J Geogr Syst (2013) 15:403–426
DOI 10.1007/s10109-012-0171-2

ORIGINAL ARTICLE

Bayesian networks and agent-based modeling approach for urban land-use and population density change: a BNAS model

Verda Kocabas · Suzana Dragicevic

Received: 1 September 2011 / Accepted: 23 May 2012 / Published online: 20 June 2012

Abstract Land-use change models grounded in complexity theory such as agent-based models (ABMs) are increasingly being used to examine evolving urban systems. The objective of this study is to develop a spatial model that simulates land-use change under the influence of human land-use choice behavior. This is achieved by integrating the key physical and social drivers of land-use change using Bayesian networks (BNs) coupled with agent-based modeling. The BNAS model, integrated Bayesian network–based agent system, presented in this study uses geographic information systems, ABMs, BNs, and influence diagram principles to model population change on an irregular spatial structure. The model is parameterized with historical data and then used to simulate 20 years of future population and land-use change for the City of Surrey, British Columbia, Canada. The simulation results identify feasible new urban areas for development around the main transportation corridors. The obtained new development areas and the projected population trajectories with the"what-if" scenario capabilities can provide insights into urban planners for better and more informed land-use policy or decision-making processes.

Keywords Agent-based models (ABMs) · Bayesian networks (BNs) · Cellular automata (CA) · Geographic information systems (GIS) · Land-use change · Population change

JEL Classification C11 · C63 · 021 · R23

V. Kocabas · S. Dragicevic (✉)
Spatial Analysis and Modeling Laboratory, Department of Geography,
Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
e-mail: suzanad@sfu.ca

V. Kocabas
e-mail: verdak@alumni.sfu.ca

1 Introduction

Agent-based models (ABMs) are well suited for land-use change modeling because they can capture human interactions and dynamic behavior in the land-use change processes (Rodrigues et al. 1998). AMBs are complex, dynamic, and disaggregate representations of individual decision-making processes (Miller et al. 2004). They model complex system relationships as dynamically interacting heterogeneous agents that may result in nonlinear dynamics in land-use change (Brown and Xie 2006; Brown et al. 2004). Consequently, ABMs have been widely used to study land-use change processes (Ligtenberg et al. 2001; Parker et al. 2003; Brown et al. 2005b; Evans et al. 2006; Ligmann-Zielinska and Jankowski 2007; Bennett et al. 2011; Smajgl et al. 2011; Robinson and Brown 2009). Despite their widespread use in land-use change modeling, ABM designs are challenged by three main issues.

The *first issue* of ABM design deals with the rule definition of the agents. The ABM structure has autonomous decision-making entities called agents, an environment in which the agents interact, and behavior rules that define the relationship between the agents and their environment (Ferber 1998). Usually, the behavior rules take the form of <if–then> functions such that: <if> a specific situation is observed <then> a specific action is implemented. In the present literature, most ABMs are rule-based (Torrens 2006). However, the large number of alternatives from which an agent has to choose together with the complexity of internal relationships between variables makes rule definition a challenge for ABM design. ABM rules can also be defined using random utility theory and multinomial logit models (MNL) (Torrens and Nara 2007; Waddell 2002; Miller et al. 2004). In these models, the variables directly affect the modeling phenomenon and may not account for secondary interactions (i.e., a given variable affects the other one, and the other affects the model output). Another method for rule definition is fuzzy logic (Graniero and Robinson 2006). The limitation of fuzzy logic is that it does not provide inference from observed data. Neural networks (NN) have also been used for ABM rule definition (Gilbert and Terna 2000; Collins and Jefferson 1992), but a limitation is that they do not provide information about the relationships between variables and their effects on the output (Mas et al. 2004). Genetic/evolutionary algorithms are another method that has been used in ABMs. In one study, Manson (2006) proposed and tested genetic programming for rule definition to represent human decision-making in land-use change. However, in genetic algorithms, the fitness function must be chosen carefully as the algorithm may not find a single solution to the problem.

The *second issue* of ABM design deals with acquiring data, describing the behavior of actors. In land-use change ABMs, the agents represent actors such as households, firms, and land owners requiring detailed actor behavior patterns for effective rule definitions. Sample surveys, participant observation, field and laboratory experiments, and companion modeling are some ways to obtain the data (Robinson et al. 2007). However, these data are not usually available or easy to collect. As a direct consequence of this calibration issue, validation of ABMs is difficult to achieve (Brown et al. 2005a; Kocabas and Dragicevic 2009).

The *third issue* of ABM design deals with the spatial implementation structure. Many ABMs are based on raster grids (An et al. 2005; Loibl and Toetzer 2003; Batty 2005). However, the irregular shapes that characterize urban landscapes are not well represented by regular raster grid cells. The idea to use irregular cells to model some geographic features was first proposed by Tobler (1984). Alternatively, some ABMs have used vector-based representations to accommodate irregular spatial structures when simulating residential dynamics (Benenson 2004), modeling urban segregation (Torrens and Benenson 2005) and gentrification (Torrens and Nara 2007). Nevertheless, these approaches do not include census-based social dynamics of human agents based on census spatial units. Benenson et al. (2002) have developed an ABM to simulate residential dynamics using spatial units that represent buildings and not census units although they have used the census statistics data.

In this research study, we use Bayesian networks (BNs) and influence diagrams (IDs) as an alternative methodology to overcome the three issues of ABMs discussed. BNs reduce the theoretical and computational challenges of rule-based agent models by using graphical probabilistic models to represent agent rules (Jensen 2001). This enhances ABM rule definitions by keeping track of decision-making interactions during model simulations. BNs provide a clear semantic interpretation of model variables with the conditional probabilities being intuitively understandable. Further, this improves understanding of variable relationships and identifies the indirectly interacting variables. BNs also have causal reasoning capabilities that provide knowledge about the decision-making of the agents. Causal reasoning deals with reasoning about actions, explanations, and preferences. Utilizing IDs together with BNs allows uncertainty to be accommodated in the agent rules as well as explicitly representing the causal reasoning behind decision-making.

In the research literature, it can be found that some studies relate BN with spatial modeling. Janssens et al. (2006) coupled BNs with decision trees in a transportation model to describe travel mode choices of individuals. de Almeida et al. (2005) have used Bayesian conditional probabilities in cellular automata (CA)-based land-use change model to calculate transition probabilities. However, the existing modeling approaches did not represent the causal relationships between model variables. Several past studies have also integrated BNs into agent-based modeling. Ma et al. (2004) applied IDs to a multiagent model of land-use decisions for land suitability analysis. Their model is designed for the behavior of retail agents making location decisions for new shopping centers. Also, their model is cell-based and does not use the learning algorithms to define the model structure and parameters. Moreover, their model has a simple network in which no inference algorithms were used to calculate the probabilities. BNs and IDs were also used in ABMs to model urban land market sales (Lei et al. 2005), which is conceptually different from a land-use and population change model.

The objectives of this research study are (a) to develop the hybrid land-use change model, called BNAS—Bayesian network–based agent system—model, that incorporates human behavior, (b) to use BNs to define agent behavior rules and IDs for decision alternative evaluation, and (c) to implement the model on the real data set composed of irregular spatial tessellations. Thus, this study develops a simplified model of urban land-use, and population density change with clear semantic interpretation of the parameters and their relationships with the aim to overcome the common "black-box" type models.

The BNAS model was implemented for the City of Surrey (British Columbia, Canada) which is undergoing extensive land and population change. Policy scenarios for interpreting the Metro Vancouver Livable Region Strategic Plan (LRSP) (MetroVancouver 2012a) were applied in the BNAS modeling procedures.

2 Methodology

The use of BNs in a cellular automata (CA) model of urban land-use change was previously developed by Kocabas and Dragicevic (2006). In addition, Kocabas and Dragicevic (2007) improved their BN-CA model by integrating IDs to better model future land-use changes in Vancouver, Canada. The research presented here extends the authors' previous BN-CA model (Kocabas and Dragicevic 2007) by integrating ABM concepts as well as BN and IDs to explicitly deal with human decision-making. The BNs and ABMs were coupled to design the agents' location decision-making capabilities. During the model simulations, agents observe their environment and implement an appropriate action using this new information. By using BNs, agents also change their behavior according to the changing conditions, which mimics the actual behavior of humans. It is difficult to identify and encode all the necessary rules for ABMs as they adapt themselves to the changing conditions. Consequently, BNs provide simplicity in rule definition through the learning algorithms. Moreover, BNs facilitate modeling even with missing data in the inputs, which is not possible with rule-based models.

In ABMs, action selection and learning from experience are also important challenges (Maes 1994). Action selection is a challenge for agents as there are many alternatives and opportunities in the selection set. IDs combine both BNs and utility theory to provide a solution for more effective action selection (Howard and Matheson 1984). In this research study, BNs together with IDs were used to develop the locational preference decision-making processes of an agent. As a result, agents are able to reason about the benefits of possible decisions using decision variables and utility functions.

Agent learning from experience is also an important stage in ABM design as it provides dynamic development of new behaviors and a long-term improvement in agent performance. Agents' learning approaches are not yet fully elaborated in the present literature although there is some work done by Manson (2006) using genetic algorithms and Janssens et al. (2006) using learning networks. In this study, BN learning algorithms are used to facilitate agents to learn from their previous actions.

In this study, the BNAS model was developed as a computational model for simulating actions and interactions of autonomous individuals such as households in an urban area. Agents are new arrivals to the study area, and they choose locations to live. Hence, the BNAS model simulates future population and land-use for the City of Surrey, British Columbia, Canada. The individual agents are assumed to be acting in self-interest. An agent will choose a location maximizing its utility which is called their "preference." Terms such as "agent's preference" and "behavior" are used in a computational context as software routines are coded to "behave" and "prefer" during simulations. While specific agent characteristics were selected to balance model complexity with real-world processes, more variables can be readily included in the analysis. The BNAS model has the following characteristics: (a) an agent learning process using BN learning algorithms, (b) fine spatial modeling units such as census units in which agents choose to live, (c) using real data in an urban environment for model implementation, (d) using census data for model validation, and (e) simulating future population density and urban land-use.

2.1 Study area

The Metro Vancouver region has experienced rapid land-use change in the last decade, and is the third largest urban region in Canada. Recent studies have predicted that rapid population growth will lead to around three million inhabitants in 2031 (MetroVancouver 2012b). Within the Metro Vancouver, the City of Surrey was chosen as a study area for the BNAS model application in order to reduce the model complexity and computational processing time. The BNAS model simulations for the City of Surrey generated six iterations with five-year temporal intervals. The land-use pattern at 2021 was predicted at the end of the model simulation. A total of six simulation maps which depict the land-use change at interval years were generated.

2.2 Drivers of land-use change

There are many land-use drivers that influence land-use decisions, so it is crucial to isolate the core land-use drivers for use in modeling land-use change. The potential land-use drivers were defined, and then the BN structure learning algorithms were executed with different configurations of drivers (i.e., BN structure learning algorithm was run with different variables to be able to compare different network structures and choose the best one to be used in the BNAS model). A single BN structure was obtained for each run. Next, by comparing the different BN structures, the important drivers affecting one or more variables in these structures were chosen for use in the model (i.e., these drivers were used as inputs to the BNAS model for designing the agent behavior). Structure learning algorithm was run for the last time to obtain the final network structure to be used in the BNAS model. There were 10 land-use drivers used in this study:

- Distances to major roads and SkyTrain stations affect accessibility of locations.
- Distances to town and employment centers were calculated using a cost–distance function. Cost–distance is the calculation of distance in terms of some measure of cost. Major road and SkyTrain station data were obtained from the Greater Vancouver Transportation Authority (Translink). Town and employment center data were obtained from Metro Vancouver.
- Population density in the neighborhood affects the use of available land.

- Accessibility to natural areas provides a measure of the quality of the natural environment in the neighborhood of a location. These quality factors included green spaces and scenic landscape views such as the ocean and rivers.
- Land and dwelling rental values capture the economic status of the location.
- The type of dwelling, the number of rooms, and whether they were rented or owned gives an indication of available resource capacity.

The data layers for the BNAS model were prepared using GIS operations such as reclassification and distance calculations from strategic point locations.

2.3 Data integration

Spatially, disaggregated census data are an important input for empirical agent-based modeling because of the emphasis on micro-level behavior (Parker et al. 2001). The BNAS model uses Canadian census data (Statistics-Canada 2012). The Canadian census data are the primary source of socioeconomic data in Canada that can be linked with land-use changes at the micro-level. These census units represent homogenous information on demographics, economics, and living conditions for urban populations.

There were two challenges related to using the Canadian census data source in BNAS. The first was a change in the size and shape of the census units between 1996 and 2001 (Fig. 1a). Enumeration areas (EA) are the smallest census spatial units for the 1996 census while the dissemination area (DA) units replaced the EA for the 2001 census. For example, the EA 59026014 has been split into the following two DAs: 59151980 and 59151977. The second challenge was the representation of population by the census units. Census data do not provide information on the population concentration or distribution over the census units (Dorling 1993; Wu et al. 2008). The census data give a single population value for each unit even though that unit may contain a mixture of residential and non-residential areas (Fig. 1a and b).

These two challenges were addressed by using a data enhancement procedure to create a common spatial unit termed proposed census spatial units (*PCSUs*). They were created by spatially intersecting the EAs, DAs, and land-use polygons for 1996 and 2001 to produce a common spatial unit. During the intersection process, all attribute values in the EAs, DAs, and land-use polygons were retained in the intersected polygons. In addition, the area of each PCSU was calculated. Next, population values were assigned to the PCSUs for 1996 and 2001 by using the dasymetric method first developed by Wright (1936) to redistribute the population values to the new PCSUs. The modified formula used to calculate the new population was:

$$NP = P_{EA(orDA)} \frac{AW_{PCSU}}{AW_{EA(orDA)}} \quad (1)$$

where NP is the new land-use-based population; $P_{EA(orDA)}$ is the population in EA or DA unit; AW_{PCSU} is the multiplication of the area and weight of PCSU; $AW_{EA(orDA)}$ is the sum of all AW_{PCSU} which fall in EA or DA unit. The result was then divided

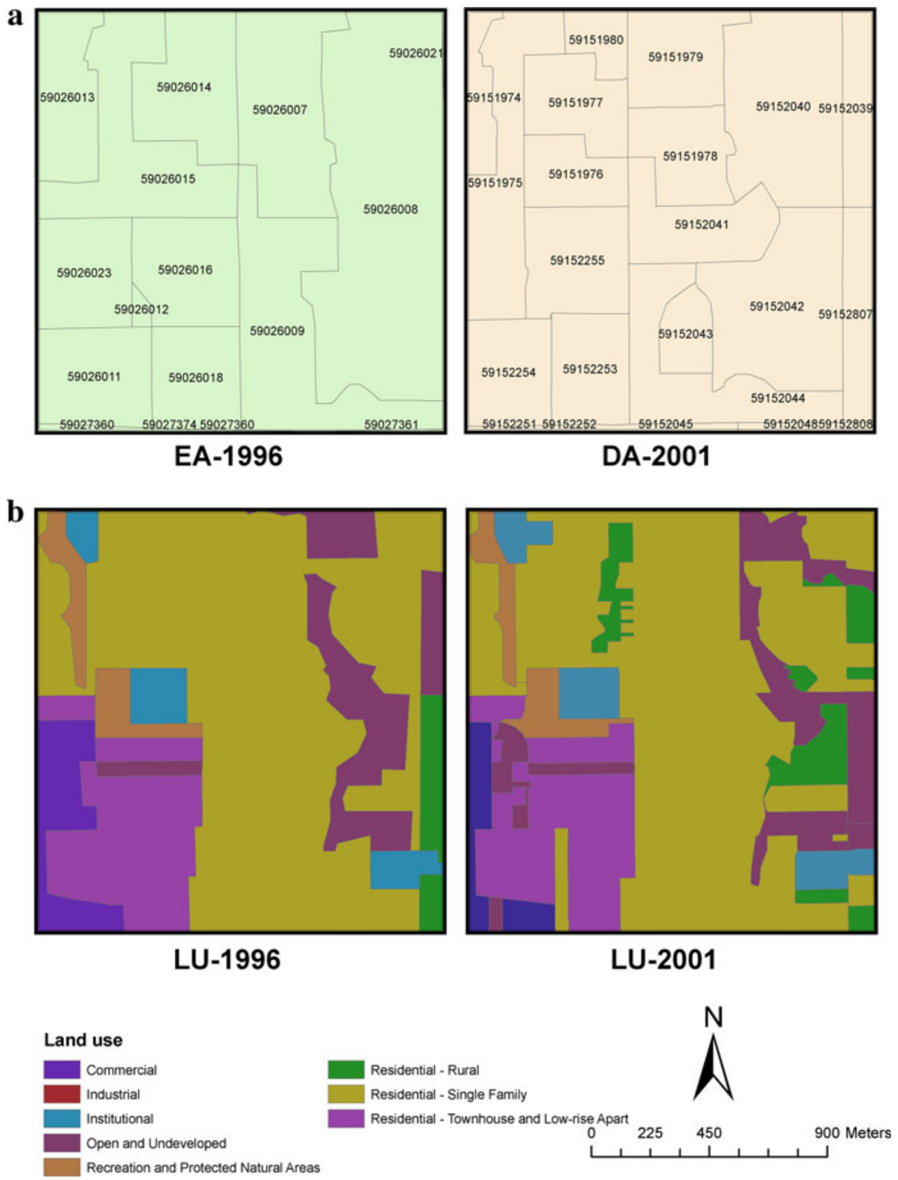


Fig. 1 Different spatial units covering the same area in Metro Vancouver **a** EA and DA **b** land-use polygons

by the aggregated weight and EA (or DA) population multiplication for the corresponding EA (or DA). The weights are important in this calculation since the land-use data set has different residential classes based on density. High-rise residential areas have more population than rural residential areas. Assigning weights to the land-use residential classes captures the different population distribution in the census spatial units (EAs or DAs). The weights are 0.4 for high-density residential, 0.3 for medium-to-high-density residential, 0.2 for medium-density residential, and 0.1 for low-density residential. Figure 2 compares the population density maps based on the 1996 and 2001 census units with the ones that are based on PCSUs.

This study used the same PCSUs (a total of 14,793 polygons) spatial units for 1996 and 2001. All the land-use change drivers were calculated using GIS tools for

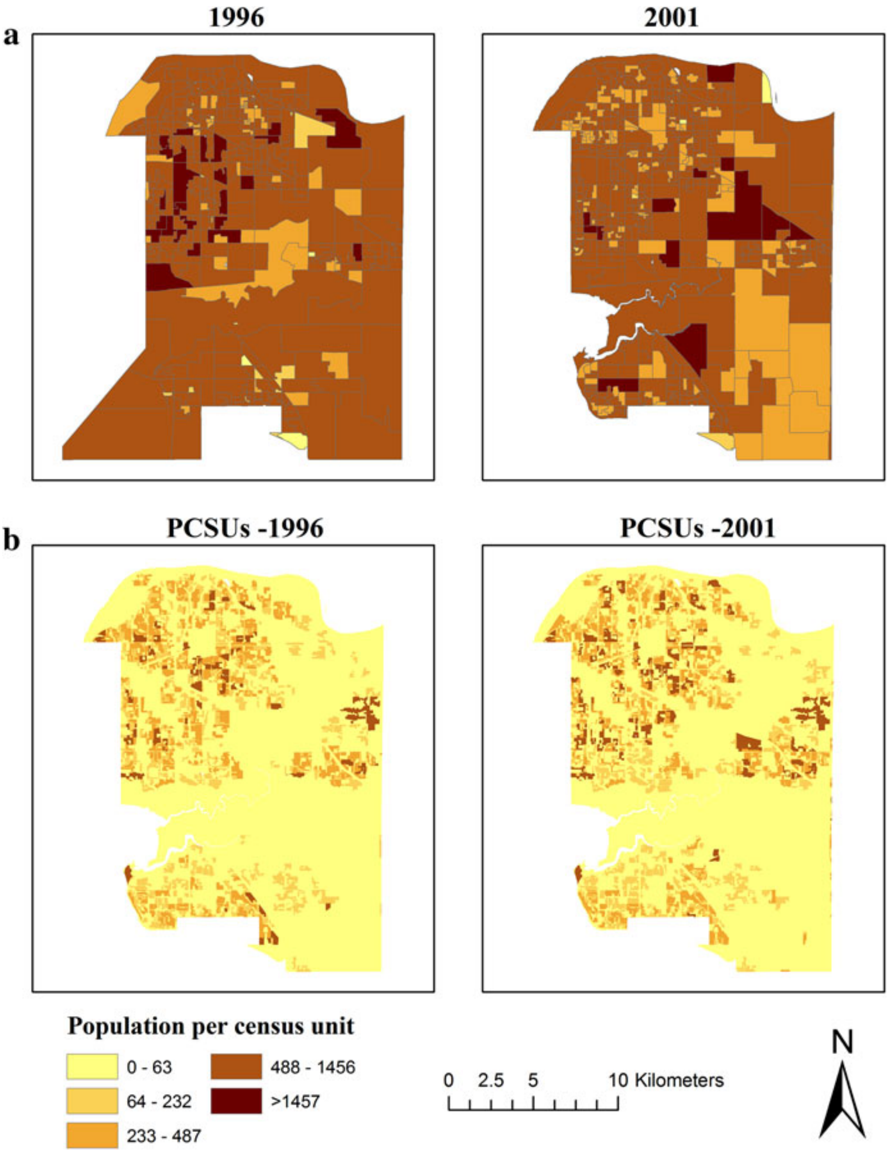


Fig. 2 Population for 1996 and 2001 **a** before and **b** after the creation of PCSUs

each PCSU and recorded as attributes of each unit. These form inputs to the agents' BNs.

2.4 Bayesian network–based agent system (BNAS) model

The BNAS model uses an object-oriented model design allowing location to be an attribute of features and facilitating the representation of methods as objects in the programming architecture. Further, the attribute of the objects can be updated asynchronously. The urban environment (*ENV*) is represented by the objects that consist of PSCUs (fixed objects that do not move across *ENV*) and agents (mobile objects that do move across *ENV*) (Torrens and Benenson 2005). PCSUs define the environment and each has attributes on the state of the environment.

The BN-based agent-based model has three sub-models (Fig. 3). The first is the BN sub-model (**A**) that implements the BN structure and parameter learning and inference algorithms. The BN structure and parameter learning algorithms find relationships between current socioeconomic characteristics of households/firms and current location characteristics of where they reside using GIS layers. The BN inference algorithm calculates the probability of choosing a particular location of an agent. The second is the agent sub-model (**B**) which structures the agents for the simulation. The third is the ID sub-model (**C**) that makes the decision whether to stay at a specific location or not. This information becomes the input for new agents

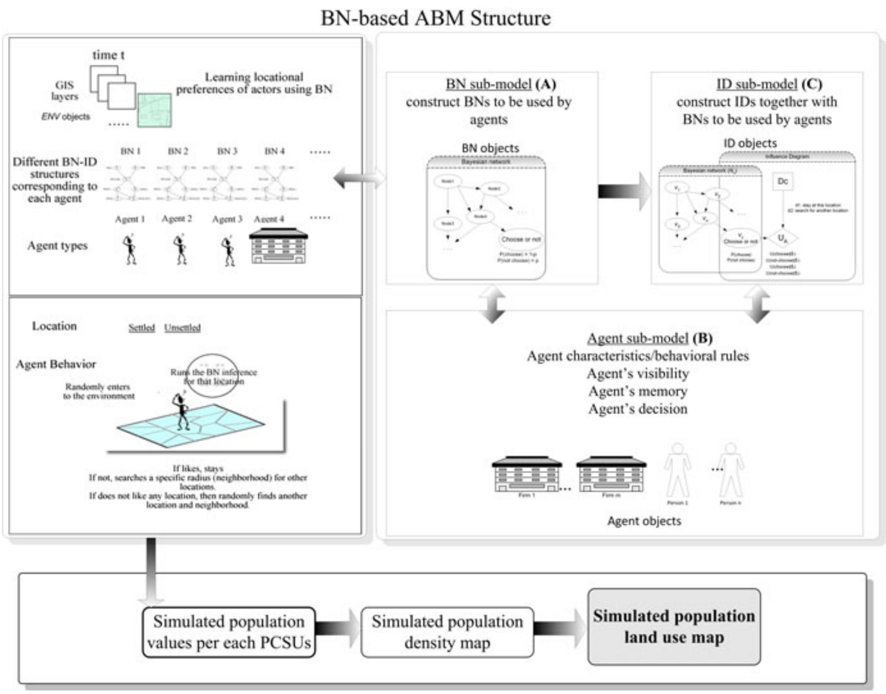


Fig. 3 BNAS structure

that represent households/firms and subsequently make decisions accordingly. These decisions lead to future population distribution patterns in an urban area and thus provide insights into future land-use change.

There are two types of agents: households and firms that are similar in the way they make location decisions. Household agents have socioeconomic characteristics affecting how they evaluate locations. As such, there are three different types of household agents: low-, middle-, and high-income agents. For example, high-income households may prefer areas with high levels of social services and high levels of environmental quality. In contrast, low-income households might prefer areas that are close to the public transport. Firms represent commercial areas. For instance, they might prefer to be close to the employment centers and high residential density areas.

*2.4.1 Bayesian network sub-model (**A**)*

A Bayesian network (N) is a pair (G, P) where G is a directed acyclic graph with a node for each variable $v_{\mathrm{i}} \in V,$ and P is a multivariate probability distribution defined for each variable v_{i} in V (Eq. (1) and Eq. (2)). A graph is called *directed* if the links in the graphs have directions. A directed graph is *acyclic* (DAG) if the graph contains no directed cycles. In this study, node and variable are used interchangeably since each variable in V corresponds to a node in a BN.

$$G = (V, E) \tag{2}$$

where $V = (v_i|1 \leq i \leq n, n > 0)$ denotes the set of nodes v_i and $E = [(v_i, v_j)|v_i, v_j \in V, i \neq j]$ denotes edges for directed graph

$$P = \{P(v_i|\pi(v_i))|v_i \in V\} \tag{3}$$

where $\pi(v_i)$ denotes the parents of $v_i \in V$ in G. A parent is a node which has a directed link to the node in question. More detailed discussions about BN concepts can be found in Pearl (1988), Jensen (2001) and Neapolitan (2003).

The advantages of using BNs in ABMs are as follows:

(a) *Capability to handle incomplete data sets:* Usually, when one of the inputs to ABMs has missing values, inaccurate prediction will result. BNs handle missing data by applying prior probabilities to estimate the missing values (Heckerman and Wellman 1995).
(b) *Employing causal relationships:* BNs can explicitly represent the causal reasoning about the preferences of agents which is beneficial in understanding the behavior of agents during simulations (Ma et al. 2007).
(c) *Facilitating the combination of prior knowledge and data*: BNs can use prior knowledge in terms of probabilities about the event being modeled. This allows for inference and optimal decision-making (Uusitalo 2007; Heckerman and Wellman 1995).
(d) *Linking multiple variables:* BNs can be used to link economic, social, or physical variables (Bromley et al. 2005), and therefore are suited to land-use change modeling.

(e) *Avoiding overfitting of training data:* BNs avoid data overfitting as they rarely generate unnecessary dependencies. In addition, they allow the use of all available data in the model training process (Heckerman and Wellman 1995).

The construction of a BN requires four items to be established. First, a set of nodes corresponding to the variables in the BN represents the most important factors of a particular event being modeled. In this study, these variables are called land-use drivers and constitute the set *V*. Figure 4 represents the BN-ID structure of BNAS. Second, each node has a set of mutually exclusive states, which are the states of the land-use drivers. Third, a set of directed links represent causal relationships between the nodes. Fourth, each node has a set of probabilities, specifying the chance that a node will be in a particular state given the state of its parents. For example, there is a conditional probability associated with choosing a specific location of an agent given the values of the land-use drivers of the location and the socioeconomic properties of the agent. These probabilities are stored in conditional probability tables (CPT) and quantify the strength of dependencies between connected variables in the BN structure (Bromley et al. 2005). This study uses different BNs (N_A) and in turn different CPTs for each agent type because of their different preferences. Individual agents are members of different agent groups (i.e., high-income household group).

Once land-use drivers are represented as variables (nodes) and their corresponding values (states), the BN structure is constructed. Although the network structure and the CPTs can be obtained using expert knowledge, it is still a challenging modeling task because there are many parameters to be learned especially in large size networks. Hence, some learning algorithms must be used to find the best

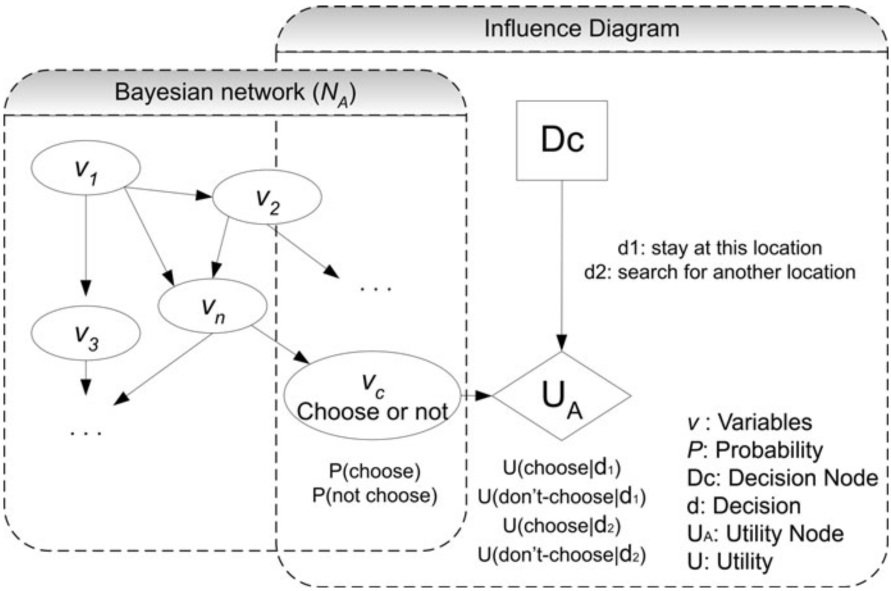


Fig. 4 BNAS model BN-ID structure

network structure that fits the observed data. Learning is a process of finding the links between a given set of variables (i.e., structure learning) and then estimating parameters for the CPT (i.e., parameter learning) using the observed data. In this study, the K2 algorithm (Cooper and Herskovits 1992) based on a Bayesian scoring method was used to learn the BN structure from the observed data. The K2 algorithm was found to be an efficient algorithm for structure learning from complete observable data since it gives better results in structure learning (Cheng et al. 2002). The algorithm exhaustively searches the DAGs space (all possible DAGs given the nodes) and finds the best network structure by recursively selecting the best set of parents for each node independently (Cooper and Herskovits 1992).

The learned BN structures for each agent type are shown in Fig. 5. Although they appear similar, there are minor differences that are important for the agent's decision-making. The variable "proximity to town centers" does not affect the decision of low-income agents, but it has a direct effect on the choices of medium- and high-income agents. The reason might be the fact that some shopping areas are expensive for low-income households. Medium- and high-income agent decisions are not influenced by the proximity to recreation, in contrast to low-income agents. Easy access to recreational areas is not required by medium- and high-income families in Metro Vancouver. Whether the houses are mostly rental or owned is

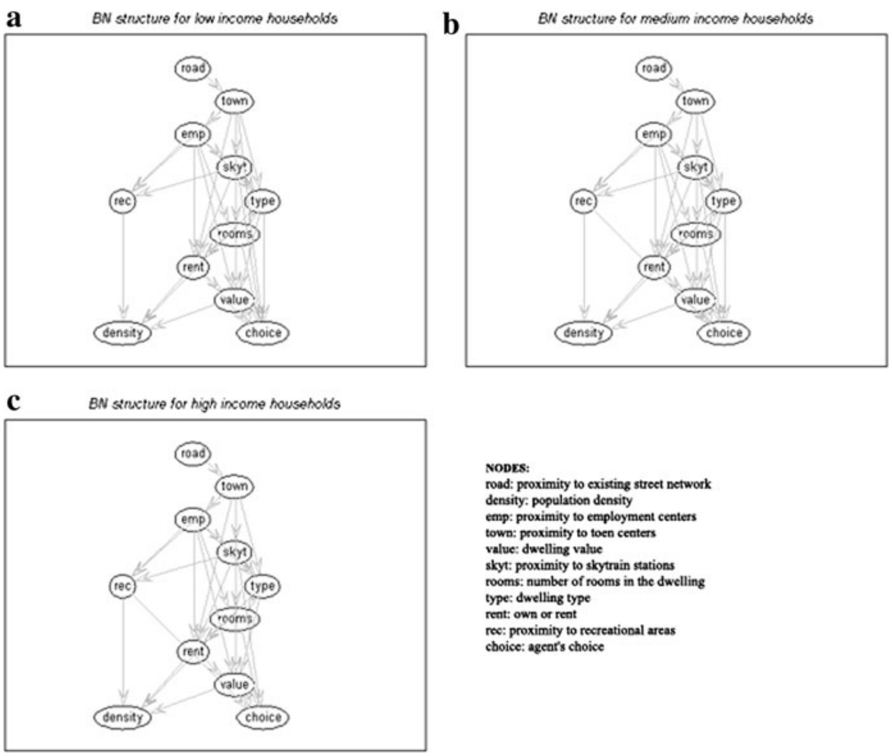


Fig. 5 Learned BN structure for **a** low-, **b** medium-, **c** high-income households

important for high- and medium-income agents, but it does not have a direct influence on the decisions of low-income agents. Also, high-income agents' decisions are affected by the dwelling types. Hence, different BN structures obtained from the learning stage of the BNAS model showed that households with different socioeconomic characteristics have different location preferences.

Each agent object stores its own socioeconomic characteristics as well as its location in the environment. All the selected land-use drivers were incorporated into the BN model as nodes. In addition to those, as a last node, the agent's decision was defined in the BN structure to obtain the probabilities of each decision: "choose" or "don't choose". Using the BN learning procedure, the structure of the networks for each type of agent (such as high-, middle- and low-income agents) was constructed, and values of CPT were estimated from the 1996 census and associated data.

Once the network structure is chosen, the CPT parameters are estimated. As in the structure learning of BN, there are several possible ways of generating estimates for the conditional probabilities in the CPT. In this study, the probabilities were derived from historical data using maximum likelihood estimation since learning these parameters from observed data can make the process easier and provide some level of objective probabilities rather than completely subjective probabilities (Neapolitan 2003). After this step, the BNs for different agent types are ready to be used in the agent sub-model (**B**).

The probability values in CPTs are different for each agent type. Each BN node has an associated CPT describing the conditional distribution of that node given different assignments of values for its parents. CPTs for each node have many parameters directly related to the number of parents. Hence, it is not possible to display all CPTs for all nodes for each agent because if the node has n parents which have m number of states, then there are m^n probabilities for each node which makes it difficult to show them here. As an example, Table 1 shows the CPT of the node "choice" given its parents, proximity to town centers, proximity to SkyTrain stations, dwelling type, no. of rooms in the dwelling, own or rent, and value (rental amount/land value) for low-income agents.

CPTs such as Table 1 only depict the prior probabilities for each variable. An advantage of Bayesian networks is the ability to compute the posterior probability distributions of the variable (or node) under consideration given that values of some other variables are known. In this case, the known states of variables can be entered as evidence in the network. In this study, the evidence are the states of variables at each iteration that is running, thus agents' observations. When evidence is entered to the BN, the states of other variables change as they are conditionally dependent (Kocabas and Dragicevic 2007), and also the probability of choosing a particular location changes.

2.5 Agent and Influence diagram sub-models (B and C)

The agent sub-model (**B**) structures the agents for the simulation. The number of agents and their types were used as inputs to the model. In the agent sub-model (**B**), agents perform observations at time t, then select and execute actions between time t and $t + \Delta t$. The outcome of these actions is reflected at time $t + \Delta t$.

Table 1 CPT for the "choice" node in the BN for low-income agents

Proximity to town centers	Proximity to SkyTrain stations	Dwelling type	No of rooms in the dwelling	Own or rent	Value	Choice	
						Don't choose	Choose
Low accessibility	Low accessibility	Single-detached house	Low	Rent	Low	0.6	0.4
Low accessibility	Low accessibility	Semi-detached house	Low	Rent	Low	0.8	0.2
Low accessibility	Low accessibility	Row house	Low	Rent	Low	1	0
Low accessibility	Low accessibility	Apartment, detached duplex	Low	Rent	Low	1	0
Low accessibility	Low accessibility	Apartment, building that has five or more storeys	Low	Rent	Low	0	1
Low accessibility	Low accessibility	Apartment, building that has fewer than five storeys	Low	Rent	Low	0	1
Low accessibility	Low accessibility	Other single-attached house	Low	Rent	Low	1	0
Low accessibility	Low accessibility	Movable dwelling	Low	Rent	Low	1	0

Different numbers of agents enter the simulation at each iteration (or time step). Population projection data were obtained from Metro Vancouver (2012a) for each time step in the simulation. Then, the total new population according to these forecasts was divided into different income levels. Subsequently, the calculated number of agents enters into the environment at each iteration.

Agents enter the environment one-by-one and select a random place (L) in the environment, for each iteration. Once in the environment, each agent randomly searches for a location and stores the information about that location. With this information, the agent runs the BN sub-model (**A**) and the ID sub-model (**C**) to create real-time decision-making in the simulation. The BN (N_A) and its nodes represent the agent's beliefs about the environment. The nodes in N_A represent the variables in BNAS model (please refer to Fig. 4). The last node in the BN structure is "*choosing a particular location* (v_c)". That particular location could be any location that is available to the agents to decide. An agent implements the inference for L. The probability of choosing L given the evidence is calculated during the inference process. The variable values of L at time t (agent's observations) are entered into the BN as evidence, $E_t = [\{v_1, v_2, \ldots, v_{i-1}\}]$, and the probability of choosing L, $P(v_c|E_t)$, at $t + \Delta t$ is obtained. Multiple algorithms exist to perform inference in BN (Lauritzen and Spiegelhalter 1988; Pearl 1988). This study used the junction tree algorithm (Jensen 2001). The algorithm first groups the nodes which are fully connected, then connects them to form a junction graph (tree). Using a message passing method, it collects probabilities from CPT children nodes and updates the probabilities at parents. Afterward, it distributes updated information to the children nodes so that new probabilities are calculated.

The ID sub-model (**C**) runs the IDs for the agents and makes the decision whether to stay at a specific location or not. An ID is a BN extended with utility nodes and decision nodes. However, d ecision and utility nodes do not have CPTs but like other nodes they have states. Decision node states represent the various decision alternatives, and utility node states represent the utility of each possible outcome of the decisions.

The utility node (U_A) in the agent BNs represents the desirability of choosing L, while the decision node (Dc) expresses the final decision of the agent. The decision is to stay (d_1) or search for another location (d_2). U_A is associated with a set of utilities $u(v_c)$, which specifies utility outcomes for the probability of choosing L For each decision (d_1 or d_2), the utilities of its outcome are multiplied by the probabilities ($P(v_c)$) that these outcomes will occur. This is termed the *expected utility* (EU) of decision d_j and given by the equation:

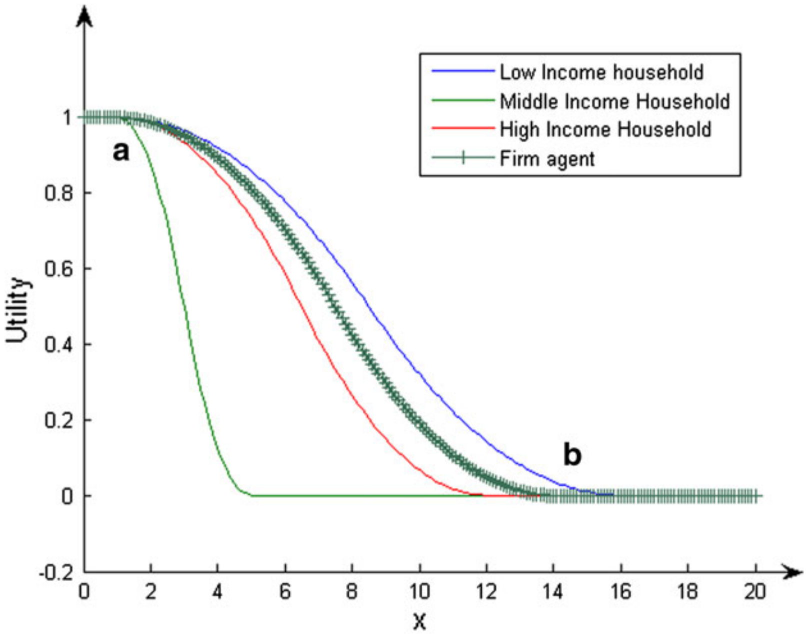
$$EU(d_j) = \sum_{j=1}^{j=n} u(v_c|d_j)P(v_c). \quad (4)$$

Figure 6 presents the utility function used in this study, where the utilities (u) depend on the cost of changing land-use state (C_{PCSU}) and the population density of the PCSU unit (D_{PCSU}). In the figure, a and b are the extreme values of the sloped portion of the utility curve. The values are set by tuning using the historical data, and each type of agent (household and firm) has different a and b values. Therefore, they have different utility outcomes for different decisions as they have different preferences in location choices.

Interpreting the Metro Vancouver Livable Region Strategic Plan (LRSP) (MetroVancouver 2012a) is a key part of constructing a policy scenario in the BNAS model. LRSP is MetroVancouver's regional growth strategy which was adopted in 1996. The primary goal of the plan is to protect the environment in the face of anticipated growth. Each land-use plan designation (e.g., green zones) may be described as a set of restrictions on development options in the BNAS model. One of the important policies that LRSP utilizes is protecting the agricultural land reserve (ALR) areas defined by the provincial government of BC. The ALR are BC lands that have the potential for agricultural production. Therefore, LRSP includes ALR areas in the green zone. In addition, it adopts a compact growth strategy to reduce the development pressure on the green zone. The green zone establishes a long-term boundary for the urban growth. Green zones are not allowed to convert to any urban category. The green zone policy has been included in the influence diagram sub-model (**C**). Cost in the utility formula increases if a location is in the green zone boundary.

After executing its ID, the agent is required to take an action according to the principle of maximum expected utility, and hence, it chooses the decision alternative that produces the highest utility.

As a result of a decision based on maximum expected utility, an agent is either "unsettled" (i.e., searching for a location) or "settled" (i.e., has found a location). If



$$u = \begin{cases} 1 & x \leq a \\ 1 - 2\left(\dfrac{x-a}{b-a}\right)^2 & a \leq x \leq \dfrac{a+b}{2} \\ 2\left(\dfrac{b-x}{b-a}\right) & \dfrac{a+b}{2} \leq x \leq b \\ 0 & x \geq b \end{cases} \qquad \text{where } x = C_{PCSU} * D_{PCSU}$$

Fig. 6 Utility functions for each agent type

the decision is to leave the location, the agent looks within a specified radius (neighborhood of that particular location) for other locations: this is called agent's vision. The neighborhood of a spatial unit consists of other spatial units that reside within a radius of one km. Although different neighborhoods could have been defined and a sensitivity analysis (Ligmann-Zielinska and Sun 2010) could have been employed on them, this will constitute future work of this study.

There is bounded rationality (Gigerenzer and Selten 2001) in agent's decision-making. Although humans in real life do not make sequential binary choices concerning their homes all the time, rational decision-making approach is not taken in agent-based modeling due to the finite computational resources. Thus, agents do not search environment and do not evaluate all possible location alternatives, but use a sequential choice of local optimizations that mimics human choice behavior. If

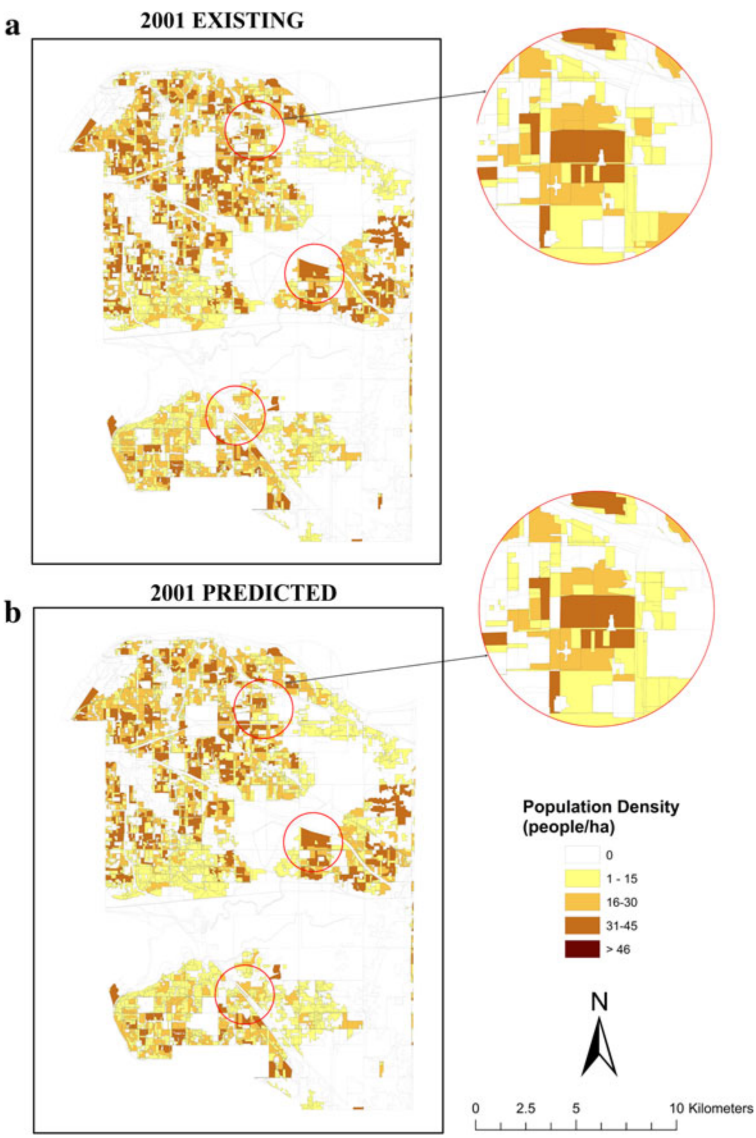


Fig. 7 **a** Actual and **b** predicted population density for the City of Surrey in 2001

the agent is not settled in one of those neighbors, it decides to leave the environment.

When an agent becomes settled, its location attribute is updated and the population density of that location increases. Hence, each new agent that enters the simulation moves until it finds a location to be settled. The vector-based CA model updates the current state of the environment ($ENV_t \rightarrow ENV_{t+\Delta t}$), such as the population density information, so that one agent affects the others present. As such, agents are influenced by nearby agents that exist in the neighborhood. The final

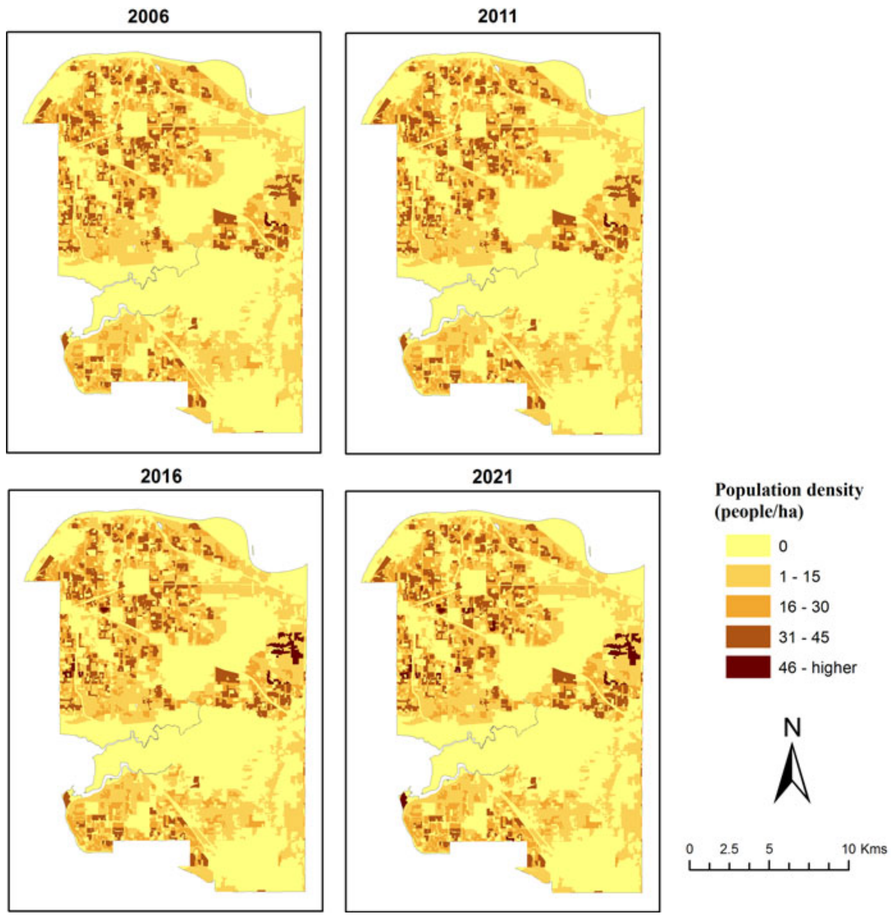


Fig. 8 Predicted population density for the City of Surrey from 2006 to 2021

result of the simulation is population values and densities per census polygon. After obtaining a population density map, the model generates a land-use map by classifying densities ranging from high- to low-residential (urban) areas. Next, the CA model uses the classified map as input to the next iteration. New agents entering the environment in the next iteration have to adapt to these changing conditions. This adaptation is implemented by the BN learning algorithms, which runs after each iteration and learns new behaviors and beliefs of the agents. This means that the agents' BNs change after each model iteration based on the behaviors of agents at the previous iteration.

According to the agent's decision, the environment changes with the agent's action. More than one agent can occupy a spatial unit. Once agents have made their decisions, the population density values for PCSUs were updated. Therefore, the final result of the simulation is population values and densities per census spatial unit. After obtaining the resultant population density map, the model generates a

Fig. 9 Simulated land-use change maps from 2006 and 2021

land-use map by classifying densities ranging from high to low-residential (urban) areas.

3 Simulation results and discussion

The BNAS model and algorithms were developed in MATLAB (Mathworks 2012) using the functions of the Bayes Net Toolbox (Murphy 2001) and Arc_Mat Toolbox (LeSage and Pace 2004; Liu and LeSage 2009) and were loosely coupled with ArcGIS (ESRI 2012). Inputs were generated with ArcGIS, and then input into the BNAS model. The model outputs the simulation results in the ArcGIS vector data format.

The BNAS model output was validated by using 1996 PCSU data as input to forecast population density in 2001. This predicted population density was compared with the actual 2001 population density for PCSUs. The actual and

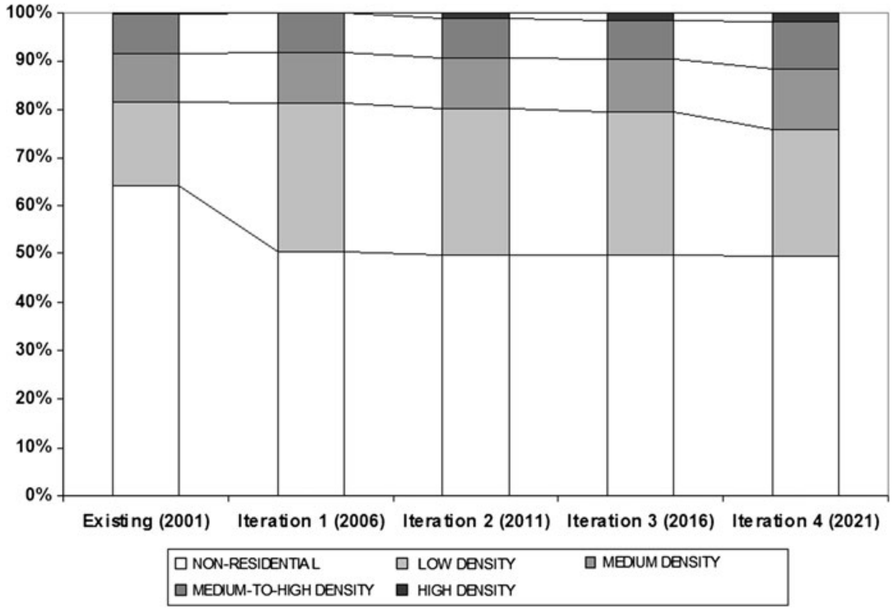


Fig. 10 Percentage change in the predicted area of population density classes for the City of Surrey from 2001 to 2021

predicted population density maps are shown in Fig. 7. The circled areas in Fig. 7 highlight particular areas where BNAS forecast population density correctly. Comparison of the actual versus predicted population per PCSU for 2001 yielded an R^2 of 0.95, indicating that BNAS predicted 2001 population density across the entire study area very well.

Population density predictions in five-year intervals from 2006 to 2021 (Fig. 8) indicate that within 20 years, high-density areas in the City of Surrey may emerge around the city center (northwest of the city) and east of the city. The results also reveal that there will be small population density increase in the southeast of the city. In addition, the predicted population density map for 2021 indicates that most change occurs around the SkyTrain stations, main transportation corridors (Fig. 9), and within the growth concentration area of the Metro Vancouver Liveable Strategic Plan. The reason that most of the change occurs within the growth concentration area can be explained by the cost variable in the utility formula as it increases if a location is in the green zone boundary.

Figure 10 depicts the amount of change in residential and non-residential land-use throughout the model iterations. In the first iteration for land-use in 2006, non-residential areas decreased by 22 % and new low-density neighborhoods appeared. After 2016, the proportion of low-density area declines sharply by 12 %, while medium-to-high-density area increases by 19 %. By 2021, there is an increase in high-density area by 67 %.

After 2016, the BNAS model simulates a significant change in low-density areas as they decrease and medium-to-high-density neighborhoods increase. There is an increase in high-density areas in the study area at the end of simulation year 2021. In addition, Fig. 10 shows that the BNAS model generates compact growth patterns as it generates high-density areas.

4 Conclusion

The BNAS model incorporates actor behavior in locational decision-making for the model transition rules using a novel approach combining BNs, IDs, GIS, and an ABM for predicting future population density and land-use change. The results showed that BNs can handle reasoning under uncertainty and therefore can mimic human reasoning. With the BNs and IDs, the agent easily converts its beliefs into actions.

The BNAS model couples BNs with agent-based approach to represent complex spatial relationships between multitudes of land-use drivers that affect land-use and population density changes. The relationships between household socioeconomic drivers and locational drivers were determined by the BNAS model through specific BN structures for different household (agent) types. These networks were extracted from real-world data by using special learning algorithms of BNs in which variables are linked together to represent dependencies between associated conditional probabilities. The results illustrate that BN learning algorithms provide simplicity in rule definition of the ABMs.

The agent decision-making was defined using causal relationships between land-use drivers. Using BNs in ABMs provides detailed knowledge about the behavior patterns of individuals, the important aspects of behavior, and the relative ease in finding correct values of the variables in the model. In addition, the BNAS model is capable of identifying the important variables and hence allows model designers to rapidly examine multiple drivers. Furthermore, the BNAS model is flexible in terms of choices of heterogeneous agents as different income groups can be easily integrated.

This study also combined an irregular spatial structure with BN-based ABMs since this offers a more realistic way of representing complex behaviors of urban land-use change. Consequently, this represents an object-oriented model design consisting of two objects: agents and irregular spatial units. Agents have attributes, which store their socioeconomic characteristics and their current location. Spatial units have location characteristics, such as distance to the town center, which are stored their attributes. As such, the BNAS design is an approach to employ census spatial units in ABMs.

Areal interpolation was used to integrate unmatched census data of the Metro Vancouver study area between different time steps. Using new derived spatial units allowed for the use of historical data to validate the model. Partial validation of the proposed modeling approach for the study area in terms of checking if the model captured the basic features of land-use change was accomplished successfully as a real urban data set were available. Please refer to Kocabas and Dragicevic (2009) for the details of the validation approach with respect to model simulation outcome comparisons.

The forecasting ability of the BNAS model is robust, and the outputs for the City of Surrey are consistent and realistic. Further the BNAS model has been tested as policy evaluator and generator as the part of the planning support system framework (Kocabas and Dragicevic 2012). However, systematic sensitivity analysis and in depth model testing would add further confidence in the model outputs once newer land-use and census data are released from official sources. Incorporating discrete choice modeling approaches or survey data as inputs to the BN would increase the demographic characteristics of the agent profile and is possible venue for model improvements.

In order to make model forecasts accurate, there is a need to integrate urban policy, population growth dynamics, and land-use change. The BNAS model addresses this need by facilitating the understanding and predicting of complex urban land-use change processes and can be used as a complementary decision support tool to guide policy makers, urban planners, and urban developers.

Acknowledgments The authors thank the Natural Sciences and Engineering Research Council (NSERC) and Social Sciences and Humanities Research Council (SSHRC) of Canada for financial support of this study. The Metro Vancouver and the Greater Vancouver Transportation Authority (Translink) provided some of the spatial data. The MATLAB software was made available by the Network Support Group, Faculty of Applied Sciences and Centre for Systems Science at Simon Fraser University. The authors are thankful to the journal Editor and the two anonymous reviewers for their valuable comments.

References

An L, Linderman M, Qi J, Shortridge A, Liu J (2005) Exploring complexity in a human-environment system: an agent-based spatial model for multidisciplinary and multiscale integration. Ann Assoc Am Geogr 95(1):54–79

Batty M (2005) Agents, cells, and cities: new representational models for simulating multiscale urban dynamics. Environ Plan A 37(8):1373–1394

Benenson I (2004) Agent-based modeling: From individual residential choice to urban residential dynamics. In: Goodchild MF, Janelle DG (eds) Spatially Integrated Social Science. Oxford University Press, New York, pp 67–95

Benenson I, Omer I, Hatna E (2002) Entity-based modeling of urban residential dynamics: the case of Yaffo, Tel Aviv. Environ Plan B 29(4):491–512

Bennett DA, Tang W, Wang S (2011) Toward an understanding of provenance in complex land use dynamics. J Land Use Sci 6(2–3):211–230

Bromley J, Jackson NA, Clymer OJ, Giacomello AM, Jensen FV (2005) The use of Hugin® to develop Bayesian Networks as an aid to integrated water resource planning. Environ Modell Softw 20(2):231–242

Brown DG, Xie Y (2006) Spatial agent-based modeling. Int J Geogr Inf Sci 20(9):941–943

Brown DG, Page SE, Riolo R, Rand W (2004) Agent-based and analytical modeling to evaluate the effectiveness of greenbelts. Environ Modell Softw 19(12):1097–1109

Brown DG, Page S, Riolo R, Zellner M, Rand W (2005a) Path dependence and the validation of agent-based spatial models of land use. Int J Geogr Inf Sci 19(2):153–174

Brown DG, Riolo R, Robinson DT, North M, Rand W (2005b) Spatial process and data models: toward integration of agent-based models and GIS. J Geogr Syst 7(1):25–47

Cheng J, Greiner R, Kelly J, Bell D, Liu WR (2002) Learning Bayesian networks from data: an information-theory based approach. Artif Intell 137(1–2):43–90

Collins RJ, Jefferson DR (1992) AntFarm: Towards simulated evolution. In: Langton CG (ed) Artificial life II : Proceedings of the Workshop on Artificial Life. Addison-Wesley, Redwood City, CA, pp 579–601

Cooper GF, Herskovits E (1992) A Bayesian Method for the Induction of Probabilistic Networks from Data. Mach Learn 9(4):309–347

de Almeida MC, Monteiro AMV, Soares GCBS, Cerqueira GC, Pennachin CL, Batty M (2005) GIS and remote sensing as tools for the simulation of urban land-use change. Int J Remote Sens 26(4):759–774

Dorling D (1993) Map design for census mapping. Cartogr J 30(2):167–183

ESRI (2012) GIS and Mapping software. http://www.esri.com. Accessed 20 May 2012

Evans TP, Sun W, Kelley H (2006) Spatially explicit experiments for the exploration of land-use decision-making dynamics. Int J Geogr Inf Sci 20(9):1013–1037

Ferber J (1998) Multi-agent systems: An introduction to distributed artificial intelligence. Addison-Wesley, Harlow

Gigerenzer G, Selten R (2001) Bounded rationality the adaptive toolbox. MIT Press, Cambridge, Mass, MIT cognet

Gilbert N, Terna P (2000) How to build and use agent-based models in social science. Mind Soc 1(1):57–72

Graniero PA, Robinson VB (2006) A probe mechanism to couple spatially explicit agents and landscape models in an integrated modeling framework. Int J Geogr Inf Sci 20(9):965–990

Heckerman D, Wellman MP (1995) Bayesian Networks. Commun ACM 38(3):27–30

Howard RA, Matheson JE (1984) Influence diagrams. In: Howard RA, Matheson JE (eds) Readings on the Principles and Applications of Decision Analysis, vol II., Strategic Decisions GroupMenlo Park, Calif, pp 719–762

Janssens D, Wets G, Brijs T, Vanhoof K, Arentze T, Timmermans H (2006) Integrating Bayesian networks and decision trees in a sequential rule-based transportation model. Eur J Oper Res 175(1):16–34

Jensen FV (2001) Bayesian networks and decision graphs. Statistics for engineering and information science. Springer, New York

Kocabas V, Dragicevic S (2006) Coupling Bayesian Networks with GIS-based cellular automata for modeling land use change. Lect Notes Comput Sci 4197:217–233

Kocabas V, Dragicevic S (2007) Enhancing a GIS cellular automata model of land use change: bayesian networks, influence diagrams and causality. Trans GIS 11(5):679–700

Kocabas V, Dragicevic S (2009) Agent-based model validation using Bayesian Networks and vector spatial data. Environ Plan B 36(5):787–801

Kocabas V, Dragicevic S (2012) Integration of a GIS-Bayesian Network agent-based model in a planning support system as framework for policy generation. URISA J 24(1):35–52

Lauritzen S, Spiegelhalter D (1988) Local computations with probabilities on graphical structures and their applications to expert systems. J Roy Stat Soc 50(2):157–224

Lei Z, Pijanowski BC, Alexandridis KT, Olson J (2005) Distributed modeling architecture of a multi-agent-based behavioral economic landscape (MABEL) model. Simul-T Soc Mod Sim 81(7):503–515

LeSage JP, Pace RK (2004) Arc_Mat, a toolbox for using ArcView shape files for spatial econometrics and statistics. Lect Notes Comput Sci 3234:179–190

Ligmann-Zielinska A, Jankowski P (2007) Agent-based models as laboratories for spatially explicit planning policies. Environ Plan B 34(2):316–335

Ligmann-Zielinska A, Sun L (2010) Applying time dependent variance-based global sensitivity analysis to represent the dynamics of an agent-based model of land use change. Int J Geogr Inf Sci 24(12):1829–1850

Ligtenberg A, Bregt AK, van Lammeren R (2001) Multi-actor-based land use modelling: spatial planning using agents. Landsc Urban Plan 56(1–2):21–33

Liu X, LeSage J (2009) Arc_mat: a Matlab-based spatial data analysis toolbox. J Geogr Syst 12(1):69–87

Loibl W, Toetzer T (2003) Modeling growth and densification processes in suburban regions–simulation of landscape transition with spatial agents. Environ Modell Softw 18(6):553–563

Ma L, Arentze T, Borgers A, Timmermans H (2004) Using Bayesian decision networks for knowledge representation under conditions of uncertainty in multi-agent land use simulation models. In: Leeuwen J, Timmermans HJP (eds) Recent Advances in Design and Decision Support Systems in Architecture and Urban Planning. Kluwer, Dordrecht; Boston, pp 129–144

Ma L, Arentze T, Borgers A, Timmermans H (2007) Modelling land-use decisions under conditions of uncertainty. Comput Environ Urban 31(4):461–476

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

Manson SM (2006) Bounded rationality in agent-based models: experiments with evolutionary programs. Int J Geogr Inf Sci 20(9):991–1012

Mas JF, Puig H, Palacio JL, Sosa-Lopez A (2004) Modelling deforestation using GIS and artificial neural networks. Environ Modell Softw 19(5):461–471

Mathworks (2012) Getting started with Matlab. http://www.mathworks.com/access/helpdesk/help/pdf_doc/matlab/getstart.pdf. Accessed 20 May 2012

MetroVancouver (2012a) Greater Vancouver regional district. http://www.metrovancouver.org/Pages/default.aspx. Accessed May 20 2012

MetroVancouver (2012b) Metro 2040 residential growth projections. http://www.metrovancouver.org/planning/development/strategy/RGSBackgroundersNew/RGSMetro2040ResidentialGrowth.pdf. Accessed 20 May 2012

Miller EJ, Douglas Hunt J, Abraham JE, Salvini PA (2004) Microsimulating urban systems. Comput Environ Urban 28(1–2):9–44

Murphy K (2001) The Bayes net toolbox for Matlab. In: Wegman EJ, Braverman A, Goodman A, Smyth P (eds) Computing science and statistics vol 33, 33rd Symposium on the interface, Costa Mesa, CA, 2001. Interface Foundation of North America, pp 331–350

Neapolitan RE (2003) Learning Bayesian Networks. Prentice Hall, Harlow

Parker DC, Berger T, Manson SM (2001) Agent-based models of land-use and land-cover change: Report and review of an international workshop. http://www.indiana.edu/~act/focus1/ABM_Report6.pdf. Accessed 20 May 2012

Parker DC, Manson SM, Janssen MA, Hoffmann MJ, Deadman P (2003) Multi-agent systems for the simulation of land-use and land-cover change: a review. Ann Assoc Am Geogr 93(2):314–337

Pearl J (1988) Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, San Mateo, CA, Morgan Kaufmann series in representation and reasoning

Robinson DT, Brown DG (2009) Evaluating the effects of land-use development policies on ex-urban forest cover: an integrated agent-based GIS approach. Int J Geogr Inf Sci 23(9):1211–1232

Robinson D, Brown DG, Parker DC, Schreinemachers P, Janssen MA, Huigen M, Wittmer H, Gotts N, Promburom P, Irwin E, Berger T, Gatzweiler F, Barnaud C (2007) Comparison of Empirical Methods for Building Agent-Based Models in Land Use Science. J Land Use Sci 2(1):31–55

Rodrigues A, Grueau C, Raper J, Neves N (1998) Environmental planning using spatial agents. In: Carver S (ed) Innovations in GIS 5. Taylor & Francis, London, UK, pp 108–118

Smajgl A, Brown DG, Valbuena D, Huigen MGA (2011) Empirical characterisation of agent behaviours in socio-ecological systems. Environ Modell Softw 26(7):837–844

Statistics-Canada (2012) Canada's national statistical agency. http://www.statcan.ca. Accessed 20 May 2012

Tobler WR (1984) Application of image processing techniques to map processing. In: Proceedings of the International Symposium on Spatial Data Handling, Zurich, IGU, pp 140–144

Torrens PM (2006) Simulating sprawl. Ann Assoc Am Geogr 96(2):248–275

Torrens PM, Benenson I (2005) Geographic automata systems. Int J Geogr Inf Sci 19(4):385–412

Torrens PM, Nara A (2007) Modeling gentrification dynamics: a hybrid approach. Comput Environ Urban 31(3):337–361

Uusitalo L (2007) Advantages and challenges of Bayesian networks in environmental modelling. Ecol Model 203(3–4):312–318

Waddell P (2002) UrbanSim—Modeling urban development for land use, transportation, and environmental planning. J Am Plan Assoc 68(3):297–314

Wright JK (1936) A method of mapping densities of population: with Cape Cos as an example. Geogr Rev 26(1):103–110

Wu S, Wang L, Qiu X (2008) Incorporating GIS building data and census housing statistics for sub-block-level population estimation. Prof Geogr 60(1):121–135