Multi-agent simulator for urban segregation (MASUS): A tool to explore alternatives for promoting inclusive cities

Flávia F. Feitosa [a,b,*], Quang Bao Le [a,c], Paul L.G. Vlek [a]

[a] Center for Development Research (ZEF), University of Bonn, Walter-Flex-Str. 3, D-53113 Bonn, Germany
[b] Earth System Science Center (CCST), National Institute for Space Research (INPE), Av. dos Astronautas, 1758, 12227-001 São José dos Campos, Brazil
[c] Natural and Social Science Interface (NSSI), Institute for Environmental Decisions (IED), Swiss Federal Institute of Technology Zurich (ETH Zurich), Universitätstr. 22, CH-8092 Zurich, Switzerland

A B S T R A C T

Urban segregation represents a significant barrier to achieving social inclusion in cities. To mitigate this problem, it is necessary to implement policies founded upon a better understanding of segregation dynamics. This paper proposes MASUS, a multi-agent simulator for urban segregation, which provides a virtual laboratory for exploring the impacts of different contextual mechanisms on the emergence of segregation patterns. We illustrate the potential of MASUS through three experiments on segregation in São José dos Campos, a medium-sized city in southeast Brazil. The first experiment compares simulated outputs with empirical data, the second exemplifies the ability of MASUS to test theories, and the third tests an anti-segregation policy. We also discuss limitations of the current version of the model, and we recommend directions for further research.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

In 2008, for the first time in history, 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.

To realize the potential role of cities in fostering development, it is essential to remove the barriers that inhibit the formation of *inclusive cities*, that is, cities capable of promoting growth with equity (UN-Habitat, 2001a). Urban segregation represents one of these barriers because it has reinforced social exclusion in the cities of the developing world (UN-Habitat, 2001b). Different types of urban segregation exist, including income and racial or ethnical segregation, depending on the contextual mechanisms within a city. This paper focuses on income segregation, which is defined as separation among the residential locations of different income groups.

Income segregation negatively impacts urban populations by contributing to the social exclusion of disadvantaged groups. Areas of poverty concentration are often associated with poor quality of built and natural environments, higher exposure to violence and diseases, and other physical issues that impose difficulties on the daily lives of poor families (Katzman & Retamoso, 2006; Préteceille, 2003; Sabatini, Caceres, & Cerda, 2001; Torres, Marques, Ferreira, & Bitar, 2003). In addition, a lack of positive relations among different social groups can increase prejudice and territorial stigmatization and can keep disadvantaged people from societal participation, thus reducing their opportunities for jobs and skill improvement (Atkinson, 2005; Briggs, 2005; Torres, 2004). Studies suggest that integration among different income groups benefits poor families (Caldeira, 2000; Jargowsky, 1997; Sabatini, 2006). Such integration, however, is unlikely to occur without introducing policies that are founded upon a better understanding of segregation dynamics and how different contextual mechanisms can influence these dynamics.

The contextual mechanisms contributing to urban segregation are many, and they vary from place to place (UN-Habitat, 2001b). Approaches to study the behavior of social groups while selecting their residential location focus on at least four different yet interdependent mechanisms: (a) personal preferences; (b) the labor market; (c) land and real estate markets; and (d) state policies and investments. The first approach, focusing on personal preferences, proposes that segregation increases because people prefer to have neighbors similar to them (Schelling, 1971). This is common among affluent families, who often seek high status or who want to protect themselves from poverty-related problems (Caldeira, 2000). The second approach proposes that the labor

* Corresponding author at: Earth System Science Center (CCST), National Institute for Space Research (INPE), Av. dos Astronautas, 1758, 12227-001 São José dos Campos, Brazil.
E-mail addresses: flavia@dpi.inpe.br (F.F. Feitosa), quang.le@env.ethz.ch (Q.B. Le), p.vlek@uni-bonn.de (P.L.G. Vlek).

market and its socio-economic impacts, such as social exclusion and inequality, are responsible for segregation and the precarious living conditions of poor families (Gerometta, Hausermann, & Longo, 2005; Lago, 2000). The third approach concentrates on land and real estate markets, and it stresses how developers and their agents stimulate competition for housing, which reinforces the self-segregation of affluent groups and excludes poor families (Abramo, 2001). Finally, the fourth approach stresses how the state 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).

Studies that seek to understand how the aforementioned mechanisms influence segregation dynamics can be challenging because segregation displays many of the hallmarks of complexity, including emergence and nonlinearity. Segregation is a macro-scale phenomenon, but it emerges from the residential choices of many individuals at the micro-level. These choices are driven by contextual mechanisms that not only influence the emergence of segregation patterns but are also influenced by them. These positive feedbacks introduce nonlinearity into the system. As a result, small differences in context or local behavior are able to produce unexpected and counterintuitive outcomes that are not equivalent to the simple sum of the constituent parts (Batty, 2008; Holland, 1998; Schelling, 1971, 1972).

The complex nature of segregation imposes difficulties on traditional modeling approaches, such as statistical modeling or classical optimization (Batty, 2005). Agent-based modeling (ABM) addresses the shortcomings of these traditional techniques by introducing individual decision-making units, called agents, that interact with each other and their environment (Batty, 2005; Gilbert, 2008). ABM offers a natural presentation of real-world processes that underlie complex social dynamics such as urban segregation (Benenson, 2004; Gilbert & Troitzsch, 1999; Miller & Page, 2007; Torrens & Benenson, 2005). Because of these advantages, ABM has been recognized as a promising approach to modeling complex segregation dynamics.

Several studies have explored segregation issues by using ABM (Benenson, Hatna, & Or, 2009; Benenson, Omer, & Hatna, 2002; Bruch & Mare, 2006; Clark & Fosset, 2008; Fossett & Senft, 2004; Schelling, 1971). The Schelling model is considered the first to demonstrate the concept of ABM and its ability to provide new insights for understanding complex phenomena (Schelling, 1971). His work inspired many other researchers, who have developed variations of the Schelling model by using alternative utility functions (Bruch & Mare, 2006; Clark, 1991; Pancs & Vriend, 2003). These variations include the following: recognizing preferences for housing and neighborhood quality (Fossett & Senft, 2004); adopting different notions of neighborhoods (Fossett & Waren, 2005; Laurie & Jaggi, 2003; O'Sullivan, Macgill, & Yu, 2003); testing different multigroup situations (Clark & Fosset, 2008); 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. One example is the ethnical segregation model for the Yaffo area of Tel Aviv, developed by Benenson and colleagues (Benenson et al., 2002, 2009). The racial segregation model of Los Angeles developed by Bruch (2006) is another example.

This paper presents an empirically based model of income segregation named MASUS – multi-agent simulator of urban segregation. The purpose of MASUS is to provide a virtual laboratory for exploring the impact of different mechanisms on the emergence of segregation patterns. We present the conceptual framework of MASUS and the model's specifications, and we illustrate its

potentiality through three experiments on segregation in São José dos Campos, a medium-sized city located in the State of São Paulo, Brazil.

The first experiment compares simulated outputs that replicate a past segregation scenario in the city with empirical data, and tests whether the model can provide an adequate representation of segregation patterns as observed in São José dos Campos. The aim of the second experiment is to demonstrate how MASUS can be used to test theoretical issues of segregation, exploring the relationship between income inequalities, seen as a product of the labor market, and the configuration of segregation patterns. The third experiment demonstrates the ability of MASUS to provide insights regarding the impact of social-mix policies.

2. Conceptual framework

The conceptual framework used as the basis for specifying the MASUS model has three main components: *urban population*, *urban landscape*, and *experimental factors*. Fig. 1 presents these components and their interrelations.

The *urban population* represents the targeted system of the MASUS model. It depicts micro and macro-level processes of the system. At the *micro-level*, household agents represent the residents of the city, which have their specific state and decision-making sub-model. At the *macro-level* of the system, the urban population emerges from the activities of household agents over space and time. The urban population is characterized by spatial and non-spatial attributes. The spatial characteristics correspond to the residential location of households belonging to different social groups, that is, the segregation pattern of the city. The non-spatial characteristics correspond to the entire aggregation of household attributes, such as the income and education levels of the population as a whole.

The *urban landscape* is the environment where household agents are situated and where they act. This environment provides a spatially explicit context for the agents' decisions about whether to move or remain in place. The urban landscape is represented as a grid of patches or cells with their own state and transitional dynamics. The cell state is described by a list of spatial variables that are relevant for the households' behavior, such as land use type, land value and infrastructure. In the MASUS model, the main dynamics of the urban landscape system are driven by four sub-models: urban sprawl, dwelling offers, infrastructure quality, and land value. These landscape processes function in parallel with the residential mobility of households. For example, once a household agent decides to move to another place, there is a change in the dwelling offers of both the previous and new locations. The *experimental factors* represent exogenous parameters and input data that can be changed to test theories and policy approaches toward segregation.

The segregation pattern of the city emerges from the individual decisions of numerous household agents. According to its decision-making sub-model, an agent may decide to act (i.e., move to another location). The agent's decision depends on its state, which is comprised of the agent's attributes, and its perceptions about different residential locations. These perceptions take into consideration the environmental characteristics (urban landscape) and the population composition of neighborhoods population (urban population) of potential locations. Household residential mobility leads to changes in the spatial arrangement of the population, and therefore, the segregation of the city (macro-level of urban population). In addition, the agents' actions also influence certain features of the urban landscape, like land value and residential offers.

MASUS CONCEPTUAL MODEL
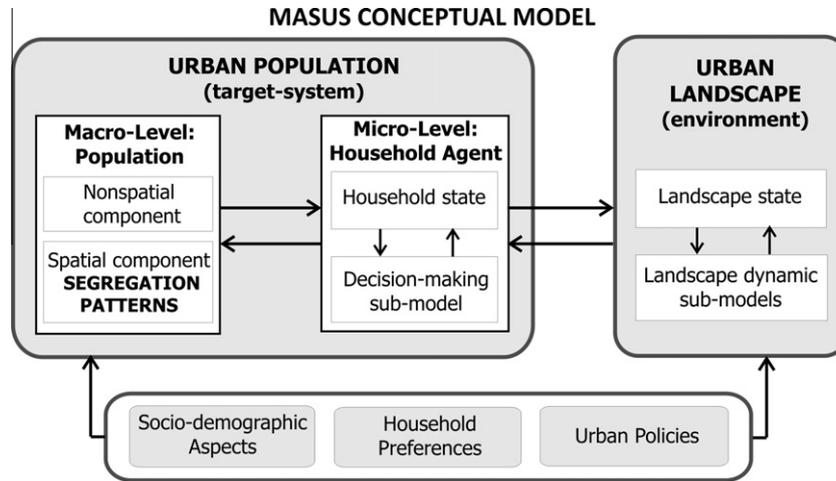
Fig. 1. The conceptual MASUS framework.

3. Model specification

Based on the conceptual model components, we have specified three modules for the operation of MASUS: URBAN-POPULATION, URBAN-LANDSCAPE, and EXPERIMENTAL-FACTOR.

3.1. URBAN-POPULATION module

The URBAN-POPULATION module is organized into three inter-related levels (Fig. 2): household agent (*HouseholdAgent*), household social group (*HouseholdGroup*), and population (*Population*).

The *HouseholdAgent* is the minimal unit of the urban population system. It is comprised of:

- the agent profile, which includes variables that are relevant to the agent's choice of location;

- the household transition sub-model (H-TRANSITION), which is an internal model of the household agent that simulates some natural dynamics of its profile;
- the agent's perception about residential locations in the city, including its own; and
- the decision-making sub-model (DECISION), which rules the agent's location choice.

The *HouseholdGroup* provides group-specific functions and parameters that shape the agent's perception of residential locations and, consequently, the agent's decision-making sub-model.

The *Population* is the collection of all household agents. It is comprised of:

- the socio-demographic state of the population (P_{soc}), including its size and other statistics;
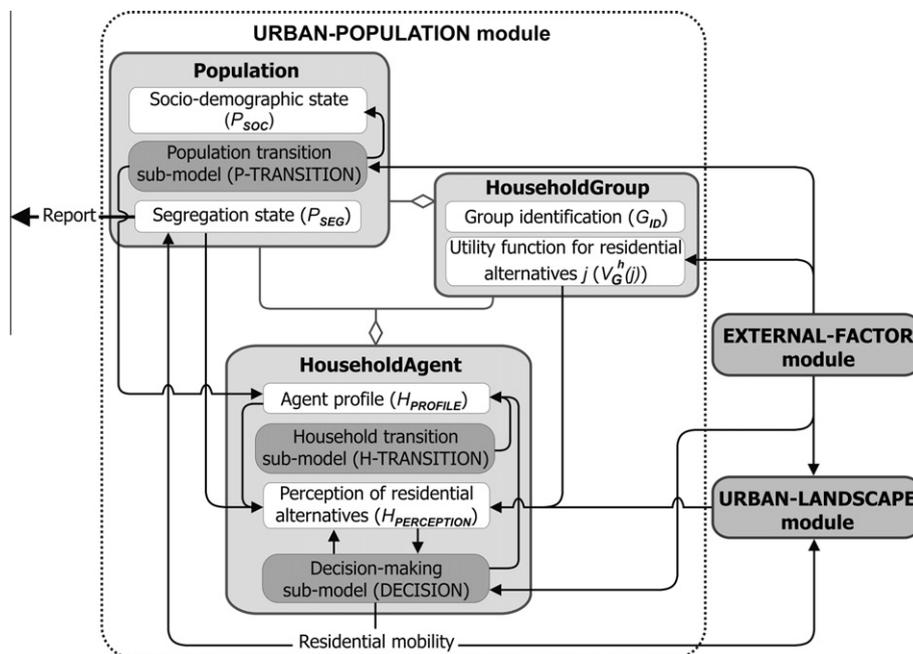
Fig. 2. Architecture of the URBAN-POPULATION module.

- the population transition sub-model (P-TRANSITION), which keeps the socio-demographic state of the population according to expected levels provided by the modeler; and
- the segregation state of the population (P_{seg}).

The dynamics of the socio-demographic state of the population (P_{soc}) are driven by sub-models belonging to the household agent level (H-TRANSITION) and the population level (P-TRANSITION). The H-TRANSITION sub-model is composed of a set of rule-based functions that perform the dynamics of some household characteristics, including income, age of the household head, and the presence of children. This sub-model also includes a function that, based on probabilities that differ according to the household profile, can 'dissolve' the household agent. This function can represent, for instance, households migrating to another city.

The transition rules executed by the H-TRANSITION are complemented by the P-TRANSITION sub-model. Because demographic prediction is not among the purposes of MASUS, the P-TRANSITION is responsible for keeping the growth and socio-demographic composition of the population according to annual control variables defined by the modeler. It creates households with profiles that meet the expected composition of the population as a whole. These new households represent migrants moving into the city as well as households that develop from demographic events that are not simulated by the H-TRANSITION sub-model, such as children leaving the home, divorce, and death of the head of the household.

The segregation state of the population (P_{seg}) is depicted by spatial indices of segregation as described by Feitosa, Câmara, Monteiro, Koschitzki, and Silva (2007). Global and local versions of the *generalized spatial dissimilarity index* ($\breve{D}$ and $\breve{d}_j$) and the *spatial isolation index* ($\breve{Q}_m$ and $\breve{q}_{m,j}$) were adopted. While global indices summarize the segregation degree of the whole city, the local indices show segregation as a spatially variant phenomenon that can be displayed in maps (Feitosa et al., 2007).

The global version of the generalized spatial dissimilarity index ($\breve{D}$) measures how the population composition of each neighborhood differs, on average, from the population composition of the whole city. The index $\breve{D}$ varies from 0 to 1 (maximum segregation) and its formula is:

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

where

$$I = \sum_{m=1}^{M} (\tau_m)(1 - \tau_m) \tag{2}$$

and

$$\breve{\tau}_{jm} = \frac{\breve{L}_{jm}}{\breve{L}_j}. \tag{3}$$

In Eqs. (1) and (2), 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 geographically-weighted proportion of group m in neighborhood j; J is the total number of areal units in the study area; and M is the total number of population groups. In Eq. (3), $\breve{L}_{jm}$ is the geographically-weighted average of the population belonging to group m in neighborhood j; and $\breve{L}_j$ is the geographically-weighted average of the population in neighborhood j. The geographically-weighted average of population data is computed using a kernel estimator, which is placed on the centroid of areal unit j. The weights are given by the choice of a distance decay function and a bandwidth parameter.

The local version of the generalized spatial dissimilarity index $\breve{d}_j$ shows how much each neighborhood contributes to the global $\breve{D}$ measure of the city. The index $\breve{d}_j$ is obtained by decomposing the global index $\breve{D}$.

The global version of the spatial isolation index of group m ($\breve{Q}_m$) measures the average proportion of group m in the neighborhood of each member of this group. The formula of the index $\breve{Q}_m$ is:

$$\breve{Q}_m = \sum_{j=1}^{J} \frac{N_{jm}}{N_m} \left(\frac{\breve{L}_{jm}}{\breve{L}_j} \right), \tag{4}$$

In Eq. (4), N_{jm} is the population of group m in areal unit j; N_m is the population of group m in the study region; and the other equation parameters are as in Eq. (3). The index ranges from 0 to 1 (maximum isolation) and its values depend on the overall population composition of the city. For example, if there is an increase in proportion of group m in the city, the value of $\breve{Q}_m$ tends to become higher. The index $\breve{Q}_m$ also presents a local version ($\breve{q}_{m,j}$), which is obtained by decomposing the global index.

The dynamics of segregation are driven by the DECISION sub-model, an internal mechanism of the *HouseholdAgent* that guides the agent's decisions and actions regarding its residential location. According to this sub-model, the household agent chooses among the following alternatives:

- stay in the current location;
- move within the same neighborhood;
- move to the same type of neighborhood, e.g., from an irregular settlement to another settlement (n locations are randomly selected); and
- move to a different type of neighborhood, e.g., from a socially diverse neighborhood to a gated and segregated settlement (m locations are randomly selected).

While selecting residential alternatives, the household agent randomly chooses locations from a valid set that excludes places without available dwellings. Because the model assumes that agents can evaluate the possibility of living in any neighborhood in the city, the selection imposes no restrictions regarding the distance between the alternative and the household's current location. This modeling decision takes into consideration that real households can acquire knowledge about several neighborhoods, including some in more distant locations, through their social contacts or other sources (e.g., newspapers).

The residential locations are landscape patches that occupy a space $100\,\text{m} \times 100\,\text{m}$, and neighborhoods are sets of landscape patches that correspond to census tracts. We classify neighborhoods in four different types: the first has a high concentration of affluent households, the second is socially diverse, the third has a high concentration of low-income households, and the fourth is similar to the third, but includes clandestine settlements (e.g., slums).

After selecting a set of residential alternatives, the second step in the DECISION sub-model is to calculate the agent's perception ($H_{perception}$), that is, to obtain the agent's utility for each selected alternative k ($V^h(k)$). The function $V^h(k)$ is a nested logit utility function that considers the agent's profile ($H_{profile}$), the environmental attributes (L_{state}), and the neighborhood population composition (P_{seg}) of alternative k. The utility function and its reference parameters are obtained from the estimation of a three-level nested logit model (see Fig. 3), which jointly models a household's mobility (first level (i)), the choice of neighborhood type (second level (j)), and the choice of neighborhood location (third level (k)). The utility function and its parameters are group-specific and, therefore, provided by the level of the household's social group (*HouseholdGroup*).

STRUCTURE OF THE NESTED LOGIT MODEL
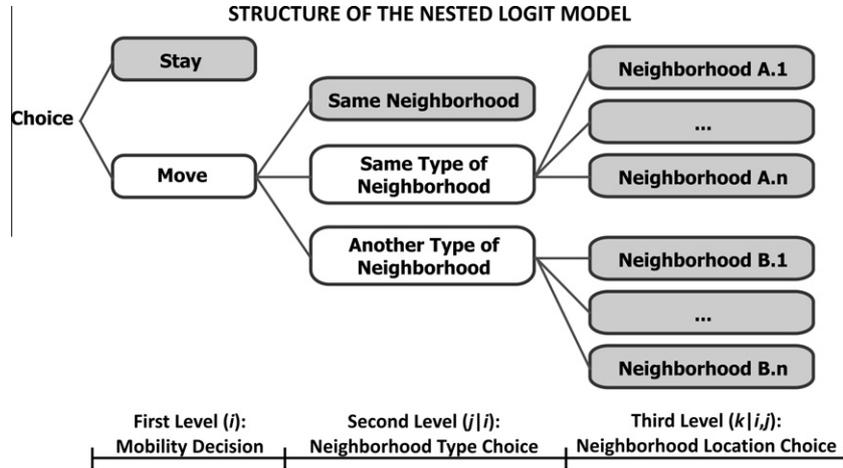
Fig. 3. Nested logit framework for the DECISION sub-model.

After obtaining $H_{perception}$, it is possible to calculate the agent's probability of choosing each alternative based on the nested logit probability equation. Considering that $X_{k|i,j}$, $Y_{j|i}$ and Z_i refer to the vectors of explanatory variables specific to 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):

$$\Pr(k) = \Pr(k|i,j) \cdot \Pr(j|i) \cdot \Pr(i) \tag{5}$$

The conditional probability $\Pr(k|i,j)$ and $\Pr(j|i)$ in Eq. (5) are the functions of the forms:

$$\Pr(k|i,j) = \frac{\exp\left(\frac{1}{\tau_{(j|i)}}(\beta'^{X_{(k|i,j)}})\right)}{\sum_n \exp\left(\frac{1}{\tau_{(m|i)}}(\beta'^{X_{(n|i,j)}})\right)} \tag{6}$$

and

$$\Pr(j|i) = \frac{\exp\left(\frac{1}{\tau_i}(\alpha'Y_{(j|i)} + \tau_{(j|i)I_{(j|i)}})\right)}{\sum_m \exp\left(\frac{1}{\tau_i}(\alpha'Y_{(m|i)} + \tau_{(m|i)I_{(m|i)}})\right)} \tag{7}$$

where $I_{j|i}$ is the inclusive value for category $(j|i)$ and $\tau_{j|i}$ is the dissimilarity 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 $\Pr(k|i,j)$:

$$I_{j|i} = \ln\left(\sum_n \exp\left(\frac{1}{\tau_{m|i}}(\beta'X_{n|ij})\right)\right) \tag{8}$$

The dissimilarity parameter $\tau_{i|j}$ provides a summary measure of the degree of correlation among alternatives in the nest $(j|i)$. If $k, l \in (j|i)$, we have:

$$\tau_{j|i} = \sqrt{1 - corr(U_k, U_l)} \tag{9}$$

where U_k and U_l are the utility functions of alternatives k and l, respectively.

The term $\tau_{j|i}I_{j|i}$ in Eq. (7) represents the expected utility that the decision maker receives from the choice among the alternatives in nest $(j|i)$.

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

$$\Pr(i) = \frac{\exp(\gamma'Z_i) + \tau_i I_i}{\sum_i \exp(\gamma'Z_i + \tau_i I_i)'} \tag{10}$$

where

$$I_i = \ln\left(\sum_m \exp\left(\frac{1}{\tau_i}\left(\alpha'Y_{(m|i)} + \tau_{(m|i)I_{(m|i)}}\right)\right)\right) \tag{11}$$

After computing the agent's probability of choosing the alternatives, the DECISION sub-model executes a Monte Carlo simulation to select one of the alternatives. The household agent then performs the action that matches the selected alternative. 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 (its location and, possibly, its tenure status) and the urban landscape state of its previous and new residential location (the number of dwelling offers in the corresponding cells).

3.2. URBAN-LANDSCAPE module

The URBAN-LANDSCAPE module is organized into two levels (Fig. 4): the entire landscape (*EntireLandscape*) and the landscape patch (*LandscapePatch*). The *EntireLandscape* plays a limited role in the model because the agent's decisions do not consider this landscape level. This happens because agents can only access
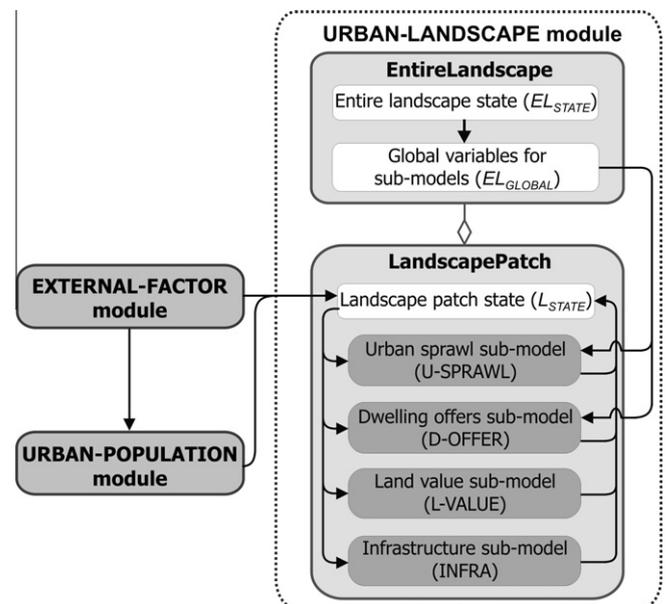
Fig. 4. Architecture of the URBAN-LANDSCAPE module.

information about some landscape portions. Nevertheless, the state of the entire landscape (EL_{state}) is relevant for simulating the dynamic of some the landscape patches' attributes.

The *LandscapePatch* is the portion of the environment that corresponds to the minimal unit of the urban landscape system. It comprises: (i) the landscape patch state (L_{state}), which includes environmental variables that are relevant, directly or indirectly, to the locational behavior of households; (ii) the urban sprawl sub-model (U-SPRAWL); (iii) the dwelling offers sub-model (D-OFFER); (iv) the land value sub-model (L-VALUE); and (v) the infrastructure sub-model (INFRA).

The *urban sprawl sub-model* (U-SPRAWL) simulates the expansion of the city's urbanized areas. This sub-model adapts the methodological approach presented by Antoni (2001). It has two phases: the first is the transition phase ("how much?"), which quantifies the sprawl; the second is the allocation phase ("where?"), which identifies the locations of the new urban patches. For the transition phase, the Markov chain is employed to estimate a global transition probability, which is used to compute how many patches convert their use from non-urban to urban. The allocation phase relies on a binary logistic regression to estimate the probability of a non-urban patch becoming urban. Based on these probabilities, the sub-model allocates the new urban patches. These urban patches represent new alternatives that can be considered by the household agents during their decision-making process. The concrete variables and functional specification of the urban sprawl sub-model are shown in Feitosa (2010: pp. 117–123).

The *dwelling offers sub-model* (D-OFFER) also has transition and allocation phases. The transition phase updates the total number of dwellings, which is the number of occupied dwellings plus a proportion θ_{stock} of this number (the housing stock). The allocation phase allocates the total number of dwellings based on two linear regression models: one model estimates the patches' loss of dwellings due to the expansion of nonresidential uses (e.g., expansion of commercial use in residential areas); the other model estimates the patches' increase of dwellings. The *land value sub-model* (L-VALUE) is based on a hedonic price model to estimate the cell's land value, while the *infrastructure sub-model* (INFRA) relies on linear regression models to estimate the infrastructure quality of each cell. A detailed specification of these sub-models can be found in Feitosa (2010: pp. 124–134).

3.3. EXPERIMENTAL-FACTOR module

The EXPERIMENTAL-FACTOR module consists of specification templates that can be set to explore the impacts of certain contextual mechanisms on segregation patterns. These contextual mechanisms include personal preferences, income inequality levels, social-mix policies, and investments in infrastructure or the regularization of settlements. The specification templates implemented in the current version of MASUS can affect the system's behavior through four pathways:

- Changing global variables of the P-TRANSITION sub-model that affect the social composition of the population. For example, it is possible to test theories about the relationship between segregation and income inequality by simulating alternative scenarios in which the income composition of the population reveals different levels of inequality (see Section 5.2).
- Changing parameters of the nested logit functions that drive the behavior of the household agents (DECISION sub-model). For example, it is possible to test whether the preferences of affluent households regarding the social composition of their neighborhoods influence segregation patterns.
- Changing the structure of the DECISION sub-model. For example, instead of adopting the structure presented in Section 3.1

for all household agents, it is possible to simulate policies of poverty dispersion in which poor households receive housing vouchers to move out of distressed neighborhoods (see Section 5.3).
- Changing the state of the environment (URBAN-LANDSCAPE). For example, it is possible to explore how investments that regularize all the clandestine settlements of the city and provide an equal distribution of infrastructure could change the segregation patterns of the city.

4. Model implementation

The MASUS model was implemented for São José dos Campos, a Brazilian city located in the State of São Paulo with an estimated population of 609,229 (IBGE, 2008). To estimate the parameters of the URBAN-POPULATION sub-models, we used the following data: a household survey of 7910 respondents (universe of 141,814 households) conducted in 2003 (NEPO, 2003) and census data from 1991 and 2000. To estimate the parameters of the URBAN-LANDSCAPE sub-models, we used environment data from 1991 and 2000. These data were obtained from different sources, including satellite images, census data, and official maps.

The parameters for the decision-making sub-model, which is the most important sub-model, were estimated using a nested logit approach. These parameters indicate the relevance of household and neighborhood variables to an agent's residential choice. The household variables include tenure status, age, income, education, size, and presence of children. The neighborhood variables include land value, dwelling offers, infrastructure, type of settlement, distance to the CBD, distance to the original neighborhood, and the proportion of residents belonging to the same social group as the household evaluating the alternatives. Full empirical calibrations of the MASUS parameters are given by Feitosa (2010: pp. 94–116).

The operational MASUS model for the city of São José dos Campos was implemented in NetLogo 4.0.4 (Wilensky, 1999). Fig. 5 presents the simulation protocol performed by the MASUS program, which includes:

(a) Set up the initial state of the system.
(b) Start the main time-loop (annual cycle):
  (i) Execute the decision-making sub-model (DECISION) for all households.
  (ii) Calculate segregation indices and other population statistics.
  (iii) Report simulated outputs (statistics, maps, and graphs).
  (iv) Update population and landscape states for the next cycle.
  (v) Update year ($^{t+1}year = {}^{t}year + 1$) and repeat the annual cycle.

5. Simulation experiments

This section presents three simulation experiments: the first validates the MASUS model regarding the fit between the simulated and real data; the second tests a theoretical aspect of segregation; and the third explores the effects of an anti-segregation policy.

5.1. Experiment 1: comparing simulation outputs with empirical data

The first experiment simulates 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. The data were obtained from the Brazilian census, LANDSAT satellite images, property advertisements in local newspapers, and additional maps provided by the local government. The initial population represents the full population of the city (106,591

MASUS PROCESS SCHEDULE
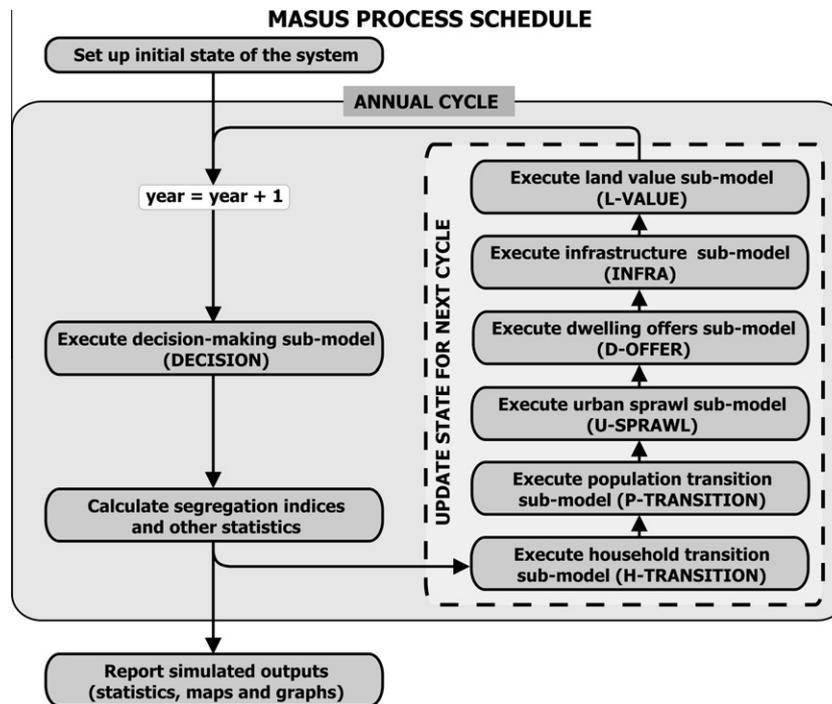
Fig. 5. Flow chart showing the main steps of the MASUS simulation process.

households in 1991), which was directly obtained from the Brazilian census. The census provides the universal microdata for this particular year.

After setting the initial state, nine annual cycles were executed, and the result was compared with real data from the year 2000. A calibration consisting of small changes in the parameters of the decision-making sub-model, originally obtained from the estimation of the nested logit models, improved the fit between the simulated and real data. Fig. 6 shows the results of the segregation indices computed for the initial state (1991), the simulated data (2000), and the real data (2000). The local segregation indices are displayed as maps, with darker colors representing higher levels of segregation. Five replications of the experiment were performed and, despite the stochastic nature of the model, all of the replications produced the same results.

In general, the simulated patterns of segregation demonstrate a good agreement with the observed pattern over time. Both show how the global dissimilarity index increased slightly from 1991 to 2000 (from 0.26 to 0.28). The maps of the local version of the index (Fig. 6b and c) suggest that this increase occurred especially in areas close to the center, toward the western region, and in the southern region. The isolation maps complement this information by showing that the higher dissimilarity in the central areas is caused by the isolation of affluent households (Fig. 6h and i), and the higher dissimilarity in the south is due to the isolation of poor households (Fig. 6e and f).

An interpretation of the global indices of isolation demands caution because the proportions of the city's social groups influence their values. During the period 1991–2000, the proportion of low-income households (up to 4 minimum wages) decreased from 0.54 to 0.51, and their spatial isolation decreased from 0.6 to 0.58 (both in the simulated and real data). These changes mean that, on average, 58% of the neighbors of a low-income household belonged to the same income group. This value is higher than the overall percentage of this group in the city (51%). The isolation indices also vary according to the neighborhood concept used for the measurement: in these experiments, a household's neighborhood comprises the area within a 700-m radius of its residence.

According to the simulated maps (Fig. 6d–f), the decrease in poverty isolation occurred mainly in areas close to the center, thus retaining a high (or higher) isolation in the outer fringes of the city. This spatial configuration increases the distance between social groups, making the city's underdevelopment less visible for the upper classes and imposing additional difficulties for the daily lives of low-income families. The poor areas in the city's outskirts, known as *periferias*, are typically clandestine. These areas are created and sold by private developers who conduct 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 has become the only alternative for many poor families. These families are excluded from the advantages of living in neighborhoods with basic infrastructure, facilities, and urban services (Torres et al., 2003). Their access to jobs, in particular, is limited because workers usually face long commutes to and from work (Caldeira, 2000).

For the isolation of high-income households, the global indices calculated for the real and simulated data presented the same trend, but with different values. The proportion of high-income households (more than 10 minimum wages) increased from 0.15 to 0.19. The isolation of these households increased from 0.33 to 0.38 according to the real data and to 0.36 according to the simulated data. It is interesting that the difference between the group proportion in the city and the isolation index is much higher for affluent households (0.19 vs. 0.38/0.36) than low-income households (0.51 vs. 0.58). This difference suggests that affluent households have a greater inclination to live in isolation from other social groups.

The local isolation maps computed for the simulated and real data (Fig. 6g–i) show that the isolation of high-income households increased in areas close to the center toward the western region. This segregation pattern is commonly observed in Brazilian cities, where the self-segregation of middle and high classes has increased. The pattern usually follows a certain direction of territorial expansion starting from the city's center (Villaça, 2001). This trend resembles the classical sector model proposed by Hoyt
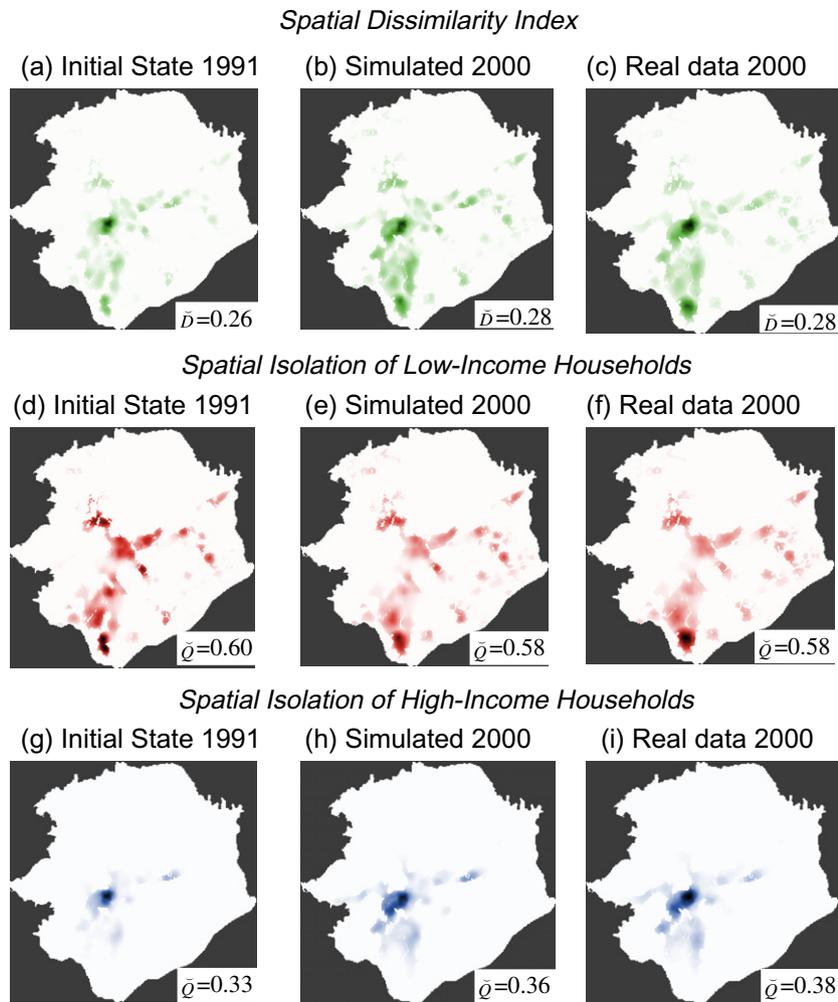
Fig. 6. Experiment 1: comparison of simulation outputs with empirical data.

(1939) because it creates a cone-shaped wealthy axis that concentrates most high-income families. For the residents of this area, the need to circulate through other parts of the city and the possibilities of confronting other realities are reduced.

5.2. Experiment 2: testing theoretical aspects of segregation

The second experiment explores the impact of different levels of income inequality, seen as a product of the labor market, on segregation patterns. In the United States, several theoretical and empirical studies advocate that income inequality promotes urban segregation (Mayer, 2001; Reardon & Bischoff, 2008; Watson, 2006). In Latin America, however, this issue has caused a controversy: while the causal relationship between inequality and segregation underlies the discourse of many researchers (Kowarick, 1979; Lago, 2000; Maricato, 1979), others advocate that this causal relationship is not necessarily true. Sabatini (2004) has criticized the argument that inequality is reflected in urban segregation, which he labeled the "mirror effect hypothesis". According to the author, the increased income inequality promoted by economic crises may also promote a backward progression in segregation. As an example, he mentioned the dispersion of elites that has occurred in many Latin American cities since the 1980s.

The aim of this experiment is to provide further insight into this debate. We compare the simulation run described in the previous section with two alternative scenarios: one where inequality increases along the simulation and another where inequality decreases. All other model specifications were kept constant. To change the inequality levels of the MASUS simulations, the user can choose templates with different settings for the global variables that control the income composition of the population (population transition sub-model).

Fig. 7 presents graphs showing the results obtained from the three simulation runs. The inequality levels and dissimilarity indices from 1991 to 2000 (Fig. 7a and b) reinforce the idea that inequality promotes segregation: once inequality increases, the dissimilarity between the income composition of the whole city and the income composition of neighborhoods also increases, and vice versa. The isolation of low-income households also varies proportionally to the inequality levels (Fig. 7c and d). It is important to mention, however, that isolation indices are sensitive to the proportion of social groups in the city. For example, once the proportion of group m increases in the city, the isolation index of group m also tends to become higher. Therefore, the increase in low-income isolation was expected in the higher-inequality case because the proportion of low-income households in the city also increased.

This expected index trend is, however, challenged by the graphs showing the isolation of high-income households and the proportion of this group in the city (Fig. 7e and f). For this reason, the results presented in these graphs are the most revealing. The low-inequality scenario presents higher proportions of affluent
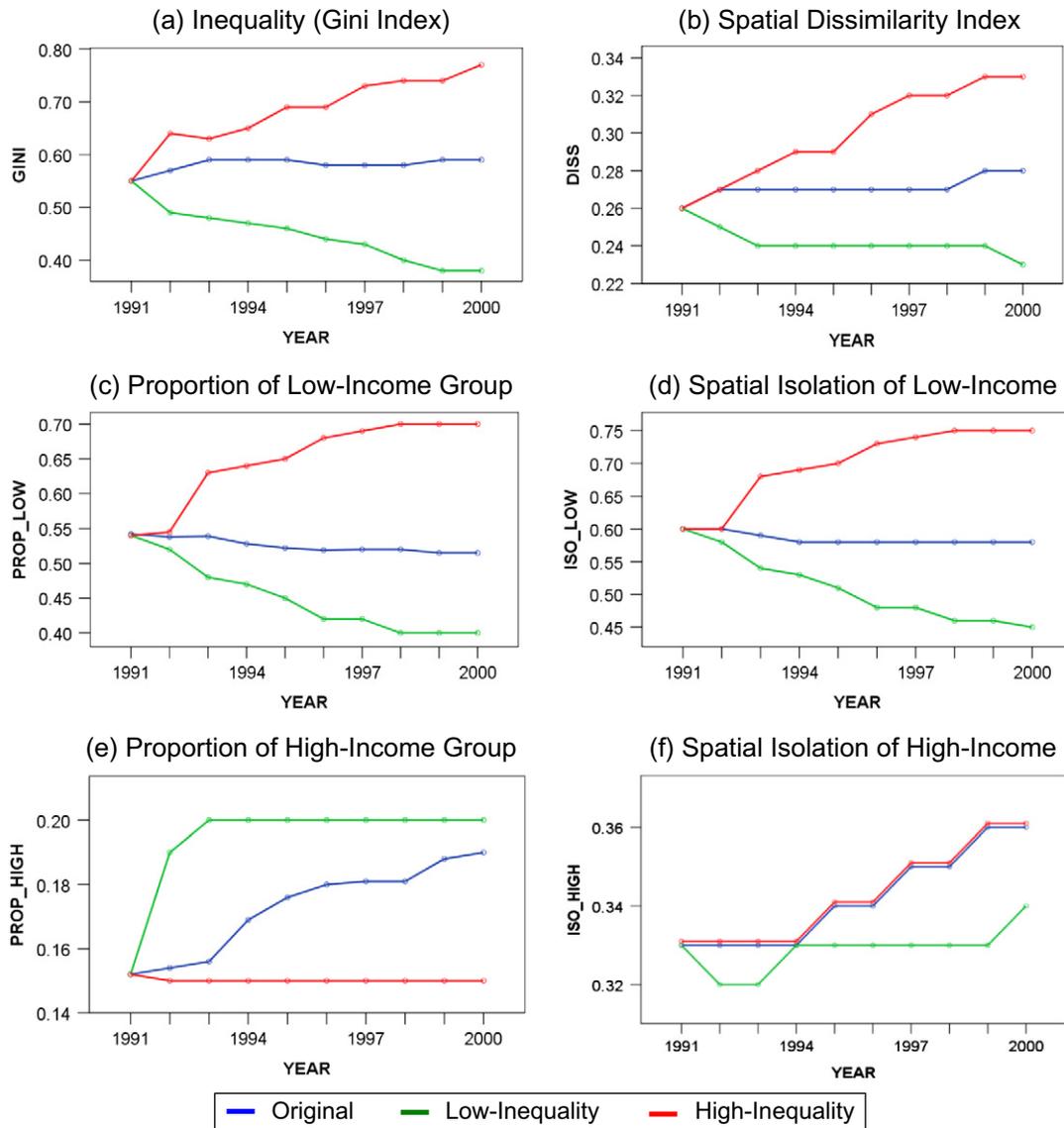
Fig. 7. Experiment 2: the relation between segregation and income inequality.

households compared to the other scenarios, but it still displays the lowest levels of isolation. This unexpected result provides an additional indication of the causal relation between income inequality and segregation, suggesting that decreasing levels of income inequality promote spatial integration among different social groups.

5.3. Experiment 3: testing an anti-segregation policy

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 strategies to promote social mixing, including the dispersal of poverty, regeneration of troubled neighborhoods, and regulations for new developments. 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 and into

middle-class neighborhoods. To do this, low-income families receive housing vouchers that are used to rent private dwellings in neighborhoods with a low poverty rate.

To test the effects of a social-mix policy based on the distribution of housing vouchers, we compare the simulation run described in Experiment 1 (Section 5.1), which replicates the original segregation dynamics of São José dos Campos from 1991 to 2000, with two alternative scenarios. These scenarios simulate the implementation of a housing program that distributes *n* housing vouchers for poor households in 1991, and they increase the number of benefits each year. The first alternative scenario distributes vouchers to 0.3% of the poor households in 1991 (200 vouchers) and progressively expands this percentage until 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 1991 (500 vouchers), and it gradually expands the program to 5.8% of the poor households in 2000 (4200 vouchers). The housing vouchers are distributed to poor households that are randomly selected from neighborhoods with a high isolation of poverty $(\tilde{Q}_{poor} > mean(\tilde{Q}_{poor}) + sd(\tilde{Q}_{poor}))$. The
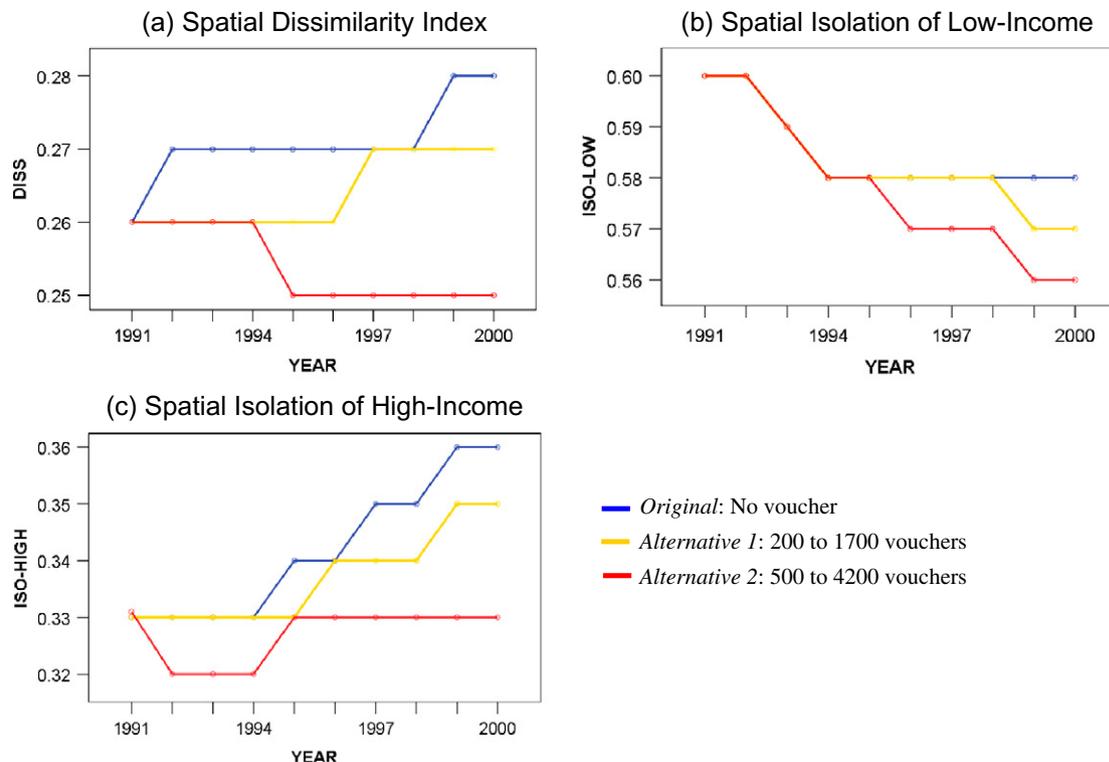
Fig. 8. Experiment 3: the impacts of a social-mix policy based on poverty dispersion.

vouchers are used to rent dwellings in neighborhoods where the isolation of low-income households is below the average ($Q_{poor} < mean(Q_{poor})$).

Fig. 8 shows the evolution of the global dissimilarity index and the isolation indices during the period 1991–2000 for three scenarios: original (no housing voucher), alternative 1 (200–1700 vouchers), and alternative 2 (500–4200 vouchers). The dissimilarity index in 2000, which in the original scenario was 0.28, changes to 0.27 in alternative scenario 2. This means that the distribution of housing vouchers to 2.3% of the poor households in the city causes a decrease of 3.5% in the dissimilarity index. In alternative scenario 3, the distribution of vouchers to 5.8% of the poor households decreases the dissimilarity index by 10.7% (from 0.28 to 0.25).

The spatial isolation index of high-income households also decreases significantly as the investment in the housing program increases. Comparing scenario 1 (baseline) with scenario 2, the distribution of housing vouchers to 2.3% of the poor households decreased the isolation of high-income households by 5.7% (from 0.36 to 0.35). In scenario 3, 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 0.36 to 0.33).

Despite these positive trends, the housing program could not promote a substantial improvement in the overall isolation level of low-income households, which is the segregation dimension that has the most harmful impacts on the lives of disadvantaged families. Comparing scenarios 1 and 2, the distribution of housing vouchers to 2.3% of the poor households decreased the isolation of low-income households by only 1.7% (from 0.58 to 0.57). Comparing scenarios 1 and 3, the distribution of vouchers to 5.8% of the poor households caused a decrease of 3.4% in the isolation of low-income households (from 0.58 to 0.56). 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 only decreased to 56% in alternative scenario 2, where 4200 housing vouchers were distributed. These numbers demonstrate the

limitations of this type of housing policy in cities where poor households represent a large share of the population. In addition, the policy was not able to promote a change in the spatial arrangement of social groups, keeping a configuration in which the city decays, socially and physically, toward its outskirts, except in the 'wealthy cone' area.

To produce a substantial change in the isolation level of poor families in São José dos Campos, 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). Further experiments simulating the dispersion of affluent families could provide additional insights into the implementation of social-mix policies. Examples of experiments able to contribute in this direction include the simulation of policies that stimulate the construction of developments for middle and upper classes in poor neighborhoods. Such policies can take advantage of tax exemption measures, concessions, changes in the norms of land use, and public investments in infrastructure and security (Sabatini, 2006).

6. Conclusions and recommendations

This paper has introduced MASUS, a multi-agent simulator for urban segregation. MASUS is a scientific tool that allows researchers to explore the impacts of different contextual mechanisms on the configuration of segregation patterns. After presenting the conceptual model and specifications of MASUS, this paper demonstrated the potentiality of the model through three types of experiments using data from São José dos Campos (SP, Brazil). These experiments demonstrate how the MASUS model can serve as a virtual laboratory that contributes to scientific and policy debates on segregation.

Nevertheless, as with any other tool that simulates a complex system, MASUS outputs must be interpreted with caution. These neither represent quantitative and accurate forecasting of segregation patterns nor provide a deterministic answer regarding the best policy approach. Instead, these simulation outputs should be considered in terms of how 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 of the factors that are relevant for the residential location choices of households. Only after such observations and deliberations is it possible to obtain insights that contribute toward structuring debates on open theoretical questions about segregation or the development of better informed anti-segregation policies.

Several considerations can be drawn regarding further research directions. In the current version of MASUS, the modeled decision-making process relies on nested logit models (NMNL) with three levels. NMNL is a joint modeling approach that has the advantage of assuming, for instance, that a household's mobility decision (move or stay) is influenced by the characteristics of the residential alternatives available in the market. However, an important drawback is that these statistical models essentially provide a statistical representation of the agent's reasoning. Therefore, the model outcomes are not likely to be robust once the agent's behavior changes. To address this issue, further research should explore the development of adaptive and learning agents.

In addition, the decision-making ssprocess encoded in MASUS does not take into consideration the past decisions of households. More empirical research should be done on the impact of these decisions on the residential choice behavior of households. The residential choice model also ignores the influence of the behaviors of neighbors on the decision process of a household. By considering this spatial component, it may 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 the area. To overcome this drawback, the use of spatially discrete choice models (as reported by Flemming, 2004) that include neighborhood effects could be explored.

Finally, there are 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 improvements in the usability of MASUS and in the design of experiments still depend on feedbacks obtained from potential users and stakeholders. Several techniques based on principles of participatory research can be explored to keep users closely involved in the development of an improved version of MASUS, including techniques such as rapid iterative development and user workshops (Ramanath & Gilbert, 2004).

References

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

Andersson, R. (2008). Neighborhood effects and the welfare state. Towards a European research agenda? *Schmollers Jahrbuch, 128*(1), 49–63.

Antoni, J. (2001). Urban sprawl modelling: A methodological approach. In *Cybergeo – 12th European Colloquium on Quantitative and Theoretical Geography*.

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

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

Batty, M. (2008). The size, scale, and shape of cities. *Science, 319*(5864), 769–771.

Benenson, I. (2004). Agent-based modeling: From individual residential choice to urban residential dynamics. In M. F. Goodchild & D. G. Janelle (Eds.), *Spatially integrated social science: Examples in best practice* (pp. 67–95). Oxford: Oxford University Press.

Benenson, I., Omer, I., & Hatna, E. (2002). Entity-based modeling of urban residential dynamics: the case of Yaffo, Tel Aviv. *Environment and Planning B: Planning and Design, 29*, 491–512.

Benenson, I., Hatna, E., & Or, E. (2009). From Schelling to spatially explicit modeling of urban ethnic and economic residential dynamics. *Sociological Methods & Research, 37*(4), 463–497.

Briggs, X. S. (2005). Social capital and segregation in the United States. In D. Varady (Ed.), *Desegregating the city*. Albany: Suny Press.

Bruch, E.E. (2006). Residential mobility, income inequality, and race/ethnic segregation in Los Angeles. *Population association of America (PAA) 2006 annual meeting program*. Los Angeles, CA.

Bruch, E. E., & Mare, R. D. (2006). Neighborhood choice and neighborhood change. *American Journal of Sociology, 112*(3), 667–709.

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

Clark, W. A. V. (1991). Residential preferences and neighborhood racial segregation: A test of the Schelling segregation model. *Demography, 28*(1), 1–19.

Clark, W. A. V., & Fosset, M. (2008). Understanding the social context of the Schelling segregation model. *Proceedings of the National Academy of Sciences of the United States of America (PNAS), 105*(11), 4109–4114.

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

Feitosa, F. F. (2010). *Urban segregation as a complex system: An agent-based simulation approach.. Ecology and development series 70*. Göttingen: Cuvillier Verlag.

Feitosa, F. F., Câmara, G., Monteiro, A. M. V., Koschitzki, T., & Silva, M. P. S. (2007). Global and local spatial indices of urban segregation. *International Journal of Geographical Information Science, 21*(3), 299–323.

Flemming, M. (2004). Techniques for estimating spatially dependent discrete choice models. In L. Anselin, J. G. M. Florax, & S. J. Rey (Eds.), *Advances in spatial econometrics* (pp. 145–168). Amsterdam: Springer.

Fossett, M., & Senft, R. (2004). *SIMSEG and generative models: A typology of model-generated segregation patterns*. Paper presented at the agent 2004 conference on social dynamics: Interaction, reflexivity and emergence, Chicago, IL.

Fossett, M., & Waren, W. (2005). Overlooked implications of ethnic preferences for residential segregation in agent-based models. *Urban Studies, 42*(11), 1893–1917.

Gerometta, J., Hausermann, H. H., & Longo, G. (2005). Social innovation and civil society in urban governance: Strategies for an inclusive city. *Urban Studies, 42*(11), 2007–2021.

Gilbert, N., & Troitzsch, K. G. (1999). *Simulation for the social scientist*. Buckingham: Open University Press.

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

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

Holland, J. H. (1998). *Emergence: From chaos to order*. Oxford: Oxford University Press.

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

IBGE (2008). *Estimativa das populações residentes segundo os municípios*. Rio de Janeiro: Brazilian Institute of Geography and Statistics – IBGE.

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

Katzman, R., & Retamoso, A. (2006). *Segregación residencial en Montevideo: Desafíos para la equidad educativa*. Paper presented at the Reunión de Expertos sobre Población y Pobreza en América Latina y el Caribe, Santiago, Chile.

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

Lago, L. C. (2000). *Desigualdades e segregação na metrópole: O Rio de Janeiro em tempos de crise*. Rio de Janeiro: Revan/Fase.

Laurie, A. J., & Jaggi, N. K. (2003). Role of 'vision' in neighborhood racial segregation: A variante of the Schelling segregation model. *Urban Studies, 40*(13), 2687–2704.

Maricato, E. (Ed.). (1979). *A produção capitalista da casa (e da cidade) no Brasil industrial*. São Paulo: Alfa-Ômega.

Mayer, S. E. (2001). *How the growth in income inequality increased economic segregation*. Chicago: The Harris School, University of Chicago.

Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems: An introduction to computational models of social life*. Princeton: Princeton University Press.

NEPO & PMSJC (2003). *Survey for urban planning instrumentation and evaluation of the housing deficit in São José dos Campos*. São José dos Campos.

O'Sullivan, D., Macgill, J. R., & Yu, C. (2003). *Agent-based residential segregation: A hierarchically structured spatial model*. Paper presented at the Agent 2003 'Challenges in Social Simulation', University of Chicago and Argonne National Laboratory, IL, USA (October 3–4).

Pancs, R., & Vriend, N. J. (2003). *Schelling's spatial proximity model of segregation revisited*. London: Department of Economics, Queen Mary University of London.

Préteceille, E. (2003). A evolução da segregação social e das desigualdades urbanas: o caso da metrópole parisiense nas últimas décadas. *Caderno CRH, 38*.

Ramanath, A. M., & Gilbert, N. (2004). The design of participatory agent-based social simulations. *Journal of Artificial Societies and Social Simulation, 7*.

Reardon, S. F., & Bischoff, K. (2008). *Does income inequality cause income segregation? Evidence from a within-race analysis*. Paper presented at the annual meeting of the American sociological association annual meeting, Boston, MA.

Rolnik, R. (1997). *A cidade e a lei: legislação, política urbana e territórios na cidade de São Paulo*. São Paulo: Studio Nobel/Fapesp.

Sabatini, F. (2004). La segregación residencial en las ciudades Latinoamericanas: Causas, posibles políticas y rol de los mercados de suelo. In F. Arenas, J. L. Coll, & R. Hidalgo (Eds.), *Los nuevos modos de gestión de la metropolización*. Santiago, Chile: Pontificia Universidad Católica de 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., & Cerda, J. (2001). Residential segregation in main Chilean cities: Tendencies from the past three decades and possible frameworks for action. *Eure-Revista Latinoamericana De Estudios Urbano Regionales, 27*(82), 21–42.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1*(1), 143–186.

Schelling, T. C. (1972). The process of residential segregation: Neighborhood tipping. In A. Pascal (Ed.), *Racial discrimination in economic life* (pp. 157–185). Lexington, MA: D.C. Heath.

Smith, A. (2002). *Mixed-income housing developments: Promise and reality*. Harvard: NeighborWorks and Joint Center for Housing Studies, Harvard University.

Torrens, P. M., & Benenson, I. (2005). Geographic automata systems. *International Journal of Geographical Information Science, 19*(4), 385–412.

Torres, H. G., Marques, E. C., Ferreira, M. P., & Bitar, S. (2003). Pobreza e espaço: padrões de segregação em São Paulo. *Revista do Instituto de Estudos Avançados, IEA, 17*(47), 13–42.

Torres, H. G. (2004). Segregação residencial e políticas públicas: São Paulo na década de 1990. *Revista Brasileira de Ciências Sociais, 19*(54), 41–56.

UN-Habitat (2001a). *Inclusive cities initiative: The way forward*. Nairobi: UN-Habitat.

UN-Habitat (2001b). *The state of the world's cities report*. Nairobi: UN-Habitat.

UNFPA (2007). *State of world population 2007: Unleashing the potential of urban growth*. New York: United Nations Population Fund – UNFPA.

Villaça, F. (2001). *Segregation in the Brazilian metropolis*. Paper presented at the International seminar on segregation in the city, Lincoln Institute of Land Policy, Cambridge.

Watson, T. (2006). *Metropolitan growth and neighborhood segregation by income*. Williamstown, MA: Williams College.

Wilensky, U. (1999). *NetLogo*. Evanston, IL: Northwestern University.

Zhang, J. (2004). Residential segregation in an all-integrationist world. *Journal of Economic Behavior and Organization, 54*, 533–550.