.9(10^-4)
	Move to the same type of neighborhood	Constant ($C_{move\ 2}$)	-2.631 ***	0.765
		Real estate offers (N_{offers})	1.9(10^-3) ***	7.1(10^-4)
		Type B neighborhood (N_B)	0.515	0.447
		Type C neighborhood (N_C)	0.446	0.610
		Infrastructure * Type C ($N_{infra} * N_C$)	0.191	0.454
	Move to another type of neighborhood	Constant ($C_{move\ 3}$)	-2.451 ***	0.676
		Real estate offers (N_{offers})	1.9(10^-3) ***	7.5(10^-4)
		Type B neighborhood (N_B)	0.308	0.261
		Type C neighborhood (N_C)	-0.432	0.743
		Infrastructure * Type C ($N_{infra} * N_C$)	0.567	0.798
3rd	Generic variables	Land price/ income ($HN_{price\ /\ inc}$)	-0.004	0.011
		Distance from original residence (N_{dist})	-11.1(10^-5)***	4.2(10^-5)
		Distance to CBD (N_{CBD})	1.9(10^-5)	1.8(10^-5)
		Prop. of middle-income families (N_{middle})	1.435 *	0.740
Dissimilarity Parameters		τ_{move} (first level)	0.752	0.284
		$\tau_{move\ 2}$ (second level)	0.292	0.113
		$\tau_{move\ 3}$ (second level)	0.453	0.196

Likelihood-ratio test for IIA ($\tau = 1$) : $\chi^2 = 20.41$ ***
Wald test: $\chi^2 = 262.54$ ***
Likelihood-ratio index (McFadden's R^2) = 0.225

***, **, and * indicate statistical significance at the 99%, 95%, and 90% levels.

Table 5.7 NMNL estimations for high-income households (N observations = 25278, N cases = 1149, choice 'stay' as the base case)

Level	*Choice*	*Variable*	*Coef.*	*Std. err.*
1st	Move	Age of the household head (H_{age})	-0.040***	0.011
		Renter (H_{renter})	2.542***	0.425
		Renter * household income ($H_{renter*income}$)	-9.4(10^{-5})	-7.5(10^{-5})
2nd	Move within the same neighborhood	Constant (C_{move1})	-2.532 ***	0.693
	Move to the same type of neighborhood	Constant (C_{move2})	-2.464 ***	0.855
		Type A neighborhood (N_A)	0.477	0.661
		Type B neighborhood (N_B)	0.062	0.495
		Kids * Type A ($HN_{kids,A}$)	-0.368	0.636
	Move to another type of neighborhood	Constant (C_{move3})	-3.457 ***	1.053
		Type A neighborhood (N_A)	-0.256	0.732
		Type B neighborhood (N_B)	1.760 ***	0.709
		Kids * Type A ($HN_{kids,A}$)	1.49 **	0.784
3rd	Generic variables	Land price/ income ($HN_{price/inc}$)	-0.084	0.053
		Real estate offers (N_{offers})	1.4(10^{-3}) ***	5.1(10^{-4})
		Distance from original neighborhood (N_{dist})	-4.9(10^{-5}) **	2.5(10^{-5})
		Distance to CBD (N_{CBD})	2.3(10^{-5})	2.9(10^{-5})
		Prop. of high-income families (N_{higher})	0.960 **	0.503
Dissimilarity Parameters		τ_{move} (first level)	0.666	0.290
		τ_{move2} (second level)	0.384	0.139
		τ_{move3} (second level)	0.552	0.213

Likelihood-ratio test for IIA ($\tau = 1$) : χ^2 = 11.0 ***
Wald test: χ^2 = 104.67 ***
Likelihood-ratio index (McFadden's R^2) = 0.206
***, **, and * indicate statistical significance at the 99%, 95%, and 90% levels.

The NMNL estimations confirmed all the first-level hypotheses, which regard the households' choice of moving or staying. The coefficients estimated for the variables *age* and *tenure status* of the household head (H_{age} and H_{renter}) are highly significant for the mobility decisions of all income groups. The sign of the coefficient for variable H_{age} indicates that an increase in age of the household head is associated with a lower probability to move. This result is consistent with the Hypothesis 1 (section 5.2.3), which states that demographic events that motivate a change of residence affect younger heads of household more often. On the other hand, the estimated coefficient for the variable H_{renter} indicating that the tenure status 'renter' increases the households' probability of moving to another residence, which is a result that corroborates Hypothesis 2.

The variable interacting the tenure status renter with the *household income* ($H_{renter*income}$) was significant for poor households. This result confirms Hypothesis 3 by showing that as the income of poor renters increases, all other variables being constant, their mobility rate decreases. This is an indication that renters with lower income are more vulnerable to constant changes in residence due to their economic constraints. As expected, as soon as the analysis shifts the focus towards renters belonging to a higher-income level, the coefficients estimated for the variable $H_{renter*income}$ are not significant.

Regarding the second level of the NMNL, which focuses on the neighborhood type choice, the coefficient of all alternative-specific constants (C_{move1}, C_{move2}, and C_{move3}) were negative and highly significant. The negative effect of these constants corroborates Hypothesis 4 by showing that, over the course of a year, households are more likely to stay in their current residence than to move. These results also provide new insights: They suggest that, for the groups with lower and higher income, the alternative 'moving to another type of neighborhood' provides a higher decrease in utility than the other moving alternatives. In other words, the poorest and the richest households are those with a higher resistance to move to another type of neighborhood.

In accordance with Hypothesis 5, the estimated coefficients for the variable N_{dist} suggest that the disutility of moving is intensified when the distance from the original place of residence increases. The coefficients of N_{dist} are negative and significant at a 99% confidence level for lower-income households and at 95% for the other income groups. Since previous estimations of model showed that N_{dist} coefficients did not vary amongst the second-level alternatives, N_{dist} was considered as a generic variable, and its coefficients were jointly estimated at the third level of the model.

The impact of the land and real estate market on households' residential decisions, which is the focus of hypotheses 6 and 7, was tested through the variables *average land price in the neighborhood divided by the head of household income*

($HN_{price/inc}$), and the number *of market offers in the neighborhood* (N_{offers}). The NMNL models did not support Hypothesis 6, which was tested through the variable $HN_{price/inc}$ and states that households prefer to spend a smaller portion of their income on housing. The coefficients estimated for variable N_{offers} corroborate Hypothesis 7, which states that new housing developments increase the attractiveness of neighborhoods. In the model for middle- and lower-income families, the N_{offers} coefficients are positive and significant for the alternatives ‘move to the same type of neighborhood’ and ‘move to another type of neighborhood’. In the model for high-income households, the variable N_{offers} could be included as generic (no difference amongst second-level alternatives), and its estimated coefficients are also positive and significant at a 99% confidence level.

The estimated coefficients for the variable *distance to CBD* (N_{CBD}) do not corroborate the hypotheses that households tend to choose the most accessible neighborhoods (Hypothesis 8). Similar results have been obtained by other researchers, such as Molin and Timmermans (2003), who advocate that accessibility can be considered less important than other neighborhood attributes in the case when people are able to afford flexible means of transportation. In the case of poor households, however, the estimated coefficients reveal that the variable N_{CBD} has a positive and significant effect on the residential choice of those families who move to another neighborhood. This suggests that recent moves of poor families have pushed them further from the center of the city. In the case of these families, it is likely that they have chosen to face increasing commuting time in exchange for lower housing prices and the possibility of ownership.

The hypothesis that households tend to choose places with a higher proportion of the neighborhood belonging to their own income group was confirmed by the NMNL models estimated for low- and high-income households. In the model for low-income households, the variable *proportion of low-income neighbors* (N_{lower}) is significant for the alternatives ‘move to the same type of neighborhood’ and ‘move to another type of neighborhood’. This suggests that poor households who decided to move to a new neighborhood have chosen places with a higher concentration of poverty. In the model for high-income households, the estimated coefficients for the variable *proportion of high-income neighbors* (N_{higher}) did not diverge amongst the second-level alternatives and, therefore, N_{higher} was included as generic variable in the final NMNL model.

The dummy variables for neighborhood types were included in the model in order to test the hypotheses 10, 11, and 12, which deal with neighborhood types that aremore likely to be chosen by different income groups (N_A and N_B for high income, N_B and N_C for middle income, N_C and N_D for low income). However, due to multicolinearity constraints, these variables were only included for the two second-level alternatives that include moving to a new neighborhood. The model estimated for poor households suggests that, when moving to another type of neighborhood, families are more likely to choose type C and D neighborhoods. This idea is supported by the coefficients of the variables N_C and N_D, which are significant and positive when estimated for this second-level alternative.

Nevertheless, the coefficients of the variable $H_{edu} * N_B$ indicate that poor households increase their chances to move to less segregated neighborhoods (type B) if their head has a higher education level. This suggests that higher levels of education can decrease the vulnerability of poor families with respect to problems associated with the concentration of poverty, like violence and discrimination. However, the hypotheses relating education and size of poor families to the chance of moving to irregular neighborhoods (type D) were not confirmed by the NMNL model. The model also did not corroborate the hypothesis regarding the likelihood of middle-class families to move to type B and C neighborhoods (Hypothesis 11).

In the case of high-income households, the coefficients estimated for the variables *type B* (N_B) and *type A interacted with presence of children* ($HN_{kids,A}$) suggest that these households are more likely to move to another type of neighborhood if the new neighborhood is classified as type B and, in case they have children, as type A. The latter result corroborates the hypothesis that affluent households with children are more likely to choose gated neighborhoods or condominiums in order to guarantee the family's safety.

5.3 Urban landscape dynamics

As observed above, the residential choice behavior of households is influenced by many aspects related to the urban landscape. Therefore, it is necessary to consider the dynamics of this system when simulating urban segregation. In the MASUS model, the main dynamics of the urban landscape system are driven by four sub-models: urban sprawl (U-SPRAWL), dwelling offers (D-OFFERS), infrastructure quality (INFRA), and land value (L-VALUE). The parameterization of these sub-models based on São José dos Campos data is presented in the following.

5.3.1 Urban sprawl

Urban sprawl is the spreading of urban developments over undeveloped land at the fringe of a city (Gillham and MacLean 2002). In the MASUS framework, urban sprawl represents the appearance of new residential areas that can be considered by the households during their decision-making process. The sub-model U-SPRAWL, which simulates the urban sprawl dynamics, consists of two phases (section 4.3.2): the *transitional phase*, responsible for quantifying the total sprawl that is expected to occur in the period $t \to t+1$, and the *allocation phase*, which indicates the location of this sprawl.

Transitional phase

The U-SPRAWL sub-model uses Markov chain to compute the annual global transition probability $P_{NU \to U}$, which accounts for the probability of a landscape patch to change from 'non urban' (NU) to 'urban' (U) (see section 4.3.2). For estimating the $P_{NU \to U}$, this study used two thematic maps of the urban areas in São José dos Campos of 1990 and 2000 (Figure 5.4). These maps were generated by Feitosa (2005) from satellite images LANDSAT-5/TM (INPE 1990) and LANDSAT-7/TM (INPE 2000).

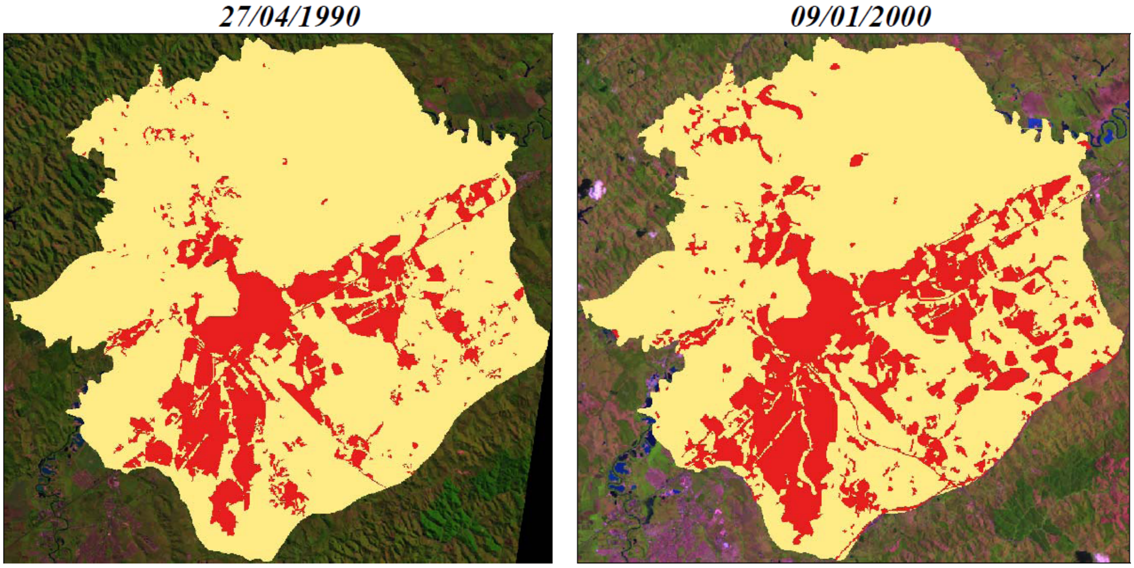


Figure 5.4 Urbanized areas in São José dos Campos in 1990 and 2000 (Feitosa, 2005).

In order to obtain parameters that are compatible with the data used in the MASUS simulations, the vector-GIS data corresponding to the urban areas in São José dos Campos in 1990 and 2000 were loaded into NetLogo 4.0.4 (Wilensky 1999). For that, a surface of landscape patches (100 m x 100 m) corresponding to the study area was defined in NetLogo (world), and the properties of the GIS data features (urban or non-urban) were imported as binary landscape patch variables *urban 1990* and *urban 2000*. Additional vector-GIS data containing areas that should be excluded from the analysis (parks, industrial and institutional areas, protected areas, rivers, etc.) were also incorporated into NetLogo. From the resulting 'NetLogo world', it was possible to perform queries about the transition of landscape-patch states during the period 1990-2000. The results of these queries were used for computing global transition probabilities for land-use change (Table 5.8).

Table 5.8 Matrix of global transition probabilities for São José dos Campos, 1990-2000 (N = 26,168 patches)

Land Use	*Non-Urban (NU)*	*Urban (U)*
Non-Urban (NU)	0.9120 (18,053 patches)	0.0872 (1,725 patches)
Urban (U)	0 (0 patches)	1 (6,390 patches)

The global transition probability for non-urban to urban (Table 5.8) covers the time period of 10 years, while each simulation cycle executed in MASUS represents the period of 1 year. For this reason, it was necessary to decompose the original transitional probability by using the principal components method in order to obtain the annual global transition probability ($P_{NU \to U}$). This procedure was conducted according to Equation 5.8 (Bell and Hinojosa 1977):

$$P^{n} = H \; V^{n} \, H^{-1} \; , \tag{5.8}$$

Where: P is the matrix of global transition probabilities, H is the eigenvector matrix, H^{-1} is the transposed eingenvector matrix, V is the eigenvalue matrix, and n is the number of steps.

The global transition probability $P_{NU \to U}$ estimated from São José dos Campos data is equal to 9.05(10^{-3}). In the U-SPRAWL sub-model, this value accounts for the annual probability of a non-urban landscape patch changing the value of its binary variable 'urban use' (L_{urban}) to 1. The $P_{NU \to U}$ is provided by the user while setting the initial state of the simulation. Once the U-SPRAWL sub-model is executed, which occurs at the end of each simulation cycle (section 4.4), the $P_{NU \to U}$ is multiplied by the total number of non-urban patches at time t, and the resulting value is equivalent to the total number of non-urban patches that will change the status to urban during the period $t \to t+1$ ($T_{new-urban}$). The $T_{new-urban}$ is stored as a global variable of the *EntireLandscape* and retrieved in the following phase of the U-SPRAWL sub-model, when the new urban cells are allocated.

Allocation phase

The allocation phase is responsible for indicating which non-urban patches convert their use to urban during the period $t \to t+1$. It relies on a binary logistic regression to compute the local transition probability of a non-urban patch becoming urban ($p_{NU \to U}$). For estimating the binary logistic regression of the U-SPRAWL allocation phase, the dependent variable is *transition to urban (1) or not (0)* ($L_{transition}$). Since data about urban areas in São José dos Campos had been already loaded into NetLogo, the dependent variable $L_{transition}$ was computed for each landscape patch that was not urban in 1990, based on the information about its state transition during the period 1990-2000. The variable $L_{transition}$ is equal to 1 in case the non-urban patch became urban during 1990-2000, and is equal to 0 otherwise.

Based on the hypothesis that the transition from non-urban to urban is more likely to occur for patches located close to urbanized areas and with a better accessibility, the variables *distance to closest urban patch* ($L_{d-urban}$), *number of urban patches within a 700 m radius* ($L_{neigh-urban}$), *number of households living within a 700 m radius* ($L_{neigh-pop}$), *distance to roads* ($L_{d-roads}$) and *distance to CBD* (L_{d-CBD}) were included in the model (Table 5.9). For each non-urban patch, the variables $L_{d-urban}$ and $L_{neigh-urban}$ were computed in NetLogo considering the location of the patches that were urbanized in 1990. To compute the variable $L_{neigh-pop}$, it was necessary to import household micro-data obtained from the Census 1991 into NetLogo (see section 6.2.1). The accessibility-related variables, $L_{d-roads}$ and L_{d-CBD}, were obtained from GIS-based calculations and then loaded into NetLogo as landscape-patch variables. To calculate the distance to roads ($L_{d-roads}$), a road map provided by the municipal government of São José dos Campos (PMSJC 2003) was used.

The variable L_{slope} was selected based on the hypothesis that *terrain slope* can represent an environmental constraint that inhibits urban occupation. The terrain slope was obtained from GIS-based calculations on a topographic map (scale 1:20000) provided by the municipal government of São José dos Campos (PMSJC 2003). Afterwards, the data were loaded into NetLogo as a landscape-patch variable. Other environmental constraints that can also inhibit urban occupation are represented by the zoning variables as *protected 1* and *protected 2* (L_{prot-1} and L_{prot-2}). The difference between the both types of protected zones is the degree of restriction on residential occupation. According to the zoning legislation, areas assigned as *protected 1* cannot be occupied, while the areas *protected 2* can have a limited occupation.

Table 5.9 Explanatory variables for the U-SPRAWL binary logistic regression

Variable	*Description*	*Expected effect*	*Source*
$L_{d-urban}$	Distance to the closest urban patch	(-)	NetLogo-based calculations
$L_{neigh-urban}$	Number of urban patches within a 700 m radius	(+)	NetLogo-based calculations
$L_{neigh-pop}$	Number of households living within a 700 m radius	(+)	Census data + NetLogo-based calculations
$L_{d-roads}$	Distance to main roads	(-)	GIS-based calculations
L_{d-CBD}	Distance to Central Business District (CBD)	(-)	GIS-based calculations
L_{slope}	Terrain slope	(-)	GIS-based calculations
L_{prot-1}	Zoning: protected area 1 (no residential occupation)	(-)	Zoning map
L_{prot-2}	Zoning: protected area 2 (allows limited residential occupation)	(-)	Zoning map
L_{res}	Zoning: residential area	(+)	Zoning map
L_{mixed}	Zoning: mixed area	(+)	Zoning map
L_{ind}	Zoning: predominantly industrial area	(+)	Zoning map
L_{social}	Zoning: area of social interest	(+)	Zoning map
L_{vacant}	Zoning: vacant urban land	(+)	Zoning map

Additional zoning variables were also included in the binary logistic model: the variable L_{res} represents *zones that are exclusive for residential use*; L_{mixed} represents *mixed zones where residential, commercial and institutional uses are allowed*; L_{ind} represents *zones that are mainly industrial, although other uses are also allowed*; L_{social} are *areas of social interest*, i.e., areas selected for social housing projects; and L_{vacant} are *areas of vacant urban land*. These variables should have a positive effect on the transition of non-urban patches to urban because they represent zones that expect some degree of residential occupation (Table 5.10). All the zoning variables were extracted from the map of São José dos Campos' Zoning Law 3721/1990 (PMSJC 1990) and incorporated into NetLogo.

Table 5.10 Estimated parameters for binary logistic model (N = 19,778)

Variable	*Coef.*	*Std. err.*
Distance to the closest urban patch ($L_{d-urban}$)	$-1.3(10^{-3})^{***}$	$1.0(10^{-4})$
Number of urban patches within a 700 m radius ($L_{neigh-urban}$)	$5.8(10^{-2})^{***}$	0.01
Number of households living within a 700 m radius ($L_{neigh-pop}$)	$1.4(10^{-3})^{***}$	$2.1(10^{-4})$
Distance to main roads ($L_{d-roads}$)	$-1.1(10^{-4})$	$0.9(10^{-4})$
Distance to CBD (L_{d-CBD})	$1.5(10^{-4})^{***}$	$2.4(10^{-5})$
Terrain slope (L_{slope})	$5.6(10^{-4})$	$3.9(10^{-4})$
Zoning: protected area 1 – no residential occupation (L_{prot-1})	2.41^{***}	0.19
Zoning: protected area 2 - allows limited residential occupation (L_{prot-2})	4.27^{***}	0.34
Zoning: residential area (L_{res})	4.63^{***}	0.25
Zoning: mixed area (L_{mixed})	3.06^{***}	0.17
Zoning: predominant industrial area (L_{ind})	8.09^{***}	1.02
Zoning: area of social interest (L_{social})	3.90^{***}	0.18
Zoning: vacant urban land (L_{vacant})	1.21^{***}	0.19
Constant	-5.56^{***}	0.24

Chi-square test: $\chi^2 = 4597.84^{***}$
Cox & Snell $R^2 = 0.207$
Nagelkerke $R^2 = 0.465$

Predicted	Percent of correct predictions $L_{transition} = 0$	$L_{transition} = 1$	Overall (N=19,778)
% Correct	90.9	68.1	89
% Incorrect	9.1	31.9	11

***, **, and * indicate statistical significance at the 99%, 95%, and 90% levels.

The goodness-of-fit for the model was assessed through the chi-square test, the Cox&Snell pseudo-R^2, the Nagelkerke pseudo-R^2, and the percent of correct prediction (Table 5.10). The chi-square test for fit resulted in a highly significant chi-square statistics (p<0.001), which rejects the null hypothesis that the model coefficients as a group are equal to zero. The Cox&Snell R^2 reflects the improvement of the full model over the intercept model, and it is equal to 0.207. Because its maximum value is less than 1, the Nagelkerke R^2 adjusts it so that the range of possible values extends to 1. The Nagelkerke R^2 for the model is equal to 0.465. The overall percent of correct predictions is equal to 89%.

With the exception of the variables L_{slope} and $L_{d-roads}$, all the explanatory variables were highly significant. The signs on the parameter estimates for $L_{d-urban}$, $L_{neigh-urban}$, $L_{neigh-pop}$, L_{res}, L_{mixed}, L_{ind}, L_{social}, and L_{vacant} support the hypotheses outlined earlier. On the other hand, the parameter estimates for *distance to CBD* (L_{d-CBD}), *protected area 1* (L_{prot-1}), and *protected area 2* (L_{prot-2}) revealed unexpected findings. The sign on the estimated coefficient for L_{d-CBD} opposes the original idea that non-urban patches close to the CBD are more likely to become urban. This finding is an indication that local policy makers in São José dos Campos have not established an effective policy to control the speculative retention of land in areas with infrastructure. This leads to unnecessary urban sprawl, which increases the need for investments in the expansion of infrastructure networks and promotes large-scale segregation of the poor.

The positive signs on the parameter estimated for L_{prot-1} and L_{prot-2} contradict the hypothesis that these protected areas are less likely to become urban. This indicates a deficiency in the state control over these areas during the period 1990-2000. In the case of L_{prot-2}, this finding is particularly important, since these areas were not supposed to have any type of human occupation.

In the U-SPRAWL sub-model, the binary logistic regression specified and estimated in this section is used for computing the local transition probability of a non-urban patch becoming urban ($p_{NU \to U}$). These non-urban patches are then sorted by $p_{NU \to U}$ in decreasing order. Following this rank, a random number evenly distributed over [0,1] is generated for each patch and compared to its $p_{NU \to U}$. In case the random number is smaller than $p_{NU \to U}$, the state of the non-urban patch is converted to urban. The U-SPRAWL sub-model repeats this procedure until the total number of new urban patches reaches the value established in the transitional phase ($T_{new-urban}$).

5.3.2 Dwelling offers

The housing stock plays an important role in the household residential dynamics, since it may discourage or motivate their choice for a certain neighborhood. In the MASUS framework, the dwelling offers (D-OFFER) sub-model is responsible for simulating the dynamics of the housing stock in a city (section 4.3.2). Like the urban sprawl sub-model, the D-OFFER has two phases: the *transitional phase*, which quantifies the overall gain and loss of dwelling offers during the period $t \rightarrow t+1$, and the *allocation phase*, which indicates the patches that gained and lost dwellings during the same period.

Transitional phase

The transitional phase computes three global variables related to the housing stock dynamics in the period $t \rightarrow t+1$: total number of dwellings in t+1 (${}^{t+1}T_{dwe}$), total number of dwelling loss during $t \rightarrow t+1$ ($T_{dwe-loss}$), and total number of new dwellings during $t \rightarrow t+1$ ($T_{dwe-gain}$). To obtain ${}^{t+1}T_{dwe}$, it is necessary to compute the housing stock of the city, which is equivalent to the proportion τ_{stock} of the occupied dwellings in t+1 (equation (4.29)). In the D-OFFER sub-model, the housing stock was considered as equivalent to 8% of the occupied dwellings in the city (τ_{stock} = 0.08). This estimation was provided by the Association of Construction Companies in Vale do Paraíba (ACONVAP), based on real estate market surveys conducted by the association. Although in reality this number varies along the years, the current version of MASUS considers it as constant.

The total number of dwelling loss during $t \rightarrow t+1$ ($T_{dwe-loss}$) is equal to a proportion τ_{loss} of the total number of dwellings in the city (equation (4.30)). Since there is no available data about the loss of residential dwellings due to the expansion of non-residential uses, this value was estimated from the 1991 and 2000 census data. For that, the population in the census tract was considered as a proxy of the housing stock in that area. The census tracts that presented a decrease in population during the period 1991-2000, which are those located close to downtown, were selected. The total household loss in these areas was considered as a proxy of the total loss of residential dwellings during the period. Decomposing it to an annual rate, 0.6% of the total dwellings are converted to non-residential uses (e.g., offices and shops) over the period of an year (τ_{loss} =0.006).

Finally, after obtaining ${}^{t+1}T_{dwe}$ and $T_{dwe-loss}$, the number of new dwellings to be created during $t \rightarrow t+1$ ($T_{dwe-gain}$) can be computed according to the equation (4.31).

Allocation phase

The allocation phase indicates where the gain and loss of dwellings will take place and, based on that, updates the landscape-patch variables *total number of dwellings* (L_{dwe}) and *dwelling offers* ($L_{dwe-offer}$). This process relies on two linear regression models: One that predicts the loss of dwellings in landscape patches (Y_{loss}), and another that predicts the gain of dwellings (Y_{gain}) (Table 5.11). For estimating these models, the annual average decrease in residents living in the landscape patch during the period 1991-2000 was considered as a proxy of Y_{loss}, while the annual average increase in residents living in the patch was considered as a proxy of Y_{gain}. The variables Y_{loss} and Y_{gain} were computed in NetLogo, after importing the household micro-data obtained from the censuses 1991 and 2000 into a NetLogo world (see section 6.2.1).

Regarding the model that estimates the loss of dwellings (Y_{loss}), the landscape-patch variables *number of households* (L_{pop}), *distance from CBD* (L_{d-CBD}), and *land value* (L_{value}) are considered in the model based on the hypothesis that the expansion of commercial use into residential neighborhoods usually happens in areas that are densely populated, close to downtown, and with high land values. Because the landscape patches have the same size, the L_{pop} is equivalent to population density.

In addition, the expansion of other land uses into residential areas should only happen in areas where non-residential uses are allowed. Thus, the zoning variables representing areas adequate for commercial, service, and industrial uses are included in the model (L_{CZ*pop}, $L_{MZ-1*pop}$, $L_{MZ-2*pop}$, $L_{ITZ*pop}$, and $L_{PIZ*pop}$). These variables were interacted with L_{pop}, assuming that dwelling losses only occur where there is a residential occupation.

Table 5.11 Descriptive statistics and sources of dependent and explanatory variables for the linear regression models for estimating the loss of dwellings (model 1) and gain of dwellings (model 2) in landscape patches (Yloss) , N = 6,247.

Variable	*Description*	*Min*	*Max.*	*Mean*	*Std. dev.*	*Model/ Expected effect*	*Source*
Y_{loss}	Loss of dwellings *dependent variable of model 1*	0	18	0.324	0.8	1 dependent	Census data + NetLogo-based calculations
Y_{gain}	Gain of dwellings *dependent variable of model 2*	0	39	0.789	2.7	2 dependent	Census data + NetLogo-based calculations
L_{d-CBD}	Distance from CBD (m)	0	13596	5299	3030	1 (-) 2 (+)	GIS-based calculations
$L_{d-roads}$	Distance from main roads (m)	0	1487	141.2	176.9	2 (-)	GIS-based calculations
L_{value}	Land value (minimum wages/m^2)	0.03	3.602	1.015	0.626	1 (+) 2 (-)	Property advertisements
L_{pop}	Number of households (population density)	1	280	16.91	14.74	1 (+) 2 (-)	Census data + NetLogo-based calculations
L_{FAR}	Floor Area Ratio (FAR)	0.02	4.00	2.49	0.74	2 (+)	Zoning map
$L_{pop\,/\,FAR}$	L_{pop} divided by the FAR	0.25	3050	8.62	50.25	1 (+)	Zoning map + NetLogo-based calculations
$L_{CZ\,*\,pop}$	Central zone multiplied by L_{pop}	0	62	0.3727	3.09	1 (+)	Zoning map + NetLogo-based calculations
$L_{MZ\,-1*\,pop}$	Mixed zone 1 (FAR = 1.3) multiplied by L_{pop}	0	81	0.3983	3.40	1 (+)	Zoning map + NetLogo-based calculations
$L_{MZ\,-2*\,pop}$	Mixed zone 2 (FAR = 3) multiplied by L_{pop}	0	164	2.41	8.44	1 (+)	Zoning map + NetLogo-based calculations
$L_{ITZ\,*\,pop}$	Industrial transition zone multiplied by L_{pop}	0	52	0.06	1.35	1 (+)	Zoning map + NetLogo-based calculations
$L_{PIZ\,*\,pop}$	Predominant industrial zone multiplied by L_{pop}	0	171	0.28	2.94	1 (+)	Zoning map + NetLogo-based calculations

Areas that, according to the zoning legislation, reach an occupation level near the saturation point are also more likely to lose residential dwellings due to the expansion of other uses. For this reason, the variable $L_{pop\,/\,FAR}$ is considered in the model. This variable divides L_{pop} by the *Floor Area Ratio* (FAR) specified for the area. The

FAR is a zoning instrument for controlling the density and size of buildings. Each zoning district has an FAR control which, when multiplied by the lot area of the zoning lot, produces the maximum amount of floor area allowable in a building on the zoning lot. Landscape patches with high values for $L_{pop/FAR}$ are closer to the saturation point established by the zoning legislation than those with low $L_{pop/FAR}$.

The F-statistic test indicates that the linear regression model is able to explain significantly the variation of the loss of dwellings (p < 0.001) (Table 5.12). The R^2 of 0.53 means that 53% of the observed variance of dwelling loss is explained by the model.

Table 5.12 Results of the linear regression model for estimating the loss of dwellings in landscape patches (Y_{loss}), N=6247.

Variable	*Unstandardized coefficient*	*Std. error*
(Constant)	-0.234^{***}	0.019
Distance from CBD (L_{d-CBD})	$-3(10^{-7})$	0.000
Land value (L_{value})	0.005	0.004
Number of households (L_{pop})	0.030^{***}	0.001
L_{pop} divided by FAR ($L_{pop/FAR}$)	0.001^{**}	0.000
Central zone multiplied by L_{pop} (L_{CZ*pop})	0.020^{***}	0.003
Mixed zone 1 (FAR=1.3) multiplied by L_{pop}	0.012^{***}	0.002
Mixed zone 2 (FAR=3) multiplied by L_{pop}	0.005^{**}	0.001
Industrial transition zone multiplied by L_{pop} ($L_{ITZ*pop}$)	0.027^{***}	0.006
Predominant industrial zone multiplied by L_{pop} ($L_{PIZ*pop}$)	0.058^{***}	0.003
F-statistic test: $F = 729.89^{***}$		
$R^2 = 0.532$		

***, **, and * indicate statistical significance at the 99%, 95%, and 90% levels.

Explanatory variables that have significant effects on dwelling loss are L_{pop} (+), $L_{pop/FAR}$ (+), L_{CZ*pop} (+), $L_{MZ-1*pop}$ (+), $L_{MZ-2*pop}$ (+), $L_{ITZ*pop}$ (+), $L_{PIZ*pop}$ (+). The directions in which these variables operate (all positive) support the hypotheses related to the importance of population density, zoning variables, and the interaction between both for the loss of dwellings, since they are able to encourage the expansion of non-residential land uses.

Regarding the linear regression model that estimates the gain in dwellings (Y_{gain}), the explanatory variables L_{pop} , L_{value} and L_{CBD} were included in the model based on the hypothesis that urban patches with low occupancy, low land prices, and distant from CBD are more likely to attract new investments in residential developments (Table 5.13). In addition, the model includes the variable *distance to main roads* ($L_{d-roads}$), based on the hypothesis that real estate developers prefer to invest in areas that can be easily accessed by roads. The *floor area ratio* (FAR) defined by the zoning legislation is also likely to influence the attractiveness of urban patches for real estate developers. Areas with high FAR can accommodate more residences, especially in those areas where the occupancy has not reached a saturated level. To test this hypothesis, the variables L_{FAR} and $L_{pop/FAR}$ were considered in the model.

Table 5.13 Results of the linear regression model for estimating the gain of dwellings in landscape patches (Y_{gain}), N=6,247.

Variable	*Unstandardized coefficient*	*Std. error*
(Constant)	0.212^{***}	0.076
Distance from CBD (L_{d-CBD})	$6.4(10^{-5})^{***}$	0.000
Distance from main roads ($L_{d-roads}$)	-0.001^{***}	0.000
Land value (L_{value})	0.008	0.006
Number of households (L_{pop})	-0.046^{***}	0.001
Floor area ratio - FAR (L_{FAR})	0.261^{***}	0.024
L_{pop} divided by FAR ($L_{pop/FAR}$)	0.001	0.001
F-statistic test: $F = 647.52^{***}$		
$R^2 = 0.483$		

***, **, and * indicate statistical significance at the 99%, 95%, and 90% levels.

The F-statistic presents a significant value ($p < 0.001$) and shows that the model is capable of explaining the increase in residential units. The R^2 of 0.48 indicates that 48% of the observed variance in the gain in dwellings can be explained by the model.

Explanatory variables that have significant effects on the dwelling gain are L_{d-CBD} (+), $L_{d-roads}$ (-), L_{pop} (-), and L_{FAR} (+). All the significant variables presented the expected effect on the model, confirming the hypothesis that an increase in residential units tends to occur in areas that are distant from downtown and less densely populated (more availability of land), close to roads (better accessibility), and with higher floor area ratio (permission for higher building density).

The D-OFFER sub-model adopts the results from both linear regression models (Table 5.12 and 5.13) to estimate the Y_{loss} and Y_{gain} of each urban patch during a simulation cycle. Since the sum of dwelling loss and gain has to meet the global values computed in the transitional phase ($T_{dwe-loss}$ and $T_{dwe-gain}$, respectively), the local transitions Y_{loss} and Y_{gain} are normalized by a factor (equations (4.34) and (4.35)). Then, new values for the landscape-patch variables *total number of dwellings* (L_{dwe}) and *dwelling offers* ($L_{dwe-offer}$) are computed according to the equations (4.36) and (4.37).

5.3.3 Infrastructure

The infrastructure sub-model (INFRA) simulates the dynamics of the landscape-patch variable *infrastructure* (L_{infra}), which is a composed index that ranges from 0 to 1, and represents the provision of water, sewage, and garbage collection. This sub-model relies on a linear regression equation that explains the *annual improvement in* L_{infra} (Y_{infra}) (Table 5.14 and 5.15). To obtain the dependent variable Y_{infra}, the difference between the infrastructure quality (L_{infra}) in 1991 and 2000 was calculated in NetLogo and divided by 9, which is the number of years covered during the period 1991-2000.

Based on the hypothesis that patches with better accessibility are more likely to achieve higher infrastructure quality, the variables *distance from CBD* (L_{d-CBD}) and *distance from main roads* ($L_{d-roads}$) were included in the model as explanatory variables and are expected to have a negative effect. Considering that the municipal government is constantly expanding the provision of water, sewage, and garbage collection, the magnitude of the infrastructure improvement (Y_{infra}) tends to be higher in those areas that had very little or no infrastructure. For this reason, the infrastructure quality index (L_{infra}) was considered in the model and it is expected to have a negative effect on Y_{infra}.

Public investments also tend to be prioritized in the most densely populated areas. To capture this aspect, the variable *number of households* (L_{pop}) was included in the model and is expected to have a positive effect. Finally, because public investments in infrastructure exclude settlements that do not belong to the 'legal city', the binary variable *irregular settlement* (L_D) was also considered and is expected to have a negative effect.

Table 5.14 Descriptive statistics and sources of dependent and explanatory variables for the linear regression model for estimating annual improvement in the infrastructure of landscape patches (Y_{infra}), N = 6,781.

Variable	*Description*	*Min*	*Max.*	*Mean*	*Std. dev.*	*Expected effect*	*Source*
Y_{infra}	Annual improvement in the infrastructure	0	0.111	0.011	0.018	Dependent variable	Census data
L_{pop}	Number of households (population density)	1	280	16.87	14.32	(+)	Census data + NetLogo-based calculations
L_{d-CBD}	Distance from CBD (m)	0	13730	5489	3149	(-)	GIS-based calculations
$L_{d-roads}$	Distance from main roads (m)	0	1341	142.9	175.65	(-)	GIS-based calculations
${}^{t-1}L_{infra}$	Infrastructure quality index in the previous year	0	1	0.848	0.262	(-)	Census data
L_D	Irregular settlements (type D)	Binary variable: 0 (6276 patches – 92.6%) 1 (505 patches – 7.4%)				(-)	Neighborhood type map

The F-statistic test indicates that the model is able to explain significantly the annual improvement in the infrastructure index (p<0.001) (Table 5.15). The R^2 of 0.841 means that 84.1 % of the variation in Y_{infra} is explained by the model. This indicates a very good fit of the model to the observed data. All explanatory variables presented significant effects on the improvement in the infrastructure index. The direction in which these variables operate support the hypotheses presented in the above.

The INFRA sub-model adopts the results of the regression model (Table 5.15) to compute the Y_{infra} of each urban patch. Based on this value, the landscape patch variable is updated after each annual cycle (${}^{t+1}L_{infra} = {}^{t}L_{infra} + Y_{infra}$).

Table 5.15 Results of the linear regression model for estimating annual improvement in the infrastructure of landscape patches (Y_{infra}), N = 6781.

Variable	*Unstandardized coefficient*	*Std. error*
(Constant)	0.078^{***}	0.001
Distance from CBD (L_{d-CBD})	$-7.7(10^{-8})^{**}$	0.000
Distance from main roads ($L_{d-roads}$)	$-8.6(10^{-6})^{***}$	0.000
Number of households (L_{pop})	$2.01(10^{-5})^{***}$	0.001
Infrastructure quality index - previous year (${}^{t-1}L_{infra}$)	-0.077^{***}	0.000
Irregular settlements - type D (L_D)	-0.017^{***}	0.000
F-statistic test: $F = 7162^{***}$		
$R^2 = 0.841$		

***, **, and * indicate statistical significance at the 99%, 95%, and 90% levels.

5.3.4 Land value

The land value sub-model (L-VALUE) is responsible for updating the landscape variable *land value* (L_{value}) after each annual cycle. For each urban patch, this sub-model calculates the land value based on a linear regression model (Table 5.16 and 5.17). The dependent variable L_{value} of the regression model represents the land value of the landscape patch (minimum wages/m^2) in the year 2000. Because there is a relation between the income of resident families and land value, the variables *proportion of low-income households* ($L_{prop-up2}$) and *proportion of high-income households* ($L_{prop-more20}$) were included in the model as explanatory variables (Table 5.16). In addition, this relation is likely to be influenced by the land-use restrictions in the neighborhood. To test this hypothesis, the variables $L_{prop-up2}$ and $L_{prop-more20}$ were interacted with the binary variables *residential zone* and *mixed zone*, and generated the following explanatory variables: *residential zone interacted with the proportion of low-income households* (L_{ZR-up2}), *residential zone interacted with the proportion of high-income households* ($L_{ZR-more20}$), *mixed zone interacted with the proportion of low-income households* (L_{ZM-up2}), and *mixed zone interacted with high-income households* ($L_{ZM-more20}$).

Table 5.16 Descriptive statistics and sources of dependent and explanatory variables for the linear regression model for estimating the land value of landscape patches (L_{value}), N = 8181.

Continuous Variables

Variable	*Description*	*Min*	*Max.*	*Mean*	*Std. dev.*	*Expected effect*	*Source*
L_{value}	Land value (minimum wages/m^2)	0.03	2.60	1.01	0.63	Dependent variable	Property advertisements 2000
$L_{prop-up2}$	Proportion of low-income households (up to 2 minimum wages)	0	1	0.275	0.173	(-)	Census 2000
$L_{prop-more20}$	Proportion of high-income households (more than 20 minimum wages)	0	1	0.104	0.189	(+)	Census 2000
L_{d-CBD}	Distance from CBD (m)	0	13730	5908	3200	(-)	GIS-based calculations
$L_{d-roads}$	Distance from main roads (m)	0	1414	160	192	(-)	GIS-based calculations
L_{infra}	Infrastructure quality index	0.14	1	0.9	0.18	(-)	Census 2000
L_{ZR-up2}	Residential zone * proportion of low-income households	0	0.48	0.0078	0.033	(-)	Zoning map + Census 2000
$L_{ZR-more20}$	Residential zone * proportion of high-income households	0	0.9	0.0499	0.121	(+)	Zoning map + Census 2000
L_{ZM-up2}	Mixed zone * proportion of low-income households	0	0.83	0.185	0.17	(-)	Zoning map + Census 2000
$L_{ZM-more20}$	Mixed zone * proportion of high-income households	0	0.9	0.05	0.12	(+)	Zoning map + Census 2000

Binary Variables

Variable	*Description*	*0*	*1*	*Expected effect*	*Source*
L_D	Irregular settlement (type D)	7124 patches (87.1%)	1057 patches (12.9%)	(-)	Neighborhood type map
$L_{enclosed}$	Enclosed settlement	7641 patches (93.4%)	540 patches (6.6%)	(+)	Field observation
L_{ZC}	Central Zone	8066 patches (98.6%)	115 patches (1.4%)	(+)	Zoning map

The F-statistic test indicates that the regression model is able to explain significantly the change in land price ($p<0.001$) (Table 5.17). Quantitatively, the R^2 shows that the model is able to explain 81.4 % of the variation in land prices, which indicates a good fit of the model to the observed data. All explanatory variables presented highly significant effects on the improvement in the infrastructure index. The direction in which these variables operate supports the hypotheses presented above. Nevertheless, the interaction between the proportion of low-/high-income families and zoning variables revealed interesting outputs. As expected, low-income households have a negative impact on land prices while high-income households have a positive impact. The model revealed, however, that these impacts are more substantial in mixed areas than in exclusively residential neighborhoods. This may happen because the land-use flexibility of mixed zones increases the ability of developers and real estate agents to stimulate competition for the best locations, and, therefore, increase prices.

Other neighborhood aspects that are expected to influence land prices and were included in the model are related to the type of settlement where the patches are located, their accessibility, and provision of infrastructure. Patches located in the *Central Business District* of the city (L_{ZC}) or in *enclosed settlements* for high and middle classes ($L_{ENCLOSED}$) are expected to present high land values. On the other hand, *irregular settlements* (L_D) are expected to present lower land values. Patches with poor accessibility, i.e., long *distance from the CBD* (L_{d-CBD}) or long *distance from the main roads* ($L_{d-roads}$), and with low *infrastructure quality* (L_{infra}) are also expected to have low prices.

Table 5.17 Results of the linear regression model for estimating annual improvement the land value of landscape patches (L_{value}), N = 6781.

Variable	*Unstandardized coefficient*	*Std. error*
(Constant)	1.398***	0.031
Proportion of low-income households ($L_{prop-up2}$)	-0.669***	0.031
Proportion of high-income households ($L_{prop-more20}$)	1.047***	0.032
Distance from CBD (L_{d-CBD})	$-5.5(10^{-5})$***	0.000
Distance from main roads ($L_{d-roads}$)	$-6.2\ (10^{-5})$ ***	0.000
Infrastructure quality index (L_{infra})	0.099 ***	0.029
Residential zone * proportion of low-income households (L_{ZR-up2})	0.408 ***	0.100
Residential zone * proportion of high-income households ($L_{ZR-more20}$)	0.039 ***	0.040
Mixed zone * proportion of low-income households (L_{ZM-up2})	-0.293 ***	0.027
Mixed zone * proportion of high-income households ($L_{ZM-more20}$)	0.432 ***	0.037
Irregular settlements - type D (L_D)	-0.475 ***	0.016
Enclosed settlement ($L_{enclosed}$)	0.178 ***	0.017
Central Zone (L_{ZC})	0.411 ***	0.027
F-statistic test: F = 2747 ***		
$R^2 = 0.814$		

***, **, and * indicate statistical significance at the 99%, 95%, and 90% levels.

6 OPERATIONAL MASUS MODEL AND SIMULATION EXPERIMENTS

Given the MASUS theoretical framework (Chapter 4) and the empirical parameters (Chapter 5), this chapter presents an operational MASUS model with a range of functions for testing theories and policies on segregation. The chapter is organized in two main sections. The first section introduces the MASUS computer program, including a brief description of its procedures and graphic-user interface, and the second illustrates the potential of the model through three sets of experiments on segregation in São José dos Campos. The first set of experiments aims at validating the model. It compares simulated data that replicates a past segregation state of the city with empirical data, and checks whether the model provides an accurate representation of the segregation patterns in São José dos Campos. The aim of the second set of experiments is to demonstrate how MASUS can be used to explore theoretical issues of segregation. Finally, the third set of experiments shows the model's ability to provide insights about the impact of anti-segregation policies.

6.1 Implementation of an operational MASUS model

The first operational MASUS model was implemented in NetLogo 4.0.4, a cross-platform multi-agent programmable modeling environment (Wilensky 1999). The program includes the following main sub-programs/procedures (in order of execution):

1. *Initialization* (complex procedure[1]): It includes routines to import datasets and to set initial global parameters. The routine to import datasets (*ImportDatasets*) uses the NetLogo's GIS extension 1.0 to load vector and raster GIS data into NetLogo. Household data are loaded and assigned to mobile agents (household agents), while data related to the urban environment are assigned to a grid of stationary agents (urban landscape patches). The routine to set initial global parameters (*SetInitialParameters*) includes lists of empirically defined parameters that are relevant to the sub-models responsible for creating new households (population transition sub-model, see section 4.3.1) and assigning new urban patches (urban sprawl sub-model, see section

[1] As in Le (2005), a complex procedure here means that it contains one or more procedures.

4.3.2). Considering the large amount of data usually required by empirically based urban simulations, *Initialization* is a time-consuming procedure, which imports a large number of external files. To simplify this process, the current operational MASUS offers an optional procedure, where the user can simply import 'world-files' that represent the state of the study area at a specific point in time (*ImportCity*). These files were created through the NetLogo's primitive 'export-world', which writes the values of all variables/parameters of the system to a single external file (worksheet format).

2. *HouseholdChoice* (complex procedure): It performs the households' residential choice according to the specifications of the decision-making sub-model (section 4.3.1). It is the most time-consuming procedure of the MASUS simulation protocol, since it computes the probability of each household to choose among different residential alternatives.
3. *ComputeSegregation*: It calculates global and local segregation indices (section 2.5) for different neighborhood scales.
4. *DrawGraphs*: It draws graphs of different population indicators.
5. *HouseholdTransition*: It updates changes in the household profile by following the specifications of the household transition sub-model (section 4.3.1).
6. *PopulationTransition*: It creates new households with profiles that meet the expected socio-demographic composition of the population as a whole. It follows the specifications of the population transition sub-model (section 4.3.1).
7. *UrbanSprawl* (complex procedure): It performs the transition from non-urban areas (patches) to urban according to the specifications of the urban sprawl sub-model (section 4.3.2).
8. *DwellingOffers* (complex procedure): It updates the number of dwellings in each urban patch according to the specifications of the dwelling offers sub-model (section 4.3.2).
9. *Infrastructure*: It updates the level of infrastructure quality of each urban patch according to the specifications of the infrastructure sub-model (section 4.3.2).
10. *LandValue*: It updates the land value of each urban patch according to the specification of the land value sub-model (section 4.3.2).

The process of developing the operational MASUS also included its verification, i.e., checking whether the program executes what is stated in the theoretical specification. The verification of the MASUS model was performed for each sub-program/procedure as well as for the simulation program as a whole. During the verification process, intermediate outputs were constantly recorded and checked step-by-step, often through comparisons with calculations done in a spreadsheet. The verification also included tests for improving the program code to achieve better performance.

6.1.1 Inputs and outputs

Inputs

Inputs for simulations with MASUS include spatial data and parameters. Spatial data for initializing the MASUS simulation can be distinguished as: vector GIS data (points, lines, and polygons), in shapefile format (.shp); and raster GIS data (grids), in ascii format (.asc). Input data about the urban environment can be provided as raster or vector polygons, while household data is provided as vector points.

The second type of input, the parameters, can be distinguished as modeler's input parameters and user's input parameters (Le 2005; Le et al. 2008). Modeler's input parameters are those that are not exposed to users. They can represent global control parameters or coefficients obtained from empirical analyses (like the coefficients presented in Chapter 5). User's input parameters are inputs that can be set to test theories and policy approaches on segregation. They can be easily modified at the graphic-user interface.

Outputs

During a simulation, the MASUS model provides different types of outputs: population statistics, segregation indices, graphs, maps, and simulated worlds. The population statistics include the total number of households, the number of households belonging to each income group, and the Gini index (an income-inequality measure). The segregation indices include global and local measures (section 2.5). The MASUS program adopts two different scales for calculating these segregation indices. The first considers that the neighborhood of a household comprises the area within a 700 m radius of the household's residence (local scale), and the second considers a 2000 m radius (large scale).

Population statistics and global segregation indices are reported and plotted as graphs. The graphs provide an overview of the change in statistics and indices along the years and can be exported as worksheet format files for further analysis. Another graph presented by the MASUS program is the Lorenz curve, which complements the information provided by the Gini index. The MASUS model uses algorithms from the Wealthy Distribution model of Wilensky (1998) for reporting the Gini index and plotting the Lorenz curve.

The local segregation indices are shown as maps, which can be displayed through command buttons, and allow users to visually identify the most/least segregated areas and how these areas changed over time. Another type of output produced by the MASUS program is the simulated world, a worksheet file containing all the information about the simulated system. The program records a simulated world after each annual cycle, allowing users to retrieve this information at any time.

6.1.2 Graphic user interface

The graphic user interface (GUI) of the MASUS model for the city of São José dos Campos is presented in Figure 6.1. Element (1) represents command buttons for initializing and starting the simulation:

- The 'Initialization' button executes the procedures to import GIS data from São José dos Campos (year 1991).
- The 'Import City 1991' button imports the NetLogo world representing São José dos Campos in the year 1991 (faster than the initialization procedure).
- The 'Import City 2000' button imports the NetLogo world representing São José dos Campos in the year 2000.
- The 'Start Simulation' button executes the procedures corresponding to the simulation annual cycle. These procedures are repeatedly executed until the user clicks the button again to stop the action.

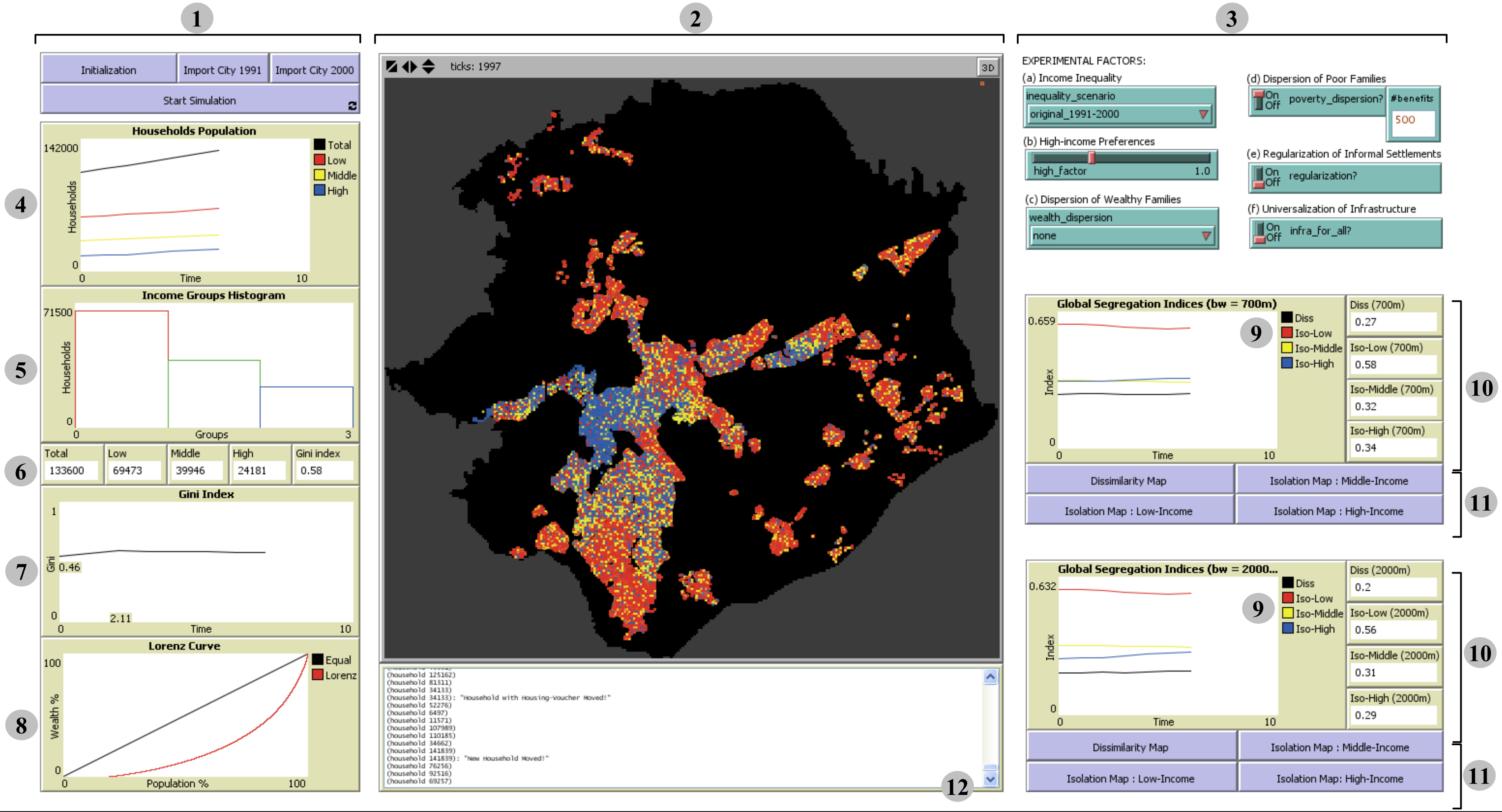
Element (2) is a graphic window that displays segregation maps or the location of households in the city. In the example, the graphic window is showing the location of low-, middle-, and high-income households (colors red, yellow, and blue, respectively) and the limits of the study area (in black). Using the mouse, the user can right-click on landscape patches and agents, and inspect the value of variables that comprise their states. To display segregation maps (e.g., Figure 6.2), the user should press the command buttons indicated as (11).

Element (3) represents the input parameters chosen by the user to test theories and policy approaches on segregation. These input parameters focus on experimental factors that address demographic aspects, personal preferences or urban policies:

- Income inequality (demographic aspect): The input parameter is provided through a chooser, where the user can select scenarios for the period 1991-2000 with original, increasing or decreasing income inequality levels, or scenarios for the period 2000-2010 with constant, increasing or decreasing inequality levels.
- High-income preferences (personal preferences): The input parameter is provided through a numeric slider, where the user can choose the value of θ_{neigh}, which is a factor that establishes the relevance of the neighborhood income composition to the decision making of affluent households. If θ_{neigh} is equal to 0, affluent households do not consider the income of their neighbors when selecting their residential locations; if θ_{neigh} is equal to 0, the preference level of affluent households for having neighbors similar to themselves is equal to the original level (calibrated from empirical data); if θ_{neigh} is equal to 3, the affluent households' preference for having neighbors with similar income is three times higher than the original level.
- Dispersion of wealthy families (urban policy): The input parameter is provided through a chooser that allows users to select a scenario where non-occupied areas close to poor neighborhoods are converted to residential developments for middle and upper classes.
- Dispersion of poor families (urban policy): The input parameters are provided through a switch (on/off) and an input box, which allow users to test policies that distribute housing vouchers for moving low-income households out from distressed areas. Once the switch is turned on, the program executes a

procedure that selects *n* poor households from areas with high levels of poverty isolation for receiving housing vouchers. The number of households who received the benefit per year is provided by the user through the input box. The selected households can only move to neighborhoods that present low levels of poverty isolation.

- Regularization of informal settlements (urban policy): The input parameter is provided through a switch (on/off). Once the switch is turned on, the program executes a procedure that converts the clandestine settlements to regular. This measure has implications for the provision of infrastructure and land value of these settlements.
- Universalization of infrastructure (urban policy): The input parameter is provided through a switch (on/off). Once the switch is turned on, the program executes a procedure that changes, for all urban patches, the value of the patch variable 'infrastructure quality' (L_{neigh}) to 1 (maximal value).



1 Command buttons for initializing and starting the simulation	5 Graph: Histogram of income groups	9 Time-series graph: Global segregation indices
2 Graphics window	6 Monitors: Number of households (total and per group) and Gini index	10 Monitors: Global segregation indices
3 User's input parameters	7 Time-series graph: Gini index	11 Command buttons for displaying local segregation indices (map) on the graphics window
4 Time-series graph: Number of households	8 Graph: Lorenz curve	12 Output area for monitoring the status of the simulation run

Figure 6.1 The MASUS graphic user interface.

The elements (4) to (8) provide population statistics during the simulation run in the form of graphs and monitors. Element (4) is a graph that presents the total number of households and the number of households belonging to each income group along the years. Element (5) is a histogram showing the number of households belonging to each income group. This histogram provides an overview of the income composition of the population. Element (6) represents monitors showing the total number of households, the number of households belonging to each income group, and the Gini index of the current year. Element (7) is a graph that illustrates the evolution of the Gini index through the years, and element (8) is a graph representing the Lorenz curve in the current year.

The elements (9) to (11) allow users to monitor global and local indices of segregation computed for different scales of neighborhoods (700 m and 2000 m). Element (9) represents graphs that show the evolution of global segregation indices through the years in two different scales. Element (10) is a set of monitors that provide the values of global segregation indices for the current year. Element (11) represents a set of command buttons for displaying local segregation indices of the current year in the graphic window (as segregation maps).

Finally, element (12) is an output area that informs the users about the status of the simulation run, i.e., which sub-model/procedure is being executed. During the execution of the decision-making sub-model, the output area also shows the identification number of households that are evaluating their residential alternatives, and whether they decided to move to another location or not.

6.2 Simulation experiments I: Comparing simulated outputs with empirical data

The first simulation experiment validates the MASUS model regarding the fit between simulated and real data. It tests whether the model provides an accurate representation of the segregation dynamics in São José dos Campos. The initial state of this simulation experiment replicates the characteristics of the city in 1991. Nine annual cycles were executed and the simulation results were compared with real data from the year 2000. Details about the inputs and outputs of the simulation are presented in the following paragraphs.

6.2.1 Initial state of the simulation

Spatial data

To reproduce relevant characteristics of São José dos Campos in 1991, GIS data from different sources were imported into NetLogo (Table 6.1) and assigned to MASUS entities (household agents or urban landscape-patches).

Table 6.1 Spatial data of São José dos Campos imported into NetLogo.

	MASUS entity	*Description*	*GIS file format*	*Source*
1	Household agent	Household locations and attributes	Vector point (shapefile)	Census 1991, universal microdata (IBGE 1991a)
2	Urban landscape patch	Urban/non-urban areas	Vector polygon (shapefile)	Satellite image LANDSAT-5/ TM (INPE 1990)
3	Urban landscape patch	Zoning law	Vector polygon (shapefile)	Zoning Law 3721/1990 (PMSJC 1990)
4	Urban landscape patch	Neighborhood types	Vector polygon (shapefile)	Neighborhood type map (section 5.2.2)
5	Urban landscape patch	Infrastructure quality	Vector polygon (shapefile)	Census 1991 (IBGE 1991b)
6	Urban landscape patch	Land value per m^2	Vector polygon (shapefile)	Property advertisements
7	Urban landscape patch	Distance to Central Business District (CBD)	Raster (ascii grid)	GIS-based calculation
8	Urban landscape patch	Distance to roads	Raster (ascii grid)	Road map (PMSJC 2003) + GIS-based calculation
9	Urban landscape patch	Terrain slope	Raster (ascii grid)	Topographic map (PMSJC 2003) + GIS-based calculation

The household microdata from the Census 1991 (no. 1 in Table 6.1) were originally provided as text files (.txt) containing information about the population of São José dos Campos and the respective dwellings, including the identification number of the census tracts where these are located. The text files were processed to create a single file containing the following information about the 106,591 households living in the study area in 1991: household identification number, household location (census tract), head of household (HoH) income, HoH education, HoH age, household size, presence of children, and tenure status. To convert this information into spatially explicit vector points, two GIS files (shapefile format) were used as auxiliary data: one representing census tracts in 1991 (IBGE 1991b), and the other representing the urban areas in 1991 (no. 2 in Table 6.1). Each line of the resulting text file, which contains the data concerning one household, was converted into a vector point located within the occupied area (urban) of the census tract to which the respective household belongs. Then, these vector point data were imported into NetLogo and assigned to household agents.

With the exception of data nos. 1 and 6, all the imported spatial data shown in Table 6.1 are presented in Chapter 5. The data about urban/non-urban areas (no. 2) is described in section 5.3.1 and used for the empirical parameterization of the urban sprawl sub-model (U-SPRAWL). The vector data containing the zones delimited according to the São José dos Campos' Zoning Law 3721/1990 (no. 3 in Table 6.1) is mentioned in sections 5.3.1, 5.3.3, and 5.3.4, since it was considered in the empirical parameterization of the U-SPRAWL sub-model, the dwelling-offers sub-model (D-OFFER), and the land-value sub-model (L-VALUE).

The definition and characterization of neighborhood types in São José dos Campos (no. 4 in Table 6.1) are described in section 5.2.2, and were used in the parameterization of the decision-making sub-model (DECISION), as well as in the sub-models infrastructure (INFRA) and L-VALUE. The infrastructure quality (no. 5 in Table 6.1) is a composed index obtained from the Census 1991 (IBGE 1991b) for each census tract. This data were also used in the empirical parameterization of the INFRA and L-VALUE sub-models.

The average land value per m^2 of neighborhoods in São José dos Campos (no. 6 in Table 6.1) was obtained from property advertisements collected from local newspapers (Vale Paraibano) dated January to December 1991. This data collection was conducted at the Municipal Archive of the City of São José dos Campos. The spatial data regarding the distance to CBD, distance to roads, and terrain slope (nos. 7 to 9 in Table 6.1) were provided as raster files. They were produced based on GIS calculations using base maps (shapefile format) provided by the Municipal Government of São José dos Campos (PMSJC 2003).

Parameters

The modeler's input parameters of the first operational MASUS model for the city of São José dos Campos (Table 6.2) include coefficients of the statistical models presented in Chapter 5, as well as global control parameters obtained from different sources, including exploratory analysis of census data.

Table 6.2 Modeler's input parameters of the first operational MASUS model for the city of São José dos Campos .

	MASUS Sub-model	*Description*	*Source*
1	Decision-making	Coefficients of nested logit models that jointly model the household's mobility, neighborhood type choice, and specific neighborhood location choice	Statistical models (section 5.2.5)
2	Population transition	Population growth rate	Census data 1991 and 2000
3	Population transition	Global controls regarding the population (household) composition in terms of income, head of household's age, tenure status, presence of kids, family size per income group, and head of household's education per income group	Census data 1991 and 2000 (global controls for years in-between the period 1991-2000 are interpolated)
5	Urban sprawl (transitional phase)	Global transition probability of converting the land use from non-urban to urban ($P_{NU \to U}$).	LANDSAT Images + Markov chain (section 5.3.1)
6	Urban sprawl (allocation phase)	Coefficients of a binary logistic regression that estimates the local transitional probability of converting a non-urban patch to urban ($p_{NU \to U}$)	Statistical model (section 5.3.1)
7	Dwelling offers (transitional phase)	Global control for the housing stock (τ_{stock})	ACONVAP real estate market survey (section 5.3.2)
8	Dwelling offers (transitional phase)	Global control for the loss of residential dwellings due to the expansion of non-residential uses (τ_{loss})	Census data 1991 and 2000 (section 5.3.2)
9	Infrastructure	Coefficients estimated for a linear regression model for the annual improvement in infrastructure	Statistical model (section 5.3.3)
10	Land value	Coefficients estimated for a linear regression model for land value	Statistical model (section 5.3.4)

The user's input parameters (Table 6.3) are presented on the MASUS' graphic-user interface as experimental factors, and can be modified by the user to test segregation theories and policy strategies. The intent of this first simulation experiment, however, was not to test theories and policies, but to reproduce the segregation dynamics of São José dos Campos during the period 1991-2000 (baseline scenario).

Table 6.3 User's input parameters for baseline scenario

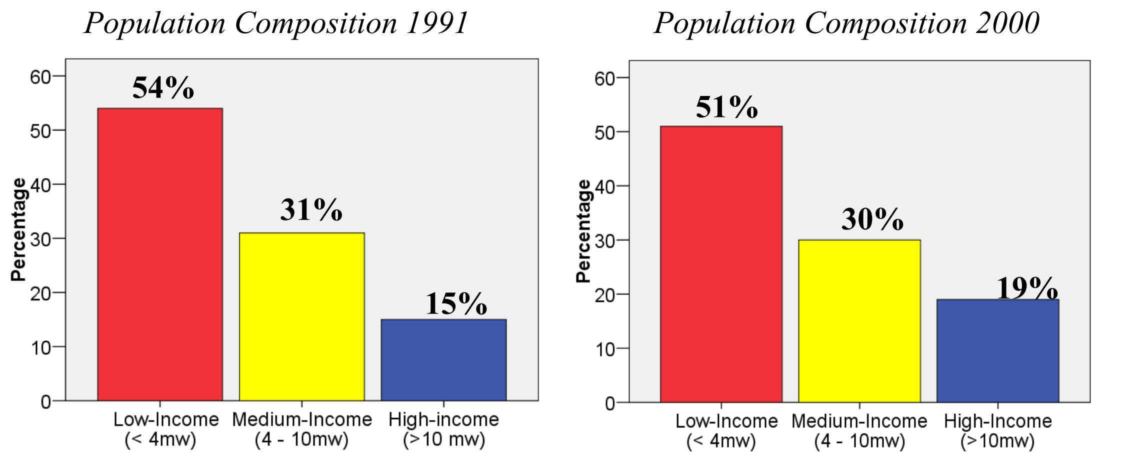
no.	*Experimental factor*	*Description*
1	*Income inequality* Factor: demographic aspects	*Chooser*: Inequality scenario = 'original 1991-2000' (scenario where the simulated income composition of the population is equal to the real one)
2	*High-income preferences* Factor: personal preferences	*Slider*: θ_{neigh} = 1 (preference of affluent households for having neighbors similar to themselves is equal to the one calibrated from empirical data)
3	*Dispersion of wealthy families* Factor: urban policies	*Chooser*: Wealthy dispersion = 'none'
4	*Dispersion of poor families* Factor: urban policies	*Switch*: Poverty dispersion? = 'off' *Input box*: # benefits = deactivated (0)
5	*Regularization of informal settlements* Factor: urban policies	*Switch:* Regularization? = 'off'
6	*Universalization of infrastructure* Factor: urban policies	*Switch:* Infrastructure for all? = 'off'

6.2.2 Results

After setting the initial state of the experiment, nine simulation annual cycles were executed in order to reproduce the segregation dynamics of São José dos Campos during the period 1991-2000. The simulated results were compared with real data from the year 2000. A calibration consisting of small changes in the input parameters of the decision-making sub-model (no. 1 in Table 6.2), originally obtained from the estimation of nested logit models (section 5.2.5), improved the fit between the simulated and real data.

Table 6.4 presents the results of global segregation indices computed for the initial state (year 1991), simulated data (year 2000, after calibration), and real data (year

2000). The indices were computed for different scales of segregation: (a) local scale, where the household's neighborhood comprises the area within a 700 m radius of its residence (no. 1 to 4 in Table 6.4), and (b) large scale, where this radius is equal to 2000 m (no. 5 to 8 in Table 6.4). The indices computed for a large segregation scale always present lower magnitude. This is because the population composition of larger neighborhoods tends to be more diverse and similar to the overall population composition of the city. To assist the interpretation of global indices, Figure 6.2 provides the population composition (income groups) of São José dos Campos for the years 1991 and 2000.



mw: minimum wages

Figure 6.2 Population composition (income groups) in São José dos Campos.

The local segregation indices computed for local and large scales are presented as maps (Figure 6.3 and 6.4), with darker colors representing higher levels of segregation. Five replications of the experiment were performed and, despite the stochastic nature of the model, all produced the same results.

In general, simulated patterns of segregation demonstrate a good agreement with the observed pattern over time. Both show how the global dissimilarity index $\breve{D}(m)$ slightly increased during the period 1991-2000 considering different scales (Table 6.4). The $\breve{D}(m)$ index compares the population composition of the whole city (Figure 6.2) with that of the neighborhoods, measuring the proportion of people who would have to move from their neighborhoods to achieve an even population distribution. The index varies from 0 to 1, where 0 stands for the minimum degree of segregation, i.e., the case when the population composition of all neighborhoods is equal to the population composition of the city.

Table 6.4 Validation experiment: global indices of segregation, computed for local (radius = 700 m) and large (radius = 2000 m) scales.

No.	*Global Segregation Index*	*Initial State* (Year 1991)	*Simulated* (Year 2000)	*Real Data* (Year 2000)
1	Spatial dissimilarity index $\breve{D}(m)$ (700 m)	.26	.28	.28
2	Spatial isolation of low-income households $\breve{Q}_{low}$ (700 m)	.6	.58	.58
3	Spatial isolation of medium-income households $\breve{Q}_{medium}$ (700 m)	.33	.32	.32
4	Spatial isolation of high-income households $\breve{Q}_{high}$ (700 m)	.33	.36	.38
5	Spatial dissimilarity index $\breve{D}(m)$ (2000 m)	.19	.21	.22
6	Spatial isolation of low-income households $\breve{Q}_{low}$ (2000 m)	.57	.56	.56
7	Spatial isolation of medium-income households $\breve{Q}_{medium}$ (2000 m)	.32	.31	.31
8	Spatial isolation of high-income households $\breve{Q}_{high}$ (2000 m)	.26	.31	.32

The local segregation indices computed for local and large scales are presented as maps (Figure 6.3 and 6.4), with darker colors representing higher levels of segregation. Five replications of the experiment were performed and, despite the stochastic nature of the model, all produced the same results.

In general, simulated patterns of segregation demonstrate a good agreement with the observed pattern over time. Both show how the global dissimilarity index $\breve{D}(m)$ slightly increased during the period 1991-2000 considering different scales (Table 6.4). The $\breve{D}(m)$ index compares the population composition of the whole city (Figure 6.2) with that of the neighborhoods, measuring the proportion of people who would have to move from their neighborhoods to achieve an even population distribution. The index varies from 0 to 1, where 0 stands for the minimum degree of segregation, i.e., the case when the population composition of all neighborhoods is equal to the population composition of the city.

Considering a local scale (no. 1 in Table 6.4), the $\breve{D}(m)$ index during the period 1991-2000 increased from .26 to .28. The $\breve{D}(m)$ computed for a larger scale (no. 5 in Table 6.4) also increased during the period (from .19 to .21/.22). The maps of the local version of the dissimilarity index (Figure 6.3 and 6.4) provide further information about this change, showing where it happened. The maps that consider a local scale of segregation (Figure 6.3) show details of these changes, while the maps considering a larger scale of segregation are more suitable for the observation of global trends (Figure 6.4)

The segregation maps suggest that the increase in dissimilarity occurred especially in areas close to the center, towards the western region, and in the southern region (Figure 6.3 and 6.4 (a-c)). The isolation maps complement this information by showing that the higher dissimilarity in central areas is caused by the isolation of affluent households (Figure 6.3 and 6.4 (g-i)), while the higher dissimilarity in the south is due to the isolation of poor households (Figure 6.3 and 6.4 (d-f)).

The interpretation of global indices of isolation demands caution, since the proportions of social groups in the city influence their values. During the period 1991-2000, the proportion of low-income households (up to 4 minimum wages) decreased from .54 to .51 (Figure 6.2). Meanwhile, their spatial isolation computed for a local scale decreased from .6 to .58 (no. 2 in Table 6.4). This means that, on average, 58% of the neighbors of a low-income household belong to the same income group. This value is higher than the overall percentage of this group in the city (51%). According to the maps, the decrease in isolation occurred mainly in areas closed to the center, keeping a high (or higher) isolation of poverty in the outskirts of the city (Figure 6.3 and 6.4 (d-f)).

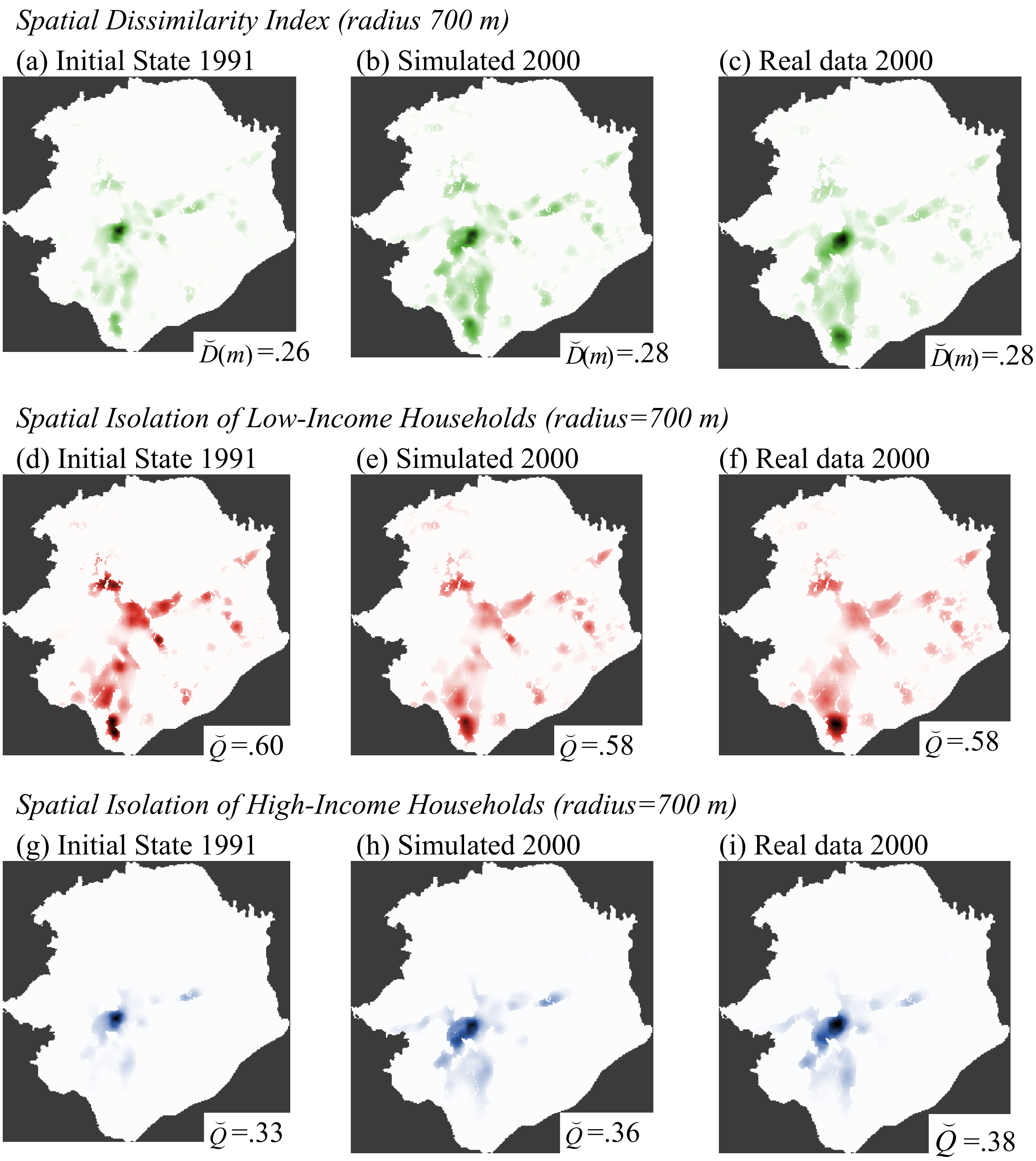


Figure 6.3 Validation experiment: local indices of segregation (local scale, radius 700 m).

In the case of the isolation of high-income households, the global indices calculated for real and simulated data presented the same trend but different values. The proportion of high-income households (more than 10 minimum wages) increased from .15 to .19 (Figure 6.2), while their isolation computed for a local scale increased from .33 to .38 according to the real data, and to .36 according to the simulated data (no. 4 in Table 6.4). It is interesting to observe that the difference between the group proportion in the city and the isolation index is much higher for affluent households (.19 vs.

.38/.36) than for low-income households (.51 vs. .58). This suggests that affluent households have a higher inclination to live isolated from other social groups. The maps of local isolation computed for simulated and real data (Figure 6.3 and 6.4 (g-i)) show that the isolation of high-income households increased in areas close to the center towards the western region, configuring a 'wealthy axis' in the city.

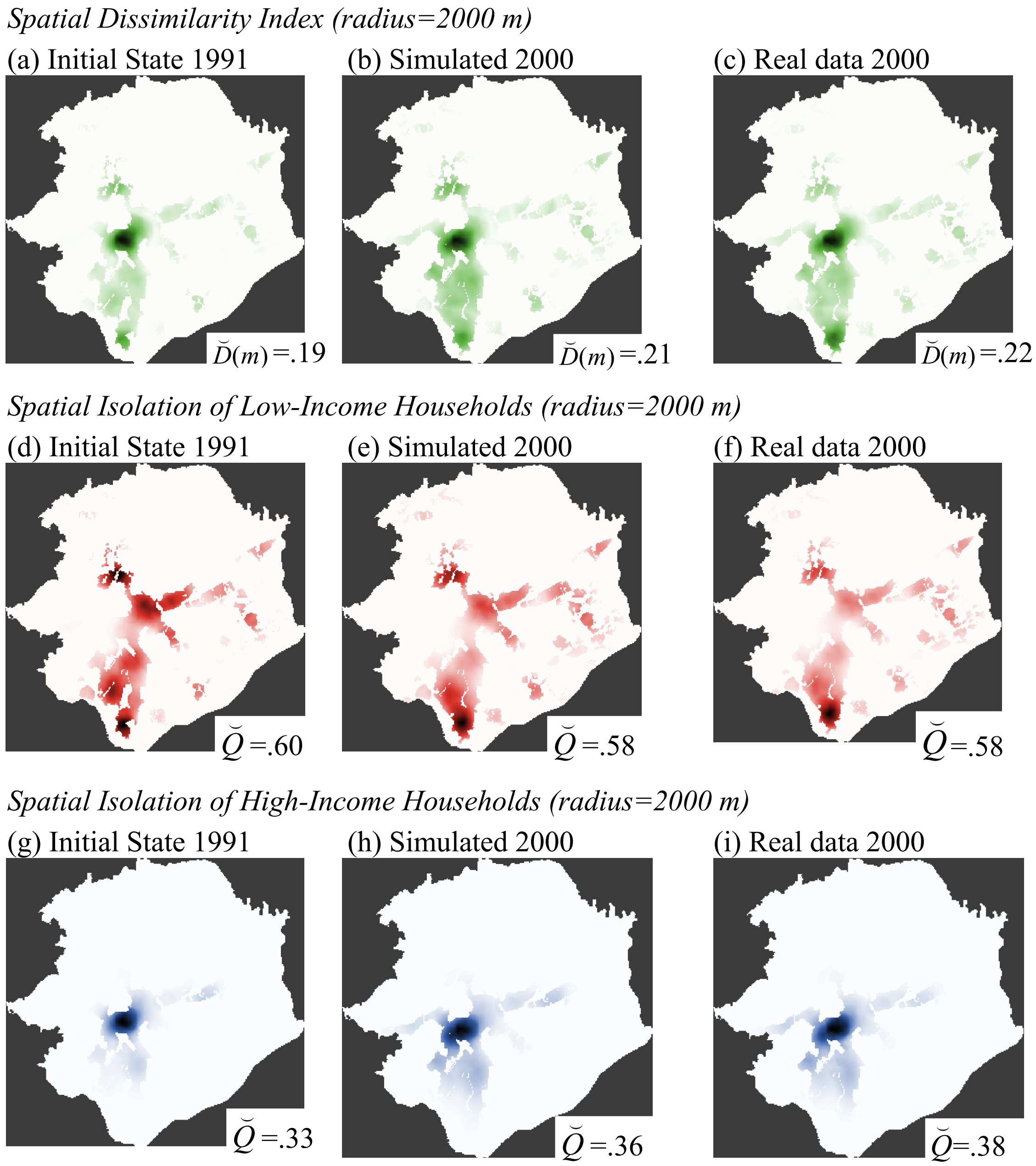


Figure 6.4 Validation experiment: local indices of segregation (large scale, radius 2000 m).

6.3 Simulation experiments II: Testing theoretical issues of segregation

This section presents simulation experiments that demonstrate the potential of the MASUS model to explore and test theoretical issues about urban segregation. Researchers have pointed out four different and complementary causal mechanisms of segregation: labor market, personal preferences, land and real-estate market, and the state (details in section 2.4). The experiments presented in this section focus on aspects regarding two of these mechanisms: income inequality, seen as a product of the labor market, and the neighborhood preferences of high-income families, seen as a sort of personal preference. Experiments concerning the remaining factors (land market and the state) are presented in section 6.4.

6.3.1 Impact of income inequality on segregation

This experiment explores the impact of different levels of income inequality on segregation patterns. In the United States, many theoretical and empirical studies advocate that income inequality promotes urban segregation (Mayer 2001; Reardon and Bischoff 2008; Watson 2006; Wheeler and La Jeunesse 2007). In Latin America, however, this issue has caused controversy. While the causal relationship between inequality and segregation underlies the discourse of some researchers (Kowarick 1979; Lago 2000; Maricato 1979b), others advocate that this is not necessarily true. Sabatini (2004) criticizes the argument that inequality is reflected in urban segregation, which he labeled as ‘mirror effect hypothesis’. According to the author, inequality and urban segregation are closely related phenomena, though one is not a simple reflection of the other. As an example, he mentions the economic crises in Latin America during the 1980’s, which increased social inequalities but, at the same time, promoted a higher proximity among different income groups in some cities (Sabatini 2004,2006).

The purpose of the experiment in the present study is to provide further insights into this debate. For this, the baseline simulation run described in the previous section is compared with two alternative scenarios for the period 1991-2000: one where inequality increases during the simulation and another where inequality decreases. All the other specifications were kept as in the baseline scenario (section 6.2.1). These experiments were repeated for the period 2000-2010, where the inequality level considered for the baseline scenario is constant. The inequality level is one of the experimental factors presented on the MASUS interface (no. 3 in Figure 6.1), where the user can choose templates with different settings for the global variables that control the income composition of the population (population transition sub-model).

Results: 1991-2000

Graphs with the global segregation indices obtained from the three simulation runs covering the period 1991-2000 were developed (Figure 6.5). The graphs showing the inequality levels and dissimilarity indices along this period (Figure 6.5 (a-b)) support the hypothesis that inequality promotes segregation: Once inequality increases, the dissimilarity between the income composition of the whole city and the income composition of neighborhoods ($\breve{D}(m)$) also increases, and vice-versa. The maps of local dissimilarity for the three scenarios (Figure 6.6 (a-c)) indicate where segregation would decrease or increase in each case.

The global isolation of low-income households ($\breve{Q}_{low}$) also varied proportionally to the inequality levels (Figure 6.5 (c-d)). The isolation maps (Figure 6.6 (d-f)) complements this information by showing that in case of higher inequality the concentration of poverty would increase mainly in the outskirts of the city. It is important to mention, however, that the increase in low-income isolation was expected in the case of higher inequality, since these indices follow the progression of the proportion of low-income households in the city. This is a natural trend of the index: once the proportion of group *m* increases in the city, the global isolation index of group *m* also tends to become higher.

The expected trend of the index is, however, challenged by the graphs showing the isolation of high-income households $\breve{Q}_{high}$ and the proportion of this group in the city (Figure 6.5 (e-f)). For this reason, the results presented in these graphs are the most revealing ones. The low-inequality scenario presents higher proportions of affluent households when compared to the other scenarios, but still displays the lowest levels of isolation. This unexpected result represents an additional indication of the causal relation between income inequality and segregation.

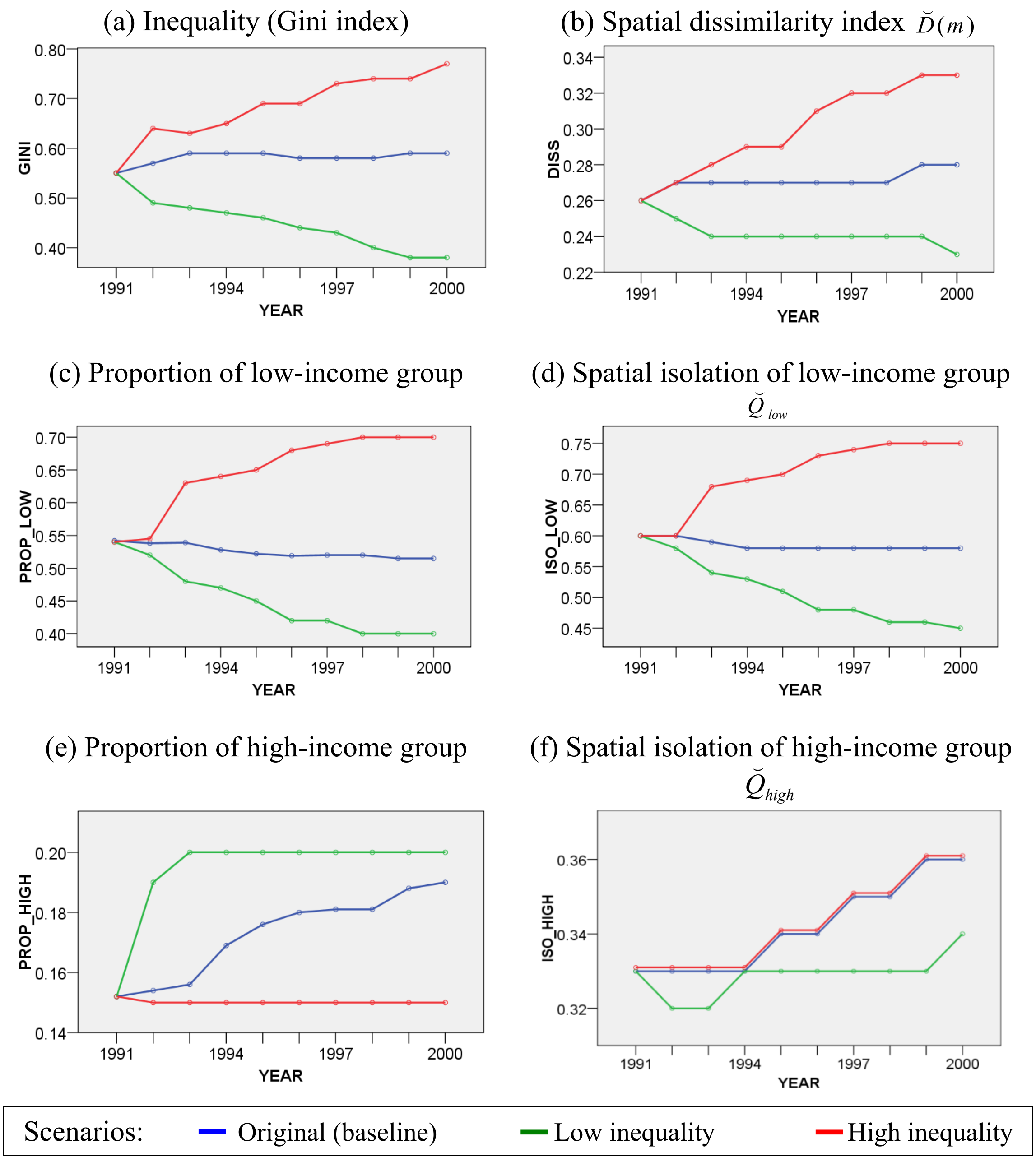


Figure 6.5 Progression of population statistics and global segregation indices (radius 700 m) for three simulation scenarios 1991-2000: original, decreasing, and increasing income-inequality levels.

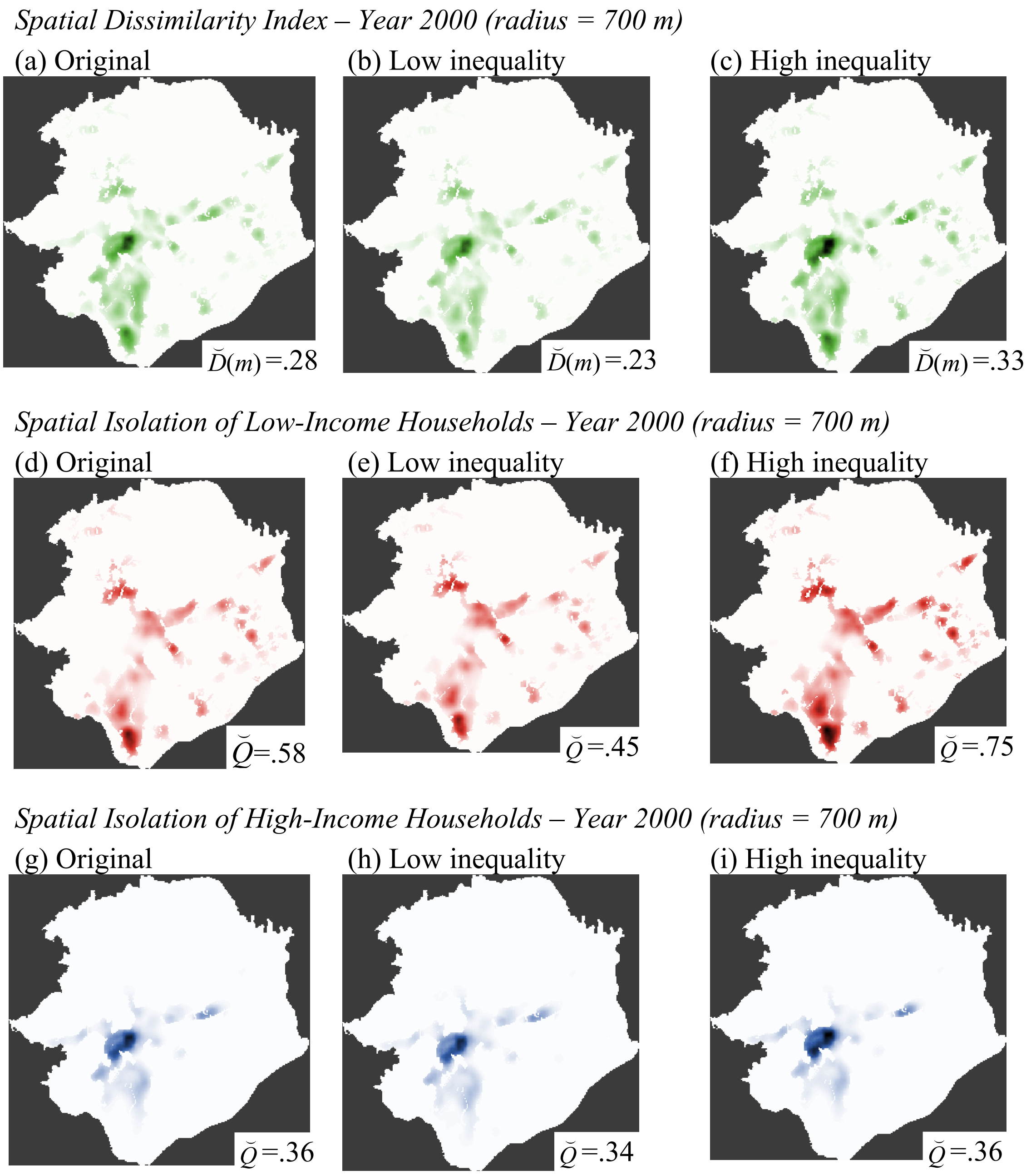


Figure 6.6 Local indices of segregation (local scale, radius 700 m) for three simulation scenarios for the year 2000: original, decreasing, and increasing income inequality levels.

Results: 2000-2010

The simulation experiment was repeated for the period 2000-2010. In this case, the baseline scenario presents constant inequality levels. This characteristic differs from the original baseline scenario 1991-2000, where the Gini index increased from .55 to .59. The purpose of this experiment was to observe whether simulations conducted from the year 2000, an initial condition that differs from 1991, support the insights obtained from the previous experiment or not.

The results observed from the graphs of global segregation indices computed for simulation scenarios 2000-2010 (Figure 6.7) are similar to the ones obtained for the period 1991-2000 (Figure 6.5). The global index of dissimilarity increases when the Gini index increases, and vice-versa. However, it is interesting to note that when inequality is kept constant (baseline scenario) the dissimilarity index continues to increase. This outcome indicates that besides inequality there are other factors promoting segregation in the baseline scenario (e.g., household preferences).

As in the inequality experiment for the period 1991-2000, the isolation indices of high-income households also demonstrate the causal relation between income inequality and segregation. The scenarios considering increasing inequality and decreasing inequality present equal levels of high-income isolation in 2010 ($\breve{Q}_{high}$ = 0.38). For both scenarios, this means that on average 38% of the neighbors of a high-income household belong to the same income group. This number, compared to the overall percentage of high-income households in each scenario, reveals that the isolation of affluent households is more significant in the high-inequality scenario, where the overall percentage of high-income households is 15%, than in the low-inequality scenario, where this percentage is 20%.

The dissimilarity maps for the year 2010 (Figure 6.8 (a-c)) show two hotspots at the central and western region of the city. These areas correspond to areas of intense isolation of high-income households (Figure 6.8 (g-i)) and demonstrate that for the three scenarios the segregation measured by the dissimilarity index is mainly enhanced by affluent households. This pattern differs from the one observed for the year 2000 (Figure 6.6 (a-c)), where areas characterized by the concentration of poor households also represent hotspots in the dissimilarity maps.

In general, the segregation maps resulting from the three inequality scenarios for the period 2000-2010 show increasing distances between social groups with contrasting income levels. Compared to the 2000 maps, the isolation maps of low-income households for the year 2010 (Figure 6.8 (d-f)) show that the isolation of this group tends to become much more peripheral and distant from high-income households, especially in the scenarios of constant and high-inequality. In the scenario of high-inequality, the isolation of poor households increases considerably in clandestine settlements located in the eastern and northern part of the city.

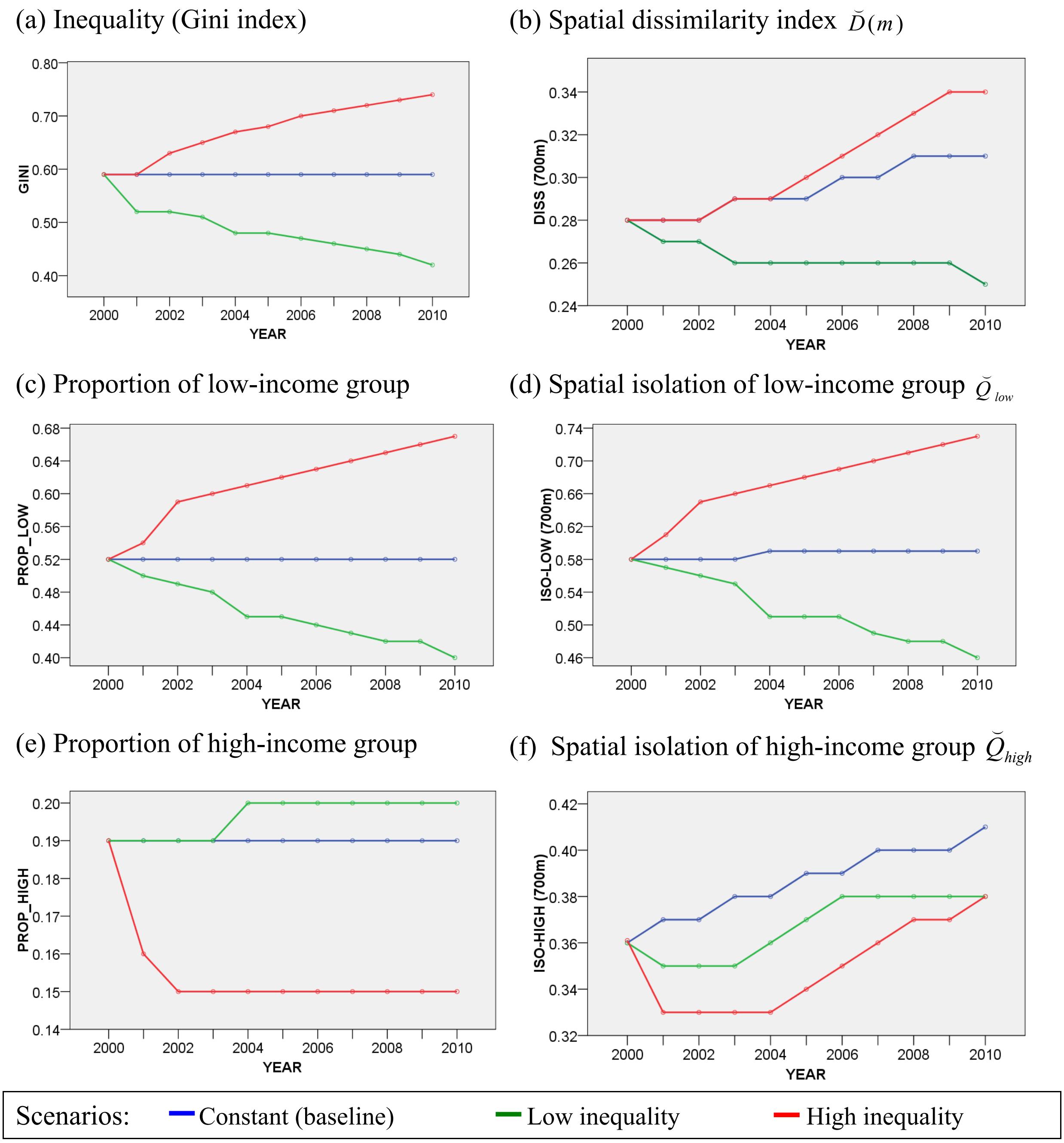


Figure 6.7 Progression of population statistics and global segregation indices (radius 700 m) for three simulation scenarios 2000-2010: constant, decreasing, and increasing income inequality levels.

The isolation maps of high-income households resulting from the three simulation runs (Figure 6.8 (g-i)) present very similar spatial patters, with affluent households strengthening their self-segregation in a wealthy axis, which extends from the center towards the west of the city. In comparison to the simulated outputs for the year 2000, these maps show a stronger presence of high-income households in the western region, which is an area with a high concentration of gated neighborhoods.

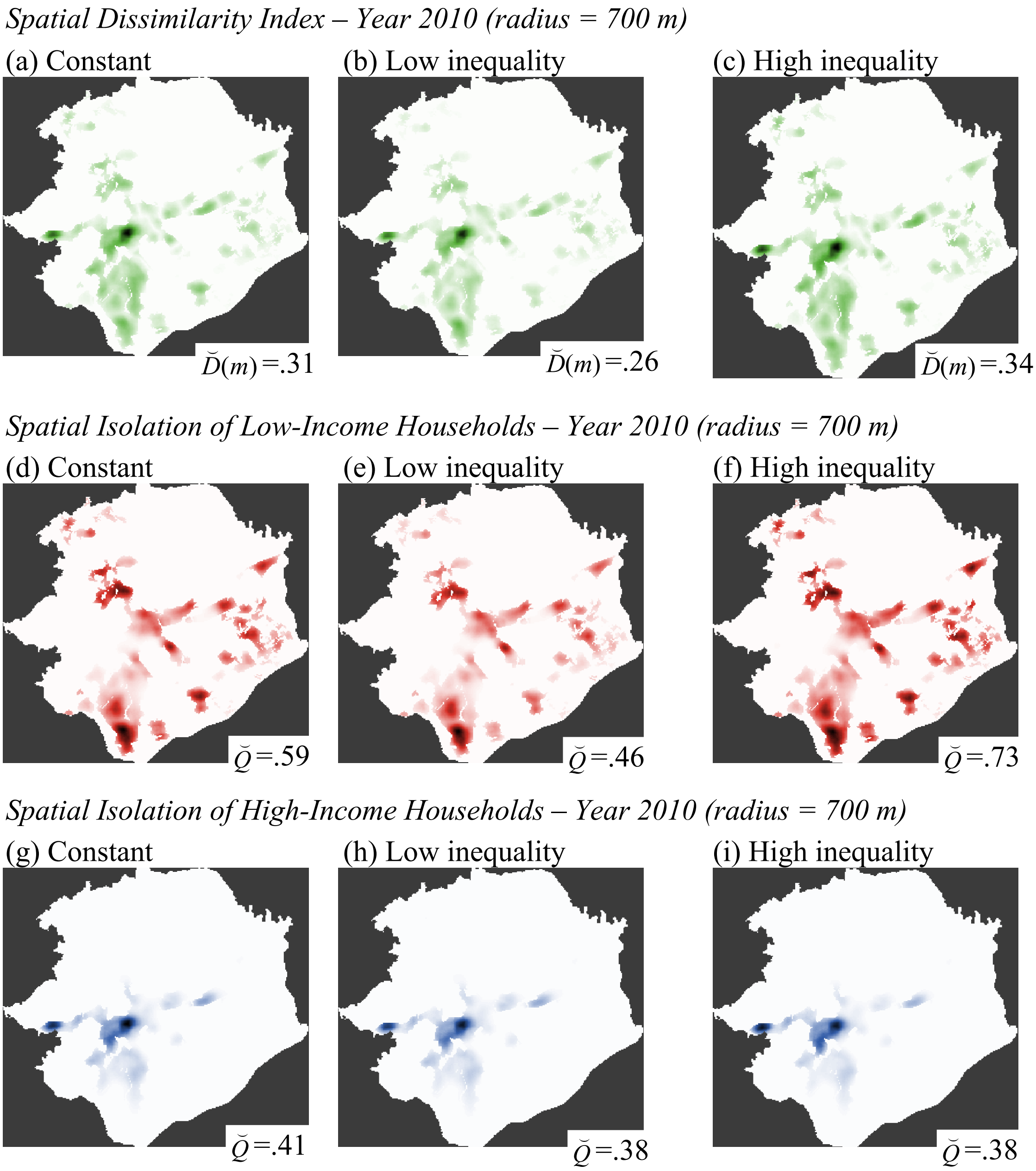


Figure 6.8 Local indices of segregation (local scale, radius 700 m) for three simulated scenarios for the year 2010: constant, decreasing, and increasing income inequality levels.

6.3.2 Impact of affluent households' residential preferences on segregation

Because people tend to prefer to live among neighbors similar to themselves, personal preferences have been commonly pointed out as being one mechanism that can increase segregation. It promotes the so-called 'self segregation', which is more common among affluent families, and results from the families' attempt to improve their quality of life and strengthen their social identities through shared values (Marcuse 2005). In Brazil, where cities are often perceived as dangerous and unmanageable, gated neighborhoods are one of the most explicit materialization of this process of self-segregation (Caldeira 2000).

Focusing on this issue, the experiment presented in this section explores how the neighborhood preferences of high-income families can influence segregation patterns. It compares the baseline scenario 1991-2000 (section 6.2) with alternatives scenarios with different user input parameter values for the high-income preference factor θ_{neigh} (Table 6.5). The factor θ_{neigh} determines the relevance of the neighborhood income composition to the residential choice of affluent families.

Global segregation indices were computed along the four simulation scenarios (Table 6.5, Figure 6.9). The simulations runs cover the period 1991-2000, and the segregation indices were computed for neighborhoods defined by a radius of 700 m and 2000 m. The graphs showing the isolation of high-income households $\breve{Q}_{high}$ (Figure 6.9 (e-f)) demonstrate that there is a linear relation between these indices and the parameter θ_{neigh}. For example, considering the $\breve{Q}_{high}$ for the year 2000, for every 1 unit increase in θ_{neigh}, there is a corresponding .05 unit increase in the $\breve{Q}_{high}$ computed with a neighborhood radius of 700 m, and a .03 unit increase in $\breve{Q}_{high}$ computed with a radius of 2000 m. The maps displaying the local version of the isolation index of high-income households $\breve{q}_{high}$ (Figure 6.10 (g-i)) show that this change in isolation spatially occurs in the center and western region of the city.

Table 6.5 Simulation scenarios of the high-income preferences experiment: values selected for the input parameter θ_{neigh}

Scenario	*Value* θ_{neigh}	*Description*
(1) Baseline	$\theta_{neigh} = 1$	Preference of affluent households for having neighbors similar to themselves **is equal** to the one calibrated from empirical data.
(2)	$\theta_{neigh} = 0$	Affluent households **do not consider** the income composition of neighborhoods when selecting their residential location.
(3)	$\theta_{neigh} = 2$	Preference of affluent households for having neighbors similar to themselves is **two times higher** than the one considered in the baseline scenario.
(4)	$\theta_{neigh} = 3$	Preference of affluent households for having neighbors similar to themselves is **three times higher** than the one considered in the baseline scenario.

This relation between θ_{neigh} and the isolation index $\breve{Q}_{high}$ seems to directly influence the segregation dimension evenness/clustering, which is measured by the dissimilarity index $\breve{D}(m)$. The index $\breve{D}(m)$ increases with an increase in θ_{neigh} (Figure 6.9 (a-b)) and, according to the maps of the local version of the index (Figure 6.10), such variation follows the spatial trends presented by the local isolation of high-income households $\breve{q}_{high}$.

The influence of the parameter θ_{neigh} on the isolation index of low-income households $\breve{Q}_{low}$ is also linear, but not substantial. Considering the $\breve{Q}_{low}$ for the year 2000 and neighborhood radius of 700 m, for every 1 unit increase in θ_{neigh}, there is a corresponding increase of only .01 unit in $\breve{Q}_{low}$. Even in the case where affluent households do not care about the income composition of neighborhoods when selecting their residential location (θ_{neigh} = 0), the isolation index of low-income households is very high ($\breve{Q}_{low}$ = .57), showing that this situation did not promote the integration between affluent and poor families, but that between affluent and middle-income households. This indicates that, independent of the affluent households' preferences regarding the income composition of neighborhoods, they still choose areas that the poorest households cannot afford (e.g., areas with higher quality/prices).

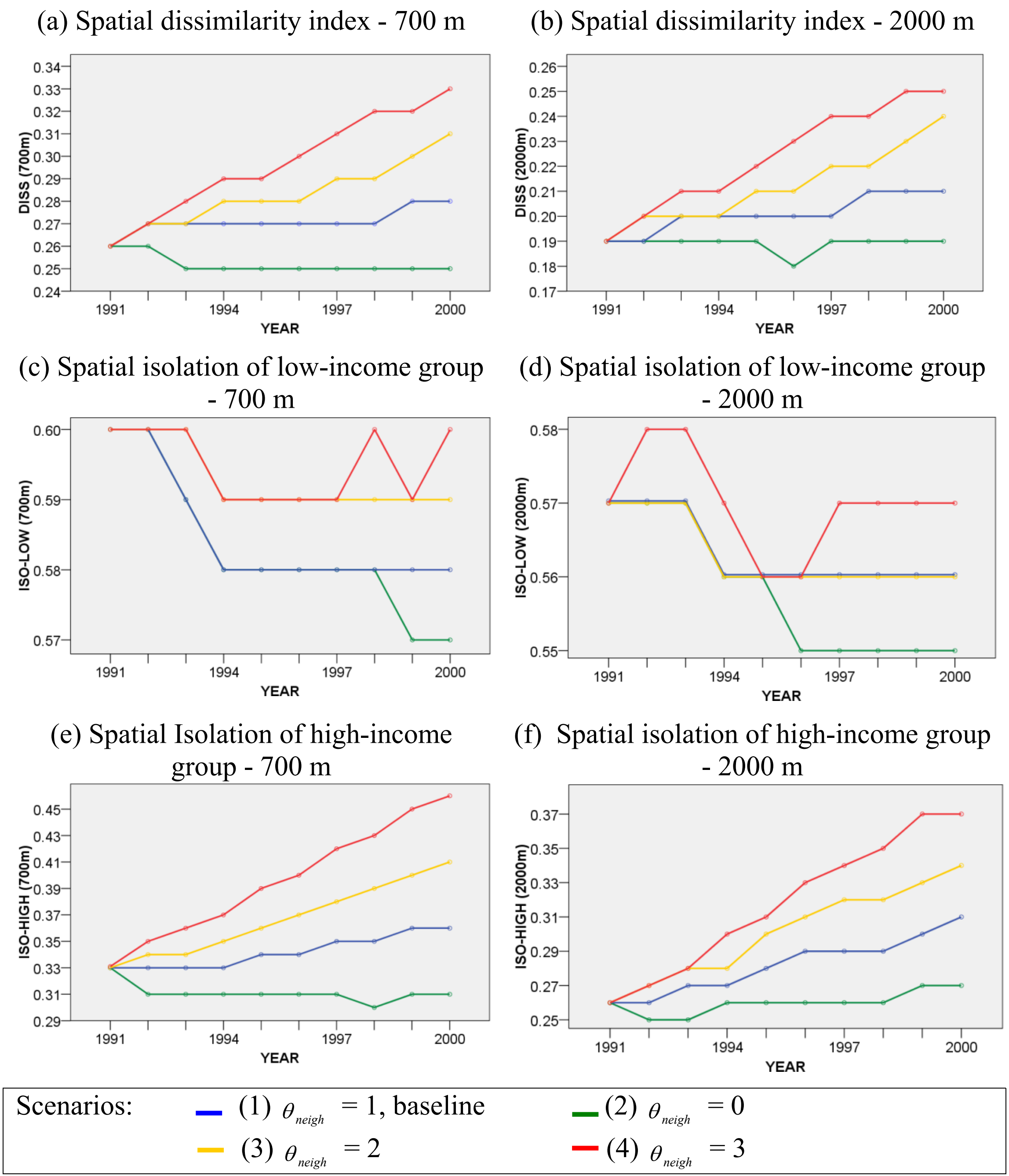


Figure 6.9 Progression of global segregation indices (radius 700 m and radius 2000 m) for four simulation runs 1991-2000 executed for the experiment with high-income households' preferences.

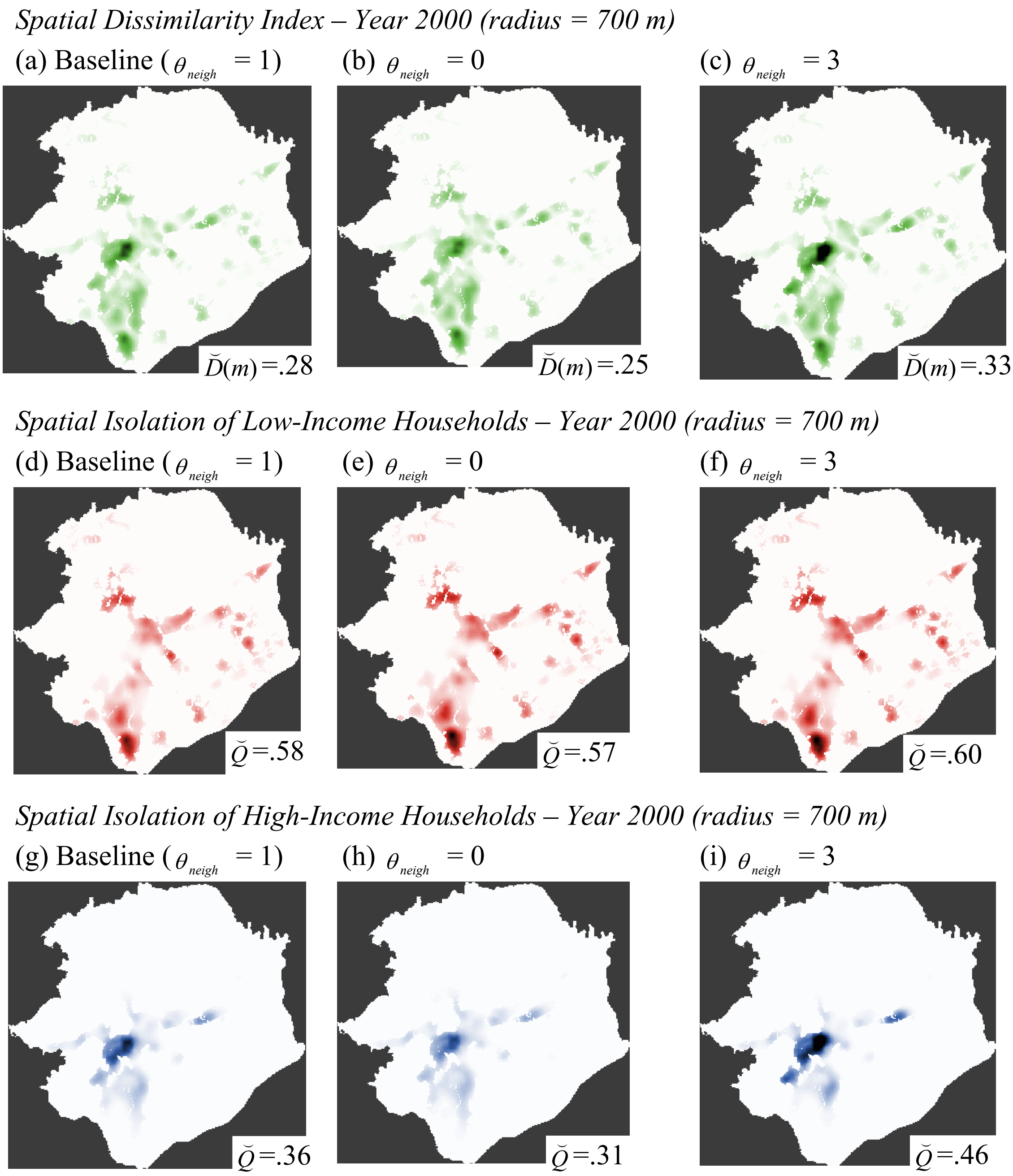


Figure 6.10 Local indices of segregation (local scale, radius 700 m) for the year 2000 resulting from three simulated runs for the experiment on high-income households' preferences: $\theta_{neigh} = 1$ (baseline), $\theta_{neigh} = 0$, and $\theta_{neigh} = 3$.

6.4 Simulation experiments III: Testing urban policies

In the United States and some European countries, the residential mix of advantaged and disadvantaged groups represents a target explicitly expressed in many scientific and policy discourses (Andersson 2008; Smith 2002). In practice, these countries have adopted different policy strategies to promote social mix, including the dispersal of poverty, regulation of land-market dynamics, and regeneration of troubled neighborhoods. The aim of the simulation experiments presented in this section is to provide new insights about the impact that different urban policies can have on segregation. Two different social-mix policy approaches are tested: one based on the dispersion of poverty, and the other on the dispersion of wealth. The first promotes integration by moving poor households out of problematic neighborhoods, while the second stimulates the construction of residential developments for middle and upper classes in poor regions of the city. In addition, a third experiment tests the impact of regularizing clandestine settlements and promoting an equitable distribution of infrastructure in the city.

6.4.1 Impact of a social-mix policy based on poverty dispersion

The experiment presented in this section tests how an anti-segregation policy based on the dispersal of poverty could impact the segregation dynamics of a Latin American city like São José dos Campos. Policies for promoting integration through the spatial dispersion of poverty focus on moving poor households out of distressed areas into middle-class neighborhoods. For this, low-income households receive housing vouchers that are used to rent private dwellings in neighborhoods with a low poverty rate.

To test the effect of a social-mix policy based on the distribution of housing vouchers, we compare the simulation run that replicates the original segregation dynamics of São José dos Campos during the period 1991-2000 (section 6.2) with two alternative scenarios. These scenarios simulate the implementation of a housing program that distributes *n* housing vouchers for poor families in 1991, and increases the number of benefits each year (Figure 6.11). The first alternative scenario distributes vouchers to 0.3% of the poor households in the year 1991 (200 vouchers) and progressively increases this percentage until the year 2000, when 2.3% of the poor households in the city are assisted by the housing program (1700 vouchers). The second alternative scenario increases the investments in the program: it distributes vouchers to 0.9% of the poor households in the year 1991 (500 vouchers) and gradually extends the program to 5.8% of the poor households in the year 2000 (4200 vouchers). The housing vouchers are distributed to poor families that are randomly selected from neighborhoods with a high isolation of poverty ($\breve{Q}_{poor} > mean(\breve{Q}_{poor}) + sd(\breve{Q}_{poor})$) and are used for renting dwellings in neighborhoods where the isolation of poor families is below the average ($\breve{Q}_{poor} < mean\ (\breve{Q}_{poor})$).

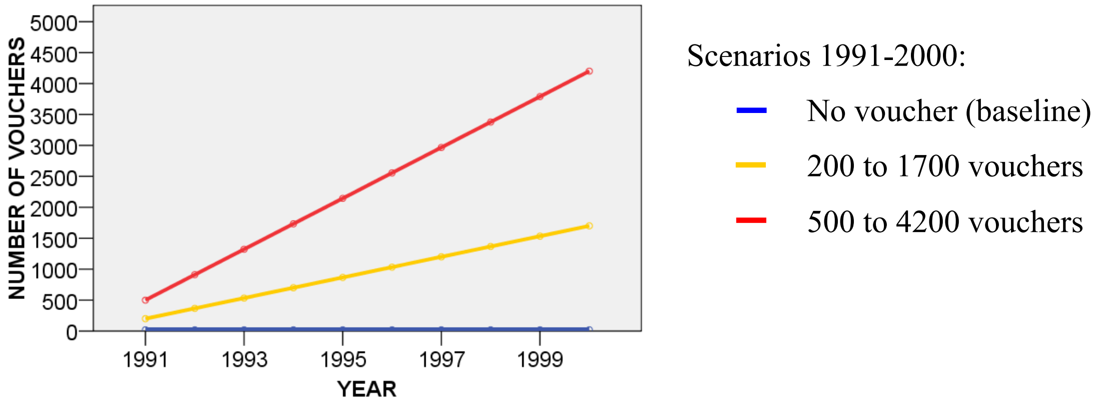


Figure 6.11 Scenarios of the experiment on poverty dispersion: number of housing vouchers distributed during the period 1991-2000.

Figure 6.12 shows the evolution of the global dissimilarity index and isolation indices during the period 1991-2000 for the three simulated scenarios: baseline (no housing voucher), alternative 1 (200 to 1700 vouchers), and alternative 2 (500 to 4200 vouchers). The dissimilarity index in the year 2000, which in the baseline scenario is equal to .28, changed to .27 in the alternative scenario 2. This means that the distribution of housing vouchers to 2.3% of the poor households in the city caused a decrease of 3.5% in the dissimilarity index. In the alternative scenario 3, the distribution of vouchers to 5.8% of the poor households decreased the dissimilarity index by 10.7% (from .28 to .25).

The spatial isolation index of high-income families also decreased significantly as the investment in the housing program increased. Comparing the baseline scenario with the alternative scenario 1, the distribution of housing vouchers to 2.3% of the poor households decreased the isolation of high-income households by 5.7% (from .36 to .35). Regarding the alternative scenario 2, the distribution of housing vouchers to 5.8% of the poor households caused a decrease of 8.3% in the isolation of high-income households (from .36 to .33).

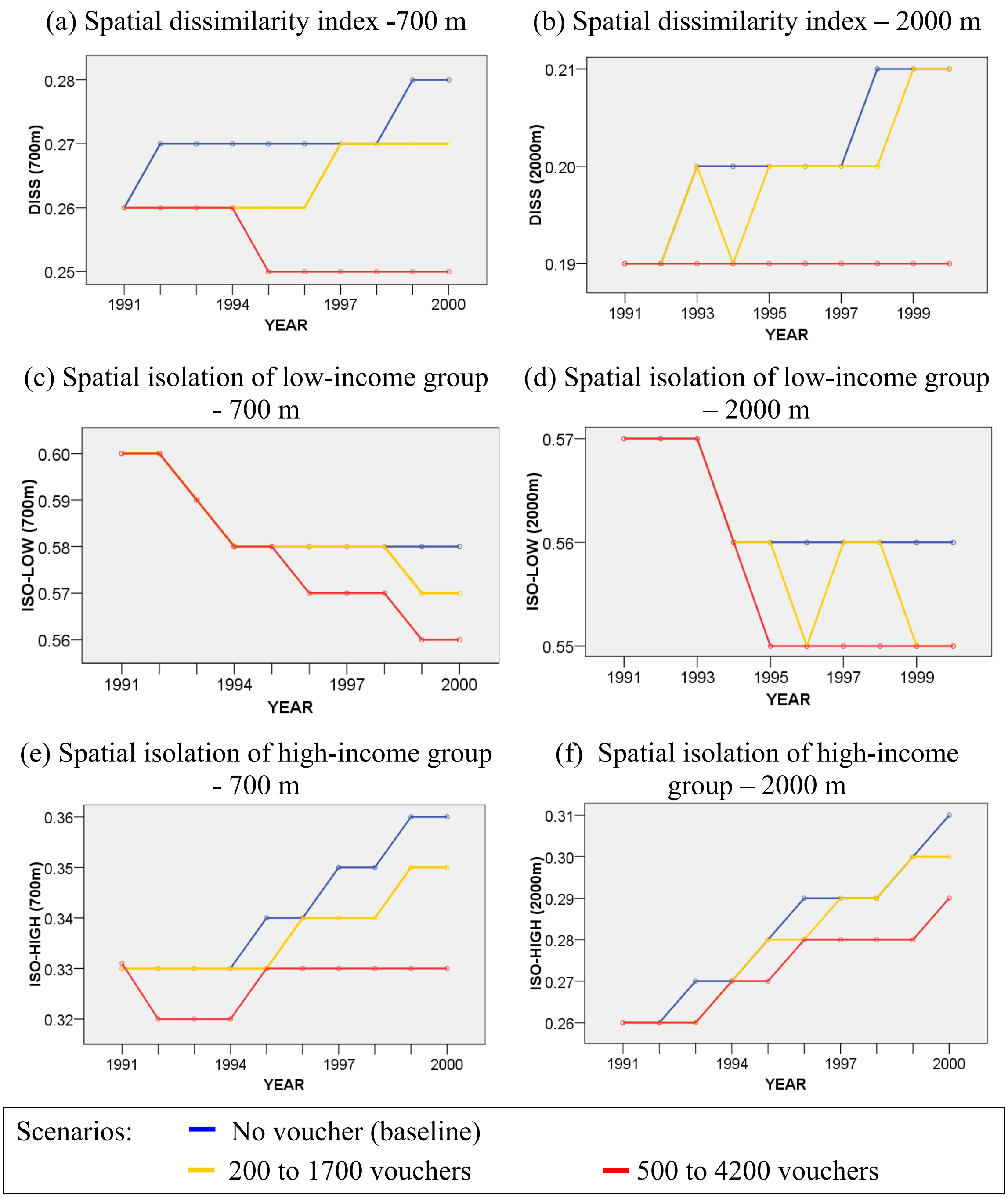


Figure 6.12 Progression of global segregation indices 1991-2000 (radius 700 m and radius 2000 m) for three scenarios 1991-2000 on poverty dispersion.

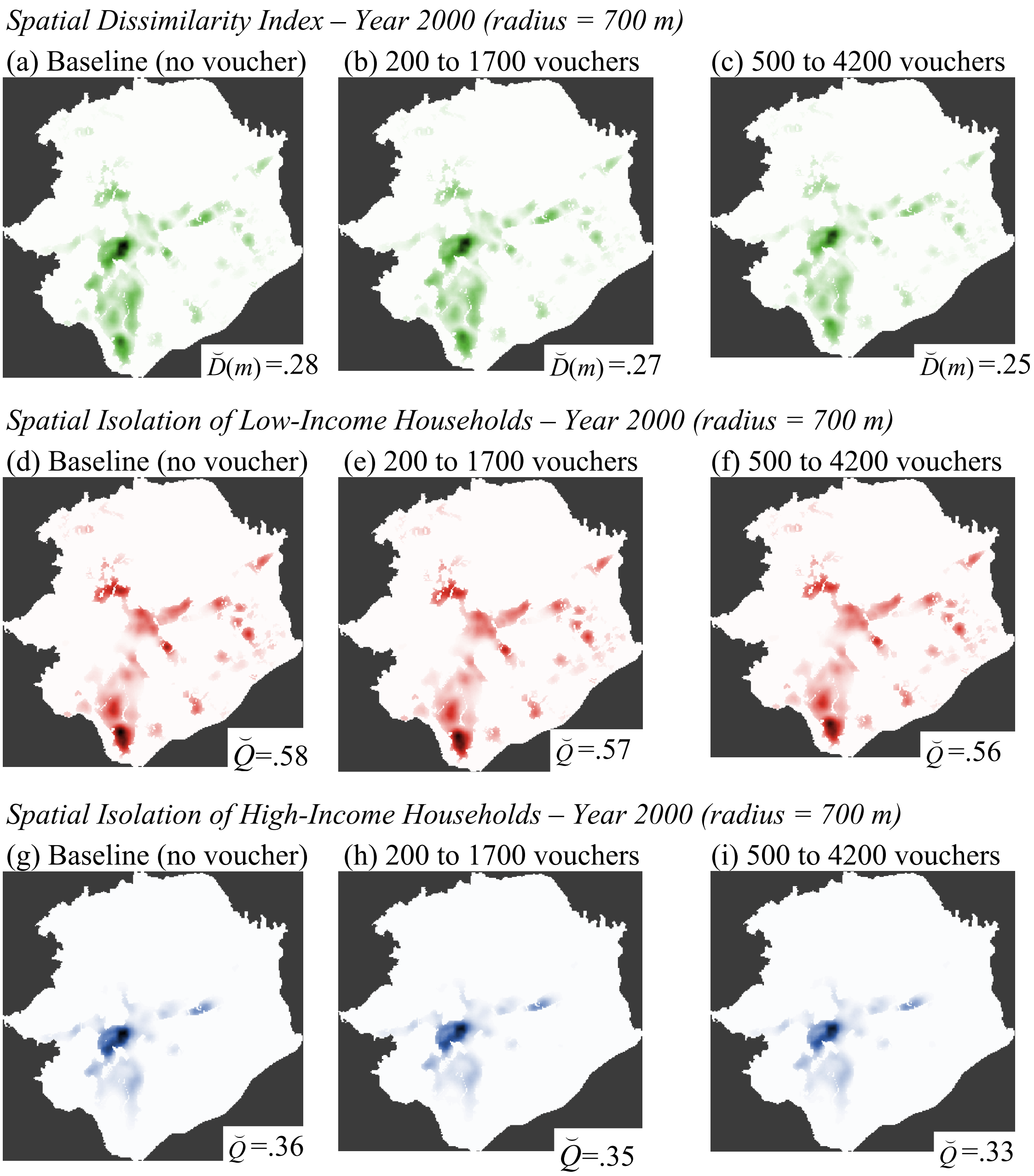


Figure 6.13 Local indices of segregation (local scale, radius 700 m) for three simulation scenarios for the year 2000: (1) baseline scenario, (2) 200 to 1700 vouchers, and (3) 500 to 4200 vouchers.

Despite these positive trends, the housing program did not lead to a substantial improvement in the overall isolation level of low-income households, which is the segregation dimension that causes the most harmful impacts on the lives of disadvantaged families. Comparing the baseline scenario with the alternative scenario 1, the distribution of housing vouchers to 2.3% of the poor households decreased the isolation of low-income households by only 1.7% (from .58 to .57). Comparing the baseline scenario with the alternative scenario 2, the distribution of housing vouchers to 5.8% of the poor households caused a decrease of 3.4% in the isolation of low-income households (from .58 to .56). This means that on average 58% of the neighbors of a poor family belong to the same income group in the baseline scenario for 2000. This percentage decreased only to 56% in the alternative scenario 2, where 4200 housing vouchers were distributed. These values demonstrate the limitation of this type of housing policy in cities where poor families represent a large share of the population. The maps of segregation support this idea (Figure 6.13), since no substantial difference can be observed between the maps produced for the different scenarios.

The experiment on poverty dispersion was also extended for the period 2000-2010, but instead of testing the impact of increasing investments on housing vouchers, the simulation runs covering this period kept the investment constant (Figure 6.14). In this case, the baseline scenario 2000-2010 (section 6.3.1) is compared with two alternative scenarios where 1700 and 4200 vouchers are distributed. The results show that this continued investment only slows down the increase in segregation, being unable to modify the segregation trends in comparison with the baseline scenario (Figure 6.15).

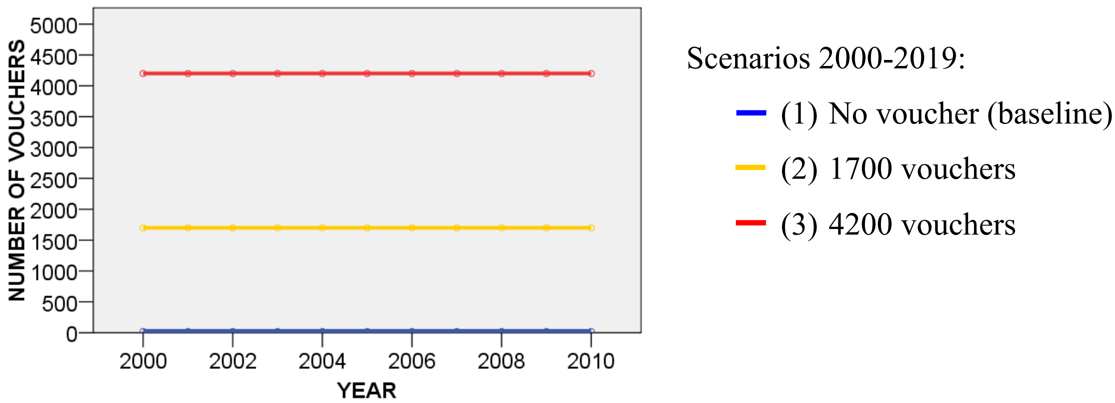


Figure 6.14 Scenarios of the experiment on poverty dispersion: number of housing vouchers distributed during the period 2000-2010.

In order to produce a substantial change in the overall isolation level of poor families, social-mix policies based on the distribution of housing vouchers would demand a massive and continuous investment. Because such investment is not realistic for cities in developing countries, different social-mix strategies should be explored. For these cities, the dispersion of affluent families may represent a more effective way to promote positive changes in segregation patterns (Sabatini 2006). Experiments focusing on this alternative will be presented in the next section.

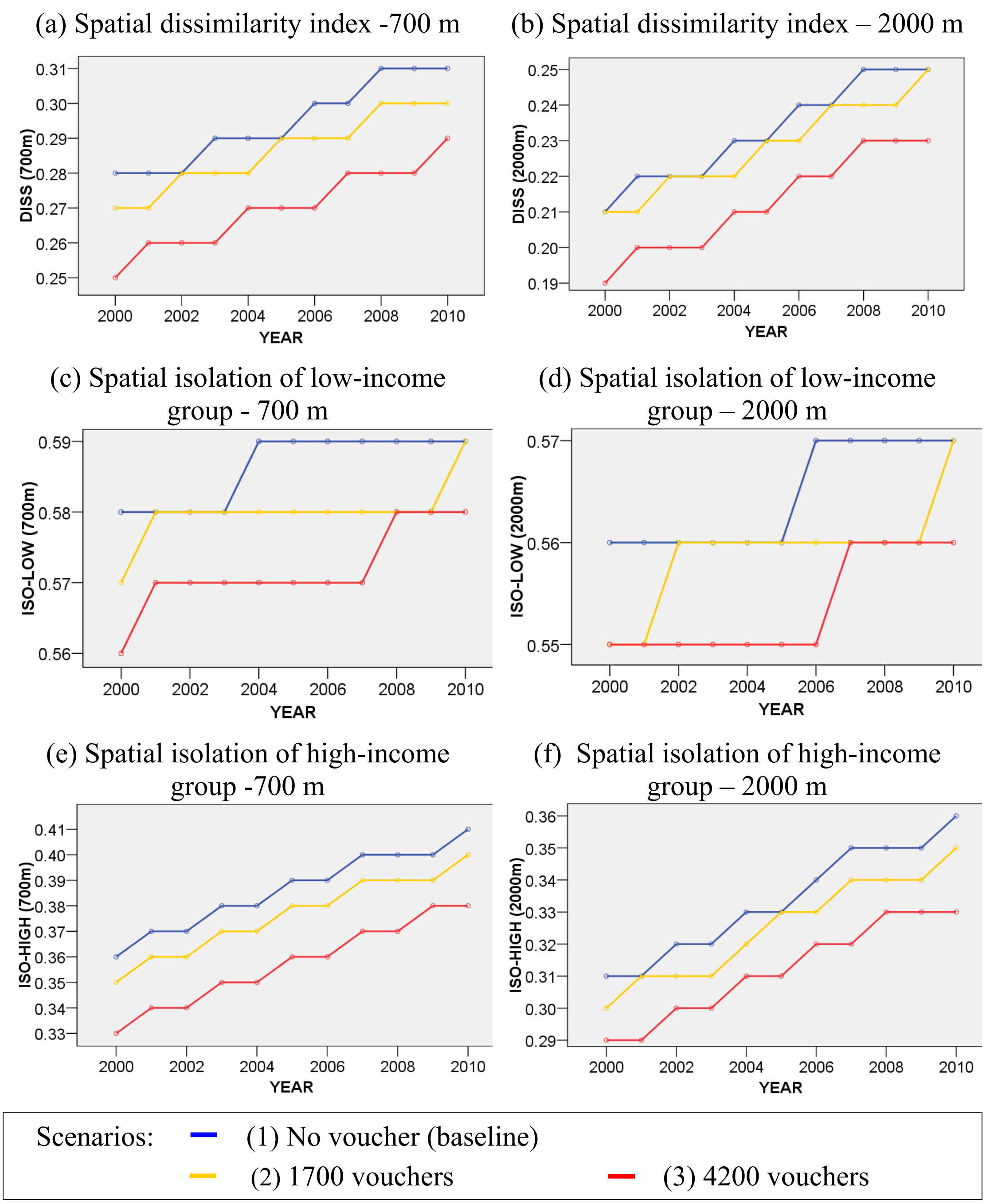


Figure 6.15 Progression of global segregation indices (radius 700 m and radius 2000 m) for three scenarios 2000-2010 on poverty dispersion.

6.4.2 Impact of a social-mix policy based on wealth dispersion

While the operation of real-estate markets is often pointed out as being an important causal mechanism of segregation, some development processes taking place in certain Latin-American cities provide indications that these markets can also promote a decrease in segregation, or a decrease in its scale. These processes include (Sabatini 2006):

1. The dispersion of condominiums for middle- and high-income families around the urban periphery, many in areas already populated by the poor. This process can promote a decrease in the scale of segregation.
2. The densification of wealthy neighborhoods, through vertical residential buildings for families of lower than average income in the area. These projects allow developers to significantly profit, and their indirect impact is to reduce urban segregation.

The experiment presented in this section considers the effective implementation of urban polices that aim at stimulating the first of these processes, i.e., the construction of residential developments for middle and upper classes in poor regions of the city. This can occur through tax exemption measures, concessions, changes in the norms of land use, and public investment in infrastructure and security (Sabatini 2006). The purpose of the experiment was to check if such initiative based on the dispersion of wealthy families would produce positive impacts on the segregation patterns of the city.

To conduct the experiment, undeveloped areas located in poor regions of the city were identified from orthophotos taken in 2000, scale 1:30000 and spatial resolution of 0.6m (PMSJC 2003). These areas were digitalized, imported into NetLogo, and classified as 'type A' neighborhoods (see section 5.2.2), i.e., settlements designed for residential occupation by middle and upper classes, with good housing quality, infrastructure, and services. A simulation run considering these new type A neighborhoods was executed for the period 1991-2010, and the results compared with the baseline scenarios 1991-2000 (section 6.2) and 2000-2010 (section 6.3.1). The graphs comparing the global segregation indices of these different 'what-if' scenarios through the years are presented in Figure 6.17.

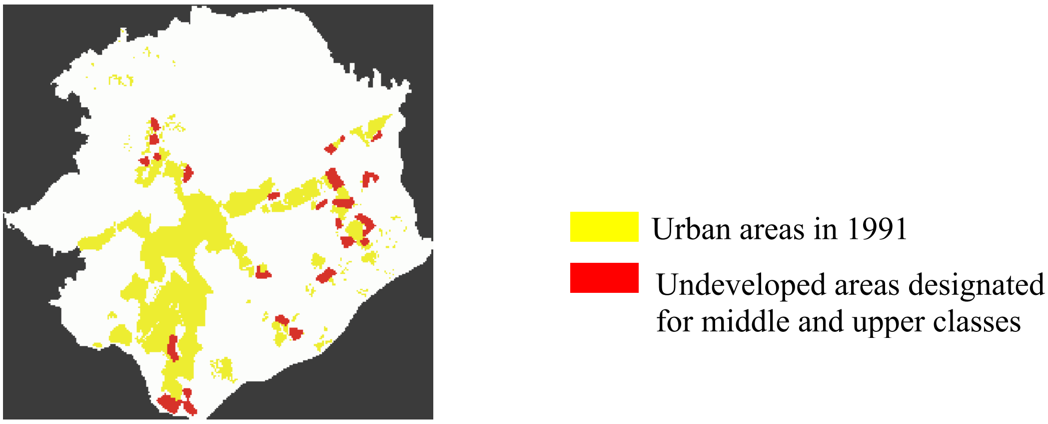


Figure 6.16 Location of new areas designated for middle and upper classes.

In general, it can be observed that the policy approach based on wealth dispersion produces long-term outcomes. The consolidation of the new areas designated for upper classes may take some years, and therefore their positive impacts on the global segregation indices become more substantial with time. This is an advantage in comparison with the poverty dispersion policy tested in the previous experiment, which demands a continued public investment for housing vouchers. As soon as this investment ceases, its positive impact on segregation cannot be sustained.

The global segregation indices presented in Figure 6.17 indicate that the dispersion of wealthy families tends to be more effective at decreasing large-scale segregation. For example, the dissimilarity index for 2010 computed for a local scale (700 m) decreases 19% when the policy based on wealth dispersion is adopted, less than the decrease of 36% that is observed when the same index is computed for a large scale (2000 m). The same occurs for the other indices: the isolation of poor households in 2010 decreases only 1.7% at the local scale, but 5.3% at the large scale, while the isolation of affluent households in 2010 decreases 17% at the local and 28% at the large scale. This outcome is another advantage in comparison with the policy based on poverty dispersion, which is more effective for decreasing local-scale segregation. Segregation at larger scales, specially the concentration of disadvantages, is considered more damaging than segregation at local scales (Sabatini et al. 2001).

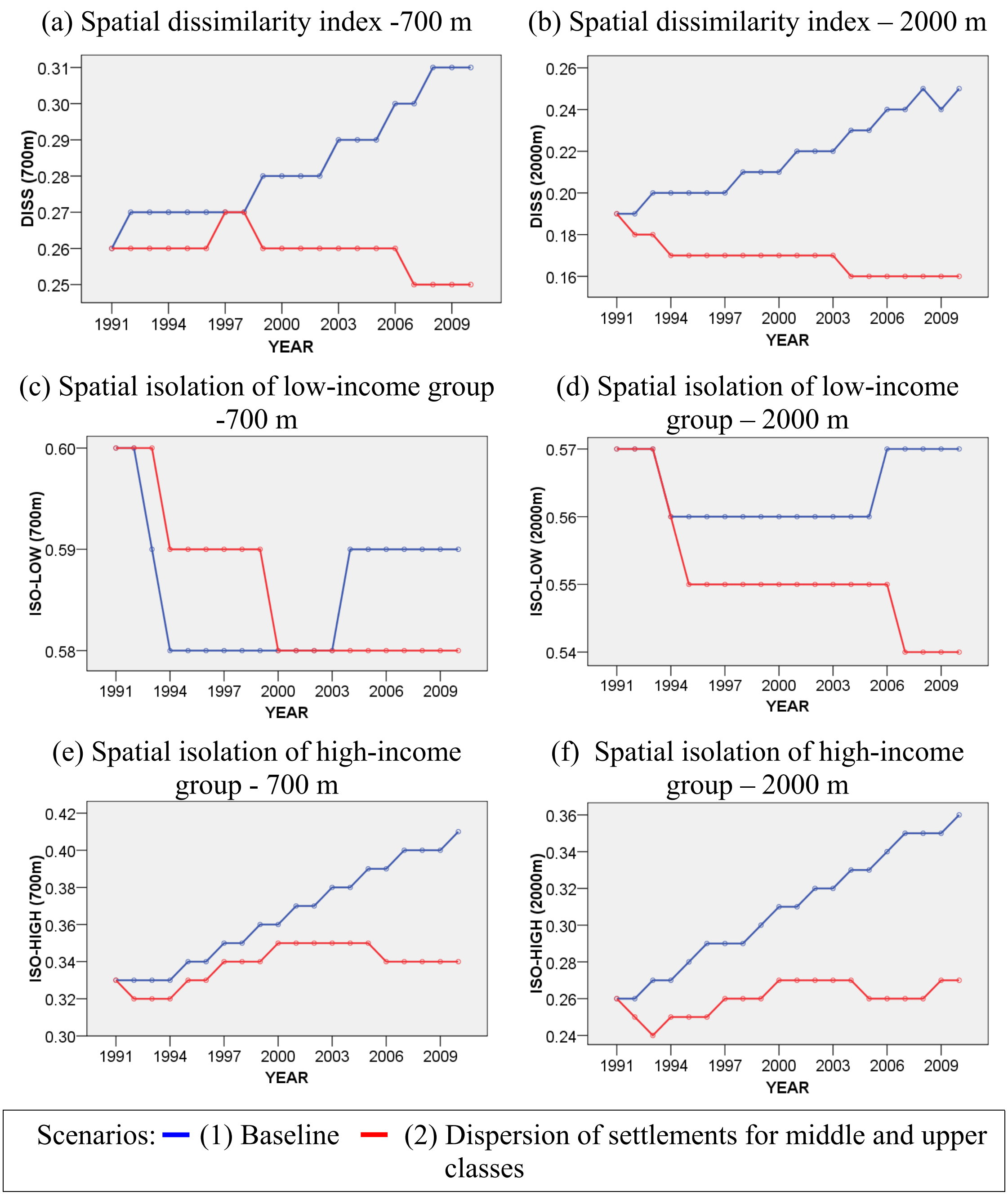


Figure 6.17 Progression of global segregation indices 1991-2010 (radius 700 m and radius 2000 m) for the scenarios of wealth dispersion

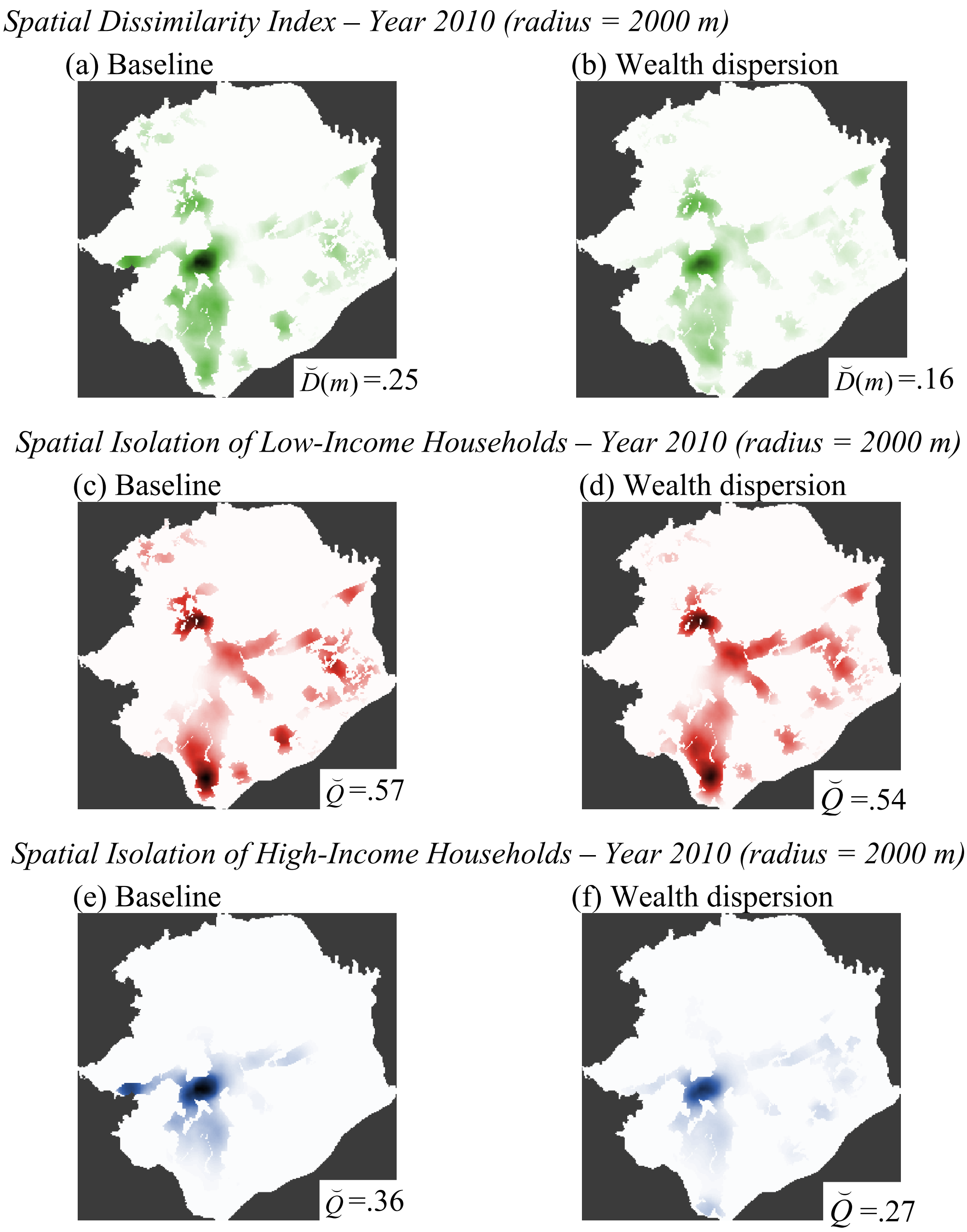


Figure 6.18 Local indices of segregation (large scale, radius 2000 m) for two simulation scenarios for the year 2010: (1) baseline scenario, and (2) wealth dispersion.

The segregation maps for the year 2010 show that the simulation of the policy on wealth dispersion was also able to modify spatial patterns of segregation in the city in a positive manner (Figure 6.18):

1. The local dissimilarity indices computed for a large scale became smoother and spread through the city (Figure 6.18 (b)).

2. The isolation of poor households became less remarkable in the outskirts of the city, particularly in the clandestine settlements located in the northern and eastern regions of the city. There was also a stronger presence of poor households in central areas of the city, where better quality of infrastructure and higher levels of accessibility can be found (Figure 6.18 (d)).
3. The isolation pattern of affluent households, which was observed in the baseline scenario as an axis starting from the central area of the city towards the western region, became spatially diffuse throughout the city (Figure 6.18 (e-f)). The trend presented in the alternative scenario is positive, since wealthy residents are then more likely to circulate through different parts of the city and increase their contact with distinct social groups and realities (Villaça 1998). In addition, poor families that end up near residential projects for upper classes often benefit in terms of employment, quality of services, and urban facilities (Sabatini 2006; Sabatini et al. 2001).

6.4.3 Impact of regularizing informal settlements and providing an equitable distribution of infrastructure

This final experiment tests whether urban policies that aim at improving the life conditions of the urban poor can influence the segregation patterns of the city. To conduct this test, the baseline simulation 1991-2000 (section 6.2) was compared with an alternative simulation started from a different initial state. This alternative initial state differs from the original in two aspects:

1. The informal settlements ('type D' neighborhoods) were regularized, i.e., reclassified as 'type C' neighborhoods.
2. The value of the landscape-patch variable *infrastructure* (L_{infra}), which is a composed index that represents the provision of water, sewage, and garbage collection, was set to 1 (maximum value) for all urban patches.

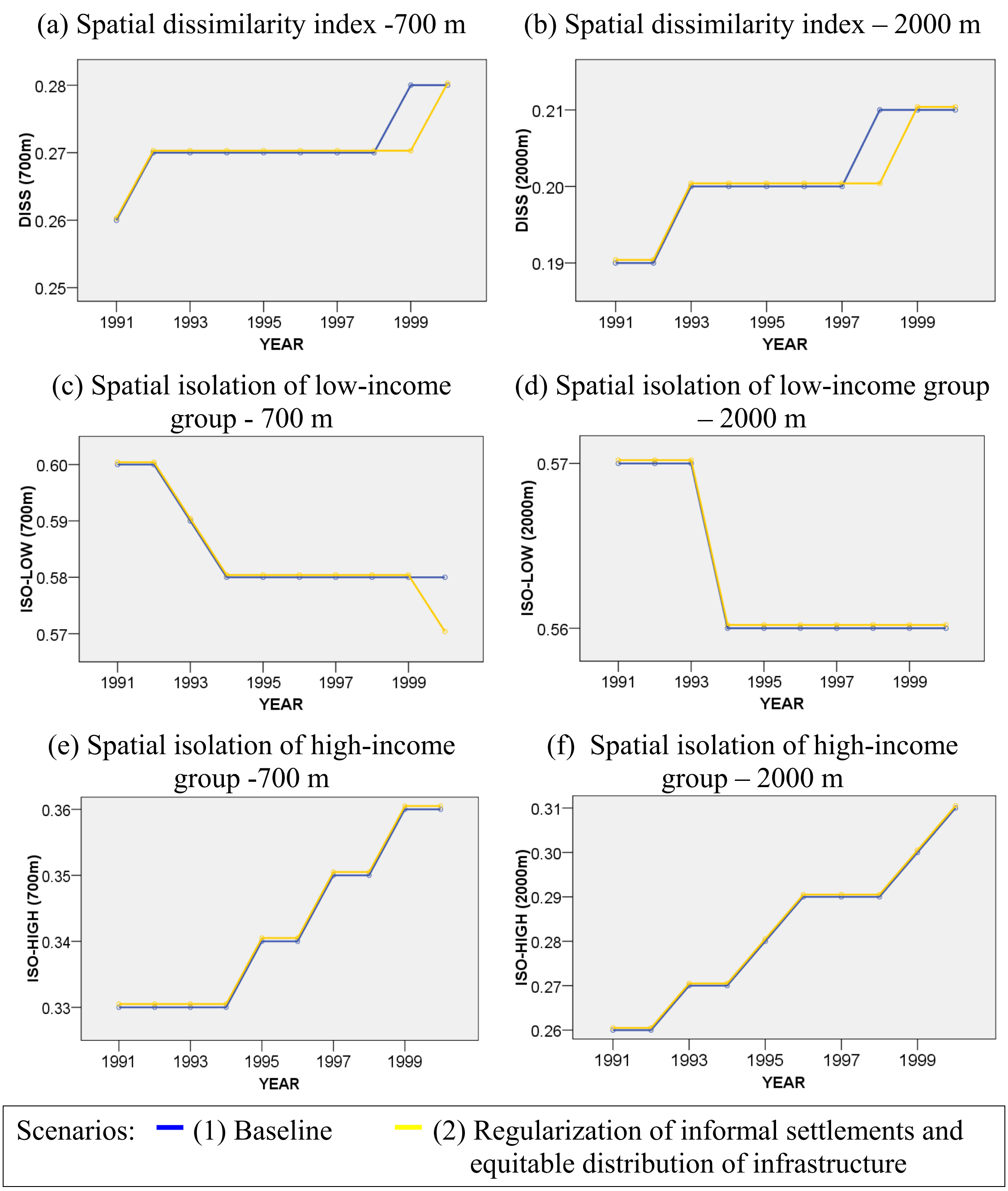


Figure 6.19 Progression of global segregation indices 1991-2000 (radius 700 m and radius 2000 m) for the scenario testing the regularization of informal settlements and equitable distribution of infrastructure

Figure 6.19 shows the global dissimilarity index and isolation indices during the period 1991-2000 computed for the baseline and the alternative scenario where informal settlements are regularized and the infrastructure is equally distributed. In general, the outcomes indicate that these investments had no significant impact on the spatial patterns of segregation in the city of São José dos Campos. However, it is important to mention that the experiment does not invalidate the merit of such policies. Despite their apparent inefficiency to improve the segregation levels, these policies provide innumerous benefits for the quality of the life of poor families and, in some cases, even contribute towards their upward social mobility.

7 CONCLUSIONS

Urban segregation has been a persistent and pervasive feature of cities. Its consequences are harmful to disadvantaged families and impose barriers regarding the achievement of social inclusion in urban areas. To overcome these negative impacts, it is necessary to implement policies founded upon a better understanding of segregation and the influence of different contextual mechanisms on its dynamics. However, studies on segregation face the challenge of dealing with a phenomenon that displays many of the characteristic hallmarks of a complex system. Segregation is a coherent and recognizable macro-structure, but emerges from local interactions able to produce unexpected and counterintuitive outcomes that cannot be defined a priori.

Following the complex systems theory mindset, this study presents an empirically based simulation model named MASUS, Multi-Agent Simulator of Urban Segregation, which enables researchers to explore the impact of different mechanisms on the emergence of segregation patterns. An agent-based simulation approach was chosen for the development of MASUS due to its suitability for addressing the methodological challenges of understanding a complex system like segregation.

MASUS provides a virtual laboratory for testing theoretical issues and policy approaches concerning segregation. It represents urban households as individual units (household agents) that interact with each other and their environment in order to decide whether or not to move to a different residential location. Within this framework, urban segregation arises as an outcome of all these complex interactions. The conceptual MASUS model includes the relevant aspects for simulating segregation in two distinct systems: the urban population system and the urban landscape system. The urban population system is the target system of the MASUS model. It is comprised of micro- and macro-levels: Household agents are considered at the micro-level, while the macro-level represents the urban population in its totality, including the residential location of households with different income levels, i.e., the segregation pattern of the city. The urban landscape system is the environment where household agents are located and provides a dynamic context for their decisions about whether to move or not. Experimental factors addressing both these systems can be modified to perform experiments aiming at exploring relevant questions about segregation.

Based on the components of the conceptual model, three modules were theoretically specified for the operational MASUS. In its essence, the urban-population module characterizes, at the micro-level, the household agent and the decision-making mechanism that rules the agent's residential location choice. At the macro-level, this module defines a population transition sub-model that keeps the socio-demographic composition of the population according to user-defined proportions. The urban-landscape module defines landscape patches, which are individual parts of the environment, and sub-models that simulate the dynamics of landscape-patche attributes that are relevant, directly or indirectly, to the locational behavior of household agents (e.g., land price and infrastructure). Finally, the experimental-factor module consists of specification templates regarding causal mechanisms of segregation that can be set to test theoretical issues and the impact of policies on the segregation patterns of an urban area.

As an empirically based simulation model, MASUS provides different levels of generalization in each of its specification levels. The conceptual framework is highly generalizable and can be applicable to distinct types of segregation in different contexts. The theoretical specification, however, cannot achieve the same level of generalization, since some specifics necessary for the MASUS implementation depend on the availability of data and empirical parameterization.

The MASUS model was first implemented for São José dos Campos, a medium-sized Brazilian city. Based on the data of this city, the model was parameterized and calibrated. Census data and a survey including the residential mobility history of 7,910 households were used to parameterize the decision-making sub-model guiding the behavior of household agents. This sub-model, which is the most important sub-model of the urban-population module, adopts an approach based on utility maximization using nested multinomial logit functions. The nested framework adopted in the specification of these functions is organized in three levels. The first level concerns the household decision about moving or staying, and focuses on how personal attributes such as age and tenure status can influence the mobility rate of different income groups. The second and third levels focus on how households assess the characteristics of potential residential locations. The second level considers the impact of these characteristics in terms of households' neighborhood type choices, while the third level concerns their general impact on the location choice, regardless of the neighborhood type.

To estimate the parameters of the urban-landscape sub-models, we used data from two distinct dates (1991 and 2000), which were obtained from different sources, including satellite images, census data, and official maps. The urban-sprawl sub-model relies on the Markov chain to compute the total number of patches converting from non-urban to urban, and binary logistic regression to estimate the probability of a non-urban patch becoming urban and to allocate the new urban patches. The dwelling offers sub-model updates the number of dwellings of a patch based on two linear regression models: One that estimates the patches' loss of dwellings due to the expansion of non-residential uses (e.g., expansion of commercial use in residential areas), and another that estimates the gain due to new investments in residential developments. The land-value sub-model is based on a hedonic price model to estimate a patch's land value, while the infrastructure sub-model relies on linear regression models to estimate the infrastructure quality of each patch.

Given the proposed theoretical framework and the parameters estimated from empirical data, the operational MASUS model was implemented in NetLogo 4.0.4, a multi-agent programmable modeling environment. This thesis presents the MASUS computer program, including details of its main sub-programs, inputs, outputs, and graphic user interface. The potential of the model is demonstrated through three different sets of simulation experiments concerning segregation in São José dos Campos: the first validates the model, the second tests theories about segregation, and the third explores the impact of anti-segregation policies.

The first set of experiments provides a retrospective validation of the MASUS model by simulating the segregation dynamics of São José dos Campos during the period 1991-2000. The initial state of the experiment replicates the characteristics of the city in 1991. Nine annual cycles were executed and the simulated outputs were compared with real data from the year 2000. In general, simulated and real data reveal the same trends, a result that demonstrates that the model is able to accurately represent the segregation dynamics of the study area.

The second set of experiments aims at demonstrating the potential of the MASUS model to explore and test theoretical issues about urban segregation. These experiments explore the impact of two mechanisms on segregation: income inequality (as a product of the labour market) and personal preferences. To test the impact of income inequality on segregation, scenarios considering different income distributions were simulated and compared. The results, sometimes unexpected, show how decreasing levels of income inequality promote the spatial integration of different social groups in the city.

Following this experiment, new tests were conducted to explore how the neighborhood preferences of high-income families could affect segregation patterns. The simulated outputs indicate a linear and positive relation between indices measuring different dimensions of segregation and the preference of affluent families for neighbors with similar income levels. This relation, however, is not substantial when considering the isolation index of poor residents. The results reveal, for instance, that the high levels of poverty isolation were maintained even in a scenario where affluent households did not take into account the income composition of neighborhoods when selecting their residential location. This level of poverty isolation probably persists because affluent families, independent of the preference regarding their neighbors, still choose to live in high-quality areas that the poorest families cannot afford.

Finally, the third set of experiments provides new insights about the impact of different urban policies on segregation. One experiment tests whether the regularization of clandestine settlements and equitable distribution of infrastructure would affect the segregation trends in the city. Despite the importance of these measures in improving the life conditions of the urban poor, the simulated outputs indicate that they had no significant impact on the segregation patterns.

In addition to this test focusing on a general urban policy, two specific social-mix policy approaches were explored: based on poverty dispersion and on wealth dispersion. The policy promoting poverty dispersion moves poor households out of distressed areas by distributing housing vouchers to be used for renting private dwellings in neighborhoods with a low poverty rate. The other policy, based on the dispersion of wealth, focuses on providing incentives for the construction of residential areas for middle and upper classes in poor regions of the city.

A comparison between the scenarios simulating these two policies reveals that poverty dispersion is the least effective strategy to promote positive changes in the segregation of developing cities. In these cities, where a substantial part of the population has a low income level, policies based on the dispersion of the poor require very high investments that tend to be unfeasible and, once the investments cease, their positive impacts on segregation are not sustained. On the other hand, the policy based on wealth dispersion was able to produce substantial and long-term improvements in the segregation patterns of the city. These improvements became more visible with time, as the consolidation of the residential developments for middle and upper classes started to began to become effective. The simulation experiments also revealed that, unlike in the poverty dispersion policy approach, the dispersion of wealth is more efficient in decreasing large-scale segregation, which is considered to have a more damaging impact on the lives of poor citizens than local-scale segregation.

7.1 Limitations and recommendations

MASUS is a scientific tool able to produce simulation scenarios that contribute to a better understanding of segregation and the impact of different mechanisms on its dynamics. Nevertheless, as with any other tool that simulates a complex system, MASUS outputs must be interpreted with caution. They do not represent quantitative and accurate forecasting of segregation patterns, nor do they provide a deterministic answer regarding the best policy approaches. Instead, these simulation outputs should be considered in terms of how the different factors of the model are related and contribute to a change in segregation dynamics. During this process, it is still important to keep in mind that no model can explicitly represent all the factors that are relevant for the residential location choice of households. Only after such observations and deliberations is possible to obtain insights that contribute towards structuring debates on open theoretical questions about segregation or the development of better informed anti-segregation policies.

Regarding the decision-making sub-model that guides the behavior of households, which represents the main ‘engine’ of the MASUS model, some conclusions can be drawn. In the current version of MASUS, this sub-model relies on nested multinomial logit (NMNL) functions, which jointly model a household’s mobility choice, neighborhood type choice, and specific neighborhood location choice. This joint modeling approach has the advantage of assuming, for instance, that the household's mobility decision (move or stay) is influenced by the characteristics of the residential alternatives available on the market. Nevertheless, an important drawback of the use of NMNL is the fact that these statistical models essentially provide a static representation of the agents` reasoning. Therefore, the model outcomes are not likely to be robust once the agents' behavior changes. To address this issue, further research should explore the development of adaptive and learning agents (Gilbert 2008; Holland and Miller 1991; Maes 1994).

Additional shortcomings of the decision-making sub-model that should be considered in an improved version of MASUS include:

1. The sub-model does not take into consideration the past decisions of households, i.e., households have no memory when deciding whether to stay in their current residence or move to a new neighborhood. More empirical research should be done about the impact of these past decisions on the locational behavior of households and a new version of the decision-making sub-model able to support these new findings should be developed.
2. The sub-model ignores the influence of the neighbors' behavior on the decision process of a household. By considering this spatial component, it could be possible to capture factors or events associated with a specific neighborhood that were not explicitly represented in the model but nevertheless influence the mobility of households living in this area. An alternative to overcome this drawback is to explore the use of spatial discrete choice models (Flemming 2004) to represent the residential choice behavior of households.
3. For simulating the segregation dynamics of São José dos Campos, a city where the number of households varied from 107,045 to 142,541 during the period 1991-2000, the decision sub-model presented high computational costs. This can become a crucial limitation for adapting the model to mega cities, e.g., the metropolis of São Paulo, which has more than 19 million inhabitants. Therefore, it is necessary to look for alternative modeling strategies able to address this shortcoming, e.g., the use of an agent-based simulation platform that prioritizes the execution speed, such as MASON or Repast (Railsback et al. 2006).

The sub-models simulating the dynamics of the urban environment also deserve further consideration. The dynamics of residential land markets, in particular, are crucial for the establishment of segregation patterns, and the simulation of these dynamics poses additional challenges that are not addressed in the current version of the MASUS model. In this version, the land-price sub-model relies exclusively on inductive models of price expectation formation based on local neighborhoods and spatial externalities, ignoring the role of competitive bidding in this process. To improve the simulation of land markets, the sub-model should be more closely linked to urban economics by combining the inductive models already used in the first version of MASUS with deductive models of bid and ask price formation, as suggested by Parker and Filatova (2008).

The measurement of segregation, which is crucial for monitoring the simulation outcomes, represents an open issue that should also be reviewed. The current version of MASUS adopts the spatial indices of dissimilarity and isolation suggested by Feitosa (2007). These indices have the advantage of presenting not only global versions that summarize the segregation degree of the whole city, but also local versions that assume the spatial variance of segregation through the study area. However, these measures are more adequate to categorical variables (e.g., race), being unable to take the original distribution of continuous variables into account. This fact represents a limitation for the use of these indices in this work, since income is a continuous variable and collapsing it into a limited number of income groups certainly causes a loss of information. There are a number of global segregation measures appropriate for continuous variables (Jargowsky and Kim 2005; Reardon et al. 2006). The rank-order information theory index, proposed by Reardon et al. (2006), is a particularly interesting measure that relies on information about the rank ordering of incomes among households and could be explored in further versions of the MASUS model. Nevertheless, further research is needed to develop local segregation measures for continuous variables, which depict segregation as a spatially variant phenomenon and can be displayed as maps.

The MASUS model is built on a framework that can be adapted to different urban realities. For that, it is necessary first to identify essential factors influencing the residential mobility of households in the study area, a task that can be done through literature review and/or exploratory analysis of empirical data. Based on these initial findings, the adaptation of the MASUS model to a new urban context will consist of:

1. Reviewing the sub-models' structures, which may involve the inclusion of different variables, the specification of new functions, or even the development of new sub-models responsible for simulating environmental aspects that are not explicitly represented in the current version of the model;
2. Performing a new parameterization and calibration based on empirical data of the study area;
3. Idealizing and conducting experiments that meet the objectives of the study and the specificities of the study area.

Finally, there is a wide range of experiments that can still be explored in MASUS. It is possible, for instance, to investigate how segregation can be affected by policies that diversify land uses or control land speculation. Nevertheless, further improvement in the usability of the MASUS and in the design of experiments still depends on feedbacks obtained from potential users and stakeholders. Several techniques based on principles of participatory research have been suggested to keep users closely involved in the model development, testing and use, including techniques such as rapid iterative development and user workshops (Ramanath and Gilbert 2004).

8 REFERENCES

Abramo P (2001) Mercado e ordem urbana: do caos à teoria da localização residencial. Bertrand Brasil, Rio de Janeiro

Abramo P (2002) Mercado imobiliário carioca: algumas conclusões. Prefeitura Municipal do Rio de Janeiro, PMRJ, Rio de Janeiro

Allen C et al (2005) Mixed tenure, twenty years on: nothing out of the ordinary. Joseph Rowntree Foundation, York

Andersson R (2008) Neighborhood effects and the welfare state: towards a European research agenda? Schmollers Jahrbuch 128(1):49-63

Atkinson R (2005) Neighbourhoods and the impacts of social mix: crime, tenure diversification and assisted mobility. Housing and Community Research Unit, ESRC Centre for Neighbourhood Research, Tasmania

Axelrod R (2003) Advancing the art of simulation in the social sciences. Jpn J Manag Inf Syst 12(3):3-16

Bar-Yam Y (1997) Dynamics of complex systems. Addison-Wesley, Reading, Massachusetts

Barbon AL (2004) Mobilidade residencial intra-urbana em grandes centros - Região Metropolitana de São Paulo - estudo de caso. Paper presented at XIV Encontro Nacional de Estudos Populacionais, ABEP, Caxambú, 20-24 Sept 2004

Batty M (2005) Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals. The MIT Press, London

Batty M (2009) Urban modelling. In: Kitchin R and Thrift N (eds) International Encyclopedia of Human Geography, Elsevier, London

Batty M, Barros J, and Alves Júnior S (2006) Cities: continuity, transformation and emergence. In: Garnsey E and McGlade J (eds) Complexity and co-evolution: continuity and change in socio-economic systems, Edward Elgar Publishing, Cheltenham

Batty M and Torrens P (2005) Modelling and predicition in a complex world. Futures 37(7):745-766

Bell EJ and Hinojosa RC (1977) Markov analysis of land use change: continuous time and stationary processes. Socio Econ Plan Sci 8:13-17

Bell W (1954) A probability model for the measurement of ecological segregation. Soc Forces 32:337-364

Ben-Akiva M (1973) The structure of travel demand models, Massachusetts Institute of Technology Cambridge, MA

Ben-Akiva M and Lerman SR (1987) Discrete choice analysis: theory and application to travel demand. The MIT Press, Cambridge, Massachusets

Benenson I, Omer I, and Hatna E (2002) Entity-based modeling of urban residential dynamics: the case of Yaffo, Tel Aviv. Environ Plann B 29:491-512

Benenson I and Torrens PM (2004) Geosimulation: automata-based modeling for urban phenomena. John Wiley & Sons, London

Bichir R (2006) Segregação e acesso a políticas públicas no município de São Paulo. Master thesis, University of São Paulo, São Paulo

Bichir RM, Torres HG, and Ferreira MP (2004) Jovens no municipio de São Paulo: explorando o efeito das relações de vizinhança. Rev Bras Est Urb Reg 6(2):53-69

Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, Oxford

Bonduki N and Rolnik R (1979) Periferia da grande São Paulo: reprodução do espaço como expediente de reprodução da força de trabalho. In: Maricato E (ed) A produção capitalista da casa (e da cidade) do Brasil industrial, Editora Alpha-Ômega, São Paulo

Borgers AWJ and Timmermans HJP (1993) Transport facilities and residential choice behaviour: a model of multi-person choice process. Pap Reg Sci 72(1):45-61

Börsch-Supan A (1990) On the compatibility of nested logit models with utility maximization. J Econometrics 43:373-388

Bresser Pereira LC (1992) A crise do Estado: ensaios sobre a economia brasileira. Nobel, São Paulo

Briggs XS (2005) Social capital and segregation in the United States. In: Varady D (ed) Desegregating the city, Suny Press, Albany

Bruch EE (2006) Residential mobility, income inequality, and race/ethnic segregation in Los Angeles. Paper presented at Population Association of America (PAA) 2006 Annual Meeting Program, Los Angeles, CA, 30 Mar -1 Apr 2006

Bruch EE and Mare RD (2006) Neighborhood choice and neighborhood change. Am J Sociol 112(3):667-709

Burgess EW (1924) The growth of the city: an introduction to a research project. Publications of the Am Sociol Rev 18:85-97

Caldeira T (2000) City of walls: crime, segregation and citizenship in São Paulo. University of California Press, Berkeley

Caldeira T (2005) Fortified enclaves: the new urban segregation. In: Jan L. and Mele C. (eds) The urban sociology reader, Routledge

Castle CJE and Crooks AT (2006) Principles and concepts of agent-based modelling for developing geospatial simulations. UCL Centre for Advanced Spatial Analysis, London

Cecchetto F and Monteiro S (2006) Discriminação, cor e intervenção social entre jovens na cidade do Rio de Janeiro (RJ, Brasil): a perspectiva masculina. Rev Est Fem 14(1):199-218

Chang C-O, Chen S-M, and Somerville T (2003) Economic and social status in the household decision-making: evidence relating to extended family mobility. Urban Stud 40(4):733-746

Clampet-Lundquist SC (2004) HOPE VI relocation: moving to new neighborhoods and building new ties. Hous Policy Debate 15(2):415-447

Clark WAV (1991) Residential preferences and neighborhood racial segregation: a test of the Schelling segregation model. Demography 28(1):1-19

Clark WAV and Davies Withers S (1999) Changing jobs and changing houses: mobility outcomes of employment transitions. J Regional Sci 39(4):653 - 673

Clark WAV and Onaka JL (1983) Life cycle and housing adjustment as explanations of residential mobility. Urban Stud 20:47-57

Claydon J and Smith B (1997) Negotiating planning gains through the British development control system. Urban Stud 34(12):2003-2022

Cole I and Goodchild B (2001) Social mix and the ‘balanced community’ in British housing policy – a tale of two epochs. GeoJournal 51:351-360

Costa L (1991) Relatório do Plano Piloto de Brasília, ArPDF, Codeplan e DePHA, Brasília

Crooks AT (2006) Exploring cities using agent-based models and GIS. UCL Centre for Advanced Spatial Analysis, London

Crooks AT (2008) Constructing and implementing an agent-based model of residential segregation through vector GIS, Centre for Advanced Spatial Analysis (UCL), London

Davies RB and Pickles AR (1985) A panel study of life-cycle effects in residential mobility. Geogra Anal 17:199-216

Dieleman FM (2001) Modelling residential mobility: a review of recent trends in research. J Hous Built Env 16:249-265

Dooley K (2002) Simulation research methods. Blackwell, London

Drogoul A, Vanbergue D, and Meurisse T (2003) Multi-agent based simulation: where are the agents? Lect Notes in Comput Sc 2581:43-49

Duncan OD and Duncan B (1955) A methodological analysis of segregation indexes. Am Sociol Rev 20(3):210-217

Epstein JM and Axtell R (1996) Growing artificial societies: social science from the bottom up. Brookings Institution Press, Washington, DC

Ettema D, Jong Kd, Timmermans H, and Bakema A (2005) PUMA: multi-agent modelling of urban systems. Paper presented at 45th Congress of the European Regional Science Association "Land Use and Water Management in a Sustainable Network Society", Vrije Universiteit, Amsterdam, 23-27 Aug 2005

Fagiolo G, Windrum P, and Moneta A (2006) Empirical validation of agent-based models: a critical survey. LEM Working Paper Series, 2006/14. Sant'Anna School of Advanced Studies, Pisa, Italy

Feins JD and Shroder MD (2005) Moving to opportunity: the demonstration's design and its effect on mobility. Urban Stud 42(8):1275-1299

Feitosa FF, Reyes J, and Zesk WE (2008) Spatial patterns of residential segregation: a generative model. Paper presented at X Brazilian Symposium on GeoInformatics, Geoinfo 2008, Rio de Janeiro, 8-10 Dec 2008

Feitosa FF (2005) Índices espaciais para mensurar a segregação residencial: o caso de São José dos Campos (SP). Master thesis, National Institute for Space Research, INPE, São José dos Campos

Feitosa FF et al (2007) Global and local indicators of urban segregation. Int J Geogr Inf Sci 21(3):299-323

Fernandes E (2006) Principles, bases and challenges of the national programme to support sustainable urban land regularisation in Brazil. In: Huchzermeyer M and Karam A (eds) Informal settlements: a perpetual challenge? Juta and Company Limited

Fernandes E (2007) Constructing the 'right to the city' in Brazil. Soc Legal Stud 16(2):201-219

Fernandes G (2005) Anhangabaú. http://www.flickr.com/photos/gaf/46596717. Cited 15 Jun 2009

Flemming M (2004) Techniques for estimating spatially dependent discrete choice models. In: Anselin L. et al (eds) Advances in spatial econometrics. Springer, Amsterdam

Ford RG and Smith GC (1990) Household life-cycle change in the urban council housing sector, 1971-81. Environ Plann A 22:53-67

Fosset M and Senft R (2004) SimSeg learning edition: overview and user's guide. Department of Sociology, Texas A&M University, College Station

Fossett M and Senft R (2004) SIMSEG and generative models: A typology of model-generated segregation patterns. Paper presented at Agent 2004 Conference on Social Dynamics: Interaction, Reflexivity and Emergence, Chicago, IL, 7-9 Oct 2004

Fossett M and Waren W (2005) Overlooked implications of ethnic preferences for residential segregation in agent-based models. Urban Stud 42(11):1893-1917

Franklin S and Graesser A (1997) Is it an agent, or just a program? A taxonomy for autonomous agents. In: Muller JP et al (eds) Inteligent agents III: agents theories, architectures, and languages. Springer-Verlag, Berlin

Fukayama F (1995) Trust: the social virtues and the creation of prosperity. Hamish Hamilton, London

Gayda S (1998) Stated preference survey on residential location choice and modal choice in Brussels. Paper presented at World Conference on Transportation Research, Antwerpen, 12-16 July 1998

Gilbert GN and Troitzsch KG (1999) Simulation for the social scientist. Open University Press, Buckingham

Gilbert N (2004) Agent-based social simulation: dealing with complexity. http://cress.soc.surrey.ac.uk/web/resources/ABSS%20-%20dealing%20with%20complexity-1-1.pdf. Cited 4 Mar 2009

Gilbert N (2008). Agent-based models. Sage Publications, London

Gilbert N and Bankes S (2002) Platforms and methods for agent-based modeling. Proceeding of the National Academy of the United States of America 99(3):7197-7198

Gilbert N and Terna P (1999) How to build and use agent-based models in social science. Mind Soc 1(1):57-72

Gillham O and MacLean AS (2002) The limitless city: a primer on the urban sprawl debate, Island Press, Washington, DC

Ginsberg M (1989) Universal planning: an (almost) universally bad idea. AI Magazine 10(4):40-44

Gouvêa LAC (1995) Brasília: a capital da segregação e do controle social. Annablume, São Paulo

Graham F and Isaac AG (2002) The behavioral life-cycle theory of consumer behavior: survey evidence. J Econ Behav Organ 48(4):391-401

Greene WH (2000) Econometric analysis. Prentice-Hall, Upper Saddle River, NJ

Hanneman RA, Collins R, and Mordt G (1995) Discovering theory dynamics by computer simulation: experiments on state legitimacy and imperialist capitalism. Am Sociol Assoc 25:1-46

Harris C and Ullman EL (1945) The nature of cities. Ann Acad Polit Soc Sci 242:7-17

Herriges J and Kling C (1996) Testing the consistency of nested logit models with utility maximization. Econ Letters 50:33-39

Holland JH (1975) Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence. University of Michigan Press, Ann Arbor

Holland JH (1998) Emergence: from chaos to order. Oxford University Press, Oxford

Holland JH and Miller JH (1991) Artificial adaptive agents in economic theory. Am Econ Rev: Papers and Proceedings of the 103rd Annual Meeting of the American Economic Association 365-370

Hoyt H (1939) The structure and growth of residential neighborhoods in American cities. Federal Housing Administration, Washington DC

Huang Y and Clark WAV (2002) Housing tenure choice in transitional urban China: a multilevel analysis. Urban Stud 39(1):7-32

Hughes PJA (2004) Segregação socioespacial e violência na cidade de São Paulo: referências para a formulação de políticas públicas. São Paulo em Perspectiva 18(4):93-102

IBGE (1991a) Censo demográfico 1991 - microdados do universo. Instituto Brasileiro de Geografia e Estatística, IBGE, Rio de Janeiro.

IBGE (1991b) Censo demográfico 1991 - resultados do universo: agregados de setores censitários. Instituto Brasileiro de Geografia e Estatística, IBGE, Rio de Janeiro.

IBGE (2007) Produto interno bruto dos municípios 2003-2006. Instituto Brasileiro de Geografia e Estatística, IBGE, Rio de Janeiro.

IBGE (2008) Estimativa das populações residentes segundo os municípios. Instituto Brasileiro de Geografia e Estatística, IBGE, Rio de Janeiro

Ihlanfeldt KR (2004) Exclusionary land-use regulations within suburban communities: a review of the evidence and policy prescriptions. Urban Stud 41(2):261-283

INPE (1990) LANDSAT-5/TM satellite image WRS 219/76, aquisition date 27-04-90. Instituto Nacional de Pesquisas Espaciais, INPE, São José dos Campos.

INPE (2000) LANDSAT-7/ETM+ satellite image WRS 219/76, acquisition date 09-01-00. Instituto Nacional de Pesquisas Espaciais, INPE, São José dos Campos.

Jakubs JF (1981) A distance-based segregation index. J Socio Plann Sci 61:129-136

Janssen MA and Ostrom E (2006) Empirically based, agent-based models. Ecol Soc 11(2):37

Jargowsky P (1997) Poverty and place: ghettos, barrios, and the American city. Russel Sage Foundation, New York

Jargowsky PA (1996) Take the money and run: economic segregation in U.S. metropolitan areas. Am J Sociol 61(6):984-999

Jargowsky PA and Kim J (2005) A measure of spatial segregation: the generalized neighborhood sorting index, University of Texas, Dallas

Jennings N, Sycara K, and Wooldridge MJ (1998) A roadmap of agent research and development. Auton Agent Multi-Ag 1:7-38

Jennings N and Wooldridge MJ (2000) Agent-oriented software engineering. In: Bradshaw J. (eds) Handbook of Agent TechnologyAAAI/MIT Press, Cambridge, MA

Kan K (1999) Expected and unexpected residential mobility. J Urban Econ 45:72-96

Katzman R and Retamoso A (2006) Segregación residencial en Montevideo: desafíos para la equidad educativa. Paper presented at Reunión de Expertos sobre Población y Pobreza en América Latina y el Caribe, CELADE/UNFPA, Santiago, Chile

Klahr D, Langley P, and Neches RT (1987) Production system models of learning and development. MIT Press, Cambridge, MA

Kleijnen JPC (1995) Verification and validation of simulation models. Eur J Popul 82:145-162

Kleijnen JPC (1999) Validation of models: statistical techniques and data availability. Paper presented at 1999 Winter Simulation Conference, Phoenix, 5-8 Dec 1999

Kleinhans (2004) Social implications of housing diversification in urban renewal: a review of recent literature. J Hous Built Environ 19:367-390

Knapp TA, White NE, and Clark DE (2001) A nested logit approach to household mobility. J Regional Sci 41(1):1-22

Kowarick L (1979) A espoliação urbana. Paz e Terra, Rio de Janeiro

Lago LC (1998) Dinâmica urbana e segregação socioespacial: a metrópole do Rio de Janeiro nos anos 80. Paper presented at XI Encontro Nacional de Estudos Populacionais, ABEP, Caxambu, 19-22 Oct 1998

Lago LC (2000) Desigualdades e segregação na metrópole: o Rio de Janeiro em tempos de crise. Revan/Fase, Rio de Janeiro

Lambin EF (1994) Modelling deforestation processes: a review. European Commission, Brussels

Laurie AJ and Jaggi NK (2003) Role of 'vision' in neighbourhood racial segregation: a variant of the Schelling segregation model. Urban Stud 40(13):2687-2704

Law AM and Kelton WD (2000) Simulation modeling and analysis. McGraw-Hill, Boston

Le QB (2005) Multi-Agent system for simulation of land-use and land cover change. A theoretical framework and its first implementation for an upland watershed in the central coast of Vietnam. Cuvillier Verlag, Göttingen

Le QB, Park SJ, Vlek PLG, and Cremers AB (2008) Land use dynamic simulator (LUDAS): a multi-agent system model for simulating spatial-temporal dynamics of coupled human-landscape system. 1. Structure and theoretical specification. Ecol Inform 3(2):135-153

Levy S (1992) Artificial life. Vintage Books, New York

Lind J (2001) Iterative software engineering for multiagent systems: the MASSIVE method. Springer, Berlin

Luco CA and Rodríguez J (2003) Segregación residencial en áreas metropolitanas de América Latina: magnitud, características, evolución e implicaciones de política. Proyecto Regional de Población CELADE/Fondo de Población de las Naciones Unidas (UNFPA), Santiago de Chile

Luger GF (2005) Artificial intelligence: structures and strategies for complex problem Solving. Addison-Wesley, Harlow

Luke S et al (2005) MASON: A new multi-agent simulation toolkit. Simulation 81:517-527

Macy MW and Willer R (2002) From factors to actors: computational sociology and agent-based modeling. Annu Rev Sociol 28:143-166

Maes P (1994) Modeling adaptive autonomous agents. Artif Life 1(1-2):135-162

Magnusson L (2006) Residential mobility on a dynamic housing market. Paper presented at the European Network for Housing Research Conference, Ljubljana, Slovenia, 2-5 July

Marcuse P (2005) Enclaves yes, ghettoes no: segregation and the state. In: Varady DP (ed) Desegregating the city: ghettos, enclaves, and inequality. State University of New York Press, New York

Maricato E (1979a) Autoconstrução, a arquitetura possível. In: Maricato E (ed) A produção da casa (e da cidade) no Brasil industrial. Alfa-Ômega, São Paulo

Maricato E (ed) (1979b) A produção capitalista da casa (e da cidade) no Brasil industrial. Alfa-Ômega, São Paulo

Marques E and Bichir R (2002) Investimentos públicos, infra-estrutura urbana e produção da periferia em São Paulo. Rev Esp Deb 42:9-30

Marques E and Torres H (eds) (2004) São Paulo: segregação, pobreza e desigualdades sociais. Editora SENAC, São Paulo

Massey DS and Denton NA (1987) Trends in the residential segregation of Blacks, Hispanics, and Asians: 1970-1980. Am Sociol Rev 52(6):802-824

Massey DS and Denton NA (1988) The dimensions of residential segregation. Soc Forces 67(2):281-315

Massey DS and Denton NA (1993) American apartheid: segregation and the making of the underclass. Harvard University Press, Cambridge

Mayer SE (2001) How the growth in income inequality increases economic segregation. The Harris School, University of Chicago, Chicago

McFadden D (1973) Conditional logic analysis of qualitative choice behaviour. In: Zarembka P. (eds) Frontiers of econometrics. Academic Press, New York, NY

McFadden D (1977) Modelling the choice of residential location. Cowlers Foundation, Yale University

McHugh K (2008) Jardim Raineri: brick and wood houses hug the hillsides http://www.prx.org/pieces/26898-brazil-tackles-old-challenges-with-new-responses. Cited 15 Jun 2009

Miller JH and Page SE (2007) Complex adaptive systems: an introduction to computational models of social life. Princeton University Press, Princeton

Minar N, Burkhart R, Langton C, and Askenazi M (1996) The Swarm simulation system: a toolkit for building multi-agent simulations, Santa Fe Institute, Santa Fe, NM

Mitchell M (1996) An introduction to genetic algorithms. MIT Press, Cambridge

Molin EJE and Timmermans HJP (2003) Accessibility considerations in residential choice decisions: accumulated evidence from the Benelux. Paper presented at 82nd Annual Meeting of the Transportation Research Board, Washington D.C., 12-16 Jan 2003

Moore DS and McCabe GP (2003) Introduction to the practice of Statistics. W. H. Freeman & Company, New York

Morgan BS (1975) The segregation of socioeconomic groups in urban areas: A comparative analysis. Urban Stud 12(1):47-60

Morgan BS (1983a) An alternative approach to the development of a distance-based measure of racial segregation. Am J Sociol 88:1237-1249

Morgan BS (1983b) A temporal perspective on the properties of the index of dissimilarity. Environ Plann 15:379-389

Morrill RL (1991) On the measure of spatial segregation. Geo Res Forum 11:25-36

Morris L (1995) Social divisions: economic decline and social structural change. UCL Press, London

Moss S (2008) Alternative approaches to the empirical validation of agent-based models. JASSS 11(1):5

Mulder CH and Hooimeijer P (1999) Residential relocation in the life course. In: Van Wissen LJG and Dykstra PA (eds) Population issues, an interdisciplinary focus. Kluwer Academic/ Plenum Publishers, New York

Musterd S and Andersson R (2005) Housing mix, social mix, and social opportunities. Urban Aff Rev 40(6):761-790

Musterd S, Ostendorf W, and De Vos S (2003) Neighborhood effects and social mobility: a longitudinal analysis. Housing Stud 18(6):877-892

Naiff LAM and Naiff GM (2005) A favela e seus moradores: culpados ou vítimas? Representações sociais em tempos de violência. Est Pesq Psic 5(2):107-119

Neter J, Kutner MH, Nachtsheim CJ, and Wasserman W (1996) Applied linear statistical models. Irwin, Chicago

Nilsson NJ (1998) Artificial intelligence : a new synthesis. Morgan Kaufmann, San Francisco

North (2006) Experiences creating three implementations of the Repast agent modeling toolkit. ACM T Model Comput S 16(1):1-25

O'Sullivan D, Macgill JR, and Yu C (2003) Agent-based residential segregation: a hierarchically structured spatial model. Paper presented at Agent 2003 'Challenges in Social Simulation', University of Chicago and Argonne National Laboratory, IL, 3-4 Oct 2003

Omer I (2005) How ethnicity influences residential distributions: an agent-based simulation. Environ Plann B 32(5):657-672

Pancs R and Vriend NJ (2003) Schelling's Spatial Proximity Model of Segregation Revisited, Queen Mary University of London. Department of Economics, London

Parker DC (2005) Integration of geographic information systems and agent-based models of land use: challenges and prospects. In: Maguire D.J. et al (eds) GIS, spatial analysis and modeling. ESRI Press, Redlands

Parker DC and Filatova T (2008) A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. Comput Environ Urban 32(6):454-463

Paviani A (ed) (1996) Brasília: moradia e exclusão. Editora UNB, Brasília

Pessoa de Souza e Silva MF (2007) Gated communities: the new ideal way of life in Natal, Brazil. Hous Policy Debate 18(3):557-576

Phe HH and Wakely P (2000) Status, quality and the other trade-off: towards a new theory of urban residential location. Urban Stud 37(1):7-35

PMSJC (1990) Lei de zoneamento 3721/90, Prefeitura Municipal de São José dos Campos, PMSJC, São José dos Campos.

PMSJC (2003) Cidade viva. Prefeitura Municipal de São José dos Campos, PMSJC, São José dos Campos.

Poole DL, Mackworth AK, and Goebel R (1998) Computational intelligence : a logical approach. Oxford University Press, New York

Popkin SJ et al (2004) The HOPE VI program: what about the residents? Hous Policy Debate 15(2):385-414

Préteceille E (2003) A evolução da segregação social e das desigualdades urbanas: o caso da metrópole parisiense nas últimas décadas. Caderno CRH 38:27-48

Preteceille E and Ribeiro LCQ (1999) Tendências da segregação social em metópoles globais e desiguais: Paris e Rio de Janeiro nos anos 80. Rev Bras Ciências Sociais 14(40):143-162

Prike M (1998) Thinking social policy into social housing. In: Hughes G. and Lewis G. (eds) Unsettling welfare: the reconstruction of social policy. Routledge, London

Psycology Wiki (2009) Emergence. http://psychology.wikia.com/wiki/emergence. Cited 11 Feb 2009

Putnam R (1995) Bowling alone: America's declining social capital. J Democr 6:65-78

Railsback SF, Lytinen SL, and Jackson SK (2006) Agent-based simulation platforms: review and development recommendations. Simulation 82(9):609-624

Ramanath AM and Gilbert N (2004) The design of participatory agent-based social simulations. JASSS 7(4)

Reardon S and O'Sullivan D (2004) Measures of spatial segregation. Sociol Methodol 34(1):121-162

Reardon SF and Bischoff K (2008) Does income inequality cause income segregation? Evidence from a within-race analysis. Paper presented at Annual Meeting of the American Sociological Association, Boston, MA, 14-17 Aug 2008

Reardon SF and Firebaugh G (2002) Measures of multigroup segregation. Sociol Methodol 32(1):33-67

Reardon SF, Firebaugh G, O'Sullivan D, and Matthews SA (2006) Measures of socioeconomic segregation. Paper presented at 29th General Conference of the International Association for Research in Income and Wealth, Joensuu, Finland, 20-26 Aug 2006

Reardon SF and O´Sullivan D (2004) Measures of spatial segregation, Pennsylvania State University

Ribeiro LCQ (2001) Segregação, desigualdade e habitação: a metrópole do Rio de Janeiro. Paper presented at IX Encontro Nacional da Associação Nacional de pós-graduação e Pesquisa em Planejamento Urbano e Regional, ANPUR, Rio de Janeiro, 25-29 May 2001

Ribeiro LCQ and Santos Junior OA (2003) Democracia e segregação urbana: reflexões sobre a relação entre cidade e cidadania na sociedade brasileira. EURE (Santiago) 29(88):79-95

Richiardi M, Leombruni R, Saam N, and Sonnessa M (2006) A common protocol for agent-based social simulation. JASSS 9(1)

Roberts N (1983) Introduction to computer simulation: the system dynamics approach. Addison-Wesley, Reading

Rocha LM and Araújo E (2008) Programa nacional de estímulo ao primeiro emprego no Rio de Janeiro: desafios para a inserção no mercado de trabalho. Paper presented at XVI Encontro Nacional de Estudos Populacionais, ABEP, Caxambu, 29 Sep - 3Oct 2008

Rodríguez J (2001) Segregación residencial socioeconómica: que és?, cómo se mide?, que está pasando?, importa? CELADE, Santiago de Chile

Rolnik R (1997) A cidade e a lei: legislação, política urbana e territórios na cidade de São Paulo. Studio Nobel/Fapesp, São Paulo

Rolnik R and Saule Jr. N (eds) (2001) Estatuto da Cidade: guia para implementação pelos municípios e cidadãos. Câmara dos Deputados, Coordenação de Publicações, Brasília

Rosenbaum JE (1995) Changing the geography of opportunity by expanding residential choice: lessons from the Gautreux Program. Hous Policy Debate 6(1):231-270

Rosenbaum JE and DeLuca S (2000) Is housing mobility the key to welfare reform? The Broking Institution, Washington DC

Rosenbaum JE, Stroh LK, and Flynn CA (1998) Lake Parc Place: a study of mixed-income housing. Hous Policy Debate(9):703-740

Rossi (1955) Why families move: A study in the social psychology of urban residential mobility. Free Press, Glencoe, IL

Russel SJ and Norvig P (2003) Artificial intelligence: a modern approach. Prentice Hall, Englewood Cliffs, NJ

Sabagh G, Arsdol MV, and Butler E (1969) Some determinants of intrametropolitan residential mobility: conceptual considerations. Soc Forces 48(1):88-98

Sabatini F (2000) Reforma de los mercados de suelo en Santiago, Chile: efectos sobre los precios de la tierra y la segregación residencial. EURE (Santiago) 26(77):49-80

Sabatini F (2004) La segregación residencial en las ciudades latinoamericanas: causas, posibles políticas y rol de los mercados de suelo. In: Arenas F et al (eds) Los nuevos modos de gestión de la metropolización. Pontificia Universidad Católica de Chile, Santiago, Chile

Sabatini F (2006) The social spatial segregation in the cities of Latin America. Inter-American Development Bank, Sustainable Development Department, Social Programs Division

Sabatini F, Caceres G, and Cerda J (2001) Residential segregation in main Chilean cities: tendencies from the past three decades and possible frameworks for action. EURE (Santiago) 27(82):21-42

Sabatini F and Salcedo R (2007) Gated communities and the poor in Santiago, Chile: functional and symbolic integration in a context of aggressive capitalist colonization of lower-class areas. Hous Policy Debate 18(3):577-606

Sabatini F, Wormald G, Sierralta C, and Peters PA (2005) Residential segregation in metropolitan Santiago, 1992-2002: geographical scale and linked social effects. In: Roberts B and Wilson R (eds) Spatial differentiation and governance in the Americas. Palgrave MacMillan, Austin

Sakoda J (1981) A generalized index of dissimilarity. Demography 18(2):245-250

Sakoda JM (1971) The checkerboard model of social interaction. J Math Sociol 1(1):119-132

Santos CNF (1980) Velhas novidades nos modos de urbanização brasileiros. In: Valladares L. (eds) Habitação em questão. Zahar, Rio de Janeiro

Sarkissian W (1976) The idea of social mix in town planning: a historical review. Urban Stud 13(3):231-246

Schelling TC (1971) Dynamic models of segregation. J Math Sociol 1(1):143-186

Schelling TC (1972) The process of residential segregation: neighborhood tipping. In: Pascal A. (eds) Racial discrimination in economic life. D.C. Heath, Lexington, MA

Schelling TC (1978) Micromotives and macrobehavior. W. W. Norton, London & New York

Schiffer SR (2001) Economic restructuring and urban segregation in São Paulo. In: Marcuse P and Kempen R (eds) Of States and cities: the partioning of urban space. Oxford University Press, London

Schut MC (2007) Scientific handbook for simulation of collective intelligence. http://www.sci-sci.org/. Cited 14 Mar 2009

Serenko A and Detlor B (2002) Agent toolkits: a general overview of the market and an assessment of instructor satisfaction with utilizing toolkits in the classroom. Michael G. DeGroote School of Business, McMaster University, Hamilton, Ontario

Silverman BW (1986) Density estimation for statistics and data analysis. Chapman and Hall, London

Silverman E, Lupton R, and Fenton A (2005) A good place for children? Attracting and retaining families in inner urban mixed income communities. Joseph Rowntree Foundation, York

Simon HA (1996) The sciences of the artificial. MIT Press, Cambridge

Sklar E (2007) NetLogo, a multi-agent simulation environment. Artif Life 13(3):303-311

Smets P and den Uyl M (2008) The complex role of ethnicity in urban mixing: a study of two deprived neighborhoods in Amsterdam. Urban Stud 45(7):1439-1460

Smith A (2002) Mixed-income housing developments: promise and reality. NeighborWorks and Joint Center for Housing Studies, Harvard University, Boston

Smith K, Brighton H, and Kirby S (2003) Complex systems in language evolution: the cultural emergence of compositional structure. Adv Complex Syst 6(4):1-22

Smolka MO (1992) Mobilidade intra-urbana no Rio de Janeiro: da estratificação social à segregação residencial no espaço. Rev Bras Est Pop 9(2):97-114

Smolka MO (2005) El funcionamento de los mercados de suelo en América Latina: conceptos, antecedentes históricos y nexos críticos. In: Basualdo JL (ed) Manejo de suelo urbano: posibilidades y desafios en el desarrollo de la ciudad de Corrientes. Lincoln Institute of Land Policy, Instituto de Vivienda de Corrientes, Corrientes, Argentina

Soares F and Soares Y (2005) The socio-economic impact of Favela-Bairro: what do the data say? Inter-American Development Bank, Washington D.C.

South SJ and Deane GD (1993) Race and residential mobility: individual determinants and structural constraints. Soc Forces 72(1):147-167

Speare A (1970) Home ownership, life cycle stage, and residential mobility. Demography 7(4):449-458

Speare A (1974) Residential satisfaction as intervening variable in residential mobility. Demography 11(2):173-188

Sposati A (1999) Exclusão social abaixo da linha do Equador. In: Veras MP (ed) Por uma sociologia da exclusão social - o debate com Serge Paugam. EDUC, São Paulo

Steels L (1997) The synthetic modeling of language origins. Evol Communication 1(1):1-34

Strassmann WP (2001) Residential mobility: contrasting approaches in Europe and the United States. Housing Stud 16(1):7 - 20

Sugai MI (2002) Segregação silenciosa: investimentos públicos e distribuição sócio-espacial na área conurbada de Florianópolis. PhD thesis, Universidade de São Paulo, São Paulo

Tesfatsion L (2003) Agent-based computational economics. Inform Sciences 149(4):262-268

Thomsen J et al (1999) A trajectory for validating computational emulation models of organizations. Comput Math Organ Th 5(4):385-401

Tisue S and Wilensky U (2004) Netlogo: a simple environment for modeling complexity. Paper presented at International Conference on Complex Systems, Boston, 16-21 May 2004

Tobias R and Hofman C (2004) Evaluation of free java-libraries for social-scientific agent based simulation. JASSS 7(1)

Todaro MP and Smith SC (2008) Economic development. Pearson Education, Essex

Torrens P (2000) How cellular models of urban systems Work (1. Theory) . CASA Working Paper Series 28. Centre for Advanced Spatial Analysis, London
Torres HG (2004) Segregação residencial e políticas públicas: São Paulo na década de 1990. Rev Bras Ci Soc 19(54):41-56
Torres HG, Ferreira MP, and Gomes S (2005) Educação e segregação social: explorando as relações de vizinhança. In: Marques EC and Torres HG (eds) São Paulo: segregação, pobreza e desigualdade. Editora do Senac, São Paulo
Torres HG and Marques EC (2001) Reflexões sobre a hiperperiferia: novas e velhas faces da pobreza no entorno metropolitano. Rev Bras Est Urb Reg 4:49-70
Torres HG, Marques EC, and Bichir RM (2006) Políticas públicas, pobreza urbana e segregação residencial. In: Cunha JMP (ed) Novas metrópoles paulistas: população, vulnerabilidade e segregação. Nepo/Unicamp, Campinas
Torres HG, Marques EC, Ferreira MP, and Bitar S (2002) Poverty and space: pattern of segregation in São Paulo. Paper presented at Workshop on Spatial Segregation and Urban Inequality in Latin America, Austin, Texas, 15-16 Nov 2002
Torres HG, Marques EC, Ferreira MP, and Bitar S (2003) Pobreza e espaço: padrões de segregação em São Paulo. Revista do IEA 17(47):13-42
Train KE (2003) Discrete choice methods with simulation. Cambridge University Press, Cambridge
Troitzsch KG (1998) Multilevel process modeling in the social sciences: mathematical analysis and computer simulation. In: Liebrand WBG et al (eds) Computer modeling of social processes. SAGE Publications, London
Troitzsch KG (2004) Validating simulation models. Paper presented at 18th European Simulation Multiconference, SCS Europe, 13-16 Jun 2004
Tunstall R and Fenton A (2006) In the mix: A review of research on mixed income, mixed tenure and mixed communities. What do we know? Joseph Rowntree Foundation, York
Turbov M and Piper V (2005) HOPE VI and mixed-finance redevelopments: a catalyst for neighborhood renewal. Brookings Institution, Washington, DC
Turok I and Edge N (1999) The jobs gap in Britain's cities: employment loss and labour market consequences. The Policy Press, Bristol
Uitermark J (2003) 'Social mixing' and the management of disadvantaged neighbourhoods: the dutch policy of urban restructuring revisited. Urban Stud 40(3):531-549
UN-Habitat (2001a) Inclusive cities initiative: the way forward. UN-Habitat, Nairobi
UN-Habitat (2001b) The state of the World's cities report. UN-Habitat, Nairobi
UNFPA (2007) State of World population 2007: unleashing the potential of urban growth. United Nations Population Fund, UNFPA, New York
Valladares LP (1999) A outra Brasília. Caderno CRH 30/31:369-381
van Kempen ET (1994) The dual city and the poor: social polarisation, social segregation and life chances. Urban Stud 31(7):995-1015
Vieira T (2005) Paraisópolis. http://www.fototucavieira.com.br. Cited 13 Jun 2009
Villaça F (1998) Espaço intra-urbano no Brasil. Studio Nobel, São Paulo
Villaça F (2001) Segregation in the Brazilian metropolis. Paper presented at International Seminar on Segregation in the City, Lincoln Institute of Land Policy, Cambridge, 26-28 Jun 2001

Waddell P (1996) Accessibility and residential location: the interaction of workplace, residential mobility, tenure, and location choice. http://www.oregon.gov/ODOT/TD/TP/docs/TMR/UrbanSim/tred96b.pdf. Cited 12 Apr 2009

Waddell P et al (2007) Modeling the interdependence in household residence and workplace choices. Transp Res Record 2003/2007:84-92

Watson T (2006) Metropolitan growth and neighborhood segregation by income, Williams College, Williamstown

Webster D (1999) Target local jobs: the missing elements in labour's social inclusion policy. New Economy 6(4):193-198

Weyns D et al (2005) Environments in multiagent systems. Knowl Eng Rev 20(2):127-141

Wheeler CH and La Jeunesse EA (2007) Neighborhood income inequality. Research Division, Federal Reserve Bank of St. Louis, St. Louis

White MJ (1983) The measurement of spatial segregation. Am J Sociol 88(4):1008-1018

Wilensky U (1998) NetLogo wealth distribution model. Center for Connected Learning and Computer-Based Modeling (CCL), Northwestern University, Evanston

Wilensky U (1999) NetLogo. Northwestern University, Evanston, IL

Wong DWS (1993) Spatial indices of segregation. Urban Stud 30(3):559-572

Wong DWS (1998) Measuring multiethnic spatial segregation. Urban Geogr 19(1):77-87

Wood T and Murray WE (2007) Participatory democracy in Brazil and local geographies: Porto Alegre and Belo Horizonte compared. Eur Rev Latin Ame Carib Stud 83:19-41

Wooldridge MJ (1997) Agent-based software engineering. Software Engineering, IEE Proceedings 144(1):26-37

Wooldridge MJ and Jennings N (1995) Intelligent agents: theory and practice. Knowle Eng Rev 10(2):115-152

Worboys M and Duckham M (2004) GIS: a computing perspective. CRC Press, Boca Raton

Wu F (2002) Complexity and urban simulation: towards a computational laboratory. Geogr Res Forum 22:22-40

Yilmaz L (2006) Validation and verification of social processes withing agent-based computational organization models. Comput Math Organ Th 12(4):283-312

Yizhaq H, Portnov BA, and Meron E (2004) A mathematical model of segregation patterns in residential neighbourhoods. Environ Plann 36:149-172

Zambonelli F, Jennings N, and Wooldridge MJ (2001) Organisational abstractions for the analysis and design of multi-agent systems. In: Ciancarini P. and Wooldridge M.J. (eds) Agent-oriented software engineering. Springer-Verlag, Berlin

Zhang J (2004) Residential segregation in an all-integrationist world. J Econ Behav Organ 54:533-550

ACKNOWLEDGEMENT

This work would not have been possible without the support from numerous people and organizations from Bonn (Germany) and São José dos Campos (Brazil).

I would like to thank the German government through the German Academic Exchange Services (DAAD) and the Ministry of Education, Science and Technology (BMBF) for providing my stipend and the funds for my fieldwork. I would like to name and thank Ms. Anke Stahl and Ms. Birgitt Skailes, from DAAD, for their continuous support.

I am also thankful to all those who contributed scientifically to this dissertation. My most sincere gratitude goes to Dr. Quang Bao Le for his guidance, continued encouragement, and trust during my research. I am very grateful for his valuable advices and constant willingness to assist me whenever I needed. I would like to extend my gratitude to Prof. Dr. Paul Vlek, my principal supervisor, for selecting me as a student and supporting all the stages of this research. I want to thank Prof. Dr. Eckart Ehlers, Prof. Dr. Klaus Troitzsch, and Prof Dr. Armin B. Cremers for their willingness to be members of the Examination Committee, as well as their valuable comments and suggestions on the dissertation. In addition, I would like to thank Dr. Guido Lüchters for his enthusiastic advices on Statistics.

In Germany, the logistical and academic support for my research was provided by ZEF and the University of Bonn. I would like to express my gratitude to the entire staff members of ZEF, especially to Dr. Günter Manske and Mrs. Rosemarie Zabel. I really want to thank Mrs. Zabel for all her extraordinary assistance during my stay in Germany. It is hard to imagine how ZEF students would be able to survive without her kindness and constant readiness to help. Furthermore, I am very grateful to my friends at ZEF, including Philine, George, Aibek, Dela, Ermias, Job, Edward, Patricia, Chian-Woei, and Daniel. Special thanks go to Alam, my office colleague, for his inspiring and daily smile.

In Brazil, the National Institute for Space Research (INPE) provided the logistical support for this study. I am very grateful to my friends at INPE, who are always making the life in São José dos Campos so special. My warmest thanks go to Dr. Antonio Miguel Vieira Monteiro, my eternal advisor, who always supported my craziest ideas and had something crazier to add. I'll never be able to express how grateful I am for all these years of constant encouragement.

My sincere thanks also go to my incredible friend and hard-working field assistant, Roberta Rosemback, with whom I shared unforgettable adventures while exploring the neighborhoods of São José dos Campos. I am also grateful to Mircea Claro, for helping me with the data, and to the GEOPRO team, for creating a fun research environment.

I would like to express my gratitude to the Municipal Government of São José dos Campos (PMSJC) and to the Population Studies Center (NEPO) of the University of Campinas (UNICAMP), for making available data that were essential to this research. For accepting my data request, I would like to thank Dr. Rosana Baeninger and Dr. Roberto Luiz do Carmo, from NEPO; Cynthia Márcia de Oliveira Gonçalo, director of the Urban Planning Department/PMSJC; and Maria Rita de Cássia Singulano, former director of the Housing Department/PMSJC. I am particularly grateful to Maria Angélica Braga de Avellar Silva and Jussara Starling de Medeiros, from the Urban

Planning Department/PMSJC, for being always so friendly, supportive, and for preparing a copy of the data I requested.

My deepest gratitude goes to Ty Frazier, for being an endless source of strength. Thank you very much for tolerating my mood swings and for making me laugh. Most of all, I would like to thank my parents, Emilson Feitosa and Vera Lúcia da Fonseca Feitosa, and my brother, Leandro Augusto da Fonseca Feitosa, for their everlasting support and incentive. Thanks for being my safe harbor, for always expressing your heartfelt happiness with my achievements, and for keeping saying “Flavinha is always making us proud”. I dedicate this thesis to you.