Retrieving Individuals' Attributes From Aggregated Dataset For Urban Micro-simulation: A Primary Exploration

LONG Ying
School of Architecture
Tsinghua University
Beijing, China
longying1980@gmail.com

SHEN Zhenjiang
School of Environment Design
Kanazawa University
Kanazawa, Japan

MAO Qizhi
School of Architecture
Tsinghua University
Beijing, China

DANG Anrong
School of Architecture
Tsinghua University
Beijing, China

Abstract—**As the traditional top-bottom based macro-simulation models cannot well adapt to spatiotemporally dynamic urban systems researches, the bottom-up micro-simulation models (e.g., multi-agent systems, MASs) based on individual data have gradually become a novel and dominating perspective for investigating urban dynamics. However, one of the factors restraining the wide application of micro-simulation models is l the poor availability of disaggregated individual data. Such situation is especially serious in China, because the individual data from the official census is not available and can only be obtained via various small-scale surveys conducted by separate units. The contribution of this paper is to propose a solution to retrieve individuals' data from aggregated dataset, e.g. official statistical yearbook, census reports and contingent questionnaire reports, for urban micro-simulation under the current sparse-individual-data context. Based on the existing multi-source official statistical date, non-official surveys, as well as general relationships among individuals' attributes (e.g., the marriage condition of a person depends on his/her age), this approach can inverse both the individuals' attributions and spatial distribution. The approach was tested in the Beijing Metropolitan Area for retrieving the population spatial distribution and their attributes using the Fifth Population Census Report of Beijing, in which the aggregated information of populations is provided. The inversion results were proved to be significantly accurate in statistical level. Using the inversed samples' as the input of the urban micro-simulation models is promising to settle their bottle-neck problems for wide application.**

Key words: micro-simulation; inversion; multi-agent system; statistical data

I. INTRODUCTION

Urban micro-simulation is facing the predicament of individual data scarcity. To tackle this condition, we proposed an approach to inverse the disaggregated samples based on existing aggregated statistical data, questionnaire surveys, as well as the dependence relationships among attributes of samples obtained through data mining or common sense. Urban microsimulation models (UMMs), firstly proposed by Orcutt et al. (1957), can relatively better remedy the defects of macro models in a bottom-top manner. UMMs adopt micro-individual, like firm, family, and even resident, as the basic unit to be described, analyzed, and simulated (Ballas & Clarke, 2000; Hanaoka & Clarke, 2007; Wu et al., 2008). Multi-agent System (MAS, or termed as Agent Based Modeling, ABM) is one of the typical and dominating approaches of UMMs at present (Parker et al., 2003). At present, the sparse individual samples limit the wide application of MAS. The attributes is the core information of individual agents in MAS (Pudney et al., 1994). For instance, the attributes of the resident samples include the age, gender, marriage, education, and career. Obtaining these original attributes data usually takes a lot of time, human, material and financial resources (van Sonsbeek et al., 2006). Nowadays in China, the original disaggregated census hold by the statistical department is not freely open to public, even academic researches. Only the aggregated statistical data can be acquired, such as the overall information of districts or industries, which are appropriate for macro models rather than MAS. Even the Statistical Law of China emended on 2009-6-27 forbids the original census samples from being accessed by public due to privacy problems. In addition, most of related researches of MAS in other countries also cannot get necessary existing individual samples (Crooks et al., 2008). The independent survey exclusively for the research is the key approach for retrieving micro samples by academic or commercial users. Therefore, the situation of data sparse restricts the further development of MAS all around the world.

Under the current sparse data context, the key issue of MAS is to take the heterogeneity attributes or preferences of individual agents into account. Many MAS for investigating land use dynamics cannot well reflect the heterogeneity of sample attributes in actor-level. Those models tend to aggregate all agents into several types due to poor individual samples, like ILUTE (Miller et al, 2004). Another solution is to obtain the acceptable range of attribute through literature

Sponsors: The National Natural Science Foundation of China (No. 50678088) and the National Project of Scientific and Technical Support Programs by the Ministry of Science & Technology of China (No. 2006BAJ14B08).

review, and assign agents' attribute value, like LUCITA model (Deadman et al., 2004). Each agent in most existing MAS can not correspond with an individual resident due to lacking of micro samples. For example, Zhang et al. (2008) used the average number of residents in the grid of 30 m * 30 m as one agent. Li and Liu (2008) regarded one regular grid as one resident agent in their MAS for residential location choice, rather than the actual number of residents within each grid, not to say agents with heterogenous attributes or preferences. Shen et al. (2009) attempted to use, respectively, 1, 2, 3, 5, and 10 residents as an agent in the shopping policy MAS (ShopSim), and find that different proportions led to considerable uncertainty to simulation results. Brown and Robinson (2006) also indicate that in MAS, the heterogeneous preferences' degree of agents affects the simulated land use pattern greatly. Therefore, facing the problem of data scarcity, most of MAS cannot actually regard one agent as one individual. Aggregating several individuals as an agent will result in uncertainty to the simulation outcome, and the excessive aggregation may lose the nature and essence of micro-simulation based MAS.

The individual sample mainly includes its spatial location and attributes. Accordingly, the sample inversion can be classified as spatial inversion and attribute inversion. With regard to the spatial inversion, the researches on mapping population using related factors can be referred, which is well researched (Langford & Unwin, 1994; Mennis, 2003; Liao et al., 2010). In these researches, population density surface is interpolated by spatial factors using the population census statistical data, and population's attributes inversion is not considered in this process. Agents' location can be retrieved using the approach for mapping population, and then the environmental attributes of agents will be identified via overlaying the agents' location with environmental layers, such as accessibilities to education facilities, work place, neighborhood similarity, and landscape quality in Robinson and Brown(2009). The environmental attributes of agents can be applied in the MAS simulation, which is a popular method for urban MASs in current data scarcity context (Crooks, 2006; Crooks, 2008; Li & Liu, 2008; Shen et al., 2009). The socio-economic attributes are lack of consideration in these MASs.

The socio-economic attributes of agents, as descriptive variables for exploring urban dynamics, are critical for urban microsimulation. For instance, the residential location choice preference depends on residents' income, education, and family size attributes. Brown and Robinson (2006) designed 5 experiments for modelling residential location choice, respectively, with different degrees of agents' attributes heterogeneity. They only took agents' preference for various environmental attributes into account, rather than residential agents' socio-economic attributes. Li and Liu (2007) initially pointed out that it was possible to define agents' attributes using aggregated statistical data. They divided all residents into six types according to the statistical data about whether a resident has children (two categories) and the income category (three categories). Each type of residential agents has different preferences on environmental factors. However, they only taken two attributes of samples into account, without providing the specific value of every sample and the relationship between attributes of samples. Hynes et al. (2008) matched field survey data with the national agriculture survey using simulated annealing algorithm, which was different from our research for inversing samples' location and attributes. Therefore, there is no related research focusing on the inversion of attributes of individual samples to our knowledge. In this paper, we aim to inverse samples data with location and attributes using aggregated data under current data sparse environment.

II. APPROACH

A. *Aggregated data as input materials*

Individual samples in MAS are the key input and the primary units for simulations. Existing aggregated data for samples inversion can mainly be divided into three types:

- Official statistical data, as the statistical description of the original samples, include the total number of samples, the distribution of attributes of samples (such as the population of each occupation, income category), and the correlation among attributes of samples (for example, samples number of the marriage status category for each age range, income range for each education category).
- Samples acquired from questionnaire or surveys, which can be used to retrieve the probability density function (PDF) of attributes and the relationships among attributes.
- General relationships among attributes, standing for the dependent relationships among attributes, common sense (such as if the age of a resident is below 18, the marriage status will be unmarried), and mined scientific research achievements.

Therefore, existing aggregated data can be summarized as the PDFs of attributes and relationships among attributes, which will be detailed in the following Section Ⅱ.B and Ⅱ.C.

B. *Variables in the model*

The attributes of samples can be divided into the continuous (age) and categorical (marriage) types. We choose the categorical type for our approach to simplify the model and speed up the inversion computation since any continuous variable can be divided into several categories. For example, the marriage attributes are divided into two discrete values: married, and unmarried, and the age attribute is divided into different age intervals like 0 to 4, 4 to 10, and 10 to 20 in year. Therefore, every optional value of an attribute can be presented as a string (e.g., "married", "0-4"); no matter it originally belongs to continuous or categorical types. The specific value of continuous attribute can be obtained via randomly selecting value from the inversed value interval, such as 3 from "0-4" for the age attribute.

In this paper, the optional values of every attribute are termed as domain, and every value is termed as domain element (a string type). The number of every domain element in the whole samples is termed as frequency number, and its proportion in the whole samples is termed as probability. For

example, the domain of marriage attribute is {married; unmarried; divorced}, in which "married" is a domain element. The corresponding frequency number is {45; 20; 35}, which indicates that 45 samples are married, 20 unmarried, and 35 divorced among 100 persons. In this way, the aggregated input is transformed into the domain, frequency number, probability, and the relationships among domain elements of attributes.

The variables needed in the samples inversion process are shown as follows. *N*: The total number of samples; *M*: The total number of attributes; *i*, $i \in \{1,2,3,...,N\}$: The ID of sample; *j*, $j \in \{1,2,3,...,M\}$: The ID of attribute; T_j : The attribute type ("STR" for the string type and "NUM" for the numerous type); $a_{i,j}$: The inversed value of attribute *j* of sample *i* (the string type); $A_{N*M} = \{a_{i,j}\}$: The inversed attribute values of all samples in a two-dimensional array; $a'_{i,j}$: The observed value of attribute *j* of sample *i* (the integral, decimal or string type); $A'_{N*M} = \{a'_{i,j}\}$: The real attribute values of all samples in a two-dimensional array; $a_{i,}$: All the attribute values of sample *i*; $a_{,j}$: All the sample values of attribute *j*; b_j : The domain of attribute *j* (set); *k*: The ID of a domain element; $b_{j,k}$: The k^{th} domain element of attribute *j*; K_j : The total number of domain elements of attribute *j*; P_j : The frequency number of all the domain elements of attribute *j* (set); $P_{j,k}$: The frequency number of the k^{th} domain element of attribute *j*; $p_{j,k} = P_{j,k} / N$: The probability of the k^{th} domain element of attribute *j*; f_j : The PDF of attribute *j*; h_j : The attribute ID that is related with attribute *j*; H_j : The set of attribute IDs that are related with attribute *j*; g_j : The relationship between attribute *j* and attributes H_j.

C. Distribution analysis

For every attribute, the type of known information is different, neither is the samples inversion method. Therefore, we classify sample attributes into several inversion types according to the known distribution types. Here we take the attribute j of N samples as an example.

- The DA distribution type with known PDF: The known PDF of the attribute j is $p(x = x_0) = f_j(x_0)$ (e.g., Gaussian, uniform), from which K_j number of domain elements are sampled as $P_{j,k} = N * p(x = b_{j,k}) = N * f_j(b_{j,k})$. Then there are $N * f_j(b_{j,k})$ number of samples which equal to $b_{j,k}$, and the DA type is transformed into the DB type as $a_{,j} = randO(\bigcup_{k=1}^{K_j} \{\underbrace{b_{j,k}, b_{j,k}, \cdots, b_{j,k}}_{N*f_j(b_{j,k})}\})$.
- The DB distribution type with known frequency numbers: The inversed values of attribute *j* for all samples can be calculated as $a_{,j} = randO(\bigcup_{k=1}^{K_j} \{\underbrace{b_{j,k}, b_{j,k}, \cdots, b_{j,k}}_{P_{j,k}}\})$, where randO is the function for randomizing order of the inversed values.
- The DC distribution type with no distribution information: The distribution of such attributes is unknown or no provided information.

D. Relationship analysis

Relationships among attributes can be established using existing known information, such as questionnaires, general knowledge, common sense, and statistical reports. Three types of relationships are established as follows.

- The functional relationship RA: The value of one attribute depends on at least one other attributes as $a_{i,j} = g_j(\bigcup_{h \in H_j} a_{i,h})$, where the function g_j can be either linear or nonlinear.
- The probability relationship RB: There is a probability relationship between attribute *j* and attribute h_j (simplified as *h*). Known the frequency numbers of both attributes, the probability relationship can be expressed by the matrix $Q_{K_h * K_j}$. q_{k_h,k_j}, as the basic element of this matrix, can be calculated using $q_{k_h,k_j} = P\{(a_{i,h} = b_{h,k_h}) \bigcap (a_{i,j} = b_{j,k_j})\}$, in which P is the probability ranging from 0 to 1 (note P is not P_j). The probability of each domain element can be calculated for each attribute of every sample using $P\{a_{i,j} = b_{j,k_j}\} = q_{k_h,k_j} * p_{j,k_j}$ $where\ a_{i,h} = b_{h,k_h}$, which takes two probabilities into account. The Monte Carlo method is used to obtain all values of attribute *j* for all samples $a_{,j}$.
- The RC relationship type with no existing or known relationship.

E. Coupling distribution and relationship to inverse samples

Since distribution and relationship types vary from attributes, different methods are developed for inversing various attributes. Totally 9 coupled types are generated according to 3 types of distributions and 3 types of relationships as illustrated in Tab. Ⅰ. VBB and VAB both adopt the results of RB to compare with those of DB, and modify the relationship RB to fit with DB or DA. VCB, VBC, VAC, and VCA directly adopt the result of RB, DB, DA, and RA, respectively. VAA, VBA, and VCC are neglected in this paper because they do not exist or need not be analyzed ("NA" in Tab. Ⅰ). Therefore, 9 coupled types analysis can be all conducted using the previous algorithms for distribution and relationship analysis.

TABLE I. THE COUPLED DISTRIBUTION AND RELATIONSHIP TYPES FOR SAMPLES INVERSION

Distribution/Relationship	RA	RB	RC
DA	VAA (NA)	VAB	VAC
DB	VBA (NA)	VBB	VBC
DC	VCA	VCB	VCC (NA)

If the value of attribute j is the numerous type (NUM), the inversed $a_{,j}$ represents a value range with $a_{,j}^{L}$ as the minimum value and $a_{,j}^{H}$ as the maximum value. Then the final inversed sample value is $a'_{,j} = \begin{cases} a_{,j}, & if\ T_j = "STR" \\ randV(a_{,j}^{L}, a_{,j}^{H}), & if\ T_j = "NUM" \end{cases}$, where randV is the function for randomly (uniform) selecting values within the value range.

F. Mapping inversed samples

The inversed samples are generally stored as attributes data in the tabular form. Mapping the already inversed samples is critical for MAS which is spatial explicit. For this, an attribute FID needed to be added to every sample to denote the unique ID of the sample's corresponding geometry in GIS, such as a point, line or polygon. All samples correspond to an entire space, and the space is formed by a number of geometries. The number of samples in each geometry can be speculated based on statistical data (e.g., the number of residents within a parcel). The attribute FID can meanwhile be inversed using the approach for non-spatial attributes by regarding FIDs of all samples as a probability distribution. On this basis, every geometry generates a point element as the spatial object for each sample randomly, and then the inversed samples are mapped via matching the samples with spatial objects. The inversed samples by our approach are stored as the point feature class in GIS, enabling the results including both the location and attributes information. Based on the location of the point for the corresponding sample, the relevant spatial attributes (such as accessibility and planning conditions) can be identified as the additional attributes for agents in MAS. However, we focus on the inversion approach of individual samples in this paper, and will not consider the relationship between spatial and non-spatial attributes of samples. We will study how to inverse individual samples with spatial attributes in next step researches.

III. A CASE STUDY IN BEIJING

We developed the Agenter (Agent Producer) model based on Geoprocessing of the ESRI ArcGIS family using Python for the proposed samples inversion approach. In this paper we demonstrate the model application for inversing population samples using the Fifth Population Census Report of Beijing in 2000 (Beijing 5th Population Census Office and Beijing Municipal Statistics Bureau, 2002), which covers the whole Beijing Metropolitan Area (BMA). The aggregated statistical data in this census report ranges from age, marriage, job, income, ethnic, education, family, residential area, birth, and death, to mitigation, which provides the input data of the Agenter model. The inversed samples to be inversed are 10,000. Actually the population in the BMA is far more than 10,000, and we use this number to test and exemplify the model application.

A. Data input

An Access database illustrated is used to store the input and output of Agenter. The "Agents" table is the inversed samples using the input tables; The "Agents_spt" feature class in Point is the mapping results of the "Agents" table. In addition, The Access query module is used to validate the VBB couple type using inversed and statistical observed frequencies as to modify the RB relationship. The AgentAttrInfo table records the attributes inventory of samples, like their distribution type, relationship type, coupling type, and data type. This table includes 18 kinds of attribute information of population samples, like age, marriage status, monthly income, and education level. The attribute PARCEL indicates the corresponding parcel's ID of population samples, and can be used for mapping the inversed results. The attribute "AID" stands for the corresponding ID of point within parcels.

The "DA" table stores the basic information on the PDFs, including the name and their parameters. For example, the attribute INCOME accords with the Gaussian distribution, whose mean value is 6,000, and standard deviation is 1,000.

The "DB_" type tables store the domain elements and their frequency number for related attributes. For the attribute AGE, its DB_ table (DB_AGE) stores every domain element in the NAME column and corresponding frequency number in the PERCENT column.

The "RA" table stores the functional relationships among attributes. The "RB" table stores the inventory of probability relationships between two attributes. For instance, the marriage status (MARRIAGE) depends on the age range (AGE), and the living space attribute (RESIA) depends on the education level (EDUCATION). The detailed dependence relationship among population's attributes can be obtained through the Fifth Population Census Report of Beijing in 2000, in which many tables describe the relationship between two attributes. That is, one attribute is divided into several intervals, the same is samples. Samples within each interval of this attribute are further divided by the other attribute interval, and the count of

samples is provided in form of a cross matrix of two attributes intervals.

The table with "RB_" records the specific probability relationship. The RB_AGE_MARRIAGE table in Tab. Ⅱ is the relationship between AGE and MARRIAGE. The column AGE is the ID of the attribute AGE, which denotes an age range. The column MARRIAGE is the probability of every kind of marriage status for the corresponding age range. For example, for people from age 25 to 29 (ID is 6), according to the DB_MARRIAGE table, the probability of people with the first marriage and having a spouse (ID is 1) is 71.9%, the probability of unmarried (ID is 2) 27%, the probability of divorced 0.6%, the probability of people remarried and having a spouse (ID is 4) 0.5%, and the probability of other marriage status is 0.

TABLE II. THE RB_AGE_MARRIAGE TABLE (PARTIAL)

ID	AGE(*h*)	MARRIAGE(*j*)
1	1	2 100%
2	2	2 100%
3	3	2 100%
4	4	1 0.3%;2 99.7%
5	5	1 15.3%;2 84.7%
6	**6**	**1 71.9%;2 27%;3 0.6%;4 0.5%**
……	……	……
17	17	1 60%;2 0.3%;3 0.5%;4 2%;5 37.2%
18	18	1 50%;3 0.2%;4 2%;5 47.8%
19	19	1 40%;4 2%;5 58%
20	20	1 30%;4 2%;5 68%
21	21	1 20%;4 2%;5 78%

In addition, in order to mapping the inversed results, the spatial distribution layer, in which all samples are located, is prepared as the input of the Agenter model. In this paper, we use the parcel dataset of Beijing (PARCELs in Fig.1) as the spatial distribution layer. Since the expected inversion samples for testing are 10,000, which is far less than the actual population number in the BMA, we only choose partial parcels in the central city to correspond with inversed samples (see the grey polygon in Fig.2) rather than those in the whole BMA.

B. Results

The first 20 items from 10,000 inversion samples are shown in Tab. Ⅲ. The mapping results stored as the point Feature Class in ESRI Personal Geodatabase of partial inversed samples are shown in Fig. 2, in which every point locating in a certain parcel labeled in black has its corresponding information and can be used as the resident agent directly for MAS simulations. Operations like spatial analysis and statistic can be conducted using the mapped inversed results.

TABLE III. THE INVERSED RESULTS (PARTIAL SAMPLES AND ATTRIBUTES)

KEY_ID	AGE	SEX	EDUCATION	PARCEL	TRAVEL
1	52	Male	Junior High School	259	Bus
2	47	Female	Technical secondary school	197	Car
3	23	Male	Junior High School	338	Car
4	53	Male	Junior High School	211	Bus
5	32	Female	Primary school	248	Car
6	47	Female	Junior High School	90	Bus
7	19	Female	Primary school	51	Bus
8	51	Male	Junior High School	117	Bus
9	30	Female	Junior High School	320	Car
10	29	Female	Junior High School	237	Car

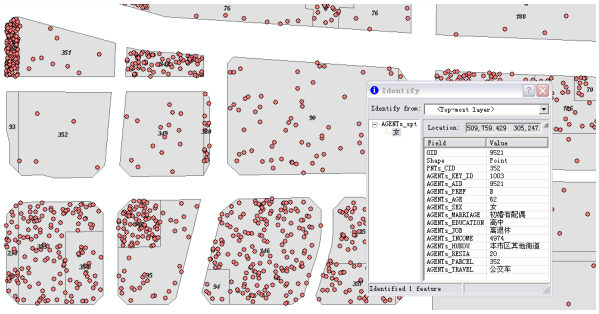

Figure 1. The spatial distribution of partial inversed samples

C. Results validation

- Different methods are used to validate the inversed results since various approaches are applied to inverse attributes with different coupling types of distributions and relationships.
- Type VAC. For example of the attribute SEX, the number of inversed male samples is 5,004 and female is 4,996, which accord to the uniform distribution. As to the attribute INCOME, we use the non-parametric test (one-sample K-S) method supposing the inversed attribute INCOME accords with a normal distribution. The mean value is 5,993.93 (6,000 as predefined), and the standard deviation is 989.99 (1,000 as predefined). The two-tailed (Asymp. Sig. (2-tailed)) significance is 0.657, and then the original hypothesis is accepted. Therefore, the attribute INCOME accords with the pre-set PDF.
- Type VBB. For instance of the validation of the attribute EDUCATION, the maximum error (21.21%) is from the education level of "junior middle school", while other education levels are well fitted. It can be overall accepted. In order to improve the degree of fitness between observed and inversed results, given that the attribute EDUCATION is calculated based on the attribute AGE, it can be achieved through adjusting the VBB relationships. The validation method of the attribute with VAB type is the same with VBB since the inversion process of DA depends on that of DB.
- VBC type. The attributes with this type of coupling strategy are determined by the aggregated statistical data, enabling it being in accordance with the original observed distribution.

- Type VCA. For the attribute "TRAVEL", the inversed samples completely correspond with the pre-defined decision tree.
- Type VCB. The attribute of this type is not involved in the experiment. Theoretically, it can be 100% in line with the pre-set dependence relationship between the attributes according to the inversion approach for the attribute of this coupling type.

In sum, the validation of attributes of the 10,000 inversed samples is in accordance with the pre-set distribution type and relationship type, as well as common sense. It is in accordance with the aggregate statistical input data, as well as the dependence relationship among attributes. The largest error rises from attributes with VBB coupling type, which is only limited in partial domain elements (like "junior middle school" of attribute "EDUCATION").

IV. CONCLUSIONS AND DISCUSSION

We proposed the Agenter approach in this paper to inverse individual samples using aggregated data and common sense. The aggregated data is used to model the distribution of attributes as well as relationship among attribute for samples inversion. Totally 9 types of inversion methods incorporating distribution with relationship are developed using the aggregated data. This approach is appropriate for the data sparse situation, especially in China. It can make full use of existing statistical information, surveys, and common sense to inverse the attribute values of samples, which is proved to be accurate in statistical level. The inversed samples can be further mapped as the input agents for multi-agent systems. This approach as a primary exploration can in some degree solve the bottleneck problem of MAS based simulations in the data scarcity context. This approach is also simple and easy to repeat in practical applications.

The inversed samples generated by our proposed approach are not the exclusive set although they obey the existing aggregated characteristics. When the inversed samples are applied in MAS simulations, we can inverse numerous pairs of samples, run MAS with each pair, and regard the mean result of all pairs of simulations as the final simulation result to reduce the uncertainty of applying the Agenter model for MAS models. In the next step, we like to retrieve the relationships among attributes through mining the observed statistical data, and update relationships for better fitness between observed and inversed data. Moreover, the relationship between samples' common attributes (like demographic attributes) and spatial attributes (like the location, area, orientation, and planning conditions) will be investigated to improve the inversed samples' quality.

REFERENCES

[1] Ballas D, Clarke G, 2000, "GIS and microsimulation for local labour market analysis" Computers, Environment and Urban Systems 24 305-330

[2] Beijing Fifth Population Census Office, Beijing Statistical Bureau, 2002, "Beijing population census of 2000" Beijing: Chinese Statistic Press (In Chinese)

[3] Brown D G, Robinson D T, 2006, "Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl" Ecology and Society 11(1) 46 [online] URL: http://www.ecologyandsociety.org/vol11/iss1/art46/

[4] Crooks A, 2006, "Exploring cities using agent-based models and GIS" CASA Working Paper No.109. Centre for Advanced Spatial Analysis, University College London

[5] Crooks A, 2008, "Constructing and implementing an agent-based model of residential segregation through vector GIS" CASA Working Paper No.133. Centre for Advanced Spatial Analysis, University College London

[6] Crooks A, Castle C, Batty M, 2008, "Key challenges in agent-based modeling for geo-spatial simulation" Computer, Environment and Urban Systems 32 417-430

[7] Deadman P J, Robinson D T, Moran E, Brondizio E, 2004, "Colonist household decision making and land-use change in the Amazon rainforest: an agent-based simulation" Environment and Planning B: Planning and Design 31(5) 693-709

[8] Hanaoka K, Clarke G P, 2007, "Spatial microsimulation modelling for retail market analysis at the small-area level" Computers, Environment and Urban Systems 31 162-187

[9] Hynes S, Farrelly N, Murphy E, O'Donoghue C, 2008, "Modelling habitat conservation and participation in agri-environmental schemes: A spatial microsimulation approach" Ecological Economics 66 258-269

[10] Langford M, Unwin D J, 1994, "Generating and mapping population density surfaces within a geographical information system" The Cartographic Journal 31 21-26

[11] Li X, Liu X, 2007, "Defining agents' behaviors to simulate complex residential development using multicriteria evaluation" Journal of Environmental Management 85 1063-1075

[12] Li X, Liu X, 2008, "Embedding sustainable development strategies in agent-based models for use as a planning tool" International Journal of Geographical Information Science 22 21-45

[13] Liao Y, Wang J, Meng B, Li X, 2010, "Integration of GP and GA for mapping population distribution" International Journal of Geographical Information Science 24(1) 47-67

[14] Mennis J, 2003, "Generating surface models of population using dasymetric mapping" The Professional Geographer 55(1) 31-42

[15] Miller J E, Hunt D J, Abraham J E, Salvini P A, 2004, "Microsimulating urban systems" Computers, Environment and Urban Systems 28 9-44

[16] Orcutt G, 1957, "A new type of socio-economic system" Review of Economics and Statistics 58 773-797

[17] Parker D C, Manson S M, Janssen M A, Hoffman M J, Deadman P, 2003, "Multi-agent systems for the simulation of land use and land cover change: a review" Annals of the Association of American Geographers 93(2) 314-37

[18] Pudney S, Sutherland H, 1994, "How reliable are microsimulation results? : An analysis of the role of sampling error in a U.K. tax-benefit model" Journal of Public Economics 53 327-365

[19] Robinson, D T, Brown D, 2009, "Evaluating the effects of land-use development policies on ex-urban forest cover: An integrated agent-based GIS approach" International Journal of Geographical Information Science 23(9) 1211-1232

[20] Shen Z, Yao X, Kawakami M, Koujin M, 2009, "Simulating the impact on downtown of large-scale shopping centre location: Integrating GIS dataset and MAS platform as a case study in Kanazawa city" Proceedings of CUPUM09 (HK)

[21] van Sonsbeek J M, Gradus R H J M, 2006, "A microsimulation analysis of the 2006 regime change in the Dutch disability scheme" Economic Modelling 23 427-456

[22] Wu B M, Birkin M H, Rees P H, 2008, "A spatial microsimulation model with student agents" Computers, Environment and Urban Systems 32 440-453

[23] Zhang H, Zeng Y, Jin X, Yin C, Zou B, 2008, "Urban land expansion model based on multi-agent system and application" Acta Geographica Sinica 63(8) 869-881 (In Chinese)