PHILIPPINE ENGINEERING JOURNAL
PEJ 2008; Vol. 29, No. 2: 57-78

MODELING LAND USE CHANGE: USING KNOWLEDGE DISCOVERY AND CELLULAR AUTOMATA IN A GIS ENVIRONMENT

Florencio Ballesteros, Jr. and Zeyuan Qui
Department of Chemical Engineering
College of Engineering, University of the Philippines Diliman
Quezon City, Philippines

ABSTRACT

We develop a parcel-based spatial land use change prediction model by coupling machine learning and interpretation algorithms such as cellular automata and decision tree in a Geographic Information System environment. We collect and process historical land use data and various driving factors that affect land use changes in Hunterdon County of New Jersey using decision tree J48 Classifier to develop a set of transition rules that illustrate the land use change processes during the period 1986-1995. Then we apply the derived transition rules to the 1995 land use data in a cellular automata model Agent Analyst® to predict the spatial land use pattern in 2004. We validate these by the actual land use in 2002. The developed decision tree-based cellular automata model has a reasonable overall accuracy of 84.46 percent in predicting land use changes. It shows a much higher capability in predicting quantitative changes (92.5%) than location changes (74.8%) in land use. With such an encouraging measure of validity, we use the model to simulate the 2011 land use patterns in Hunterdon County based on the actual land uses in 2002. We build two scenarios: the "business as usual" scenario and the "policy" scenario (with imposed government policy). The simulation results show that successfully implementing current land use policies such as down-zoning, open space, and farmland preservation could prevent 973 agricultural and 870 forest parcels (a total of 2,856 hectares) from future urban encroachment in Hunterdon County during the period 2002-2011. It becomes a significant policy instrument for government to reckon with.

KEY WORDS: land use change, cellular automata, decision tree, parcel, geographic information system, J48 Classifier, Agent Analyst, Hunterdon County

I. INTRODUCTION

Land use in the United States and many other parts of the world has been experiencing rampant changes over the last several decades because of the impact of social and economic changes in society (Ojima, Galvin, & Turner, 1994; Irwin & Geoghegan, 2001). In turn, land use changes bring about significant social and economic changes, often seriously affecting human health and the natural environment. Urban sprawl provides a perfect example of this phenomenon. Wealth, growth, and dependence on automobiles in traveling across great distances result in the spreading of urban development into adjacent rural areas, blurring rural-urban interface; this is called urban sprawl (Cieslewicz, 2002). Urban sprawl has been linked to health and environmental hazards such as water and air quality deterioration, congestion, increased risks for cardiovascular diseases and stroke, obesity and other diseases associated with being overweight (Ewing, et al., 2008; Patz, Campbell-Lendrum, et al., 2005; Jackson, 2003; Frumkin, 2002).

Correspondence to: Department of Chemical Engineering, University of the Philippines Diliman, Quezon City PHILIPPINES 1101; Email address: balleteros@gmail.com

Received: May 5, 2009
Revised: June 15, 2011
Accepted: June 28, 2011

Due to the serious effects of urban sprawl, substantial research has been conducted to predict how land uses and the interfaces between urban and rural (and/or between high intensive agriculture and non-commercial farms) will shift and change over time so that these impacts can be anticipated and addressed appropriately through policy intervention (Briassoulis, 2000; Veldkamp and Verburg, 2004). Since such changes result from the interplay of complex socioeconomic and biophysical processes, they are virtually impossible to duplicate through experiments or analysis by empirical observation (Baker, 1989; Walker, 2003; Veldkamp and Verburg, 2004). One approach to solve this difficulty is through land use change modeling (Baker, 1989; Briassoulis 2000; Chen et al., 2002). Appropriately calibrated computer-based models can provide a systematic and accurate way to predict future land use changes (Verburg, Schot, et al., 2004). Aside from forecasting, models can be used to explore land use system response to policy interventions through "what-if" scenarios. Predictions of land use changes based on scenario analysis are frequently used in the context of policy making affecting local and regional issues (Klostermann, 1999; van Ittersum, Rabbinge, & van Latesteijn, 1998). At a larger, more far-reaching scale, this method can have vital use in the pressing issues that come with global climate change (Riebsame et al., 1994; Pielke et al., 1999; Kalnay, 2003; Reid et al., 2004; Ramankutty et al., 2002; Salmun and Molod, 2006).

Land use change modeling has reached sophistication because of substantial advances in spatial science and technology. Verburg, Schot, et al. (2004) reviewed the theory, rationale, and implementation of past land use models and identified four research priorities for future land use models, which are also echoed by others in this field of research. First, a future modeling approach should better address the multi-scale characteristics of land use systems including scale dependency, scale-up effects, and interactions of processes operating at different scales (Briassoulis, 2000). Second, the future modeling approach should be an integration of multiple disciplines in methodology, evaluation of the impacts caused by land use changes, and analysis of urban-rural interaction (Lambin et al, 2000). Along this line, Geoghegan, et al. (1998) called for land use change models that can integrate the social sciences with geographic and ecological models - "socializing the pixel and pixelizing the social". Third, the future modeling approach should pay explicit attention to the temporal dimension of land use change, influence of non-linear pathways of change, feedbacks, and time-lags. Fourth, more sophisticated methods should be developed to avoid subjectivity bias from the experts, and to better assess and quantify neighborhood effects in land use change models (Torrens, 2000; Verburg, de Nijs, et al., 2004;).

Cellular automata (CA) has emerged as an effective tool that addresses these concerns in modeling land use changes at both regional and municipal levels (Landis, 1994; Engelen, et al., 1995; White, et al., 1997; Clarke and Gaydos, 1998; Batty, Xie et al., 1999; White and Engelen, 2000; Li and Yeh, 2000; Yeh and Li, 2001; Wu, 2003; Cheng and Masser, 2004). (Detailed discussion on CA below). Most land use maps are prepared through the conventional per-pixel land use classification of satellite images using the spectral signature of each pixel. The use of the regular pixel grids makes these models easy to integrate with raster Geographic Information Systems (GIS). Although regular grids are technically convenient in CA-based land use change models, cadastral parcels are the most ideal spatial unit of analysis in land use change modeling (Landis and Zhang, 1998a, 1998b; Irwin and Geoghegan, 2001; Allen and Lu, 2003). This is because stakeholder behaviors such as purchasing, selling, and developing land are made and observed at the parcel level. It is also at the parcel level that most land use policies such as zoning are crafted and implemented. Furthermore, parcels contain socio-economic information through the municipal tax assessment database – information which can be used in the evaluation of land use changes (Wu et al, 2007).

Stevens et al. (2007) developed iCity, a CA-based land use change model that utilizes irregularly shaped parcels as the unit of analysis to evaluate urban development in Saskatoon, Saskatchewan, Canada. But there are limitations in the use of iCity. First, the model allows only two classification types for parcels: urban and non-urban uses. Second, the weights assigned to various influencing factors to derive land use conversion rules, are dependent on expert knowledge, and as such, prone to subjectivity bias as discussed by Verburg, Schot et al. (2004).

This study aims to extend the capability of the parcel-based iCity model developed by Stevens et al. (2007) to predict land use changes. First, it uses the more realistic and finer land use classification types instead of a simple, dichotomous, urban/non-urban classification scheme. Second, it addresses complexity in defining driving factors and neighborhood effects when using parcels as the unit of analysis and applies an innovative data-mining scheme called decision tree (DT) to elicit land use transition rules. DT is a machine learning and interpretation algorithm which therefore avoids the subjectivity bias that usually becomes manifest when expert knowledge is used (Li & Yeh, 2004). The machine learning and interpretation approach for deriving transition rules does not require extensive quantitative skills and can be better appreciated by non-technical users such as stakeholders and land use change decision-makers. This study implements the coupled DT and CA-based land use change model through a GIS-based Agent Analyst model Recursive Porous Agent Simulation Toolkit (RePAST) (North et al., 2005). This study applies the resulting innovative land use change model to Hunterdon County, New Jersey, where dramatic land use changes have taken place during last 3 decades.

II. METHODS: MODEL DEVELOPMENT

CA is a collection of cells that evolves through a number of discrete time steps according to a set of rules based on the states of neighboring cells (Chopard and Droz, 1998). CA has five principal elements, namely - cell state, lattice, neighborhood, time, and transition rule (Chopard and Droz, 1998). Cell state represents one of finite cell conditions. Lattice refers to the space or a group of cells in which the CA exists and evolves over time. The neighborhood comprises the localized region of a CA lattice. Transition rules specify how cells change from one state to another based on the cell's state and neighborhood conditions (For a thorough discussion of CA, the reader is referred to Toffoli and Margolus, 1987).

The introduction of CA in geographical applications was catalyzed by Tobler (1979) who first recognized the advantages of CA in solving geographical problems. Following Tobler's pioneering work, other researchers applied CA to urban land use change modeling (White and Engelen, 1993; Batty and Xie, 1994; Clarke and Gaydos, 1998; Wu and Webster, 1998). Several noticeable examples of CA land use models are the Research Institute for Knowledge Systems (RIKS) of the University of Maastricht, a decision support tool for the management of river basins, coastal zones, cities and regions (Engelen et al., 1995), and an urban land use change prediction model with the input requirements: Slope, Land use map, Excluded areas from urbanization, Urban areas, Transportation map, and Hill shading area (SLEUTH) (Clarke et al., 1997). While most land use change models use regression models to elicit the transition rule, CA-based land use change models introduce new methods to elicit the transition rules. Wu (1996) introduced the use of fuzzy logic to simulate different scenarios that resulted from the implementation of different urban development policies. Li and Yeh (2002) and Pijanowski et al. (2002) used artificial neural networks to elicit or "learn" patterns from driving factors of land use change and calculate a conversion probability for a given cell based on those factors. There are also some CA-based land use models that incorporate stochastic land use change processes by developing stochastic transition rules (Ward et al., 2000; de Kok et al., 2001; and Guan et al., 2005).

In this study we aim to develop a CA-based land use change model using a parcel-level data to predict future land use changes and to apply the model to Hunterdon County, New Jersey. In this application, we represent cells by parcels; these parcels comprise the county which serves as the lattice. We consider the following driving factors: present land use type, percentages of land use types in the neighborhood, distances to the nearest urban center, to major roads, and to streams, soil suitability for development, slope, parcel size, and wetland area within a parcel if any exists. The details regarding the cell states, neighborhood, and driving factors that affect the cell states are discussed in the next section on Study Area and Data Development.

This paper begins with a discussion of two modules that implement the land use change prediction model: a DT module that generates the transition rule from the driving factors, and a CA module that predicts future land use change using the derived transition rules. It then proceeds with a discussion of the methods used to evaluate the model's accuracy.

Decision Tree: J48

A transition rule defines how driving factors determine the evolution of a cell from one state to another. DT is used to generate the transition rules for the land use change model. In geographical studies, DT is a data mining technique for classification (Moore et al., 1991; Meyer et al., 2001; Speybroeck et al., 2004, Wu et al., 2007). A DT structure entails a series of yes/no questions in which the sequence of the questions depends on the answers given in the previous question. Applied to land use/cover classification, the specific questions take on values equivalent to land attributes. Its sequence eventually determines the appropriate land use/cover classification (Aalders and Aitkenhead, 2006). DT has been used to elicit the transition rules to predict land use changes (McDonald and Urban, 2006; Liu et al., 2007).

In the present model we use the DT algorithm J48, a WEKA implementation of the latest public release (Version 8) of C4.5, a standard decision tree algorithm that is widely used for practical machine learning (Witten and Frank, 1999). J48 was developed by the Machine Learning Group of the University of Waikato, New Zealand. The chief function of J48 is the classification of a sample dataset whose data instances have a set of attributes. The J48 algorithm operates by recursively splitting sample data input based on attribute values to produce a tree that preferably generates just one branch. The first attribute to be chosen by this algorithm is designated as the root of the tree. The instances in the training sample are then split among branches based on their attribute values. If the values are continuous, then each branch takes a certain range of values; otherwise a new attribute feature (node) is then chosen and the process is repeated for the remaining instances. The process stops at a terminal node when the classification of a branch is pure (i.e., it contains only instances of a certain class). As to what attribute will be used for a given split, the choice is based on the attribute with the largest value for information gain (Quinlan, 1996; Goodman and Smyth, 1988). The final decision tree generated by J48 contains various paths from the root to the terminal node. Each path from the root of the tree down to the terminal node where a classification is shown can be translated into a transition rule; this is subsequently used by the CA module.

Creating transition rules through DT is more helpful than using statistical regressions in cases where: (1) there is a large number of variables to predict land use changes (Pal and Mather, 2002 and 2003; Speybroeck et al., 2004); (2) non-linear relationships exist between variables in the data (Razi and Athappilly, 2005); and (3) the underlying relationship between dependent and independent variables is not known (Pal and Mather, 2002). The ANN approach has similar advantages over statistical regression methods; however, it is not intuitive to policy makers and land use planners because of its black box nature. DT is a white box model and is more easily interpreted compared to a neural network wherein the derivation of the results is not explained or readily available (Breiman et al., 1984; Quinlan, 1996; Li and Yeh, 2004). Decision tree processing also entails shorter processing/computing time (Perner et al., 2001; Pal and Mather, 2002; Razi and Athappily, 2005).

Cellular Automata: Recursive Porous Agent Simulation Toolkit (RePAST)

A CA module subsequently uses the transition rules derived using J48 to evaluate how land parcels are converted from their current land uses to their future land uses. Here we choose Agent Analyst/RePAST to implement the CA process. RePAST was developed by a group of researchers from the University of Chicago and Argonne National Laboratory (North et al., 2005). Through Agent Analyst, users can create, edit, and run RePAST model within the ArcGIS® environment (Groff, 2007). This graphical user interface allows the modeler to create agents, schedule simulations, visualize the ArcGIS® layers, and specify the behavior and interactions of the agents. Aside from having the spatial analysis capabilities of ArcGIS, Agent Analyst/RePAST has two outstanding features relevant to this study. First, the model has provisions that allow the modification of agent properties, agent behavioral equations, and model properties during run time. Second, it has libraries for genetic algorithms and neural networks, including the ability to handle irregular grids or vector data as a model component.
Utilization of the transition rules in RePAST is implemented as the step function of the Agent Analyst model. First, RePAST reads from an array of all spatial data represented in the transition rule and stores the attributes into a featureList. Subsequently, an instance of LTCMain is created and assigned to the variable named ltc. The classify method of the ltc variable is then invoked with featureList as an input parameter. In this manner, the program feeds into DT the attributes placed inside featureList. After processing, DT will then make a prediction (and thus return a value) on the next land usage based on the attribute values stored in featureList. The value returned by the method is then stored as the new land use type.

The transition rules embedded in the CA model can also include government regulatory policies on land use changes. Examples of these policies are regulations or limitations on the conversion of agricultural land to developed land, steering of urban developments to sites where the soils are considered low value for agriculture, or designation of zoning laws. Following the calibration of the CA land use change model, we then use the parameter estimates to simulate future growth patterns arising from the implementation of planned land use policies.

Accuracy Assessment

The assessment of land use prediction accuracy relies on the use of a confusion matrix, which is simply a cross tabulation of the predicted land use classes in a prediction map against the same land use classes in a reference map. Suppose there are K classes of land uses. Information in the rows correspond to the land use classes of the reference map, while those in the columns show the land uses predicted from the land use change model. The element in the confusion matrix, n_{ij}, where i, j = 1, ..., K, represents the number of instances in reference class i that are predicted to class j. The diagonal element of the matrix, where i = j, represents correct predictions. Overall accuracy is calculated as the sum of the diagonal elements divided by the total instances. Although overall accuracy is useful, it does not give much information about the accuracy of the individual land use classes predicted; these are usually evaluated by producer's accuracy, user's accuracy, errors of commission, and errors of omission which show how well each of the classes in the reference map are predicted.

The overall prediction accuracy using the error matrix is recognized as an overestimation since the method used to calculate it does not account for agreements that would have occurred by chance (Lillesand and Kiefer, 2000). Another way to assess prediction accuracy is to use the Cohen's Kappa Index; it is a measure that takes into account overestimates of the computed percentage of correct values caused by agreements made by pure chance (Cohen, 1960; Foody, 1992; Monserud and Leemans, 1992; Pontius and Cheuk, 2006). This index ranges between 0 and 1 and is interpreted as the proportionate reduction in error achieved by the model being evaluated as compared with the error of a completely random prediction model.

Pontius (2000) argued that Cohen's Kappa Index also has limitations in assessing prediction accuracy. Specifically, the index does not give information about location and quantification errors. Quantification error occurs when the number of parcels or cells for a given land use type in the predicted map is different from the reference map. Location error occurs when the predicted land use type of a given parcel is different from that in the reference map. Pontius (2000) further derived two variants of Cohen's Kappa Index, namely: $K_{location}$, and $K_{quantity}$, that measure the accuracy in predicting location and quantity, respectively.

With these considerations, the overall accuracy, error of commission, error of omission, we use Kappa Index and its two variants to evaluate the performance of the land use change prediction model.

III. STUDY AREA AND DATA DEVELOPMENT

We applied the land use change prediction model developed in this study to Hunterdon County, New Jersey. Hunterdon County is one of 21 counties in New Jersey. It encompasses 1134 square kilometers of the western portion of the State. It ranks eighth among New Jersey's counties in terms of land area and has 26 municipalities classified into townships, boroughs, and towns. As shown in Figure 1, the County is traversed from east to west by the I-78 interstate highway designed to carry traffic between regions of the state and to serve as a corridor between Port Newark/Liberty Airport and other points westward. Accessibility between municipalities and adjoining counties is provided by a network of county and municipal roads that include Routes 12, 31, 202, and 517.

Hunterdon County is home to approximately 129,000 people (NJDLWD, 2006). Its population grew by 87 percent between 1970 and 2004 making it the third fastest growing county in New Jersey. Its economic growth is boosted by its proximity to high growth areas in the state where firms like Exxon, Foster Wheeler, and Merck established their corporate offices there during the 1980s and 1990s. Hunterdon is one of six counties situated in an extensive growth area known as "the wealth belt" characterized by high property values, high population, plenty of jobs and high personal income" (Hughes and Seneca, 2006). Such a trend is accelerated by the "ratables chase" policy in New Jersey that encourages local governments to permit more development to maintain the low property tax rate and to finance their public service requirements (e.g. sewer, solid waste collection, etc) (HCPB, 2007).

Despite being in the "wealth belt" area, Hunterdon County is still considered to be a mostly rural and suburban county. To maintain its rural and agricultural character, some communities in the county implemented the large-lot zoning that requires at least 0.8-, 2- and 4-hectare (two-, five- or ten-acre) lots for residential homes. Various land use policies have also been implemented in the county to restrict land use development toward smart growth and environmental protection. Notable examples are: Open space preservation, farmland preservation, and purchase of development rights. A group of 13 townships, towns, and boroughs in Hunterdon County are partly or completely within the Highlands Preservation Area. Future land use development in this group of 13 will be subject to much more stringent restrictions enforced by the New Jersey Highlands Water Protection and Planning Act (NJDEP, 2005).

In applying the land use change prediction model to Hunterdon County, we used three sets of land use/cover GIS layers of the county corresponding to the years 1986, 1995 and 2002 which we downloaded from the NJDEP website. These land use/cover data were compiled from aerial photography and Landsat satellite images; each of them was classified into 6 categories as agriculture, barren lands, forest, urban, wetlands and water modified from the Anderson Classification System (Rutgers CRSSA, 2004). The model uses all six land use classes. We obtained the land parcel GIS layer in the county from the Office of GIS, Hunterdon County, New Jersey. We likewise obtained other spatial data such as the digital elevation model (DEM), soil, streams, protected open space and preserved farmland, census, and location of urban centers and roads, streets and major highways. All these data sets in digital format are available in the NJDEP Bureau of GIS and/or the Hunterdon County Office of GIS.

Allen and Lu (2003) suggested three criteria in selecting an appropriate set of driving factors for modeling land use change. First, the factors must include all physical, economic, demographic, and social factors that affect all types of land use change. Second, they must have spatial attributes. Third, they must reflect the properties and characteristics of the parcel. Jiao and Boerboom (2003) grouped various driving forces into five categories namely: neighborhood, accessibility, suitability, policy, and socio-economic factors. Following those principles, this study considers the following driving factors: parcel size, the land use of the parcel, the distribution of land uses in the neighborhood of a parcel in terms of percentages, the area of wetlands within a parcel, the distances to the nearest streams, roads, and urban centers, slope, and the number of soil restrictions for urban development. By overlaying all the spatial data over the parcel map, we generated a parcel-based data on the driving factors that affect land uses. While the methods for deriving some of these driving factors are straightforward, there are difficulties in assigning a single land use to a parcel and defining the neighborhood of a parcel. These are discussed exclusively below.

Development of Parcel-based Land Uses

The land use change prediction model requires a single land use type assigned to each land parcel. However, after overlaying the land use and the land parcel layers, a parcel may contain multiple land user/covers. Hence, we developed a classification to assign a single land use to a parcel to develop the parcel-based land uses in the county for 1986, 1995, and 2002. This scheme requires the conduct of initial tests on each parcel for any agricultural land present. If the agricultural land is over 45 percent in a parcel, we classify the parcel as agricultural lands. This threshold of 45 percent is based on the percentage of agricultural land in all parcels in 2002, and has a mean of 19 percent with a standard deviation of 26 percent.

For a parcel with less than 45 percent of agricultural land, we compare the urban area in the parcel to a threshold value of 0.5 hectares (1.2 acres). This represents a typical house footprint in the region including the area occupied by house, driveway, patio, pool, etc. (NJWSA, 2003). Thus, we classify a parcel as urban if its urban area is greater than 0.5 hectares and is located in a residential, commercial or industrial zone and the parcel size is less than 4 hectares (10 acres). If the urban area of a parcel is less than 0.5 hectares, but represents more than 45 percent of the parcel area, we also classify the parcel as urban. For a parcel that fails to be classified as agriculture and urban, we determine its final designation by the dominant land uses in the parcel.

Definition of the Neighborhood

The usual neighborhood configurations of CA models use the von Neumann or the Moore patterns that assume a lattice composed of regularly shaped cells or grids. Since the land parcels used as the unit of analysis in this study are irregularly shaped, we have to define an alternative neighborhood configuration accordingly. In the parcel-based CA model on urban sprawl iCity, Stevens et al. (2007) used three neighborhood configurations to estimate attractiveness scores for residential land parcels and transition rules for commercial and industrial parcels. The three types of neighborhood configurations are: (A) an adjacency neighborhood that includes all parcels having a common edge with the central parcel; (B) a distance neighborhood that includes parcels that fall completely or partially within a certain distance of buffer from the edge of the central parcel; and (C) a clipped distance neighborhood that includes all parcels that fall completely and portions of the parcels that are partially within a certain distance of buffer from the edge of the central parcel.

In this study, we define the neighborhood of a parcel by an external buffer with a thickness of 145 meters around the edge of a parcel. We base this on the average size of all parcels in Hunterdon County. This neighborhood is very similar to but different from the iCity configuration (C) as the central parcel itself is excluded.

We overlaid the original land use GIS layers with the buffer to identify the percentages of different land use types within the buffer, which we consequently use as the driving factors in the land use change prediction model to determine the future land use of the parcel. We wrote a script in ArcView Avenue® Scripting Language to define the neighborhood and to calculate the percentages of land use types within each neighborhood.

Figure 2 illustrates the steps for applying the land use change prediction model in Hunterdon County, New Jersey. We derived transition rules using 1986 and 1995 re-classified parcel-based land use data and the discussed driving factors in the DT module. We then used the derived transition rules in the CA module to predict land use changes from 1995 to 2004 using the parcel based land use data in 1995. We validated the model by comparing the predicted land use pattern in 2004 to the parcel-based land use data for 2002. Since no land use map for 2004 was available, we used a 2002 map as reference. By using the 2002 land use data as a reference to validate the prediction for 2004, we assumed that changes within this two-year period would not be considerable. We used the array of accuracy measurements discussed previously to compare the predicted land use in 2004 to the actual land use in 2002 and to evaluate the performance of the model.

We used the validated model to predict the land use changes from 2002 to 2011 based on the land use data of 2002. We modeled two scenarios: A baseline scenario, which is a "business as usual" situation wherein policy interventions by the government are not included in the modeling process and another scenario that incorporates government policy interventions such as down-zoning and the preservation of farmlands and open space. With the down-zoning reflected in the 2001 NJDEP Water Quality Management Planning Rules, the lot size for residential development in non-sewered areas will have to be greater than 1.3 hectares. The model implements down-zoning and the two other land preservation policies by adopting a spatial constraints approach (Swenson and Franklin 2000; Schneider and Pontius 2001). CA uses this approach by implementing a set of transition rules that prohibits the conversion of the parcels that are classified as preserved farmland, or open space. Likewise, CA does not allow the conversion to urban lands of areas that do not conform to development requirements (i.e., greater than 1.3 hectares).

IV. RESULTS

Transition Rules Derived from 1986 to 1995 Land Uses

We converted the DT generated from J48 to the transition rules in the subsequent CA modeling. Here are some significant transition rules. The first split in the decision tree is current land use type. Although most urban parcels remain urban, they could be converted to other uses such as agriculture and forest. This occurs when the parcels are large, contain a big portion of wetlands, have three or more soil restrictions for urban development, have lower percentage of urban and barren land, have higher percentage of water, agricultural, forest, wetlands in the neighborhood, and are farther away from highways and urban centers with steeper slope.

Agricultural parcels could be converted into urban, forest, barren and wetlands depending primarily on their neighborhood land use distribution and parcel size. Small agricultural parcels tend to get converted into urban uses when there is high percentage of urban land in their neighborhood. Conversion from agriculture to forest can occur to parcels with steep slopes where there are severely restricted soils for development and when the percentage of forest in their neighborhoods is high. Agricultural parcels with a significant amount of barren land in their neighborhood have the potential of becoming barren. Conversion to wetlands usually occurs in large agricultural parcels that have significant wetlands.

Forest parcels with a high percentage of urban land and a low percentage of barren land in their neighborhood are usually converted to urban use. This type of conversion tends to occur in the case of small forest parcels. Forest parcels can be converted to agriculture or wetlands when there is a significant presence of agricultural land already in their neighborhood or wetlands within these parcels.

The actual amount of wetlands in a wetland parcel usually determines its future status; parcels with a large amount of wetlands within it always remain as wetlands. However, wetland parcels with smaller amounts of wetlands have higher likelihood of being converted into urban in a high urban neighborhood, or they become barren lands if the parcel displays three restrictions to urban development. Barren parcels can be developed into urban lands or remain as barren. Parcels classified as water, or artificial and natural lakes usually stay as water.

We evaluated the accuracy of these transition rules by testing the derived transition rules against a testing dataset. For example, to predict the land use class in 1995, we fed all the attributes of an instance in the dataset (except the 1995 land use class) into the decision tree that consists of all the transition rules. We repeated this process for all the instances in the dataset. We computed accuracy by dividing the total number of correctly predicted instances by the total number of instances in the dataset. In predicting the 1995 land uses in from the 1986 land uses, we obtained an accuracy of 81.4 percent when two-thirds of the randomly selected land parcels in the county was used. The accuracy increased to 85 percent when we applied the bootstrap method to the remaining one-third of parcels in that period.

Predicted Land Use Changes for the Period 1995-2004

Table 1 presents the predicted land use changes during the period 1995-2004. The model predicted that the urban areas would increase from 17,878 hectares (33,866 parcels) in 1995 to 32,470 hectares (39,386 parcels) in 2004 while the area of other land uses would decrease (except for water which remains the same). Forest conversion is predicted to be the biggest contributor to urban development in that period, followed by agriculture and wetlands. The total forest lost to urban development is forecasted at 10,018 hectares (3,306 parcels). 2,440 hectares (1,434 parcels) of agricultural lands would be given away to urban development. There are 2,078 hectares (739 parcels) of wetlands lost to urban development, which is quite significant since this amount represents almost a third of the county's wetlands in 1995. Forest loss would partially offset the reforestation of 144 hectares from wetlands (35 parcels), agriculture (24 parcels), and urban parcels (1 parcel).

Model Validation Using the Actual Land Uses in 2002

Table 2 presents the confusion matrix in terms of the number of parcels (the upper panel) and of the area (the lower panel) using the predicted land use distribution for 2004 and the actual land uses in 2002. As shown in the upper panel of the table, the overall prediction accuracy in terms of the number of parcels, computed as the sum of the agreements in the diagonals divided by the total number of parcels, is 84.46 percent. Similarly, the overall prediction accuracy is 80.92 percent in terms of the total acreage as shown in the lower panel of Table 2. These measurements are comparable to the values reported in the literature. Li and Yeh (2004) reported an overall accuracy of 82 percent using a DT-based CA for predicting land use change in an urbanizing city in Southern China. Allen and Lu (2003) developed a multinomial logistic land use change model with the parcel as the unit of analysis and achieved an overall accuracy of 80.76 percent in terms of number of parcels.

Table 2 also shows the error of omission and the error of commission. Omission error varies across land use categories. There are large errors of omission for barren, wetlands and water, but they account for a portion of the county. On the other hand, urban and agriculture, two major land use categories, have low errors of omission. Similar observations are also found for the errors of commission. Table 2 also shows that agricultural and urban lands are consistently underpredicted while water and wetlands are overpredicted in terms of both the total numbers of parcels and acreage.

We evaluated the agreement between the predicted land use distribution for 2004, and the actual land uses in 2002 by Cohen's Kappa Index and its variants to account for agreements by pure chance. The calculated Kappa Index of 0.644 indicates that the two patterns are in a moderate agreement based on Congalton (2001) and Landis and Koch (1977).

We calculated two variants of the standard Cohen's Kappa Index to evaluate the agreement between the predicted land use distribution for 2004 and the actual land uses in 2002: Klocation and Kquantity following Pontius (2000). Klocation is equal to 0.748, which indicates that the model has good capacity to specify location correctly. Kquantity is equal to 0.925, which indicates the model has excellent capacity to specify quantity correctly. These numbers also suggest that the model is better at predicting quantitative changes than location changes in Hunterdon County.

Table 3 compares the sizes of predicted and actual parcels that convert to other uses. The model underpredicts the sizes of those reforested parcels. The actual reforested parcels are generally larger than predicted, with greater standard deviations for both agricultural and wetland parcels. The actual parcels converted to urban use are smaller than those predicted by the model. However, the differences between the average sizes of predicted and actual converted parcels are much smaller; they range from 0.35 hectares for agricultural parcels to 1.92 hectares for forest parcels. As for the urban parcels converted to barren lands, the sizes of parcels that actually changed are larger than those predicted by the model.

Comparison of Two Future Land Use Change Scenarios

Table 4 presents a land use change pattern from 2002 to 2011 under the baseline scenario. There are a total of 3,435 hectares (2,361 parcels) of non-urban lands converted to urban uses which constitute 95 percent of all converted parcels during the 2002-2011 period. In this scenario, urbanization comes primarily at the expense of agriculture and forest land, each contributing 1,701 hectares (1,027 parcels) and 1,288 hectares (896 parcels) respectively. During the projection period, wetlands suffer an additional loss of 375 hectares (119 parcels), among which 303 hectares (99 parcels) are converted to urban and 72 hectares (20 parcels) to forest land. The total number of forest parcels nevertheless declines even if 142 hectares (74 parcels) become reforested, among which, 69 hectares (53 parcels) come from agriculture and 72 hectares (20 parcels) come from wetlands. There are no other land use parcels that are converted into agriculture and wetlands.

Table 5 presents the predicted land use change pattern during 2002-2011 under the policy scenario. We consider three land use policies in the policy scenario, namely: down-zoning, preservation of farmlands, and preservation of open spaces. Results from the policy scenario indicate that successful implementation of these land use policies could slow down the process of urbanization. Under the policy scenario, only 474 hectares (54 parcels) of agricultural lands and 61 hectares (26 parcels) of forest are converted to urban use. Compared to the baseline scenario as previously discussed, the policy scenario could protect a total of 2,856 hectares of non-urban lands from urban development. The protected non-urban lands include 300 hectares of wetlands, 1,229 hectares of agricultural lands, 101 hectares of barren, and 1,226 hectares of forest. A total of 1,097 agricultural parcels are converted to other uses in the baseline scenario whereas only 124 parcels are converted to urban (54 parcels), forest (53 parcels) and barren land (17 parcels) in the policy scenario. As for forest parcels, 898 parcels are converted to other land use types in the baseline scenario whereas only 28 parcels are converted in the policy scenario.

There are four likely outcomes when comparing the converted and not converted parcels under both the baseline and policy scenarios: (1) converted parcels predicted by both scenarios; (2) not converted parcels predicted by both scenarios; (3) converted parcels predicted by the baseline scenario but not by the policy scenario and (4) converted parcels predicted by the policy scenario but not by the baseline scenario. It is not surprising that the majority of parcels fall under the Case 2. Case 4 does not occur. Figure 3 presents the spatial distributions of the parcels in Cases 1 and 3. The parcels under Case 1, i.e. those converted under both scenarios, are primarily located in Lebanon and Franklin Townships. The parcels under Case 3 are mostly located on the western and southern portion of the county. Under Case 3, five farmland parcels were preserved, distributed among Raritan Township with three parcels and Delaware and East Amwell Townships having one parcel each. Readington Township has the most number of preserved open space at 8 parcels followed by Lebanon and Bethlehem townships with six and four parcels respectively. We can attribute this to strong local land use regulations that are in effect. Clinton, Raritan and Union Townships have a number of preserved farm parcels as well.

V. SUMMARY AND CONCLUSIONS

This study developed a DT-based CA model to predict future land use changes with parcel-level data. We validated the model in Hunterdon County, New Jersey using historical land use changes which we applied to predict future land use changes under two scenarios: baseline scenario ("business as usual") and policy scenario (imposed government policy). The model extends the classical raster-based CA model by defining the modeling space as a collection of irregularly shaped geographic objects (represented by land parcels) and defining the transition rules using a knowledge discovery algorithm DT. The model defines the neighborhood of each parcel as an external buffer around the boundary of the parcel, which is an improvement over the ways of defining the neighborhood in existing CA-based land use change models. A DT elicits land use patterns from a large set of driving factors and is free from the subjectivity biases often encountered in expert knowledge-based methods. The DT approach also offers the convenience of incorporating the land use policies such as down-zoning, open space and farmland preservation in simulating future land use changes.

The coupled DT-based CA model reasonably predicted the land use changes in Hunterdon County, New Jersey, where substantial land use changes have taken place in the last three decades. Using the historical land use changes during the period 1995-2002 as a reference, the model achieves an overall accuracy of 80.92 in terms of the total areas and of 84.46 percent in terms of the total number of land parcels. The Cohen's Kappa Index, the conventional statistics for comparing similarity of two spatial patterns, yields a 0.64. We calculated two variants of the Kappa Index to evaluate the model's ability to correctly predict location and quantity; they are 0.748 and 0.925, respectively. Such results indicate that the model has a higher capability to predict quantitative changes than location changes in land use.

Simulations of future land use scenarios using the coupled model we developed indicate that the current land use policies such as down-zoning, open space, and farmland preservation could successfully prevent 2,856 hectares of non-urban lands from future urban development in Hunterdon County during the period 2002-2011. This study defines the neighborhood of a parcel by a 145-meter buffer around the boundary of the parcel. A sensitivity analyses using the 55-meter and 221-meter buffers show that the definition of the neighborhood has no significant impact on the model's prediction accuracy.

The coupled model demonstrates the feasibility and effectiveness of using parcel-level data in land use change modeling. However, there are still challenges that need to be addressed in future land use change modeling that uses parcel-level data. First, the model assumes a single land use type for each parcel. Such is a challenging task to do over a land use map compiled from satellite images and/or aerial photography especially when the study area is too large for detailed field verification. As discussed previously, some land use classes are overestimated, while others are underestimated. The accuracy of assigning the correct land uses would have significant impacts on the overall accuracy of the modeling.

Although the transition rules on the derived parcel-based land uses achieve reasonable prediction accuracy, the overall accuracy can be further improved through better accuracy in assigning a single land use type to a parcel based on the current land use data; this can be derived from aerial and remote sensing imagery.

Second, the model assumes that the parcel boundary stays the same during the modeling process; this is not realistic. A parcel itself may evolve over time when it gets divided into several smaller parcels or consolidated with other parcels to form industrial estates as experienced in urban development. Incorporating the dynamic changes of the parcel boundaries may help explain the parcel size differences between the actual and predicted land use changes as presented previously. Future land use change models using the parcel-level data could address these challenges.

VI. ACKNOWLEDGEMENT

This research was completed as part of a Ph.D. dissertation by Florencio Ballesteros, Jr. at the New Jersey Institute of Technology where Zeyuan Qiu served as the dissertation advisor. The research was partially supported by a grant from the New Jersey Water Resources Research Institute. The authors gratefully acknowledge the assistance of the Bureau of Geographical Information System of the New Jersey Department of Environmental Protection and the Division of GIS of Hunterdon County for providing the spatial data and parcel maps used in the study. The critical comments from the Ph.D. dissertation committee members, Dr. Richard Lathrop of Rutgers University, Dr. Dan Van Abs of the New Jersey Highlands Council, and Dr. Maurie Cohen and Dr. Dan Watts of the New Jersey Institute of Technology led to the improvement of the quality of this research, and are thus, very much appreciated.

VII. REFERENCES

1. Alders, I. H., Aitkenhead, M. J., 2006. Agricultural census data and land use modeling. Computers, Environment and Urban Systems. 30(6), 799-814.
2. Allen, J., Lu, K., 2003. Modeling and prediction of future urban growth in the Charleston region of South Carolina: A GIS-based integrated approach. Conservation Ecology 8,(2), [online] URL: http://www.consecol.org/vol8/iss2/art2/.
3. Baker, W. L., 1989. A review of models of landscape change. Landscape Ecology, 2(2), 111-133.
4. Batty, M., Xie, Y., 1994. From cells to cities. Environment & Planning B: Planning & Design. 21(Celebration Issue), 531-548.
5. Batty, M., Xie, Y., Sun, Z.,1999. Modeling urban dynamics through GIS-based cellular automata. Computers, Environment and Urban Systems. 23(3), 205-233.
6. Breiman, L., Friedman, J. H., Olshen, R., Stone, C. J., 1984, Classification and regression trees. Wadsworth Statistics/Probability Series. Wadsworth Advanced Books and Software, Belmont, CA, 358 pp.
7. Briassoulis, H., 2000. Analysis of Land Use Change: Theoretical and Modeling Approaches, the Web Book of Regional Science. Retrieved September 8, 2007, from http://www.rri.wvu.edu/WebBook/Briassoulis/contents.htm
8. Chen, J., Gong, P., He, C., Luo, W., Tamura, M., Shi, P., 2002. Assessment of the urban development plan of Beijing by using a CA-based urban growth model. Photogrammetric Engineering and Remote Sensing. 68(10), 1063-1071.
9. Cheng, J., Masser, I., 2004. Understanding spatial and temporal processes of urban growth: Cellular automata modeling. Environment and Planning B: Planning and Design. 31(2), 167-194.
10. Chopard, Bastien, and Michel Droz. Cellular automata modeling of environmental systems. The Cambridge University Press. 1998.
11. Cieslewicz, David J., "The Environmental Impacts of urban sprawl," Urban Sprawl: Causes, Consequences, & Policy Responses, 1st ed., The Urban Institute Press, Washington, DC, 2002, pp. 23-38.
12. Clarke, K. C., Gaydos, L. J.,1998. Loose-coupling a cellular automaton model and GIS: Long-term urban growth prediction for San Francisco and Washington/Baltimore. International Journal of Geographical Information Science. 12(7), 699-714.

13. Clarke, K. C., Hoppen, S., Gaydos, L., 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco bay area. Environment and Planning B: Planning and Design. 24(2), 247-261.
14. Cohen, J., 1960. A coefficient of agreement for nominal scales, Educational and Psychological Measurement. 20(1), 37-46.
15. Congalton, R. G., 2001. Accuracy assessment and validation of remotely sensed and other spatial information. International Journal of Wildland Fire. 10(3-4), 321-328.
16. De Kok, J., Engelen, G., White, R., Wind, H. G., 2001. Modeling land-use change in a decision-support system for coastal-zone management. Environmental Modeling and Assessment. 6(2), 123-132.
17. Engelen, G., White, R., Uljee, I., Drazan, P., 1995. Using cellular automata for integrated modelling of socio-environmental systems. Environmental Monitoring and Assessment. 34(2), 203-214.
18. Ewing, R., Schmid, T., Killingsworth, R., Zlot, A., Raudenbush, S., "Relationship Between Urban Sprawl and Physical Activity, Obesity, and Morbidity," Urban Ecology, Springer US, 2008, pp. 567-282.
19. Foody, G. M., 1992. A fuzzy sets approach to the representation of vegetation continua from remotely sensed data: An example from lowland heath. Photogrammetric Engineering & Remote Sensing. 58(2), 221-225.
20. Frumkin, H., 2002. Urban sprawl and public health. Public Health Reports. 117 (3), 201-207
21. Geoghegan, J., Pritchard, L., Ogneva-Himmelberger, Y., Chowdhury, R., Sanderson, S., and Turner II, B. 1998. 'Socializing the Pixel' and 'Pixelizing the Social' in Land-Use and Land-Cover Change. In: Liverman, D., Moran, E., Rindfuss, R., Stern, P., (eds.) People and Pixels: Linking Remote Sensing and Social Science. Committee on the Human Dimensions of Global Environmental Change, National Research Council. Washington, D.C. National Academy of Science Press, pp. 51-69.
22. Goodman, R. M., Smyth, P.,1988. Decision tree design from a communication theory standpoint. IEEE Transactions on Information Theory. 34(5.1), 979-994.
23. Groff, E. R., 2007. 'Situating' simulation to model human spatio-temporal interactions: An example using crime events. Transactions in GIS. 11(4), 507-530.
24. Guan, Q., Wang, L., Clarke, K. C., 2005. An artificial-neural-network-based, constrained CA model for simulating urban growth. Cartography and Geographic Information Science. 32(4), 369-380.
25. HCPB. 2007. Hunterdon County Planning Board. Growth Management Plan. Retrieved March 19, 2008, from http://www.co.hunterdon.nj.us/pdf/hcpb/2007GrowthManagmentPlan.pdf
26. Hughes, J.W., Seneca, J.J., 2006. New Jersey's New Economy: Growth Challenges, Rutgers Regional Report, Issue Paper No. 25, July 2006, Edward J. Bloustein School of Planning and Public Policy. Retrieved December 8, 2007, from http://www.policy.rutgers.edu/news/reports/RRR/RRR_July_2006.pdf
27. Irwin, E. G., Geoghegan, J., 2001. Theory, data, methods: Developing spatially explicit economic models of land use change. Agriculture, Ecosystems and Environment, 85(1-3), 7-23.
28. Jackson, L.E., 2003. The relationship of urban design to human health and condition. Landscape and Urban Planning. (64) 4, 191-200.
29. Jiao, J., Boerboom, L., 2003. Transition Rule Elicitation Methods for Urban Cellular Automata Models in Innovations. In: Van Leeuwen, J.,Timmermans, H., (eds.), Design & Decision Support Systems in Architecture and Urban Planning, Springer Netherlands, 502 pp.

30. Kalnay, E., Cai, M., 2003. Impact of urbanization and land-use change on climate. Nature. 423(6939), 528-531.
31. Klosterman, R.E., 1999. The "What if" collaborative planning support system. Environment and Planning, B: Planning and Design, 26, 393-408.
32. Lambin, E. F., Rounsevell, M. D. A., Geist, H. J., 2000. Are agricultural land-use models able to predict changes in land-use intensity? Agriculture, Ecosystems and Environment. 82(1-3), 321-331.
33. Landis, J. R., Koch, G. G., 1977. A one-way components of variance model for categorical data. Biometrics. 33(4), 671-679.
34. Landis, J. D., 1994. Future tense. Planning. (APA), 60(2), 22-25.
35. Landis, J., Zhang, M., 1998a. The second generation of the california urban futures model. part 2: Specification and calibration results of the land-use change submodel. Environment and Planning B: Planning and Design. 25(6), 795-824.
36. Landis, J., Zhang, M., 1998b. The second generation of the California urban futures model. part 1: Model logic and theory. Environment and Planning B: Planning and Design. 25(5), 657-666.
37. Li, X., Yeh, A. G., 2000. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. International Journal of Geographical Information Science. 14(2), 131-152.
38. Li, X., Yeh, A. G., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. International Journal of Geographical Information Science. 16(4), 323-343.
39. Li, X., Yeh, A. G., 2004. Data mining of cellular automata's transition rules. International Journal of Geographical Information Science. 18(8), 723-744.
40. Lillesand, T. M., Keifer. R. W., 2000. Remote Sensing and Image Interpretation. John Wiley and Sons, New York, NY, USA, 228 pp.
41. Liu, X. P., Li, X., Yeh, A. G., He, J. Q., Tao, J., 2007. Discovery of transition rules for geographical cellular automata by using ant colony optimization. Science in China, Series D: Earth Sciences, 50(10), 1578-1588.
42. McDonald, R. I., Urban, D. L., 2006. Spatially varying rules of landscape change: Lessons from a case study. Landscape and Urban Planning. 74(1), 7-20.
43. Meyer, D., Wylie, B., Robbins, C., Vierling, L., Kozak, P., 2001. Decision tree classification of multidate AVIRIS data for mapping woodland encroachment into the great plains. Retrieved July 24, 2007, from ftp://popo.jpl.nasa.gov/pub/docs/workshops/01_docs/2001Meyer_web.pdf
44. Monserud, R. A., Leemans, R., 1992. Comparing global vegetation maps with the kappa statistic. Ecological Modelling. 62(4), 275-293.
45. Moore, D. M., Lees, B. G., Davey, S. M., 1991. A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. Environmental Management. 15(1), 59-71.
46. NJDEP, 2001. New Jersey Department of Environmental Protection. Water Quality Management Planning Rules. Retrieved December 8, 2007, from http://www.state.nj.us/dep/rules/adoptions/adopt_080707a.pdf
47. NJDEP, 2005. New Jersey Department of Environmental Protection Guidelines for the Highlands Protection and Planning Act. Retrieved March 11, 2008, from http://www.nj.gov/dep/highlands/faq_info.htm

48. NJDLWD, 2006. New Jersey Department of Labor and Workforce Development. Northern Regional Community Fact Book. Hunterdon County Edition. Retrieved April 25, 2008, from http://www.wnjpin.state.nj.us/OneStopCareerCenter/LaborMarketInformation/lmi12/hunfct.pdf
49. NJWSA, 2003. Watershed Model for Watersheds of the Spruce Run Reservoir. New Jersey Water Supply Authority; Retrieved December 8, 2007, from http://www.njwsa.org/WPU/SRI/Publications/WatershedModel/Section2DataDevelopment.pdf
50. North, M.J., Howe, T.R., Collier, N.T., Vos, R.J., 2005. Repast Simphony Development Environment, In: Macal, C.M., North, M.J., Sallach, D. (eds.), Proceedings of the Agent 2005 Conference on Generative Social Processes, Models, and Mechanisms, ANL/DIS-06-1, co-sponsored by Argonne National Laboratory and The University of Chicago.
51. Ojima, D.S., Galvin, K.A., Turner, B.L.II, 1994. The global impact of land-use change. Bioscience 44 (5), 300-304
52. Pal, M., Mather, P. M., 2002. A comparison of decision tree and back propagation neural network classifiers for land use classification. t1 503-505.
53. Pal, M., Mather, P. M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment. 86(4), 554-565.
54. Patz, J.A., Campbell-Lendrum, D., Holloway, T. & Foley, J., 2005. Impact of regional climate change on human health. Nature. 438, 310-317
55. Perner, P., Zscherpel, U., Jacobsen, C., 2001. A comparison between neural networks and decision trees based on data from industrial radiographic testing. Pattern Recognition Letters. 22(1), 47-54.
56. Pielke Sr., R. A., Walko, R. L., Steyaert, L. T., Vidale, P. L., Liston, G. E., Lyons, W. A. et al., 1999. The influence of anthropogenic landscape changes on weather in south Florida. Monthly Weather Review, 127(7), 1663-1672.
57. Pijanowski, B. C., Brown, D. G., Shellito, B. A., Manik, G. A., 2002. Using neural networks and GIS to forecast land use changes: A land transformation model. Computers, Environment and Urban Systems. 26(6), 553-575.
58. Pontius Jr., R. G., 2000. Quantification error versus location error in comparison of categorical maps. Photogrammetric Engineering and Remote Sensing. 66(8), 1011-1016.
59. Pontius Jr., R. G., Cheuk, M. L., 2006. A generalized cross-tabulation matrix to compare soft-classified maps at multiple resolutions. International Journal of Geographical Information Science. 20(1), 1-30.
60. Quinlan, J. R., 1996. Improved use of continuous attributes in C4.5. Journal of Artificial Intelligence Research. 4, 77-90.
61. Ramankutty, N., Foley, J. A., Norman, J., McSweeney, K., 2002. The global distribution of cultivable lands: Current patterns and sensitivity to possible climate change. Global Ecology and Biogeography. 11(5), 377-392.
62. Razi, M. A., & Athappilly, K. (2005). A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. Expert Systems with Applications. 29(1), 65-74.
63. Reid, R. S., Thornton, P. K., McCrabb, G. J., Kruska, R. L., Atieno, F., Jones, P. G., 2004. Is it possible to mitigate greenhouse gas emissions in pastoral ecosystems of the tropics? Environment, Development and Sustainability. 6(1-2), 91-109.
64. Riebsame, W. E., Meyer, W. B., Turner II, B. L., 1994. Modeling land use and cover as part of global environmental change. Climatic Change. 28(1-2), 45-64.
65. Rutgers CRSSA, 2004. Rutgers University Center for Remote Sensing and Spatial Analysis. New Jersey Land Cover Mapping: 1972, 1984, 1995, 2001. Retrieved April 25, 2008, from http://crssa.rutgers.edu/projects/lc/

66. Salmun, H., Molod, A., 2006. Progress in modeling the impact of land cover change on the global climate. Progress in Physical Geography. 30(6), 737-749.
67. Schneider, L. C., Pontius, G. Jr., R., 2001. Modeling land-use change in the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems and Environment. 85(1-3), 83-94.
68. Speybroeck, N., Berkvens, D., Mfoukou-Ntsakala, A., Aerts, M., Hens, N., Van Huylenbroeck, G. et al., 2004. Classification trees versus multinomial models in the analysis of urban farming systems in central Africa. Agricultural Systems. 80(2), 133-149.
69. Stevens, D., Dragicevic, S., Rothley, K., 2007. iCity: A GIS-CA modelling tool for urban planning and decision making. Environmental Modelling and Software. 22(6), 761-773.
70. Story, M., & Congalton, R. G., 1986. Accuracy assessment: A user's perspective. Photogrammetric Engineering & Remote Sensing. 52(3), 397-399.
71. Swenson, J. J., Franklin, J., 2000. The effects of future urban development on habitat fragmentation in the Santa Monica mountains. Landscape Ecology. 15(8), 713-730.
72. Tobler, W. R., 1979. Cellular geography. Philosophy in geography. 379-386.
73. Toffoli, Tommaso., and Margolus, Cellular automata machines: a new environment for modeling. MIT Press, 1987.
74. Torrens, P. M., 2000. "How cellular models of urban systems work". CASA Working Paper 28. University College London, Centre for Advanced Spatial Analysis. Retrieved September 8, 2007, from http://www.casa.ucl.ac.uk/working_papers/paper28.pdf.
75. Van Ittersum, M.K., Rabbinge, R., van Latesteijn, H.C., 1998. Exploratory land use studies and their role in strategic policy making. Agricultural systems. 58 (3), 309-330.
76. Veldkamp, A., & Verburg, P.H., 2004. Modeling land use change and environmental impact.Journal of Environmental Management. 72 (1-2), 1-3.
77. Verburg, P. H., Schot, P. P., Dijst, M. J., Veldkamp, A., 2004. Land use change modeling: current practice and research priorities. GeoJournal. 61(4), 309-324.
78. Verburg, P. H., de Nijs, T. C. M., van Eck, J. R., Visser, H., de Jong, K., 2004. A method to analyse neighbourhood characteristics of land use patterns. Computers, Environment and Urban Systems. 28(6), 667-690.
79. Walker, R., 2003. Evaluating the performance of spatially explicit models. Photogrammetric Engineering and Remote Sensing. 69(11), 1271-1278.
80. Ward, D. P., Murray, A. T., & Phinn, S. R. (2000). A stochastically constrained cellular model of urban growth. Computers, Environment and Urban Systems, 24(6), 539-558.
81. White, R., Engelen, G., 1993. Cellular automata and fractal urban form: A cellular modelling approach to the evolution of urban land-use patterns. Environment & Planning A. 25(8), 1175-1199.
82. White, R., Engelen, G., Uljee, I., 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. Environment and Planning B: Planning and Design. 24(3), 323-343.
83. White, R., Engelen, G., 2000. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. Computers, Environment and Urban Systems. 24(5), 383-400.
84. Witten, I.H. and Frank, E., Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufman, San Francisco, 1999.
85. Wu, F., 1996. A linguistic cellular automata simulation approach for sustainable land development in a fast growing region. Computers, Environment and Urban Systems. 20(6), 367-387.
86. Wu, F., 2003. Simulating temporal fluctuations of real estate development in a cellular automata city. Transactions in GIS. 7(2), 193-210.

87. Wu, F., Webster, C. J., 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. Environment and Planning B: Planning and Design. 25(1), 103-126.
88. Wu, S., Silván-Cárdenas, J., Wang, L., 2007. Per-field urban land use classification based on tax parcel boundaries. International Journal of Remote Sensing. 28(12), 2777-2801.
89. Yeh, A. G., Li, X., 2001. Measurement and monitoring of urban sprawl in a rapidly growing region using entropy. Photogrammetric Engineering and Remote Sensing. 67(1), 83-90.

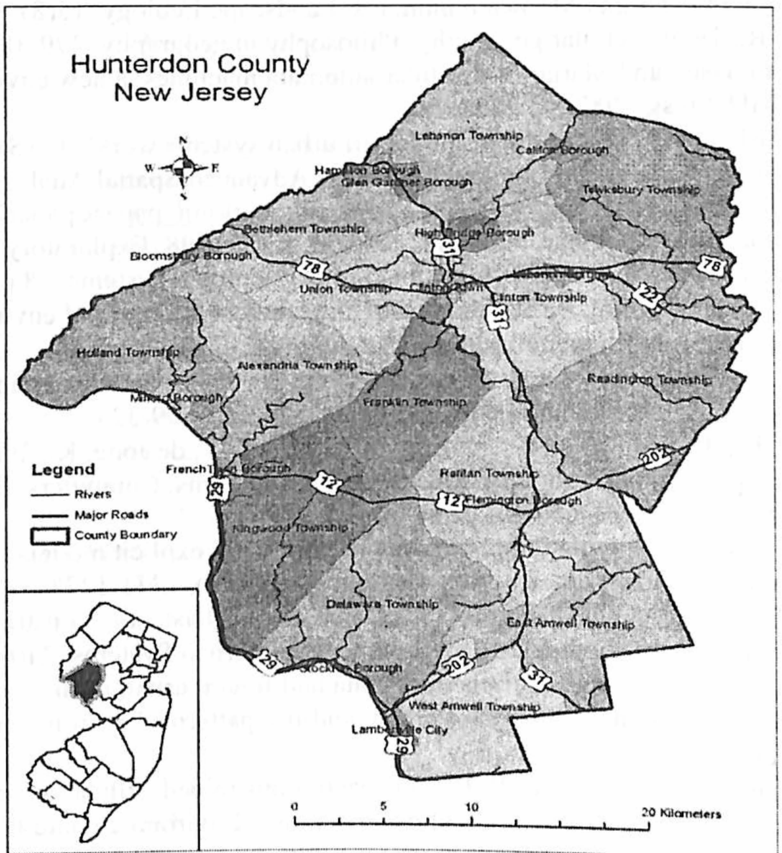

Figure 1. An overview of Hunterdon County, New Jersey

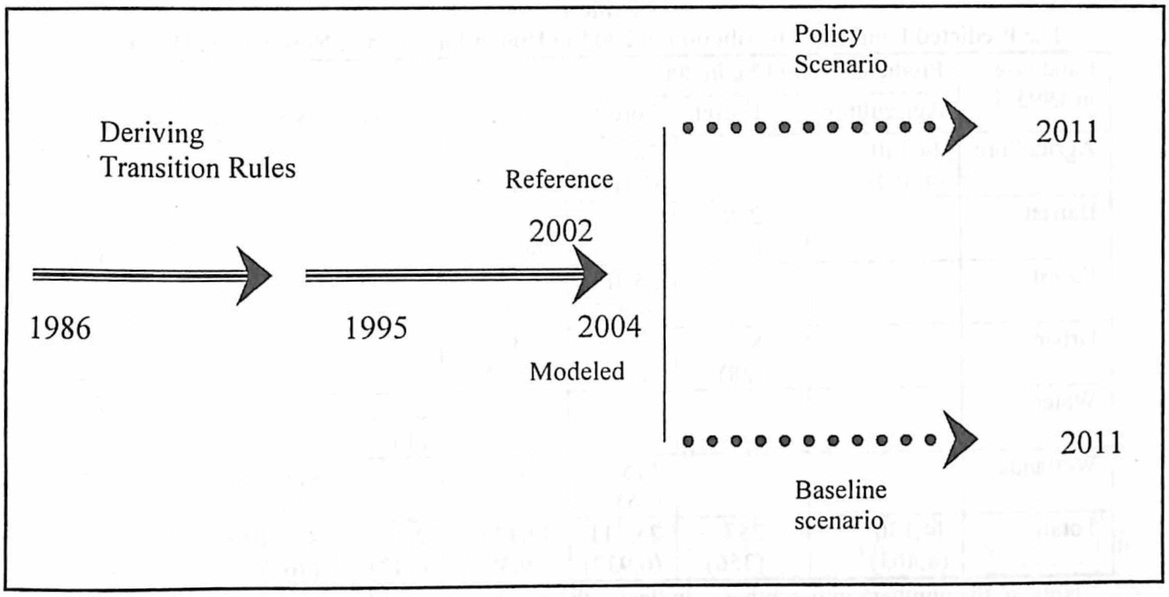

Figure 2. The schematic presentation for implementing the land use change prediction model in Hunterdon County, New Jersey

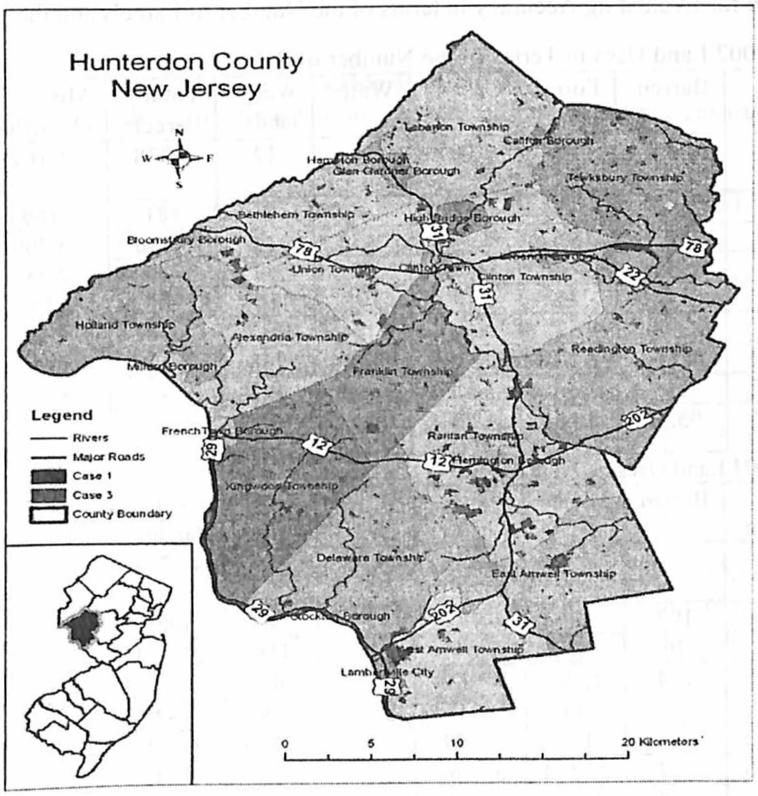

Figure 3. Spatial Distribution of the Converted Parcels under the Baseline and Policy Scenarios during the period 2002-2011

Table 1.
The Predicted Land Use Distribution in 2004 in Hunterdon County, New Jersey, Hectares[a]

Land Use in 1995	Predicted Land Use in 2004						
	Agriculture	Barren	Forest	Urban	Water	Wetlands	Total
Agriculture	44,130 (4,463)		28 (24)	2,440 (1,434)			46,598 (5,921)
Barren		249 (318)		64 (80)			313 (398)
Forest			25,367 (6,872)	10,018 (3,306)			35,384 (10,178)
Urban		8 (38)	1 (1)	17,870 (33,827)			17,878 (33,866)
Water					2,722 (142)		2,722 (142)
Wetlands			115 (35)	2,078 (739)		4,331.88 (763)	6,525 (1,537)
Total	44,130 (4,463)	257 (356)	25,511 (6,932)	32,470 (39,386)	2,722 (142)	4,331.88 (763)	109,420 (52,042)

Note: a. the numbers in parentheses indicates the number of parcels.

Table 2.
Confusion Matrix for Evaluating Accuracy in terms of the Number of Parcels and the Total Area

2002 Predicted 2002 Land Uses in Terms of the Number of Parcels

Land Uses	Agri-culture	Barren	Forest	Urban	Water	Wet-lands	Total Parcels	Mis-classified	Error of com-mission,%
Agri-culture	3,171	-	98	1,057	-	12	4,338	1,167	26.90
Barren	97	15	15	52	-	2	181	166	91.71
Forest	270	9	5,787	3,385	13	32	9,496	3,709	39.06
Urban	912	332	955	34,241	17	68	36,525	2,284	6.25
Water	4	-	25	40	105	14	188	83	44.15
Wetland	9	-	52	611	7	635	1,314	679	51.67
Total Parcels	4,463	356	6,932	39,386	142	763	52,042	-	-
Misclassified	1,292	341	1,145	5,145	37	128	-	8,088	-
Error of com mission, %	28.95	95.79	16.52	13.06	26.06	16.78	-	-	-

2002 Predicted 2004 Land Uses in Terms of the Total Area, Hectares

Land Uses	Agri-culture	Barren	Forest	Urban	Water	Wet-lands	Total Parcels	Mis-classified	Error of com-mission,%
Agri-culture	40,264	-	490	2,829	-	130	43,714	3,449	7.89
Barren	372	107	131	77	-	2	689	582	84.47
Forest	1,942	46	23,629	9,651	10	156	35,444	11,814	33.33
Urban	1,400	104	1,069	17,929	17	85	20,597	2,669	12.965
Water	25	-	33	34	2,691	38	2,820	130	4.61
Wetland	127	-	159	1,948	4	3,919	6,156	2,238	36.35
Total Parcels	44,130	257	25,511	32,468	2,722	4,332	109,420	-	-
Misclassified	3,865	150	1,881	14,540	32	413	-	20,881	-
Error of com-mission, %	8.76	58.52	7.37	44.78	1.16	9.54	-	-	-

Table 3.
The Comparison of the Size of the Converted Parcels by Land uses, Hectares

	1995-2002 Predicted Conversions				1995-2002 Actual Conversions				Difference	
Change	Mean	Std Dev	Max	Min	Mean	Std Dev	Max	Min	Mean	Std Dev
To Forest										
Agri-culture	1.17	0.27	2.29	0.146	5.77	4.53	107.81	0.004	-4.6	-4.26
Wetland	3.29	1.33	14.86	0.765	5.06	3.43	59.92	0.267	-1.77	-2.10
To Urban										
Agri-Culture	1.70	1.43	82.05	0.004	1.35	1.49	96.42	0.020	0.35	-0.06
Barren	0.80	0.27	3.64	0.020	0.42	0.26	6.28	0.008	0.38	0.01
Forest	3.03	3.21	160.29	0.001	1.11	1.81	160.29	0.004	1.92	1.40
Wetland	2.81	2.52	101.10	0.008	1.10	0.41	7.36	0.012	1.71	2.11
To Barren										
Urban	0.20	0.10	0.81	0.008	1.55	0.01	1.59	6.302	-1.35	0.09

Table 4.
The Predicted Land Use Change Patterns During 2002-2011 in the Baseline Scenario, Hectares[a]

Predicted Land Use in 2011

Land Use in 2002	Agriculture	Barren	Forest	Urban	Water	Wetland	Total
Agriculture	42,347 (3,366)	11 (17)	69 (53)	1,7023 (1027)			44,130 (4,463)
Barren		116 (17)	141 (339)				257 (356)
Forest		2 (2)	24,221 (6,034)	1,288 (896)			25,511 (6,932)
Urban		7 (31)	1 (1)	32,462 (39,354)			32,470 (39,386)
Water					2,722 (142)		2,722 (142)
Wetland			72 (20)	303 (99)		3,956 (644)	4,331 (763)
Total	42,347 (3,366)	136 (67)	24,363 (6,108)	35,897 (41,715)	2,722 (142)	3,956 (644)	109,420 (52,042)

Note: a. The numbers in parentheses indicates the number of parcels.

Table 5.

The Predicted Land use Change Pattern during 2002-2011 in the Policy Scenario, Hectares[a]

2002-2011	Agriculture	Barren	Forest	Urban	Water	Wetland	Total
Agri-culture	43,576 (4,339)	11 (17)	69 (53)	474 (54)			44,130 (4,463)
Barren		217 349		40 (7)			257 (356)
Forest		2 (2)	25,447 (6,904)	61 (26)			25,511 (6,932)
Urban		7 (31)	1 (1)	32,462 (39,354)			32,470 (39,386)
Water					2,722 (142)		2,722 (142)
Wetland			72 (20)	3 (2)		4,257 (741)	4,332 (763)
Total Area	43,576 (4,339)	237 (399)	25,590 (6,978)	33,040 (39,443)	2,722 (142)	4,257 (741)	109,420 (52,042)

Note: a. The numbers in parentheses indicates the number of parcels.