Agent-Based Land Change Modeling of a Large Watershed: Space-Time Locations of Critical Threshold

Wenwu Tang¹ and Jianxin Yang²

¹Department of Geography and Earth Sciences and Center for Applied Geographic Information Science, University of North Carolina at Charlotte, 302 McEniry, 9201 University City Blvd., 28223 Charlotte, NC, United States

²Department of Land Resource Management, School of Public Administration, China University of Geosciences (Wuhan), 388 Lumo Road, Hongshan District, Wuhan 430074, China

Correspondence should be addressed to wenwutang@uncc.edu

Journal of Artificial Societies and Social Simulation 23(1) 15, 2020

Doi: 10.18564/jasss.4226 Url: <http://jasss.soc.surrey.ac.uk/23/1/15.html>

Received: 02-02-2019 Accepted: 01-01-2020 Published: 31-01-2020

Abstract: Land use and land cover change has been recognized to have significant environmental impacts in a watershed, such as regulation of water quality. However, the identification of potential regions that are sensitive to land change activities for the protection of water quality poses a grand challenge particularly in a large watershed. These potential regions are often associated with critical thresholds in terms of, for example, water quality. In this study, we developed an agent-based land change model to investigate the relationship between land development activities and water quality in eight North Carolina counties that cover the lower High Rock Lake Watershed area. This agent-based model, which is empirically calibrated, is used to identify space-time locations of those regions at critical thresholds of water quality in this study area. Our experimental results suggest that land development as a form of system stress is of pivotal importance in affecting water quality at sub watershed level and the state transition of water quality. The agent-based model developed in this study provides solid support for investigations on the impact of land development under alternative scenarios in a large watershed.

Keywords: Agent-Based Model, Land Use and Land Cover Change, Critical Threshold, Water Quality, North Carolina

Introduction

- 1.1 Land use and land cover change (LULCC) has been playing a critical role in driving landscape dynamics of watershed ecosystems. These ecosystems are often coupled human and natural systems (Liu et al. 2007a,b) in which anthropogenic activities or policies may lead to alternative consequences in terms of watershed conditions (Haidary et al. 2013). The condition of a watershed such as water quality is inherently related to land cover patterns (e.g., impervious cover, forest, and farmland) that have been modified by spatial processes of urbanization, deforestation, or agriculture activities. A set of models, represented by impervious cover model (Kauffman & Brant 2000; Schueler 1994) and forest cover model (Gerow et al. 2015, 2016), have been proposed to study the relationship between land cover and water quality at the watershed level. A threshold effect of percent land cover (e.g., impervious cover or forest cover) on water quality at the watershed level has been identified in the literature. Watersheds within this threshold are at a tipping point that may transit to other equilibriums in terms of water quality (good or poor) in response to alternative land management practices or policies. A number of empirical studies have reported this threshold effect, which may be specific to watersheds of interest (Schueler et al. 2009). For example, a range of 10-25% of percent urban cover is a generic critical threshold based on the impervious cover model (Schueler 1994). 37-48% of percent forest cover was identified as a critical threshold of water quality specific to the High Rock Lake Watershed in North Carolina, USA (Gerow et al. 2015, 2016).
- 1.2 Threshold effects in terms of water quality at (sub) watershed level have been well established in the literature. Most of the studies are based on the use of historic or present data (land cover and indicators of water quality) to

identify those regions within the threshold (Haidary et al. 2013; Schueler et al. 2009). The identification of these potential regions in a watershed is very important for the regulation of water quality, watershed management practices, and landscape conservation. Once these regions are degraded (transition to a different equilibrium), the restoration process may be difficult (or even infeasible), time- and cost-consuming particularly for those regions that are beyond the critical threshold (e.g., densely urbanized or heavily deforested; often irreversible) (Schueler 1994). However, the identification of where and when those potential regions in a watershed may experience threshold due to, for example, future land development or alternative policy intervention poses a grand challenge. The processes of LULCC per se and the way that LULCC affect water quality in a watershed are complex, nonlinear, and uncertain. For example, a time lag effect may exhibit in terms of the impact of LULCC on watershed-level water quality (Meals et al. 2010): the impact of LULCC on water quality may not be immediately observed. It needs a period of time (depending on indicators of interest; see Meals et al. 2010 for water quality to respond to the land development or management practices. Further, the processes associated with LULCC may not be stationary over time. Complex dynamics in these processes further complicate the identification of space-time locations of those potential regions within thresholds.

- 1.3 To identify these potential regions experiencing water quality threshold and their response to land development and policy intervention often requires the use of land change simulation modeling. Agent-based models (ABMs) of land systems are such a spatiotemporal simulation approach for their support in the representation of complex pattern-process interactions and scenario analysis (Epstein & Axtell 1996; Grimm et al. 2005). As one of the approaches recommended for land change modeling (Council 2014), ABMs are a bottom-up simulation approach that uses agents as building blocks to represent interactions among actors (e.g., stakeholders or land owners) or with their environments. Agents make explicit decisions on their actions based on the contextual information that they perceive from their situated environments that are often spatially heterogeneous. The explicit representation of actor decision-making and actor-environment interactions in ABMs allows for the simulation of land developments that are often complex and uncertain (Parker et al. 2003). A number of agent-based land change models have been proposed and detailed review conducted over the past two decades (An et al. 2005, 2014; Groeneveld et al. 2017; Matthews et al. 2007; Ng et al. 2011; O'Sullivan et al. 2016; Parker et al. 2003), which offers insights into the development of agent-based model in this study.
- 1.4 Therefore, the goal of this study is to present an agent-based land change simulation model to identify the space-time locations of potential regions experiencing critical threshold of water quality in a watershed in response to land change activities. Our study area includes eight North Carolina (NC) counties that cover the lower High Rock Lake Watershed region. This study area is representative of coupled human and natural systems in which LULCC plays an important role in driving and interpreting complex ecosystem dynamics. Land-related activities such as land development and agricultural activities render the ecosystem of this region vulnerable facing a series of issues in terms of, for example, water supply and water quality. These issues may further lead to degradation in quality of life or public health concern in this study region. Thus, such an agent-based land change simulation model is needed that allows us to investigate the potential impact of land-related activities or policies on the landscape (water quality here). The agent-based modeling approach is preferred here over other simulation approaches (e.g., cellular automata or system dynamics). This is because the agent-based modeling approach provides an integrative modeling platform that can combine with other approaches as needed and, more importantly, support the explicit representation of individual-level decision-making processes of actors (e.g., preferences of land developers or owners here) and their interactions. The remainder of this article is organized in the following manner. First, we introduce the study area and data used in this study. Second, we focus on the design of the agent-based land change model. Then, we present experimental results and discussion. This article ends with conclusion of the study.

Study Area and Data

- 2.1 Our study region (see Figure 1) includes eight counties in North Carolina, USA, which cover the lower High Rock Lake Watershed (HRLW) area. The total area of our study region is 10,823 km² with a total population of 1.78 million in the year of 2017. This study region is located within the Piedmont area of North Carolina, with elevation varying from 99m to 528m. Percentage of urban land cover in the study region is 18.55%, 19.56%, and 20.09% in year 2001, 2006, and 2011. This study region experienced 5.4% of increase in percent urban cover from 2001 to 2006, and this rate becomes 2.7% from 2006 to 2011 (due to the impact of economic recession).

Figure 1: Map of the study area (HRLW: High Rock Lake Watershed; the study area includes 8 counties in North Carolina, covering the lower HRLW region).

2.2 Three years (2001, 2006, and 2011) of land cover data used in this study were retrieved from U.S. National Land Cover Database (NLCD). These land cover data were available at a spatial resolution of 30 m by 30 m. We further aggregated these land cover data into five major types: urban (including developed areas with low, medium and high intensity, and open space), farmland (including pasture and cropland), forest (including deciduous, evergreen, and mixed forest), natural (including shrub, scrub, grassland, and barren land), and water (including open water and wetlands) (see Figure 1 and Table 1). Data of road network, cities or towns and their populations, and county boundaries were obtained from US Census website. Stream network and watershed data (12 digit Hydrologic Unit Code was used) was from the USDA National Hydrography Dataset. Elevation data is from NASA SRTM data. All the vector GIS data were rasterized using a 30 m $\times$ 30 spatial resolution. The landscape size is 4,378 $\times$ 4,746 in terms of number of rows and columns of the raster GIS data.

Land cover	Year	Davie	Forsyth	Davidson	Cabarrus	Randolph	Rowan	Guilford	Iredell	Lower HRLW
Farmland	2001	239.16	171.94	354.75	231.57	528.08	424.70	421.67	536.47	2908.36
	2006	229.64	159.88	337.37	212.63	524.12	408.30	403.10	516.38	2791.43
	2011	228.23	154.88	335.26	208.72	517.00	405.92	394.90	508.46	2753.38
Natural	2001	49.20	54.83	115.80	68.12	139.36	100.63	57.93	93.56	679.43
	2006	48.27	56.86	112.29	68.36	148.44	101.42	62.31	97.41	695.36
	2011	54.22	66.43	140.65	78.02	209.40	118.24	77.89	111.46	856.32
Forest	2001	326.94	435.07	748.78	425.74	1149.07	609.31	649.14	633.34	4977.39
	2006	326.93	427.41	749.90	408.80	1138.89	606.02	631.98	628.83	4918.78
	2011	322.05	412.92	719.85	396.68	1071.52	588.92	604.69	611.57	4728.21
Urban	2001	63.57	389.36	206.93	200.32	206.33	183.18	530.91	227.11	2007.72
	2006	65.89	405.33	215.85	226.03	211.21	188.47	562.19	241.85	2116.81
	2011	66.34	415.31	219.67	232.63	215.19	191.27	580.14	253.74	2174.29
Water	2001	12.91	17.62	42.49	18.10	21.51	38.72	43.51	55.48	250.33
	2006	21.05	19.34	53.34	28.03	21.70	52.32	43.60	61.49	300.86
	2011	20.94	19.28	53.32	27.79	31.25	52.18	45.55	60.73	311.04

Table 1: Land cover data for each county and the entire region over years (unit: km²)

2.3 The High Rock Lake in our study region have been facing a water quality issue due to the nutrient loads from, for example, land use and management activities. This water quality issue in this region has received considerable attention from government agencies such as US Environmental Protection Agency and NC Forest Service. NC Forest Service developed a forest cover model to evaluate the relationship between forest cover and water quality in the HRLW region (Gerow et al. 2015, 2016). Benthic macroinvertebrates are aquatic organisms (most

of them are aquatic insects) that are sensitive to the conditions of their living environments, and have been extensively used as indicators of water or habitat quality (Mandaville 2002). A series of metrics based on structural or functional characteristics of benthic macroinvertebrate communities have been developed to evaluate water conditions (Barbour et al. 1999). Based on the empirical analysis of five years of land cover data and benthic macroinvertebrate data, a range of 37-48% of percent forest cover was identified as the threshold of water quality for the HRLW region (Gerow et al. 2015, 2016). Sub watersheds (or catchments) with percent forest cover higher than 48% (or lower than 37%) correspond to good (or poor) water quality indicated by benthic macroinvertebrate metrics. Any sub watersheds with the percentage of forest cover within this range are at the tipping point with fair water quality that may be shifted to either good or poor states. Our overall study region is at this tipping point (fair water quality) as its percent forest cover varies from 45.99% down to 43.69% from 2001 to 2011. Figure 2 shows the maps of sub watershed-level water quality based on forest cover model for 2001, 2006, and 2011. Of more interest is to identify where those regions at sub watershed level are at the tipping point particularly in the future and how the conditions of these sub watershed regions may change in response to LULCC. This drives the development of the agent-based land change model in this study.

Figure 2: Map of water quality at sub watershed level based on forest cover model (a: year 2001; b: year 2006; c: year 2011; sub watersheds with fair water quality are at the tipping point).

Methods

Design of the agent-based model

- 3.1 Figure 3 illustrates the architecture of the agent-based land change model in this study. The agent-based model was designed using the demand-allocation framework, which is commonly used in the land change modeling domain (Aquilué et al. 2017; Liu et al. 2017; Verburg et al. 2002). The demand-allocation land change framework consist of two modules: demand and allocation. The demand module controls the quantity of land development activities in a top-down manner and the allocation module determines where these activities will occur. The demand module can be linked to, for example, system dynamics model (Lauf et al. 2012; Liu et al. 2017) or a population density approach (McGarigal et al. 2018; Meentemeyer et al. 2013) to estimate land change demand. In this study, the land change demand in each county each year is the area of land that will be converted.
- 3.2 The allocation module that determines where land development will occur is driven by the interactions of two types of agents (Mustafa et al. 2017): land owners and land developers. Land owners and land developers are situated within each county. These agents interact with each other on the transactions of lands for development. Land owners decide whether to sell their lands to developers for residential development. Land developers are in charge of decision making in terms of where a residential development event will occur, the area of the development, and how to realize the development. Both land owners and land developers are hypothetical agents (Matthews et al. 2007; Valbuena et al. 2009). Through interactions between land owners and land developers, the land change allocation process is implemented in a bottom-up manner.

3.3 We documented this ABM using ODD protocol (Grimm et al. 2010) in Appendix. Please refer to the ODD protocol of the ABM for more detail. In this article, we focus on discussing the two types of agents: land owners and land developers.

Figure 3: Framework of the agent-based land change model.

Land owner agents

3.4 Land owners are in charge of management of their lands including farmlands, forest, and natural. Land owners make decisions on whether to sell their lands to developers or not (i.e., two actions). This decision is driven by land owner's perception on surrounding environments. We used a distance-decayed model to represent land owner's decision on selling their lands for development – i.e., the utility function of land owner agents (Lauf et al. 2012; McGarigal et al. 2018; Tang & Jia 2014). Land owner's decision on selling their lands can be regarded as a form of spatial interactions in which distance plays an important role (Gould 1969; Taylor & Openshaw 1975). This is fundamentally related to Tobler's First Law of Geography: "Everything is related to everything else, but near things are more related than distant things" (see Tobler 1970). In other words, these interactions are spatially dependent and may lead to the diffusion or movement of, for example, innovation, culture, and settlements (Gould 1969; Hagerstrand 1968). Spatial dependence in these interactions exhibits distance decayed effect. Thus, in this study we used the distance-decayed model, similar to the concept of information field (Morrill & Pitts 1967), to explore the spatial diffusion of land owner's decision on selling lands. Distance to nearest urban is a proxy variable here that represents the influence of existing urban area on the decision-making process of land owners in terms of selling lands.

3.5 The utility function of a land owner agent is formulated using Equation 1:

$$u_{owner}(i, j, k) = e^{-\alpha(i, j, k)} e^{\beta(i, j, k) * d} \quad (1)$$

where $u_{owner}(i, j, k)$ is the utility function of land owner agent i in county j with respect to the land cover type k of his/her land parcel ($k = 1$ for forest/natural; $k = 0$ for farmland). The percentage of natural land in our study region is low (6.3, 6.4 and 7.9% for 2001, 2006, and 2011; see Table 1). Thus, in this study, we combine forest and natural cover together to represent land owners' preference on land transactions (forest and natural cover information is separated when deriving percent in forest cover from simulation results). $\alpha(i, j, k)$ is the coefficient that controls magnitude of the utility of the land owner agent (increase in α will lower the magnitude of the utility). $\beta(i, j, k)$ is the coefficient that is associated with the steepness of the distance-decayed function representing the utility of the agent. The higher the value of β is, the faster the utility of the land owner agent will decay with distance. d is the distance of the land owned by the agent to the nearest urban area (associated with contextual information that an agent can perceive). Thus, a collection of utility functions of land owner

agents constitutes the collective knowledge in terms of decisions on land transactions for urban development. The pool of utility functions of land owner agents in county j , noted as $D(j)$, can be formulated as:

$$D(j) = \{(\alpha_{ijk}, \beta_{ijk}) \mid \alpha_{ijk} \in [\alpha_{ljk}, \alpha_{hjk}], \beta_{ijk} \in [\beta_{ljk}, \beta_{hjk}]\} \quad (2)$$

3.6 Utility functions of land owner agents can be calibrated by fitting with empirical land conversion data in each county in terms of distance to nearest urban land cover. In this study, we used the land conversion data from 2001-2006 to calibrate the utility functions of land owner agents in each county. Table 2 and Figure 4 report the calibration results of distance-decayed coefficients for each county. The calibrated distance-decayed model for land owner agents in each county of our study region is normalized — i.e., the utility of a land owner agent is the probability that the agent will sell his/her land for residential development. In this study, we use the 95% confidence intervals of the fitted coefficients to represent the pool of decision rules of land owner agents, which becomes Equation 3:

$$D(j) = \{(\alpha_{ijk}, \beta_{ijk}) \mid \alpha_{ijk} \in CI(\hat{\alpha}(j, k)), \beta_{ijk} \in CI(\hat{\beta}(j, k))\} \quad (3)$$

where $\hat{\alpha}(j, k)$ and $\hat{\beta}(j, k)$ are the fitted coefficients of the distance decayed model for land owner agents in county j with respect to land cover type k . $CI(\cdot)$ denotes the confidence interval of the corresponding variables. The decision pool of land owner agents represents the heterogeneity of agent decision-making. We assume each land unit (cell here) is managed by a land owner agent, whose utility function is randomly drawn from the decision pool of land owner agents in the county. Every time when a land owner agent is contacted by land developers for land transaction, the land owner agent will make decisions on whether to proceed on land transactions based on his/her utility function. This land transaction process is stochastic (a random probability is drawn and compared with the utility value of the land owner agent).

County	Forest/natural to urban conversion			Farmland to urban conversion		
	α	β		α	β	
Davie	1.7765 (1.159-2.394)	0.0065 (0.005-0.008)		0.927 (0.303-1.551)	0.0123 (0.010-0.015)	
Forsyth	0.9164 (0.331-1.501)	0.0114 (0.010-0.013)		1.1802 (0.918-1.442)	0.0107 (0.010-0.011)	
Davidson	1.2761 (0.967-1.585)	0.0081 (0.007-0.009)		1.0732 (0.621-1.525)	0.0092 (0.008-0.010)	
Cabarrus	1.1836 (0.811-1.556)	0.0075 (0.007-0.008)		1.5109 (1.209-1.812)	0.0085 (0.008-0.009)	
Randolph	1.2610 (0.806-1.716)	0.0082 (0.007-0.009)		1.4565 (1.030-1.883)	0.0068 (0.006-0.008)	
Rowan	1.6682 (1.144-2.193)	0.0064 (0.005-0.008)		1.1050 (0.468-1.742)	0.0091 (0.007-0.011)	
Guilford	1.3883 (1.154-1.622)	0.0069 (0.006-0.007)		1.9971 (1.778-2.216)	0.0050 (0.005-0.005)	
Iredell	1.3257 (0.934-1.718)	0.0076 (0.007-0.009)		1.0430 (0.563-1.523)	0.0118 (0.010-0.013)	

Table 2: Distance-decayed coefficients for each county in the study region.

Figure 4: Distance decayed models of land owner decision-making at county level (A: forest/natural to urban conversion; B: farmland to urban conversion).

Land developers

3.7 Land developers lead on converting lands into residential area (i.e., urban land cover). In this model, we used a patch-based approach to represent the conversion of lands — each land development event by a land developer leads to the occurrence of a new patch of residential area. This patch-based approach has been used in the literature for land change simulation at a fine spatial scale (Aquilué et al. 2017; Li et al. 2017; Liu et al. 2014; McGarigal et al. 2018; Meentemeyer et al. 2013). The behavior of a land developer for land conversion consists of three major steps: 1) search for the initial location of residential development, 2) determine the area of the residential development, and 3) generate a patch of land development to realize the land conversion.

3.8 The first step of decision making associated with land development for a land developer agent is to determine where to acquire a land for residential development. Specifically, a land developer agent will randomly inspect a land that is feasible for residential development (i.e., farmland or forest/natural). The land developer decides if a land will be chosen for development based on the development probability of the land derived from his/her utility function (discussed next). A land with higher development probability is more likely to be selected. In this study, those areas that are water bodies, wetlands, or protected areas are excluded (i.e., these areas are infeasible or prohibited for land conversion). Once a land is selected, the land developer agent decides whether he/she will acquire the land for residential development using his/her utility function. The decision of a land developer on whether to acquire a land for residential development or not is a binary variable (two actions; 1: acquire; 0: do not acquire). This decision is dependent on the composite influence of a set of driving factors. In this study, the utility function of a land developer agent to acquire a land for development is based on a logistic regression model, which is represented as in the following equations (Equations 4 and 5.):

$$u_{\text{developer}}(i, j) = (1 + e^{-y(i, j)})^{-1} \quad (4)$$

$$y(i, j) = \beta_0(j) + \beta_1(j)x_1 + \dots + \beta_m(j)x_m \quad (5)$$

where $u_{\text{developer}}(i, j)$ is the utility of a land developer agent i in county j . $x_1 - x_m$ are the driving factors (m : the number of driving factors). $\beta_1 - \beta_m$ are coefficients for the corresponding driving factors. β_0 is the intercept of the logistic regression model.

3.9 In this study, the factors that drive the decision of land developers include elevation, slope, distance to cities, distance to major roads, distance to streams, stream density, and development pressure (see Figure 5). Neighborhood size for the calculation of development pressure (the percentage of urban cover) is 35 by 35 cells. A search radius of 500 meters was used to derive stream density. For each county, the size of samples to fit the logistic regression model is 2,000 (1,000 for urban, 1,000 for nonurban) and stratified random sampling scheme was used. Land developer agents in a county share the same utility function calibrated using empirical land change data in the county from 2001 to 2006. Figure 6 shows the map of utility values (development probability) of land developers in the study region.

3.10 Once the land developer decides to acquire the land, he/she will contact the owner of the land (a land owner agent). The land owner will decide whether to sell lands to the developer or not. Only the land owner agent agrees on land transaction, the land developer can launch residential development (i.e., conversion into urban land cover type) and this land is used as the initial location of the new patch of residential development.

Figure 5: Map of driving factors of the agent-based land change model for the study region.

Figure 6: Map of the utility function of land developers in the study region.

- 3.11** The second step of land conversion for a land developer agent is to determine the area of the new patch of development. The area of the new patch of development is randomly drawn from empirical area distribution of new patches of development from 2001 to 2006. In this study, we use the observed patch of change of the entire study region from 2001 to 2006 to derive the empirical distribution that a land developer agent will use to determine the area of new patch of development.
- 3.12** Once the initial location and area of residential development are determined, the land developer agent will generate the new patch of development. A patch growing mechanism is used to form the new patch: neighboring lands (cells) are recursively added into the new patch until the patch area is reached. The probability that a neighboring land cell is added into the development patch is the composite utility that the land developer agent decides on land acquisition multiplied by the land owner agent's utility on land transactions at that cell. A first-order von Neumann neighborhood rule is used to search for neighboring cells during the process of forming a new patch of development.

Implementation

- 3.13** The code of this agent-based model was written in C/C++ programming language. All GIS data were processed in ESRI ArcGIS Pro version 2.1.3. All distance-based variables were calculated using the Euclidean distance tool in

ArcGIS. All the raster GIS data were imported into the agent-based model — i.e., a loose coupling of agent-based modeling with GIS (Brown et al. 2005b). The ABM was deployed on a high-performance computing cluster (total number of CPUs: 2,060; Redhat Linux-based) for acceleration of Monte Carlo runs of the ABM. The task scheduler for high-performance computing on the Linux cluster is PBS/TORQUE. We used Shell scripts to automate the execution of the ABM and the post-processing of simulation results.

Experiments and Discussion

- 4.1** Our agent-based land change model allows us to explore the impact of LULCC activities on water quality in the study region. In this section, we design two experiments to explore the impact of macro- and micro-level LULCC activities on the spatiotemporal pattern of water quality at sub-watershed scale. The first experiment focuses on macro-level LULCC activities by the manipulation of annual land development demand. In the second experiment, land owners' preference on land transactions was varied to represent LULCC activities at micro or individual level. We first discuss performance metrics and model calibration and validation before we present the design and results of the two experiments.

Performance metric

- 4.2** We used bivariate similarity metrics that are based on confusion matrix to evaluate the agreement between simulated and reference land cover patterns. A variety of similarity metrics based on confusion matrix have been developed in the literature. These similarity metrics include percent correct match, figure of merits, Kappa coefficients, producer accuracy, and user accuracy (Fielding & Bell 1997; Pontius Jr 2002). While these metrics are limited to categorical variables and the cell level (instead of object level), they provide support for the evaluation of locational agreement between simulation outcome and observed patterns. Further, these similarity metrics only provide an overall evaluation for a study region of interest based on cell-level comparison. Thus, often time, a study region is partitioned into a collection of sub-regions (e.g., using lattice with coarser spatial resolutions) when using these metrics to evaluate the agreement between simulation outcome and reference patterns (Kok et al. 2001; Pontius Jr 2002). In this study, we used two percent correct match indices (pcm_1 and pcm_2) that are based on confusion matrix (Pontius et al. 2008; Pontius Jr & Millones 2011; van Vliet et al. 2016), shown in the following equations (Equations 6, 7).

$$pcm_1 = \frac{n_a}{n_a + n_b} \quad (6)$$

$$pcm_2 = \frac{n_a + n_d}{n_a + n_b + n_c + n_d} \quad (7)$$

where n_a is the number of observed cells of change simulated as cells of change (i.e., hits). n_b is the number of observed cells of persistence simulated as change (false alarms). n_c is the number of observed cells of change simulated as persistence (misses). n_d represents the number of observed cells of persistence that are simulated as persistence (correct rejections). In general, larger values of these accuracy indices indicate higher agreement between simulated patterns and reference patterns. However, it has been recognized in the literature that model accuracies that focus on change (e.g., pcm_1 here, figure of merit, or kappa) tend to be low if the study region of interest is dominated by persistence (Pijanowski et al. 2005; Pontius et al. 2008; Pontius Jr et al. 2004). As reported in Pontius et al. (2008), a positive relationship exists between model accuracy (figure of merit in their study) and observed net land change. A multi-resolution approach (e.g., using lattice at coarse spatial resolutions) has been suggested to cope with this issue (Pontius Jr. et al. 2004; Pontius Jr et al. 2004).

Model calibration and validation

- 4.3** Model calibration and validation are important steps of land change simulation modeling (Brown et al. 2005a; van Vliet et al. 2016). In general, there are two approaches for model calibration and validation: partitioning of temporal duration or spatial extent of a study system into calibration set (period or area) and validation set that do not overlap with each other (Pontius Jr. et al. 2004). A model is instantiated using data in the calibration set for the specific study region of interest. Simulation outcome of the calibrated model is compared with data in the validation set to ensure the model has sufficient predictive power (instead of being overfitted to calibration data) and thus is valid for experimentation use. Typically, temporal partitioning is often used for land

change modeling given the availability of land cover data and related data over time. The simulation period of a model is divided into calibration period and validation period. In this study, the calibration period is from 2001–2006, and the validation period is from 2006 to 2011. The utility functions (the distance-decayed model) of land owners were calibrated using the land conversion data from 2001–2006 in each county. The utility function of land developers in a county was calibrated using logistic regression based on empirical land change data in the county from 2001 to 2006. We used the area of land conversion from 2001 to 2006 in each county to represent the annual land development demand during the calibration phase. Please see the ODD protocol of our ABM in the Appendix for more detail.

- 4.4** Figure 7 shows the maps of simulated land cover patterns for year 2006 and 2011. These maps depict the spatial patterns of persistence and change of land cover types. From these maps, we can see that most of farmland and forest land cover remain persistent over time — i.e., the landscape of our study region is dominated by persistence. The simulation accuracy of pcm_1 for 2001–2006 (calibration accuracy) is 4.32% on average (with a standard deviation of 0.43%; based on 100 Monte Carlo runs). The corresponding simulation accuracy in terms of pcm_2 is averaged at 96.90% (std: 0.02%).

Figure 7: Map of simulation results of land cover patterns for year 2006 and 2011.

- 4.5** We ran our simulation model to 2011 for validation purpose. Annual land change demand from 2006 to 2011 was based on the empirical land change quantity during this period. The validation accuracy of pcm_1 is 6.56% on average (std: 0.49%). The averaged pcm_2 is 96.26% (std: 0.03%) for the validation period. The simulation accuracy in terms of pcm_1 (pcm_2) is low (high) because land change quantity during the two simulation periods only occupies about 1–1.5% of the study area and most of the regions do not experience land change (i.e., high-level persistence; see Figure 7). Increase in simulation accuracy in terms of pcm_1 is partially attributed to the occurrence of more land development from 2006 to 2011 (Pontius et al. 2008). Figure 8 depicts the receiver operating characteristic (ROC; see Pontius Jr & Schneider 2001 and Fielding & Bell 1997) plot of our agent-based model for calibration and validation runs (100 runs for each). The ROC plot suggests that while our model accuracy in terms of pcm_1 is low, the model performance is acceptable by comparing with the performance due to chance (the performance of all model runs for calibration and validation is higher than chance performance in the ROC plot).

Figure 8: ROC plot of model performance for calibration and validation of our agent-based model (A: ROC plot; B: zoomed-in plot of ROC curve by adjusting the maximum of horizontal and vertical axis to 0.1; specificity = $n_d/(n_b + n_d)$; sensitivity = $n_a/(n_a + n_c)$; for notation, please refer to Equation 6 and 7).

4.6 While low net land change or high-level persistence explains low simulation accuracy of our model (Pontius et al. 2008; Pontius Jr et al. 2004), of more interest is the variation of simulation accuracy across our study region. Thus, we extracted sub region-level accuracy using a grid with coarser spatial resolution (3km by 3km here). Figure 9 shows the spatial distribution of averaged accuracy (pcm_1) for calibration and validation phases at sub region level. As we could see, the averaged simulation accuracy varies from 0-17.66% for 2001-2006 (calibration period), and can reach up to 26.14% for the simulation period of 2001-2011 (validation period). For rural areas (e.g., Davie, Randolph, Rowan Counties), simulation accuracies are low (even reach zero hits because the development probability calibrated using historical data in these rural areas is low). High simulation accuracies tend to concentrate in urban areas (e.g., those in Cabarrus, Forsyth, Guilford, and Iredell Counties). Those urban areas typically have high net land change, which is likely to have reasonably high accuracies as indicated in the literature (Pontius et al. 2008). The accuracy assessment results suggest that our ABM provides reasonable support for the simulation of land development in our study region (most developments typically concentrate in urbanized areas) and thus empowers the exploration of the relationship between LULCC and water quality.

Figure 9: Map of simulation accuracy for the agent-based land change model (a: year 2006; b: year 2011; 3km by 3km grid was used).

Experiment of land development demand at macro level

4.7 In this experiment, we aim to explore the impact of alternative land development demand on the water quality at the sub watershed level in our study region. We use the annual land development demand derived from empirical land change data from 2001 to 2006 as a baseline scenario. We vary the annual land development demand at the county level by 20%, 60%, 100%, 140%, and 180%. Correspondingly, we have five treatments (T_1 , T_2 , T_3 , T_4 , and T_5) for this experiment. For each treatment, the agent-based simulation model ran from 2001 to 2040. Each treatment was repeated 100 Monte Carlo runs. These treatments were designed to represent different land development scenarios. The first two treatments are designed to represent the scenarios of slow land development, and treatment T_4 and T_5 for the scenarios of rapid land development.

4.8 Figure 10 shows the maps of water quality patterns at sub watershed level for the five treatments. Table 3 reports the number of sub watersheds falling within the three types of water quality states (good, fair, and poor). Figure 11 illustrates the time series of percent in forest cover over years for four sub watersheds. A sub watershed belongs to good or poor water quality states only when it remains in the same state for all 100 simulation runs (otherwise, it belongs to the category of fair water quality — i.e., within the threshold). As we could observe, the number of sub watersheds with good water quality decreases (from 56 to 34) when land development demand increases. In other words, more sub watersheds (from 71 to 93) fall within the critical threshold (fair water quality). The average number of years that these critical sub watersheds remain its state is 31.79, 29.06, 25.66, 23.91, and 21.75 for the five treatments. It shows that increment in the quantity of land development tends to reduce the length of the time that critical sub watersheds remain in the same state.

Figure 10: Map of simulation results of water quality patterns at the sub watershed level for the first experiment (varying land development demand by: A: 20%; B: 60%, C: 100%, D: 140%, E: 180%; labels over sub watersheds in the maps indicate the averaged time of transition — i.e., the average number of years staying within the critical threshold).

Water quality	T1	T2	T3	T4	T5
Good	56	49	42	38	34
Fair	71	78	85	89	93
Poor	32	32	32	32	32

Table 3: Distance-decayed coefficients for each county in the study region.

Figure 11: Time series of percent in forest cover over years for the first experiment (T1-T5: five treatments corresponding to varying land development demand by 20, 60, 100, 140, and 180%; four sub watersheds were used with IDs of 25, 104, 110, and 153).

Experiment of land owner preference at micro level

- 4.9** The second experiment is to evaluate the influence of land owner's preference on water quality in the study area. In this model, the land owners' preference is associated with the decision pool of utility functions characterized by distance-decayed models (see Equation 1). Distance-decayed coefficients (α_{ijk} , β_{ijk}) in the decision pool with respect to conversion from forest or natural lands to urban were systematically varied by adding a specific amount of change ($\sigma * r(\alpha_{ijk})$, $\sigma * r(\beta_{ijk})$), where $r(\cdot)$ is the range of the corresponding coefficients (see Equation 3). Utility functions of land owner agents on farmland to urban conversion remain unchanged. We design five treatments (T_1 - T_5) by assigning σ to 1.0, 0.5, 0, -0.5, and -1.0 correspondingly. We ran our model from 2001 to 2040 for each treatment. The number of Monte Carlo runs for each treatment is 100. Varying distance-decayed coefficients will modify the shape of the distance-decayed curves: high coefficients will render that the magnitude of the utility becomes lower and the distance-decayed curve becomes steep. In other words, land owner agents with high distance-decayed coefficients tend to have low preference on selling their forest or natural lands for residential development.
- 4.10** Table 4 reports the simulation results in terms of the number of sub watersheds at different states. Figure 12 shows the maps of water quality patterns across the five treatments. Figure 13 illustrates change in the percentage of forest cover of four sub watersheds. In general, we could observe a decreasing pattern in terms of the number of sub watersheds with good water quality when distance-decayed coefficients that evaluate land owner agents' preference on land development become lower (corresponding to high utility values of selling their forest land). The average number of years that critical sub watersheds (fair water quality) remain their states is 26.41, 26.36, 25.90, 26.04, and 25.96 for the five treatments. The number of years that a sub watershed stays within the critical threshold tends to be shortened as distance-decayed coefficients become lower (from 26.41 years to 25.96 years). This means that land owner agents with higher preference on selling their forest lands for development tend to accelerate the state transition of critical sub watersheds (from fair to poor water quality).

Discussion

- 5.1** The quantity of land development functions as a top-down mechanism that drives LULCC in a study region. Variation in land development demand (e.g., due to change in economy or population) is of pivotal importance in affecting the sub watershed-level water quality in our study region. As increase in land development demand, more sub watersheds with good water quality will transit to the critical threshold state, as indicated by the first experiment. Further, the length of the time that critical sub watersheds remain their states is reduced in response to increased land development demand. This is because land developments at sub-watershed level will generally occur more in response to increased annual demand for land development at county level. More land development events consume more forest lands, leading to the detrimental effect on water quality (see Figure 11). In particular, if we remain the land develop demand same as that between 2001 and 2006 (baseline scenario), within 25.66 years from 2001 (i.e., around year 2026) all sub watersheds in our study region will tend to shift their states (from fair to poor water quality) in general. However, this transition time will be delayed if policies or regulations that control the number of land development events are adopted by, for example, government agencies. In particular, experimental results (the first experiment) suggest that gain in the delayed time of transition due to regulated land development (6.13 years for 80% of reduction in development demand) is higher than loss in the transition time resulting from increased land development (3.91 years for 80% increment in demand). This implies the importance of land conversation policies in the protection of water quality in watershed ecosystems.

Figure 12: Map of simulation results of water quality distribution at the sub watershed level (A-E: five treatments in which distance-decayed coefficients for land owners' utility function were varied; labels over sub watersheds in the maps indicate the averaged time of transition — i.e., the averaged number of years staying within the critical threshold).

Water quality	T1	T2	T3	T4	T5
Good	46	45	45	45	44
Fair	81	82	82	82	83
Poor	32	32	32	32	32

Table 4: Number of sub watersheds with different states of water quality in experiment 2.

Figure 13: Time series of percent in forest cover of four sub watersheds for the second experiment (T1-T5: five treatments in which distance-decayed coefficients for land owners' utility function were varied; four sub watersheds were used with IDs of 25, 104, 110, and 153).

- 5.2** Human decision-making processes and subsequent choices often play a fundamental role in driving the landscape dynamics of watershed ecosystems that are modified by human actors. Our experimental results (the second experiment) demonstrate this role of human decision-making. Change in coefficients of utility functions of land owners (based on distance-decayed models) correspond to a continuum from organic land development (high values) to spontaneous development (low values) (Liu et al. 2017, 2014; Torrens 2006). Varied preference of land owners may lead to alternative landscape patterns in terms of watershed conditions such as water quality here. This suggests the necessity of environmental education in terms of water quality protection for decision-makers involved in land development (Giri & Qiu 2016). Environmental education, possibly together with incentives from governments or institutional efforts on land conservation, will increase the awareness of decision makers in the potential negative effect of land development on the water quality of residents' living environments. This will warrant the sustainability of watershed ecosystems shaped by land development.
- 5.3** Further, the critical threshold of water quality driven by LULCC in this study is similar to (or can be explained as) relaxation phenomena (Malanson 2008), in which time-delayed effect exists between the occurrence of an event (e.g., disturbance, stress, or land development here) in a system and response of the system to that event (e.g., change in water quality in this study). The temporal transition patterns of system states are so-called relaxation paths, which are driven by the strength of the event (land development in this study) and the capability that the system adjusts its states to equilibriums (different water quality states in this study). The time that a system stays within the relaxation phase is relaxation time. In our model, the relaxation time of a sub watershed is the number of years it will maintain the state of fair water quality. Figure 14 depicts the distribution of the relaxation time of a specific sub watershed in our study region for the two experiments. As we could observe, the relaxation time of this specific sub watershed tends to be shortened as land development demand (macro level) increases or land owner agents have more preference on land transactions for development (micro level). This shortened relaxation time pattern can be observed for many sub watersheds with initial state of fair water quality (see Figure 10 and 12). However, for sub watersheds in which initial water quality state is good, the relaxation time tends to be delayed in response to increased land development demand and higher agent preference on land transactions. This is because it takes some time for sub watersheds to transit from the good water quality state to fair and this amount of time is shortened for higher development demand or higher preference of land owner agents. As a result, the relaxation time, which corresponds to the number of years sub watersheds stay at the fair water quality state, is longer for higher development demand or higher individual-level preference. Thus, we can see that the agent-based land change model together with the forest cover model of water quality provides substantial support for the exploration of the relaxation effect in terms of the impact of LULCC on water quality in this study. This will be of great help in terms of developing a promoted understanding of the spatiotemporal complexity of watershed ecosystems.

Figure 14: Box plots of number of years that a sub watershed stays within the water quality threshold (A: Experiment 1; B: Experiment 2; sub watershed ID: 110).

Conclusion

- 6.1** In this study, the agent-based land change model was developed to enable investigations on the relationship between LULCC and water quality in eight NC counties. Our study area covers the lower High Rock Lake Watershed region. The agent-based model was designed within a demand-allocation framework, including modules of land demand and spatial allocation. Two types of agents, land owners and land developers, were used to represent the decision-making processes involved with land development. Land owners have preference on selling their lands for land development, which is represented using distance-decayed modeling. Land developers interact with land owners and decide where and how to realize their needs of land development. Interactions between land owners and land developers produce a complex landscape of land development, which inherently drives spatiotemporal patterns of water quality at sub-watershed level in our study region.
- 6.2** The contribution of this article is the use of agent-based modeling approach to investigate the space-time locations of critical threshold of water quality driven by land use and land cover change at various decision levels (macro and micro). Critical threshold or threshold effect is a representative property of complex coupled human and natural system (Liu et al. 2007a). However, studies on where and when threshold effects may occur are inadequate in the literature. The work reported in this article has unique contribution to this thread, particularly, the study of space-time locations of critical threshold within the context of land use and land cover change. The water quality of sub watersheds in this study is represented by a forest cover model, which uses a threshold of 37-48% of percent forest cover that classifies sub watersheds into different states (three: good, fair, and poor water quality). Spatiotemporally explicit land development modifies land cover patterns, which consequentially lead to the transition of water quality in sub watersheds into different states. In particular, those sub watersheds within the threshold are of more interest in terms of the protection of water quality. These sub watersheds experience transient dynamics (Malanson 1999, 2008) that serve as a buffer state of transition between good and poor water quality (most of the time, from good to poor water quality). Our experiments indicate that the length of the transition phase, also known as relaxation time, is influenced by land development at different decision scales (macro-level: land change demand; micro-level: land owner preference). The agent-based land change model enables us to conduct alternative what-if scenario analysis that helps identify complex spatiotemporal patterns of sub watersheds experiencing critical thresholds. Further, the agent-based modeling approach is necessary (and preferred over other modeling approaches such as cellular automata or system dynamics models) because of its ability to supporting explicit representation of decision-making processes of individuals (land owners and land developers here) that are often heterogeneous and uncertain. This allows us to investigate the impact of individual-level behavior on macro-level system response (watershed-level water quality here).
- 6.3** Our agent-based land change model provide efficacious support for the exploration of complex relationship between LULCC and water quality in watershed-level ecosystems. Future work may focus on the following threads. First, we will apply this ABM to other study regions which may have alternative ecoregions (e.g., coastal plain or mountain) to further examine its utility in representing LULCC and associated spatiotemporal patterns of watershed conditions. Second, our current ABM has limitations in the representation of adaptive behavior of decision- or policy makers. Thus, in the future work, we plan to integrate machine learning approaches with

the agent-based model so as to identify optimal or adaptive land conservation strategies, which will greatly support policy makers to prioritize regions in need of, for example, water quality protection. Machine learning approaches such as evolutionary algorithms, neural networks, swarm intelligence, and reinforcement learning hold great promise for the representation of adaptation in land systems (Bone & Dragičević 2010; Tang 2008) and will be used to explore the influence of adaptation in decision making of individuals and policy making of government agencies (e.g., for local government planning) on ecosystem-level response. Further, machine learning approaches can be of great help in the calibration and validation of agent-based modeling (e.g., use of neural networks for calibration of land developer's decision instead of logistic regression; see Pijanowski et al. 2005). Third, our ABM is an integration of agent-based approach, cellular automata, and distance-decayed models together with a set of parameters and input data. In future studies, we will conduct sensitivity analysis and uncertainty analysis to evaluate the relationship between uncertainties in model input or structure and uncertainties in model outcome. This will be of great help in minimizing the occurrence of possible structural mistakes and misspecification of our ABM (O'Sullivan et al. 2016). Fourth, the agent-based land change modeling is computationally demanding, in particular, for the experimentation of scenario analysis. Thus, a parallel computing solution of the agent-based land change model will be developed to accelerate model execution by leveraging high-performance computing power from cutting-edge cyberinfrastructure (NSF 2007; Tang & Jia 2014; Tang et al. 2011).

Model Documentation

This model was implemented in C++, based on the modification of an agent-based simulation platform, GAIASP (Tang 2008). The code of the model is available online here: <https://github.com/wenwu-tang/agentlcm>.

Appendix: The Overview, Design concepts, Details (ODD) Protocol

In this supplementary material, we document our agent-based model (ABM) using the ODD protocol by Grimm et al. (2010).

Purpose

Our ABM aims to simulate land changes in a large watershed (the lower High Rock Lake Watershed, HRLW, in North Carolina, USA), thus to identify the space-time locations of potential regions (sub watersheds) that may experience critical threshold of water quality.

Entities, state variables, and scales

There are two types of agents in the model: land owners and land developers. Both land owners and land developers are hypothetical agents situated within each county. Also, we assume each land unit (grid cell here) that is covered by farmland, forest, or natural land is managed by a land owner agent, and can be acquired by a land developer for residential development. Land developers are in charge of decision making in terms of where, how, and how many to acquire land for development based on his/her utility function represented by a logistic regression model. The land owners decide whether to sell their lands for development based on his/her perception on surrounding environments when contacted by a land developer. The utility function of a land owner agent to sell land is represented using a distance-decayed model.

The grid cells in the watershed serve as spatial context where land owners and land developers interact with each other. The state variables of land owners, land developers, and grid cells are listed in Table 5.

One time step in the model represents one year in reality. We ran the model for 40 years which covers three periods: 2001-2006 for parameters calibration, 2006-2011 for model performance validation, and 2011-2040 for scenario-based future projection. The spatial resolution of the raster grid is 30 m × 30 m, resulting in 4,378 rows and 4,746 columns in the landscape of interest.

Symbol	Description (unit)	Value
Grid cell		
k	Land cover type	Farmland, forest, natural, water, urban
(x, y)	Coordinate of the grid cell (m)	
x_1	Elevation (m)	
x_2	Slope (%)	
x_3	Stream density (m/m ²)	
x_4	Distance to stream (m)	
x_5	Distance to major road (m)	
x_6	Distance to city (m)	
x_7	Development pressure	
Land owner agent		
i	Identifier of the land owner agent	
j	Identifier of county that land owner agent located in	[1, 8]
$\alpha(i, j, k)$	Coefficient that controls magnitude of the utility of the land owner agent i in county j with respect to the land cover type k of his/her land parcel	
$\beta(i, j, k)$	Coefficient that is associated with the steepness of the distance-decayed function representing the utility of land owner agent i in county j with respect to the land cover type k of his/her land parcel	
$u_owner(i, j, k)$	Utility of land owner agent i in county j with respect to the land cover type k of his/her land parcel	[0, 1]
Land developer agent		
$B_{0(j)}, B_{1(j)}, \dots, B_{m(j)}$	Coefficients for the corresponding driving factors in county j	
$u_developer(i, j)$	Utility of a land developer agent i in county j	

Table 5: Overview of the state variables that characterize the agents and their situated environments in the ABM

Process overview and scheduling

The time step is one year in our ABM. The model starts from 2001 and run into 2006 for model calibration, to 2011 for model validation, and to 2040 for future projection under different scenarios. The overall processes in the ABM are presented in Figure 15. These processes mainly include 1) acquire land for residential development, which includes four major steps: development initialization, negotiation with land owner, demand estimation, and patch growing; 2) sell land for residential development; and 3) estimation of water quality.

Figure 15: The overall processes of the ABM.

Design concepts

Basic principles

The model is the spatial representation of the urban development process in real-world, which involves the decision-making of land developers and land owners, as well as their direct and indirect interactions. In this study, we focus on the land conversion process from farmland, forest, and natural land to urban land. A land developers decide on whether to acquire a feasible land for residential development based on his/her utility function which is calculated from a set of driving factors using a logistic regression model. A land owner can be contacted by a land developer, and make his/her own decision on whether to sell land for development based on his/her utility function which is a distance-decayed abstraction of the surrounding environments.

Emergence

The spatial pattern of residential development in the lower HRLW, especially in each sub watershed, emerges from the interactions between land developer agents and land owner agents. The results of residential development in the previous period update the surrounding environment of agents, thus influencing their decision-making in subsequent steps.

Adaptation

The utility of land developer is an adaptive trait changing along with consecutive residential development activities in surrounding environments. So does the utility function of the land owners.

Objectives

Land owners closer to existing urban areas have higher utility of land conversion, thus more likely to sell their lands. Also, lands with more residential development in surrounding environment are more likely to be converted by land developers.

Learning

The ABM involves no learning attributes.

Prediction

The ABM involves no prediction attributes.

Sensing

Land developer agent can sense the state variables of grid cells that he/she intends to convert for estimating his/her utility. Land owner can sense the location and land cover type of the land that he/she manages, as well as the surrounding residential development, thus to calculate distance to nearest development areas in order to estimate his/her utility of selling land.

Interaction

Land developer has to directly contact a land owner to ask whether the owner will sell his/her land for residential development. Second, the spatial pattern of residential development emerging from interactions between land developers and land owners can influence their decision-making, which indicates indirect interactions between agents through the shared environment.

Stochasticity

First, there is a collection of utility functions of land owner agents in a county, which represents the heterogeneity of agent decision-making. For a specific land owner agent, his/her utility function is randomly drawn from the decision pool of land owner agents in the county. Second, the land selling process is stochastic. A random probability is drawn and compared against the utility value of the land owner agent to decide whether he/she will sell the land. Third, a land developer agent conducts a random inspection to find a land that is feasible to initialize a residential development activity. Fourth, the area (or size) of the new patch generated by land developer is randomly drawn from empirical distribution of area of new development patches from 2001 to 2006. Fifth, during the patch growing process, whether a neighboring cell will be added into the new patch is determined stochastically, based on a composite utility that is the multiplication of the utility of a land developer and that of a land owner with respect to the neighboring cell.

Collectives

A patch which indicates an urban land parcel in real-world is a collective of grid cells, whose area is randomly drawn from observed distribution of new patches generated by historical development activities.

Observation

In each time step, the model observes the total amount of lands that are converted to urban in each county. At the end of the simulation, the model records the percent of forest and urban cover in each sub watershed to estimate the state of water quality.

Initialization

The ABM started from 2001 and ran to 2006 for model calibration and to 2011 for model validation. The total area of our study region is 10,823 km² with size of 4,378 × 4,746 in terms of number of rows and columns. The region is administrated by 8 counties and is divided into 159 sub watersheds. Each grid cell covered by farmland, forest, or natural is assumed to be managed by a land owner agent. The collective of utility functions (the distance-decayed model) of land owners was calibrated using the land conversion data from 2001-2006 in each county. Land developers lead on the residential development process in each county. The utility function (the logistic regression model) of land developers in a county was calibrated using empirical land change data in the county from 2001 to 2006. The factors used to fit the logistic regression model include elevation, slope, distance to cities, distance to major roads, distance to streams, stream density, and development pressure. Neighborhood size for the calculation of development pressure is 35 by 35 cells. A search radius of 500 meters was used to derive stream density. The size of samples to fit the logistic regression model is 2,000 (1,000 for urban, 1,000

for nonurban) in each county and stratified random sampling scheme was used. We use the observed patch of change of the entire study region from 2001 to 2006 to derive the empirical distribution that a land developer agent will use to determine the area of new patch of development. We used the area of land conversion from 2001 to 2006 in each county to represent the annual land development demand during the calibration phase.

Input data

Three years (2001, 2006, and 2011) of land cover data were used to calibrate and validate the ABM, including the utility functions of land developers and land owners, and the structure of the ABM. Data of road and stream networks, locations of cities or towns, and DEM were used to calculate the factors that impact decisions of land developer agent. The spatial extent of HRLW, the sub watersheds in this region, and the county boundaries were from the USDA National Hydrography Dataset. The land cover data were retrieved from U.S. National Land Cover Database (NLCD). Elevation data is from NASA SRTM data.

Submodels

(1) Acquire land for residential development:

The process of acquiring land for residential development consists of four major steps:

1) Development initialization

The land developers initialize the acquisition of land for residential development by searching for an initial location that is feasible under the guidance of his/her utility function. The utility function was derived through a logistic regression analysis between historical residential development activities and a set of driving factors, which can be represented as in the following equations:

$$u_developer(i, j) = (1 + e^{-y(i, j)})^{-1} \quad (8)$$

$$y(i, j) = \beta_0(j) + \beta_1(j)x_1 + \cdots + \beta_m(j)x_m \quad (9)$$

where $u_developer(i, j)$ is the utility of a land developer agent i in county j . $x_1 - x_m$ are the driving factors (m : the number of driving factors). $\beta_1 - \beta_m$ are coefficients for the corresponding driving factors. β_0 is the intercept of the logistic regression model.

Land developer agents in a county share the same utility function calibrated using empirical land change data in the county from 2001 to 2006. Specifically, a land developer agent inspect a land that is feasible for conversion in a stochastic manner (based on the development probability of the land). In this study, those areas that are water bodies, wetlands, or protected areas are excluded from development.

2) Negotiation with land owner

Once the land developer selects a feasible initial location (a grid cell) for development, he/she has to contact the land owner dwelling in this cell to negotiate about the selling of the land. The land owner will decide whether to sell the land for residential development based on his/her utility. If the land owner agrees to sell the land, then the land developer will initialize a new patch of residential development with specified area from this location.

3) Demand estimation

If the land developer and the land owner reach an agreement on initializing a residential development activity, the land developer will specify the amount of land conversion, i.e. the area of the new patch. In this study, the area of the new patch is randomly drawn from a pool of patch size derived from the observed patches of change of the entire study region from 2001 to 2006 (calibration period).

4) Patch growing

A patch growing mechanism is used to form the new patch with specified area: neighboring lands (cells) are iteratively added into the new patch until the specified patch area is reached, or the annual demand of residential development in a county is satisfied. Similarly, each time a new cell is to be included in the new patch, the land developer has to negotiate with the land owner. In other words, the probability that a neighboring land cell is added into the development patch is the composite utility that the land developer agent decides on land acquisition multiplied by the land owner agent's utility on land transactions at that cell.

(2) Sell land for residential development

The land owner decides whether to sell his/her land (including farmland, forest, and natural in this study) during the negotiation with land developer according to the utility from land transaction. The utility function is represented by a distance-decayed model, which is formulated as follow:

$$u_owner(i, j, k) = e^{-\alpha(i, j, k)} e^{\beta(i, j, k) * d} \quad (10)$$

where $u_{owner}(i, j, k)$ is the utility function of land owner agent i in county j with respect to the land cover type k of his/her land parcel ($k = 1$ for forest/natural; $k = 0$ for farmland). $\alpha(i, j, k)$ is the coefficient that controls magnitude of the utility of the land owner agent (increase in α will lower the magnitude of the utility). $\beta(i, j, k)$ is the coefficient that is associated with the steepness of the distance-decayed function representing the utility of the agent. d is the distance of the land owned by the agent to the nearest urban area. In this study, we assume that each land owner has his/her unique utility function, thus all land owners in the watershed generate a collection of utility functions. The pool of utility functions of land owner agents in county j , noted as $D(j)$, can be formulated as:

$$D(j) = \{(\alpha_{ijk}, \beta_{ijk}) \mid \alpha_{ijk}[\alpha_{ljk}, \alpha_{hjk}], \beta_{ijk}[\beta_{ljk}, \beta_{hjk}]\} \quad (11)$$

This collective of utility functions of land owner agents in each county can be calibrated by fitting the distance-decayed function using the empirical land conversion data in terms of distance to nearest urban land cover. Here we used the land conversion data from 2001-2006 in each county to calibrate the utility functions, and the 95% confidence intervals of the fitted coefficients was used to represent the pool of decision rules of land owner agents, which becomes Equation 3:

$$D(j) = \{(\alpha_{ijk}, \beta_{ijk}) \mid \alpha_{ijk} \in CI(\hat{\alpha}(j, k)), \beta_{ijk} \in CI(\hat{\beta}(j, k))\} \quad (12)$$

where $\hat{\alpha}(j, k)$ and $\hat{\beta}(j, k)$ are the fitted coefficients of the distance decayed model for land owner agents in county j with respect to land cover type k . $CI(.)$ denotes the confidence interval of the corresponding variables. During each negotiation process between the land owner and the land developer, the land owner agents randomly draw an utility function for the pool, and compare the his/her utility with a random probability. If the utility is larger, then he/she will sell the land. Otherwise, he/she will not.

(3) Estimation of water quality:

The water quality in the HRLW was related to the percent of forest cover in a sub watershed. The relationship was established based on a empirical analysis of five years of land cover data and benthic macroinvertebrate data. Sub watersheds with percent forest cover higher than 48% (or lower than 37%) correspond to good (or poor) water quality. Any sub watersheds with the percentage of forest cover within the range of 37-48% are at the tipping point with fair water quality that may be shifted to either good or poor states.

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. *Annals of the Association of American Geographers*, 95(1), 54–79
- An, L., Zvoleff, A., Liu, J. & Axinn, W. (2014). Agent-based modeling in coupled human and natural systems (CHANS): Lessons from a comparative analysis. *Annals of the Association of American Geographers*, 104(4), 723–745
- Aquilué, N., De Cáceres, M., Fortin, M.-J., Fall, A. & Brotons, L. (2017). A spatial allocation procedure to model land-use/land-cover changes: Accounting for occurrence and spread processes. *Ecological Modelling*, 344, 73–86
- Barbour, M. T., Gerritsen, J., Snyder, B. D. & Stribling, J. B. (1999). Rapid bioassessment protocols for use in streams and wadeable rivers: Periphyton, benthic macroinvertebrates and fish (vol. 339). US Environmental Protection Agency, Office of Water Washington, DC
- Bone, C. & Dragičević, S. (2010). Simulation and validation of a reinforcement learning agent-based model for multi-stakeholder forest management. *Computers, Environment and Urban Systems*, 34(2), 162–174
- Brown, D. G., Page, S., Riolo, R., Zellner, M. & Rand, W. (2005a). Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science*, 19(2), 153–174
- Brown, D. G., Riolo, R., Robinson, D. T., North, M. & Rand, W. (2005b). Spatial process and data models: Toward integration of agent-based models and GIS. *Journal of Geographical Systems*, 7(1), 25–47
- Council, N. R. (2014). *Advancing Land Change Modeling: Opportunities and Research Requirements*. National Academies Press

- Epstein, J. M. & Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, DC: Brookings Institution Press
- Fielding, A. H. & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1), 38–49
- Gerow, T. A., Jones, D. G. & Tang, W. (2015). Assessment of forest cover in the High Rock Lake Watershed. Paper presented at the Water Resource Research Institute 2015 Annual Conference, Raleigh, NC, USA
- Gerow, T. A., Jones, D. G. & Tang, W. (2016). Assessing the relationship between forests and water in the High Rock Lake Watershed of North Carolina. In C. E. Stringer, K. W. Krauss & J. S. Latimer (Eds.), *Headwaters to Estuaries: Advances in Watershed Science and Management. Proceedings of the Fifth Interagency Conference on Research in the Watersheds. March 2-5, 2015, North Charleston, South Carolina*, (p. 302). Asheville, NC: US Department of Agriculture Forest Service, Southern Research Station
- Giri, S. & Qiu, Z. (2016). Understanding the relationship of land uses and water quality in twenty first century: A review. *Journal of Environmental Management*, 173, 41–48
- Gould, P. R. (1969). Spatial diffusion. Resource Paper No. 4. Washington, DC: Association of American Geographers
- Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J. & Railsback, S. F. (2010). The ODD protocol: A review and first update. *Ecological Modelling*, 221(23), 2760–2768
- Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., Thulke, H.-H., Weiner, J., Wiegand, T. & DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, 310(5750), 987–991
- Groeneveld, J., Müller, B., Buchmann, C., Dressler, G., Guo, C., Hase, N., Hoffmann, F., John, F., Klassert, C., Lauf, T., Liebelt, V., Nolzen, H., Pannicke, N., Schulze, J. & Weise, H. (2017). Theoretical foundations of human decision-making in agent-based land use models – A review. *Environmental Modelling & Software*, 87, 39–48
- Hagerstrand, T. (1968). *Innovation Diffusion as a Spatial Process*. Chicago, IL: University of Chicago Press
- Haidary, A., Amiri, B. J., Adamowski, J., Fohrer, N. & Nakane, K. (2013). Assessing the impacts of four land use types on the water quality of wetlands in Japan. *Water Resources Management*, 27(7), 2217–2229
- Kauffman, G. J. & Brant, T. (2000). The role of impervious cover as a watershed-based zoning tool to protect water quality in the Christina river basin of Delaware, Pennsylvania, and Maryland. *Proceedings of the Water Environment Federation*, 2000(6), 1656–1667
- Kok, K., Farrow, A., Veldkamp, A. & Verburg, P. H. (2001). A method and application of multi-scale validation in spatial land use models. *Agriculture, Ecosystems & Environment*, 85(1-3), 223–238
- Lauf, S., Haase, D., Hostert, P., Lakes, T. & Kleinschmit, B. (2012). Uncovering land-use dynamics driven by human decision-making – a combined model approach using cellular automata and system dynamics. *Environmental Modelling & Software*, 27, 71–82
- Li, X., Gong, P., Yu, L. & Hu, T. (2017). A segment derived patch-based logistic cellular automata for urban growth modeling with heuristic rules. *Computers, Environment and Urban Systems*, 65, 140–149
- Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., Pell, A. N., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C. L., Schneider, S. H. & Taylor, W. W. (2007a). Complexity of coupled human and natural systems. *Science*, 317(5844), 1513–1516
- Liu, J., Dietz, T., Carpenter, S. R., Folke, C., Alberti, M., Redman, C. L., Schneider, S. H., Ostrom, E., Pell, A. N., Lubchenco, J. et al. (2007b). Coupled human and natural systems. *AMBIO: A Journal of the Human Environment*, 36(8), 639–649
- Liu, X., Liang, X., Li, X., Xu, X., Ou, J., Chen, Y., Li, S., Wang, S. & Pei, F. (2017). A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects. *Landscape and Urban Planning*, 168, 94–116
- Liu, X., Ma, L., Li, X., Ai, B., Li, S. & He, Z. (2014). Simulating urban growth by integrating landscape expansion index (LEI) and cellular automata. *International Journal of Geographical Information Science*, 28(1), 148–163

- Malanson, G. P. (1999). Considering complexity. *Annals of the Association of American Geographers*, 89(4), 746–753
- Malanson, G. P. (2008). Extinction debt: Origins, developments, and applications of a biogeographical trope. *Progress in Physical Geography*, 32(3), 277–291
- Mandaville, S. (2002). Benthic Macroinvertebrates in Freshwaters: Taxa Tolerance Values, Metrics, and Protocols. *Citeseer*
- Matthews, R. B., Gilbert, N. G., Roach, A., Polhill, J. G. & Gotts, N. M. (2007). Agent-based land-use models: A review of applications. *Landscape Ecology*, 22(10), 1447–1459
- McGarigal, K., Plunkett, E. B., Willey, L. L., Compton, B. W., DeLuca, W. V. & Grand, J. (2018). Modeling non-stationary urban growth: The SPRAWL model and the ecological impacts of development. *Landscape and Urban Planning*, 177, 178–190
- Meals, D. W., Dressing, S. A. & Davenport, T. E. (2010). Lag time in water quality response to best management practices: A review. *Journal of Environmental Quality*, 39(1), 85–96
- Meentemeyer, R. K., Tang, W., Dorning, M. A., Vogler, J. B., Cuniffe, N. J. & Shoemaker, D. A. (2013). FUTURES: Multilevel simulations of emerging urban – Rural landscape structure using a stochastic patch-growing algorithm. *Annals of the Association of American Geographers*, 103(4), 785–807
- Morrill, R. L. & Pitts, F. R. (1967). Marriage, migration, and the mean information field: A study in uniqueness and generality. *Annals of the association of American Geographers*, 57(2), 401–422
- Mustafa, A., Cools, M., Saadi, I. & Teller, J. (2017). Coupling agent-based, cellular automata and logistic regression into a hybrid urban expansion model (HUEM). *Land Use Policy*, 69, 529–540
- Ng, T. L., Eheart, J. W., Cai, X. & Braden, J. B. (2011). An agent-based model of farmer decision-making and water quality impacts at the watershed scale under markets for carbon allowances and a second-generation biofuel crop. *Water Resources Research*, 47(9)
- NSF (2007). Cyberinfrastructure vision for 21st century discovery. Report of NSF Council. Available at: http://www.nsf.gov/od/oci/ci_v5.pdf
- O’Sullivan, D., Evans, T., Manson, S., Metcalf, S., Ligmann-Zielinska, A. & Bone, C. (2016). Strategic directions for agent-based modeling: Avoiding the YAAWN syndrome. *Journal of Land Use Science*, 11(2), 177–187
- Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J. & Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers*, 93(2), 314–337
- Pijanowski, B. C., Pithadia, S., Shellito, B. A. & Alexandridis, K. (2005). Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. *International Journal of Geographical Information Science*, 19(2), 197–215
- Pontius, R. G., Boersma, W., Castella, J.-C., Clarke, K., de Nijs, T., Dietzel, C., Duan, Z., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C. D., McConnell, W., Sood, A. M., Pijanowski, B., Pithadia, S., Sweeney, S., Trung, T. N., Veldkamp, A. T. & Verburg, P. H. (2008). Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science*, 42(1), 11–37
- Pontius Jr, R. G. (2002). Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering and Remote Sensing*, 68(10), 1041–1050
- Pontius Jr., R. G., Huffaker, D. & Denman, K. (2004). Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling*, 179(4), 445–461
- Pontius Jr, R. G. & Millones, M. (2011). Death to kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15), 4407–4429
- Pontius Jr, R. G. & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems & Environment*, 85(1-3), 239–248

- Pontius Jr, R. G., Shusas, E. & McEachern, M. (2004). Detecting important categorical land changes while accounting for persistence. *Agriculture, Ecosystems & Environment*, 101(2-3), 251–268
- Schueler, T. R. (1994). Performance of grassed swales along east coast highways. *Watershed Protection Techniques*, 1(3), 122–123
- Schueler, T. R., Fraley-McNeal, L. & Cappiella, K. (2009). Is impervious cover still important? Review of recent research. *Journal of Hydrologic Engineering*, 14(4), 309–315
- Tang, W. (2008). Simulating complex adaptive geographic systems: A geographically aware intelligent agent approach. *Cartography and Geographic Information Science*, 35(4), 239–263
- Tang, W. & Jia, M. (2014). Global sensitivity analysis of a large agent-based model of spatial opinion exchange: A heterogeneous multi-gpu acceleration approach. *Annals of the Association of American Geographers*, 104(3), 485–509
- Tang, W., Wang, S., Bennett, D. A. & Liu, Y. (2011). Agent-based modeling within a cyberinfrastructure environment: A service-oriented computing approach. *International Journal of Geographical Information Science*, 25(9), 1323–1346
- Taylor, P. J. & Openshaw, S. (1975). Distance decay in spatial interactions. Paper presented at the *Concepts and Techniques in Modern Geography* conference
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(sup 1), 234–240
- Torrens, P. M. (2006). Simulating sprawl. *Annals of the Association of American Geographers*, 96(2), 248–275
- Valbuena, D., Verburg, P. H., Bregt, A. K. & Ligtenberg, A. (2009). An agent-based approach to model land-use change at a regional scale. *Landscape Ecology*, 25(2), 185–199
- van Vliet, J., Bregt, A. K., Brown, D. G., van Delden, H., Heckbert, S. & Verburg, P. H. (2016). A review of current calibration and validation practices in land-change modeling. *Environmental Modelling & Software*, 82, 174–182
- Verburg, P. H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V. & Mastura, S. S. (2002). Modeling the spatial dynamics of regional land use: The CLUE-S model. *Environmental Management*, 30(3), 391–405