

Modeling Historical Land Use Changes at A Regional Scale: Applying Quantity and Locational Error Metrics to Assess Performance of An Artificial Neural Network-Based Back-Cast Model

A. Tayyebi^{1,2,3,*}, A. H. Tayyebi⁴, B. K. Pekin^{3,5}, H. Omrani⁶, and B. C. Pijanowski³

¹University of California-Riverside, Center for Conservation Biology, 900 University Ave, Riverside, CA 92521, USA

²University of California-Riverside, Department of Botany and Plant Sciences, 900 University Ave, Riverside, CA 92521, USA

³Purdue University, Department of Forestry and Natural Resources, West Lafayette, IN 47907, USA

⁴University of Houston, Department of Civil and Environmental Engineering, Houston, TX 77204, USA

⁵San Diego Zoo Global, Institute for Conservation Research, 15600 San Pasqual Valley Road, Escondido, CA 92027, USA

⁶Urban Development and Mobility Department, Luxembourg Institute of Socio-Economic Research, Luxembourg

Received 1 May 2015; revised 9 October 2015; accepted 5 December 2015; published online 3 February 2017

ABSTRACT. Land-use legacies, the effects of past land use changes (LUCs) on current social and biophysical processes, can persist for hundreds to thousands of years. Although spatial and temporal data are currently available at a continental scale, they are limited for investigating LUC legacies. The limitations of historical data include a lack of temporal coverage, inaccessibility at coarse resolution, availability only for few land use classes and accessibility for only specific regions. Despite the limitation in data availability, there is an urgent need to develop a data-intensive model that can back-cast multiple historic LUCs at regional scales. We developed a back-cast model for generating historic land use maps with multiple land use classes at a regional scale using a high performance computing (HPC) platform. We trained and tested the model using Retrofit Land Cover Change data between 2001 and 1992 in a backward manner at 1 km resolution for three land use categories (urban, forest and agriculture) in the Ohio River Basin (ORB) of United States. We also developed a calibration metric to assess quantity and locational errors for multiple LUCs simultaneously. Results showed that the range that the model underestimated and overestimated the quantity of LUCs was -0.05% to +0.11%. The persistence (over 95%) and location (over 80%) accuracies of multiple LUCs were quantified. We then simulated multiple LUCs annually between 2001 and 1980 using 2001 as the base year across the ORB. We describe how the output of our back-cast model can be coupled with other environmental models to assess the impact of land use change on ecosystem services.

Keywords: back-casting multiple land use changes, regional scale, high performance computing, quantity and location error, environmental management

1. Introduction

Land use legacies, or the effects of past land uses on current social and biophysical processes (Feurdean et al., 2009) have been shown to impact stream hydrology (Pijanowski et al., 2007; Tayyebi et al., 2015) and water quality (Utz et al., 2009; Liu and Tong, 2011), climate change (Yang et al., 2014; Tayyebi and Jenerette, 2016), soil properties and agricultural yields (Islam and Weil, 2000; Tayyebi et al., 2016a and 2016b), and biodiversity (Pekin and Pijanowski, 2012). Decision makers have recognized that land-use legacies continue to influence ecosystem structure for decades to centuries (Foster et al., 2003; Azari et al., 2015). Consequently, recognition and knowledge of

these historical changes is crucial to the sustainable management of ecosystems.

Land use change (LUC) models have been used to produce future and past land use maps for the past thirty years (Verburg et al., 2002). LUC models often incorporate separate modules to determine the quantity of LUC and the location of cells in the map that experience LUC. There are many similarities in forward and backward LUC modeling such as determining the location and the quantity of each land use class for the given area, and training and testing the model (Ray and Pijanowski, 2010). As a result, LUC forecasting models are often re-engineered to simulate historical land uses (Hill and Aspinall, 2000). Some researchers have used these back-cast models to produce maps showing the historical spatial distribution of different land uses at local scale. For example, Pijanowski et al. (2007) coupled an artificial neural network (ANN) based land-use back-cast model to a groundwater transport model to examine how historical land uses might influence water quality (Ray et al., 2012). In other study, Pontius et al. (2001) simulated the location of forest disturbance in Costa Rica in a backward ma-

* Corresponding author. Tel.: +1 765 412 1591; fax: +1 608 262 3322.

E-mail address: amin.tayyebi@gmail.com (A. Tayyebi).

nner for several decades.

Large-scale back-cast LUC data can provide insights into regional patterns of LUC (Reid et al., 2010) as well as LUC trends effects on a variety ecosystem services (Jia et al., 2014). However, accessing spatially explicit data on past LUC trends is not always straightforward and spatial and temporal land use data across a variety of regions are still rare (Kumar et al., 2013). There are a variety of reasons for these limitations. First, although satellite images and aerial photos are important source of information for historical LUC modeling, aerial photographs are available since the late 1930s (Pauleit et al., 2005) and civilian satellite images (Yang and Lo, 2002) date back to the early 1970s. In addition, processing imagery at regional scale (satellite images after 1970s and aerial photos after 1930s) to identify land use classes is exceedingly time consuming and expensive (Vogelmann et al., 1998).

Second, although a few studies exist on back-casting and forecasting LUC at regional scale (Waisanen and Bliss, 2002), the investigation of back-casting LUC at regional and fine scales is limited. Previous back-cast studies used coarse-resolution data mostly at the county scale (Theobald, 2001). This is due to the fact that recorded (i.e., quantity) data are generally only available at county scale. For example, Waisanen and Bliss (2002) and Theobald (2001) analyzed existing county-level data to assess the impact of past LUCs on ecosystem services. There are also similar studies for forecasting LUC based on coarse-resolution data at a regional scale. For example, the CLUE model was modified (CLUE-S) to forecast regional LUC in Europe using empirically quantitative relations between land use and its driving forces (Veldkamp and Verburg, 2004). In a similar study, urbanization at regional scales in the northeastern United States was recently simulated with the SLEUTH-3R model (Jantz et al., 2010).

Third, most LUC analyses have been conducted only for discrete time periods (e.g., Brown et al., 2005) or only for a small region (e.g., Ray and Pijanowski, 2010). For example, Brown et al. (2005) examined LUC trends using housing and farm data for each decade during the latter half of the 20th century to understand the trend of LUCs. Hammer et al. (2004) used housing data for each decade from 1940 to 1990 to show the housing density change in United States. Lepczyk et al. (2007) quantified hotspots of housing changes in United States over a 60-year time (1940 ~ 2000). Ray and Pijanowski (2010) showed that one location could pass through several land use classes over the span of a century in a regional watershed in Michigan.

The Land transformation model (LTM) has been one of the commonly used LUC models during last two decades. LTM integrates GIS and artificial neural networks to back-cast LUC (Pijanowski et al., 2007). However, previous back-cast applications of LTM model were limited in their ability to simulate multiple LUCs at the same time. Thus, each land use class was trained separately and simple what-if rules were applied to determine actual land use transitions (Ray and Pijanowski, 2010). ANN based models that forecast LUC have recently been re-configured to simulate multiple land use classes simultaneously

for future scenarios (Tayyebi and Pijanowski, 2014). Other new developments in land use forecast models include the generation of future land use maps over regional scales using high performance computing (HPC) (Pijanowski et al., 2014) and the use of hierarchical subcomponents that are responsible for calculating the quantity of LUC within relevant boundary units (Tayyebi et al., 2013). The main objective of this study is to combine all of these elements to develop a back-cast model for generating annual historical land use maps with multiple land use classes at regional scale using a HPC platform. The back-cast model presented in this study is different from previously developed back-cast models in several ways. First, the new model provides separate suitability maps for multiple LUCs simultaneously according to the number of the desired outputs. Second, the quantity of LUC is adjusted for individual sub-regions within the larger study area (i.e., county boundaries here). Third, our back-cast model is integrated with HPC to run at regional scale. Finally, the back-cast model can simulate annual land use to capture multistep LUCs across time.

Other challenges in modeling LUC is that within a given region multiple LUCs occur simultaneously (Tayyebi and Pijanowski, 2014; Song et al., 2015). For example, it is quite common for some areas to be converted from agriculture to urban while nearby forests are converted to agriculture (Brown et al., 2009). However, most calibration metrics have been developed for binary outcome and unbiased calibration metric for multiple LUCs is lacking. Nonetheless, multiple LUCs can be evaluated using contingency tables. One way to do this is to convert multiple LUCs into numerous binary LUCs that are evaluated using binary calibration metrics (Pontius et al., 2011). Thus, it is required to calibrate each land use class separately. Alternatively, binary LUC calibration metrics can be modified to assess multiple LUCs. As a second objective of this study, we took the latter approach, since it minimizes the total number of calibration runs that need to be executed compared to the former approach, and developed a metric to simultaneously characterize model performance, quantity and locations errors, for multiple LUCs.

2. Developing A Back-Cast Model at Regional Scale

The back-cast model developed in this study proceed through a series of phases. In phase 1, an initial data preparation component that is conducted in advance by a GIS professional (Figure 1, Item 1). In phase 2, the locations of multiple LUCs are determined based on the ANN algorithms (Figure 1, Item 2). In phase 3, users are required to introduce LUC scenarios for back-casting based on their understanding for the given study area (Figure 1, Item 3). The back-cast LTM has also defaults LUC scenarios that users can take advantage of them. Phase 4 involves choosing the scale or boundary to divide the data at regional scale to the smaller pieces (Figure 1, Item 4). In phase 5, the quantity of LUC for each LUC scenario is calculated and then used to transition cells according to their suitability separately within each spatial units (Figure 1, Item 5). Phase 6 involves running the back-cast LTM at regional

scale using high performance computing, suitability maps generated from model for each spatial units, LUC scenarios and quantity of LUCs (Figure 1, Item 6). Phase 7 involves using the new developed accuracy metrics in this study to evaluate the goodness of back-cast model (Figure 1, Item 7). The last step involves back-casting multiple LUCs for regional planning (Figure 1, Item 8).

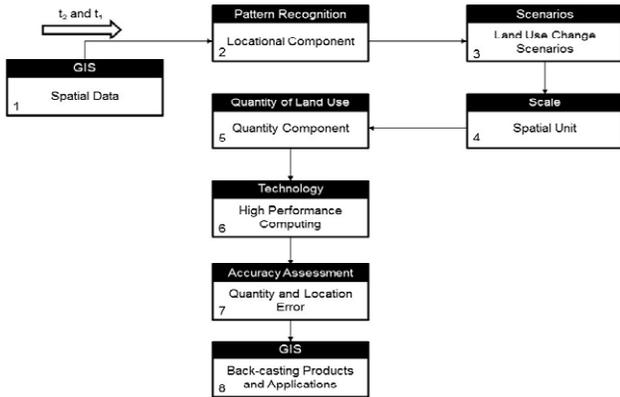


Figure 1. The main components of the developed back-cast model.

2.1. Location of Land Use Change

ANNs are machine-learning techniques capable of capturing the non-linear processes behind land transformations (Pijanowski et al., 2002). The ANN based approach (Figure 1, Item 2) incorporates socio-environmental factors, usually derived from GIS and remote sensing data (Figure 1, Item 1), to back-cast LUCs (Pijanowski et al., 2002). The back-cast ANN uses two reference land use maps separated in time in a reverse manner (t_2 and t_1 instead of t_1 and t_2). In contrast to the forward model, the factors or drivers that impact LUC in the past should be in t_2 instead of t_1 and the output of the model is LUC from t_2 to t_1 . Depending on the land use transition, the model generates suitability maps. The suitability map, calculated with the ANN component, shows the probability of cells in the region for LUC. The cells with higher probability are more likely to convert to LUC (Pijanowski et al., 2009). The conversion of suitability map to a LUC simulation map occurs according to the quantity of LUC in backward manner.

The multi-layer perceptron used in this paper typically includes three different layers, the input, hidden, and output layer (Figure 2). Neurons in the hidden layer assign the relationships between the input and output nodes and pass the result through a non-linear activation function to the output neuron or adjacent

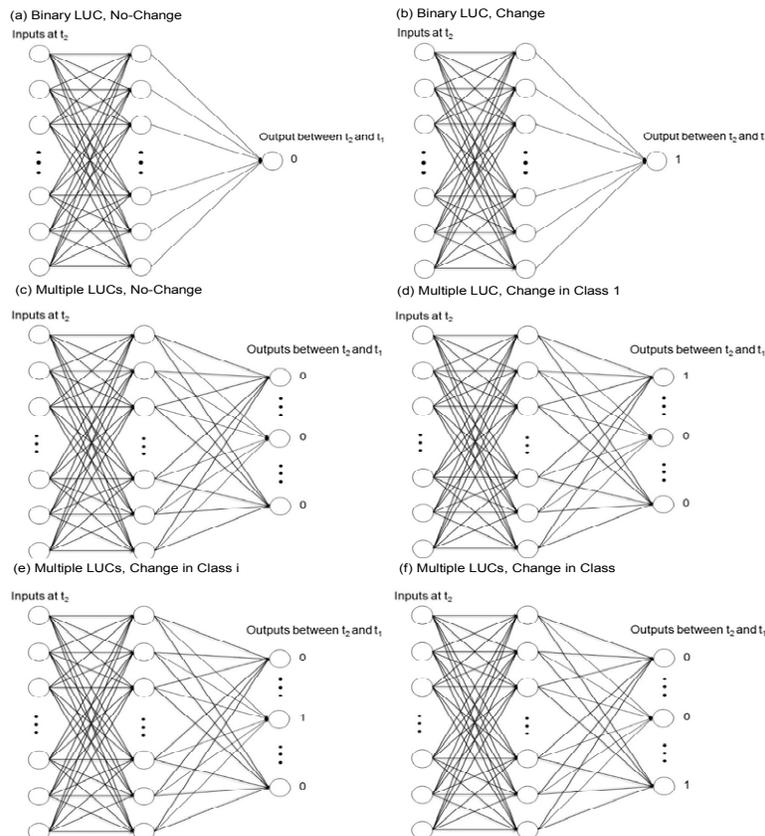


Figure 2. The difference between the model structure and coding scheme of the original back-cast ANN (a and b) and the modified back-cast ANN for simulating multiple LUCs (c to f).

neuron of the corresponding hidden layer. Figures 2a and 2b show the nature of original back-cast ANN with a single output layer. Output with 0 values show cells that were determined to be no-change while output with 1 values show cells that changed between two times in back-ward manner. The original version of the back-cast ANN followed two main changes to back-cast multiple LUCs (Figures 2c to 2f). First, we modified the original structure of the back-cast ANN, which was developed for binary LUC modeling (Figures 2a and 2b) where the number of nodes in output layer is equal to the number of the desired output (Figures 2c to 2f). The advantage of the modeling approach is that it provides separate suitability maps for each outcome according to the number of the desired outputs. Second, we used a combination of k -binary numbers to represent k -category attributes, each associated with one of the transitions. A k -class pattern classification problem can be implemented into a single ANN architecture with k outputs (Figures 2c to 2f). In order to show the state of transition for each land use class, only one of the k numbers in the output layer need to be coded as one while the others stayed zero (Figures 2d to 2f). On the other hand, when all outputs have 0 value, it represents cells that were selected as no-change (Figure 2c).

2.2. Land Use Change Scenarios for Back-Casting

Urbanization, a key LUC, is defined as the expansion of urban land uses, including residential, industrial and commercial (Radeloff et al., 2005). Urbanization can occur in sub-urban areas, including conversion of land in agriculture and forest. Agricultural areas have widespread effects on various ecosystem services as well since agricultural land area in the United States declined during the 20th century (Vesterby and Krupa, 1997). The decline can be attributed to both appropriation of agricultural land for other uses (i.e., urbanization) and abandonment of agriculture on poor quality land (Maizel et al., 1998). Agricultural decline has also led to significant afforestation in rural areas (Brown et al., 2005).

For a better LUC simulation between multiple land use classes, it is necessary to define LUC scenarios (Figure 1, Item 3). It is common to lose urban areas in a backward simulation because urbanization increases in a forward manner due to technological advances, immigration from the countryside, and population growth. Thus, it is to be expected that urban areas decrease in a back-cast land transformation simulation. Many agricultural lands have been converted into urban areas (e.g., agricultural loss) because agricultural land is usually conveniently located in the periphery of urban areas and forest areas have been converted into agricultural lands (e.g., agricultural land use gain) to meet the demand for agricultural goods (Nassauer et al., 2014; Tayyebi et al., 2014a, b). Thus, for a given area where agriculture, forest and urban can be gained or lost in a backward manner (Figure 3).

Agricultural gain: Agricultural gain quantity is less than urban loss quantity (Figure 3, Item 1), and urban cells at t_2 are the first candidates to turn into agriculture. The urban loss suitability map exhibits the locations of urban cells expected to go to agriculture first, while the rest of the urban loss goes to

forest (Figure 3, Item 1; urban to agriculture and urban to forest transition). Thus, urban loss is equal to agriculture and forest gain in this case. However, if the total quantity of urban cells in t_2 cannot satisfy the quantity of agriculture gain, the rest of the cells (e.g., forest cells first) are ranked based on the urban loss suitability map and turn into agriculture cells until the total number of agriculture gain cells is met. Thus, agriculture gain is equal to urban and forest loss in this case (Figure 3, Item 2; urban to agriculture and forest to agriculture transition).

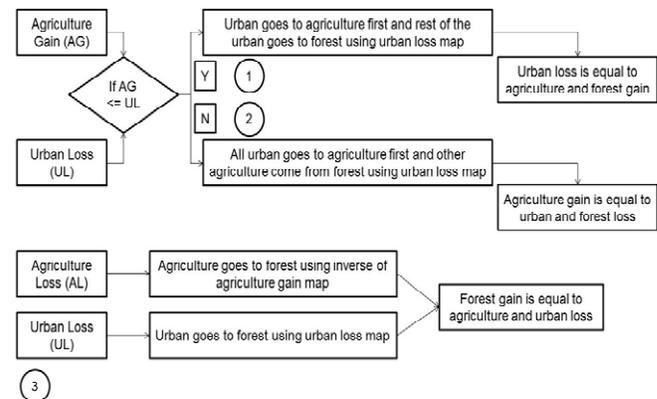


Figure 3. Land change scenarios used for back-casting multiple land use changes.

Agricultural loss: Agriculture cells at t_2 are the first candidates to turn into forest (Figure 3, Item 3; agriculture to forest transition), and the agriculture gain suitability map decides the locations of those agriculture cells that should go to forest first. The urban cells at t_2 could also convert into forest (Figure 3, Item 3; urban to forest transition), with the urban loss suitability map deciding the locations of the urban cells that go to forest first. Thus, forest gain is equal to agriculture and urban loss in this case. These model variants are exclusive, and, thus, conflicts resulting from multiple classifications are prevented (Tayyebi and Pijanowski, 2014). It should be noted that consecutive difference between two decades determine the quality of loss or gain for agriculture, urban and forest land use classes.

2.3. Estimating Quantity of Land Use Change

While historical information is available to be integrated into the back-cast model (e.g., census data and agriculture data), forward LUC modeling requires estimation of the future LUC quantity (Rhemtulla et al., 2007). However, incorporating historical information into back-cast models may not be easy since historical information are often summarized at different scales. For example, United States Census file databases on housing as a proxy of urban quantities are reported every ten years at block scale (US Bureau of the Census, 2001). United States Department of Agriculture historical land in farms data (U.S. Department of Agriculture, 1997) as a proxy of farm areas are typically reported every five years at county scale. In addition, some historical data are collected periodically and users might

be interested in creating land use maps at time periods outside the historical data exist.

Since historical quantities of urban (US Bureau of the Census, 2001) and agriculture (U.S. Department of Agriculture, 1997) land use data are available at block and county scale, respectively, we used county boundaries (Figure 1, Item 4; Figure 4) to define and adjust the amount of land use class changes. The housing and land in farms data at political units were then scaled to the number of cells within each political unit that were needed for each time step (Ray and Pijanowski, 2010). We did two separate back-cast simulations: 1) For the back-cast simulation at t_1 , we fixed the quantity of LUC for each land use category by comparing the reference maps between t_2 and t_1 (Tayyebi et al., 2014) and 2) For annual times between t_2 and t_1 as well as annual times before t_1 , we used external quantity sources (United States Census file databases and United States Department of Agriculture historical land in farms) to determine the quantity of urban and agriculture areas. Since both external quantity sources are collected periodically, we used cubic splines to fit the housing and land in farms to estimate the annual quantity of urban and agriculture areas between t_2 and t_1 as well as before t_1 (Pijanowski et al., 2007).

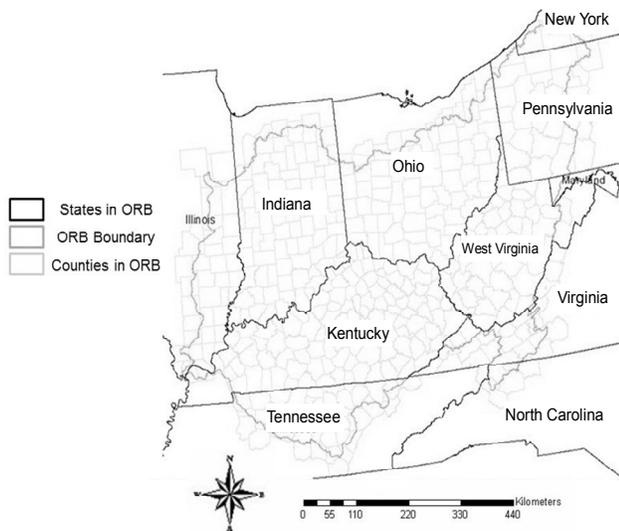


Figure 4. Using county boundaries in back-cast model to determine quantity of land use changes.

2.4. Running Big Data Simulations Using A High Performance Computing Platform

Since we back-cast multiple LUCs at regional scale (e.g., multiple states), we deal with a big dataset, and a large number of input and output files in calibration and testing run (Huang and Chang, 2003). Running LUC models at regional scale requires re-engineering the model so that it may handle larger datasets (Silva and Clarke, 2002; Clarke, 2003; Jantz et al., 2010). We redesigned and ran our back-cast model at a sub continental scale with fine resolution using a new architectural framework in Windows-based HPC cluster environment (Pija-

nowski et al., 2014). In the new architecture of our back-cast model, HPC allows for the handling of large data sets in terms of size and quantity of files and integrates tools that are executed using a variety of scripting languages (Figure 1, Item 6). For example, HPC allows for job-scheduling management of parallel computing resources and tasks, which distributes computational and file allocation across computational resources (CPUs, drives, memory, and software components). In this study, each state is a job and counties within associated state are considered as tasks. HPC uses XML format per job to identify jobs and associated tasks within each job. For example, the Ohio River Basin (ORB) in this study contains 11 jobs, each in separate XML (ORB includes 11 states of USA), for a total of 456 tasks (ORB includes 456 counties within 11 states). The number of tasks within each job depends on the numbers of counties within each state. After submitting jobs in XML format to HPC, the queue manager schedules each job to run on the queue and enable users to monitor progress. Since our back-cast model produces large number of files, we used a standard numbering system from the Federal Information Processing Systems (a two-digit code) and counties (a three-digit code) to track the inputs and outputs of the simulation. Therefore, every county in our study area is given a five-digit integer value (e.g., 55025 for Dane County, Wisconsin; for example slope_55025.asc).

2.5. Complexity of Back-Cast Model

The core of our back-cast model is ANN algorithm. Complexity of ANN can be tracked in three parts (Figure 2): (1) Number of hidden layers: The hidden layers in ANN are responsible to find non-linear patterns in data. The number of hidden layers are arbitrary while most of the previous studies suggested using one hidden layer is enough to produce satisfactory results. More hidden layers increase the complexity of the ANN. (2) Number of neurons within each layer: Each layer consists of series of neurons. The number of neurons in input layer depend on number of inputs (spatial drivers here) while the number of neurons in output layer depend on the number of outputs (land transitions between two times). Number of neurons in hidden is arbitrary although there are a lot of studies in literature suggested that it is the best that the number of neurons in hidden layers be equal the number of neurons in input layer or $2 \times n + 1$ number of neurons in input layer to produce the best results. Since more neurons within layers increase the number of unknown parameters in ANN, this will increase the complexity of the ANN. (3) Number of connections: The connections among neurons across input, hidden and output layers define the topology of ANN. The ANN model used in this study is one of the most complicated since it is a fully connected network which means all neurons in input layer are connected to all neurons in hidden layer, and all neurons in hidden layer are connected to all neurons in output layer. More connections between neurons across layers increase the number of unknown parameters, so this will increase the complexity of the ANN.

3. Developing Quantity and Location Calibration Metric

Model calibration includes two steps (Tayyebi et al., 2014): 1) building a model which is usually called a “training run”, and 2) evaluating a model which referred to as a “testing run”. LUC modelers typically divide the entire data to two parts. The first part, used for training, usually uses a smaller portion of the dataset for training the model. The second part, or testing, uses the rest of the data to assess the model performance. The generated model from the training run was applied to the entire dataset in t_2 to simulate LUC in t_1 using a fixed quantity of change between two reference times. The evaluation approach was used to assess the goodness-of-fit of the back-cast model against reference data in t_1 . Quantity and location error assessments characterize the model’s ability for back-casting multiple LUCs (Figure 1, Item 7).

Table 1a. Contingency Table Comparing the Observed Land Use Map in t_2 with the Observed Land Use Map in t_1

		Observed Map in t_1				
		C ₁	C ₂	...	C _n	Total
Observed Map in t_2	C ₁	O ₁₁	O ₁₂	...	O _{1n}	$\sum_{i=1}^n O_{1i}$
	C ₂	O ₂₁	O ₂₂	...	O _{2n}	$\sum_{i=1}^n O_{2i}$

	C _n	O _{n1}	O _{n2}	...	O _{nn}	$\sum_{i=1}^n O_{ni}$
	Total	$\sum_{i=1}^n O_{i1}$	$\sum_{i=1}^n O_{i2}$...	$\sum_{i=1}^n O_{in}$	

*C shows the land use categories in both land use maps while n shows the numbers of land use classes. O shows the number of cells that remain or change between two maps.

Table 1b. Contingency Table Comparing the Simulated Land Use Map in t_1 with the Observed Land Use Map in t_2

		Simulated Map in t_1				
		C ₁	C ₂	...	C _n	Total
Observed Map in t_2	C ₁	S ₁₁	S ₁₂	...	S _{1n}	$\sum_{i=1}^n S_{1i}$
	C ₂	S ₂₁	S ₂₂	...	S _{2n}	$\sum_{i=1}^n S_{2i}$

	C _n	S _{n1}	S _{n2}	...	S _{nn}	$\sum_{i=1}^n S_{ni}$
	Total	$\sum_{i=1}^n S_{i1}$	$\sum_{i=1}^n S_{i2}$...	$\sum_{i=1}^n S_{in}$	

*C shows the land use categories in both land use maps while n shows the numbers of land use classes. S shows the number of cells that remain or change between two maps.

Table 1c. Quantity Error Calculation

		Simulated Minus Observed Map in t_1			
		C ₁	C ₂	C _n	Total
Observed	C ₁	O ₁₁ - S ₁₁	O ₁₂ - S ₁₂	O _{1n} - S _{1n}	$\sum_{i=1}^n (O_{1i} - S_{1i})$
	C ₂	O ₂₁ - S ₂₁	O ₂₂ - S ₂₂	O _{2n} - S _{2n}	$\sum_{i=1}^n (O_{2i} - S_{2i})$

	C _n	O _{n1} - S _{n1}	O _{n2} - S _{n1}	O _{nn} - S _{nn}	$\sum_{i=1}^n (O_{ni} - S_{ni})$
	Total	$\sum_{i=1}^n (O_{i1} - S_{i1})$	$\sum_{i=1}^n (O_{i2} - S_{i2})$	$\sum_{i=1}^n (O_{in} - S_{in})$	

Due to the large size of data and difficulty of presenting error distribution at a regional scale, we used contingency tables to summarize the quantity and locational errors (Figure 1, Item 7). Here, we developed quantity and locational metric for multiple land use classes simultaneously. We used the following two preparation steps to achieve this goal. First (Table 1, Item A), we used a contingency table to compare the reference maps (we called Observed_Change map after here) between t_2 and t_1 to map the persistence and LUC, and second (Table 1, Item B), and then we used a contingency table to compare the reference map in t_2 with the simulated map in t_1 (we called Simulated_Change map after here) to quantify the persistence and simulated LUC.

For quantity error assessment, we subtracted the Simulated_Change map from Observed_Change map to estimate the quantity of error between simulated and reference changes. This is similar to subtracting Item B from Item A in Table 1 (in table format, Item C). We then divided each category of the resulting map or the resulting table by the total number of cells to show the percentage of each group (persistence and change). The negative and positive values within the contingency table show the percent of underestimate and overestimate of the back-cast model, respectively.

For our locational error assessment, we compared Item A with Item B in Table 1 (in table format), or in other words Observed_Change with Simulated_Change map (in map format), to calculate the locational error using the tabulate area function in GIS (Table 1, Item D). We divided each category of the resulting map or the resultant table by total number of cells to show the percentage of each group. After model calibration, since the quantity of land use are available for each year between t_2 and t_1 , the generated model from training run applied to the entire dataset in t_2 to simulate LUC for other times between t_2 and t_1 (Figure 1, Item 8).

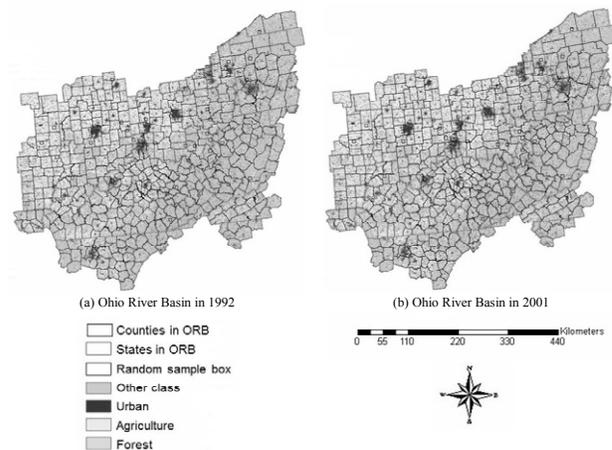


Figure 5. Land use maps of Ohio River Basin in (a) 1992 and (b) 2001. Stratified random sampling boxes for training of the back-cast ANN.

Table 1d. Locational Error Calculation

		Changed Map between Observed Map in t_2 and Simulated Map in t_1												
		C ₁₁	C ₁₂	...	C _{1n}	C ₂₁	C ₂₂	...	C _{2n}	...	C _{n1}	C _{n2}	...	C _{nn}
Observed Change Map between t_2 and t_1	C ₁₁	P ₁₁₁₁	P ₁₁₁₂	...	P _{111n}	P ₁₁₂₁	P ₁₁₂₂	...	P _{112n}	...	P _{11n1}	P _{11n2}	...	P _{11nn}
	C ₁₂	P ₁₂₁₁	P ₁₂₁₂	...	P _{121n}	P ₁₂₂₁	P ₁₂₂₂	...	P _{122n}	...	P _{12n1}	P _{12n2}	...	P _{12nn}

	C _{1n}	P _{1n11}	P _{1n12}	...	P _{1n1n}	P _{1n21}	P _{1n22}	...	P _{1n2n}	...	P _{1nn1}	P _{1nn2}	...	P _{1nnn}
	C ₂₁	P ₂₁₁₁	P ₂₁₁₂	...	P _{211n}	P ₂₁₂₁	P ₂₁₂₂	...	P _{212n}	...	P _{21n1}	P _{21n2}	...	P _{21nn}
	C ₂₂	P ₂₂₁₁	P ₂₂₁₂	...	P _{221n}	P ₂₂₂₁	P ₂₂₂₂	...	P _{222n}	...	P _{22n1}	P _{22n2}	...	P _{22nn}

	C _{2n}	P _{2n11}	P _{2n12}	...	P _{2n1n}	P _{2n21}	P _{2n22}	...	P _{2n2n}	...	P _{2nn1}	P _{2nn2}	...	P _{2nnn}

	C _{n1}	P _{n111}	P _{n112}	...	P _{n11n}	P _{n121}	P _{n122}	...	P _{n12n}	...	P _{n1n1}	P _{n1n2}	...	P _{n1nn}
	C _{n2}	P _{n211}	P _{n212}	...	P _{n21n}	P _{n221}	P _{n222}	...	P _{n22n}	...	P _{n2n1}	P _{n2n2}	...	P _{n2nn}

	C _{nn}	P _{nn11}	P _{nn12}	...	P _{nn1n}	P _{nn21}	P _{nn22}	...	P _{nn2n}	...	P _{nnn1}	P _{nnn2}	...	P _{nnnn}

*Calculating the number of cells that correctly predicted and misplaced as locational error by the back-cast model. C with indexes show the number of possible persistence and land use transitions within two maps. P shows the number of cells that have been correctly predicted as persistence or change (P along diagonal with blue color) and the number of cells that have been misplaced by back-cast model (P along off-diagonal with orange color) between two change maps.

4. Implementation of Back-Cast Model for Ohio River Basin

4.1. Study Area and Land Use Data Processing

Data used to build and test the back-cast ANN covered the entire Ohio River Basin (ORB) which included counties that wholly or partially fall within the basin borders (Figure 5). The ORB includes 456 counties in Illinois, Indiana, Ohio, Pennsylvania, New York, Maryland, West Virginia, Virginia, South Carolina, Kentucky, and Tennessee, the size of the area including counties is 515,818 km² while the ORB covers 421,962 km². We employed the retrofit National Land Cover Database (NLC-D) (Fry et al., 2008) with 30 m resolution to obtain land use data for the ORB from 2001 and 1992 (Tayyebi et al., 2013). We then reclassified the National Land Cover Database in both times from Anderson Level 1 to a coarse, four core land use cover classes (“urban”, “forest”, “agriculture” and “other class”). To accomplish this, the four developed (i.e., urban) National Land Cover Database classes were aggregated to create a single urban land use class. All forest, shrubland, and herbaceous vegetation classes were combined to create a single natural vegetation cover class, referred to as “forest” from here on. The agriculture land use class included cultivated crops. Because barren, open water, wetlands and shrubland are very minor land use classes in the ORB and much of it does not change (open water, barren), we collapsed these into an umbrella “other class” to focus on the reporting of the spatial-temporal dynamics of the major land use classes (urban, agriculture and forest) located in the ORB. The change from 2001 to 1992 was used as output in the back-cast model.

Several spatial drivers were obtained in GIS by processing the land use layers in 2001 (Table 2). The Euclidean distance tool in ArcGIS was used to create separate raster maps that stored in each cell: the distance from the nearest urban, forest and agriculture cell. ArcGIS Focal statistics was used to calcu-

late the density of each main land use class around the central cell in 10, 50 and 250 m windows. Slope was calculated from DEM using the GIS Spatial Analyst tool. These spatial drivers (obtained for the base year of 2001) were used as input in the back-cast model.

Table 2. The Drivers Used for the Back-Cast Model in 2001 and the Rationale for Including Them in the Simulation

Driver	Description
Distance to nearest road	Distance to road has been found to be one of the strongest drivers of urbanization in the United States
Distance to nearest town and distance to county capitals	People live and work near towns and proximity to cities, towns and villages strongly influences urbanization
Slope	Built environment cannot occur on steep slopes; crops are difficult to manage large scale using mechanized management. Generally, slopes > 8% are not farmed in the US
Distance to nearest urban cell	Previous urban cells are well known to create new urban cells in future time steps because infrastructure for urban likely exists
Density of urban within a fixed window size	Urban tends to fill in once a certain density of this use is reached (e.g., in 90 m, 250 m, 1 km)
Density of agriculture within a fixed window size	Large homogeneous agricultural plots are more sustainable over time (e.g., in 90 m, 250 m, 1 km)
Distance to nearest surface water body	People like to place built structures (e.g., houses) next to lakes and rivers and are thus drivers of urbanization

4.2. Training and Testing Run

Previous studies on modeling LUC at a national scale suggests that performing a calibration run across various spatial

scales (Tayyebi et al., 2013) is required to assess the accuracy of the model properly. For example, Verburg et al. (2002) and Pijanowski et al. (2014) showed that 1 km is an appropriate spatial scale to forecast multiple LUCs at continental scale since it is the proper scale to minimize the simulation error. Therefore, before training and testing, we resampled spatial predictors in 2001 and LUC map from 30 m to 1 km. For the training run, the back-cast ANN topology includes four outputs that have been coded using series of 1 s and 0 s and the following steps: 1) the cells that experience urban to agriculture change as the first model outcome, 2) the cells that experience urban to forest change as the second model outcome, 3) the cells that experience agriculture to forest change as the third model outcome, 4) the cells that experience forest to agriculture change as the fourth model outcome, and 5) other cells that experience other types of transitions. Our back-cast model ignores missing data and masks them out from further analysis. Before using data for training run in ANN, our back-cast model normalizes data using mean and standard deviation in each layer. Therefore, abnormal data are not used for the training run. We selected a training data set comprising 10% of entire dataset due to the large area covered in ORB using stratified random sampling to train the back-cast model which was then evaluated using the testing data set (other 90% of entire dataset).

We then used historical land in farms data (from 1980s to 2000s) and housing statistics (from 1980s to 2000s) to determine the quantity of agricultural and urban land use that should be allocated to each county for annual years between 2001 ~ 1980. The generated back-cast ANNs from the training run were applied to the entire dataset in 2001 to simulate LUC to 1992 using four suitability maps and the fixed quantity amounts. We then compared the simulated LUC map in 1992 with the reference map in 1992 to estimate the locational and quantity error of back-cast model. The tested back-cast model then applied to the entire dataset in 2001 to simulate LUC for annual years between 2001 ~ 1980 using four suitability maps and estimated quantity data.

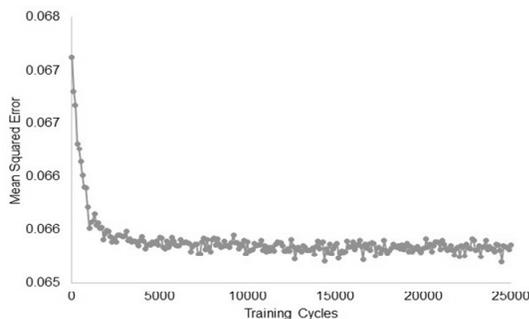


Figure 6. The training run of the back-cast ANN against mean squared error.

5. Results

5.1. Training Run and Historical Quantity Change

We conducted multiple training cycles with the back-cast

model to identify a training cycle that would generate model results that deviated to an acceptable extent from observed values. Mean squared error starts around 0.068 and drops rapidly through 5000 cycles (Figure 6). We halted the training at 25,000 cycles where the mean squared error reached a stable minimum of 0.066.

The trends of urban quantities for each state are very similar to one another and decrease smoothly in the number of urban cells from 2001 to 1980, respectively. While the trends of agriculture quantities for each state are different. For some of the states, the number of agriculture cells increase from 2001 to 1980, although the rates of these increases differed across states. For some other states, the number of agriculture cells are stable from 2001 to 1980. The portions of Ohio and Indiana within ORB have the greatest number of urban and agriculture cells. In contrast, the portion of Maryland within ORB has the smallest number of urban and agriculture cells. The cubic splines fit the observed agriculture data well (with $R^2 > 0.95$).

Table 3a. Contingency Table Quantifying the Observed Land Use Change for Entire ORB between 2001 and 1992

Percent (%)	Observed Map in 1992					
	Other Class	Urban	Agriculture	Forest	Total	
Observed Map in 2001	Other Class	1.9230	0.0097	0.0917	0.1295	2.1540
	Urban	0.0038	8.4965	0.1931	0.2111	8.9045
	Agriculture	0.0157	0.0510	36.6410	0.6832	37.3909
	Forest	0.0217	0.0163	0.4177	51.0949	51.5507
	Total	1.9642	8.5736	37.3435	52.1188	

Table 3b. Contingency Table Comparing Observed Map in 2001 and Simulated Map in 1992 for Entire ORB

Percent (%)	Simulated Map in 1992					
	Other Class	Urban	Agriculture	Forest	Total	
Map in 2001	Other Class	2.1539	0.0000	0.0000	0.0000	2.1539
	Urban	0.0000	8.5080	0.1895	0.2070	8.9045
	Agriculture	0.0000	0.0000	36.7584	0.6325	37.3909
	Forest	0.0000	0.0000	0.3956	51.1551	51.5507
	Total	2.1539	8.5080	37.3435	51.9946	

Table 3c. Quantity Error Calculation for Entire ORB

Quantity Error (%)	Simulated Map in 1992 Minus Observed Map in 1992					
	Other Class	Urban	Agriculture	Forest	Total	
Map in 2001	Other Class	0.2310	-0.0097	-0.0917	-0.1295	0.0000
	Urban	-0.0038	0.0115	-0.0036	-0.0041	0.0000
	Agriculture	-0.0157	-0.0510	0.1174	-0.0507	0.0000
	Forest	-0.0217	-0.0163	-0.0221	0.0602	0.0000
	Total	0.1898	-0.0655	0.0000	-0.1242	0.0000

*The negative and positive values in table show that back-cast model underestimated and overestimated those transitions, respectively.

5.2. Quantity Error Assessment

Urban, agriculture and forest land use classes occupied 8.90, 37.39 and 51.55% of the landscape respectively in the ORB in 2001 (Table 3a), and 8.57, 37.34 and 52.12% in 1992, respectively (Table 3a). 98.14% (1.92% in other class, 8.49% in urban, 36.64% in agriculture and 51.09% in forest) of the landscape have not changed from 2001 to 1992 (Table 3a). The model simulated the proportions of urban, agriculture and forest land use classes at 8.50, 37.34 and 51.99% in 1992 respectively (Table 3b).

5.2.1. Entire Ohio River Basin

The model overestimated the quantity of urban persistence by 0.01% in 1992 (Table 3c) and underestimated the transition from other classes, agriculture and forests to urban by 0.01, 0.05 and 0.02%, respectively (Table 3c). In total, the back-cast model underestimated the quantity of urban areas in 1992 by 0.06% (Table 3c).

The model also overestimated the quantity of agriculture persistence by 0.11% in 1992 (Table 3c), and underestimated transition from other class, urban and forest to agriculture by 0.09, 0.003 and 0.02%, respectively (Table 3c). In total, the back-cast model neither underestimated nor overestimated the quantity of agriculture areas in 1992 (Table 3c) since we fixed the quantity of agriculture change for the entire ORB. Forest persistence was overestimated by 0.06%, and changes from other classes, urban and agriculture in 2001 to forest in 1992, was underestimated by 0.13, 0.004 and 0.05%, respectively. In total, the back-cast model underestimated the quantity of forest areas in 1992 by 0.12% (Table 3c). Therefore, the range that the back-cast model underestimated and overestimated the quantity of multiple land use changes were -0.05 and 0.11% for entire ORB (only for urban, agriculture and forest), respectively.

tively.

The back-cast model assumed that when the cells converted to urban in forward manner, they would stay urban forever (Rittenhouse et al., 2012). Therefore, we did not incorporate transitions from any classes to urban from 2001 to 1992 and the back-cast model underestimated these transitions. While we fixed the quantity for urban and agriculture areas from 2001 to 1992, few overestimations or underestimations are due to the fact that forest transitions depend on other land use transitions occurring simultaneously in the landscape since there is not available online data source (such as those sources exist for agriculture and urban areas) to determine the quantity of historical forest areas. Change in quantity of agriculture cells reflects the agriculture abandonment this region experienced during the last 10 years, as well as the urbanization that occurred in the forward direction.

5.2.2. County Scale

Little LUC has occurred overall in the ORB between 2001 ~ 1992. However, this does not indicate to what extent different counties in the ORB transitioned from one land use class to another. Around 10% of counties in ORB underestimated urban persistence (Figure 7, Item a). Counties overestimated urban persistence were mostly the same counties that underestimated urban to agriculture transition (Figure 7, Item b). Around 70% of counties in ORB underestimated urban to agriculture transition where these counties were distributed across within ORB (Figure 7, Item b). In addition, the counties that underestimated urban persistence (Figure 7, Item a) were the same counties that overestimated urban to forest transition (Figure 7, Item c). Around 40% of counties in ORB have underestimated urban to forest transition.

All counties in ORB overestimated agriculture persistence

Table 3d. Comparison of the Simulated Change and Observed Change between 2001 and 1992

		Changes between Observed Map in 2001 and Simulated Map in 1992								
		11	22	23	24	33	34	43	44	Total
Observed Change between 2001 and 1992	11	1.9230	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.9230
	12	0.0097	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0097
	13	0.0917	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0917
	14	0.1295	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1295
	21	0.0000	0.0035	0.0001	0.0002	0.0000	0.0000	0.0000	0.0000	0.0038
	22	0.0000	8.4817	0.0135	0.0013	0.0000	0.0000	0.0000	0.0000	8.4965
	23	0.0000	0.0121	0.1708	0.0102	0.0000	0.0000	0.0000	0.0000	0.1931
	24	0.0000	0.0107	0.0051	0.1953	0.0000	0.0000	0.0000	0.0000	0.2111
	31	0.0000	0.0000	0.0000	0.0000	0.0154	0.0003	0.0000	0.0000	0.0157
	32	0.0000	0.0000	0.0000	0.0000	0.0506	0.0004	0.0000	0.0000	0.0510
	33	0.0000	0.0000	0.0000	0.0000	36.6305	0.0105	0.0000	0.0000	36.6410
	34	0.0000	0.0000	0.0000	0.0000	0.0619	0.6213	0.0000	0.0000	0.6832
	41	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0217	0.0217
	42	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0160	0.0163
	43	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3622	0.0555	0.4177
44	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0331	51.0618	51.0949	
Total		2.1539	8.5080	0.1895	0.2070	36.7584	0.6325	0.3956	51.1551	

^aCodes 1, 2, 3 and 4 represent other class, urban, agriculture and forest, respectively. Codes (e.g., 22) with repeated digits imply persistence while other digits (e.g., 34) show transition.

varied from 0.001 to 2.88% (Figure 7, Item d). While all counties in ORB underestimated agriculture to forest transitions varied from 0.001 to 2.70% (Figure 7, Item e). The counties that overestimated agriculture persistence (Figure 7, Item d) were the same counties that underestimated agriculture to forest (Figure 7, Item e). The percentage of overestimate in agriculture persistence was around same as the percentage underestimated in agriculture to forest transition. Around 90% of counties in ORB overestimated forest persistence (Figure 7, Item f). These were the same counties that underestimated forest to agriculture transition (Figure 7, Item g). The percentage overestimated forest persistence was around same as that underestimated forest to agriculture transition.

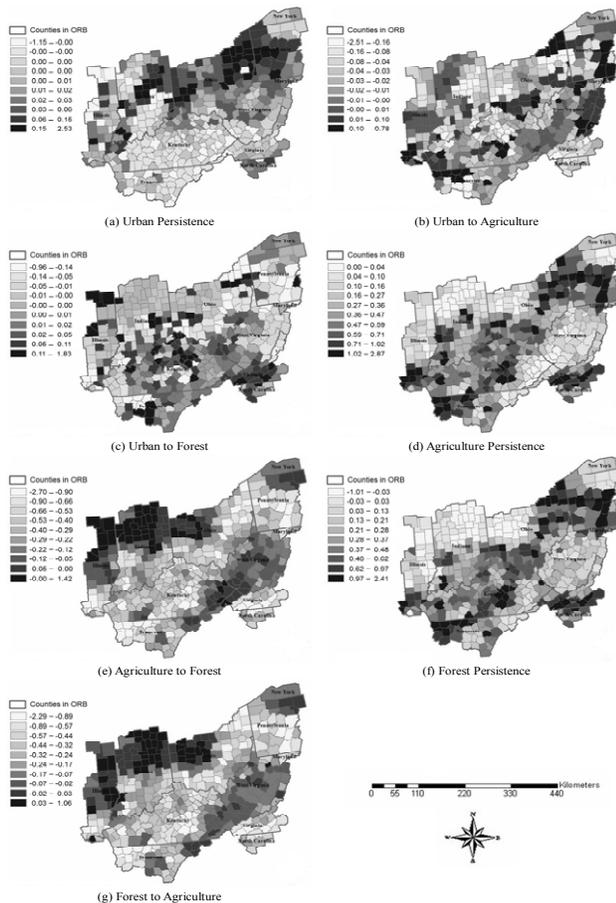


Figure 7. Spatial distribution of quantity errors within Ohio River Basin for persistence and land use transitions at county scale.

5.3. Location Error Assessment

5.3.1. Entire Ohio River Basin

The back-cast model simulated the location of persistence and LUCs for the entire ORB at 1km resolution well. The back-cast model simulated the persistence of three dominant land use classes well (8.48% of 8.49% for urban; 36.63% of 36.64% for agriculture; 51% of 51.09% for forest). On the other hand, the

back-cast model misplaced 0.01% of agriculture persistence as agriculture to forest transition, 0.03% of forest persistence as forest to agriculture transition, 0.013 and 0.001% of urban persistence as urban to agriculture transition and urban to forest transition between 2001 ~ 1992, respectively (Table 3d). In general, the ability of the back-cast model to simulate persistence for each land use class was over 95% since around 95% of the landscape have not changed between two times.

The back-cast model simulated the location of LUCs for the entire ORB at 1 km resolution relatively well. For urban loss (codes 23 and 24 along rows in table 3d), the back-cast model simulated 0.17% (row and column 23) of 0.19% for urban to agriculture change (row 23 and columns 22 ~ 24) and 0.19% (row and column 24) of 0.21% for urban to forest change (row 24 and columns 22 ~ 24) correctly. Results showed that the model misplaced 0.0121% of urban to agriculture change (row 23 and column 22) and 0.0107% of urban to forest change (row 24 and column 22) as urban persistence. Similarly, for the agriculture to forest change, the back-cast model simulated 0.62% of 0.68% these changes (row and column 34) correctly. The model misplaced others, 0.06%, as agriculture persistence (row 34 and column 33). Finally, for the forest to agriculture transition, the back-cast model simulated 0.36% of 0.41% for this change (row and column 43) correctly. The model misplaced others, 0.05%, as forest persistence (row 43 and column 44). In general, the ability of back-cast model to simulate LUCs was over 80% since only less than 5% of the landscape have changed between two times.

5.3.2. County Scale

The model performed slightly better in simulating persistence than LUC. Forest persistence with 95% agreement was the most accurate, followed by agriculture persistence with 93% agreement and urban persistence with 89% agreement. Urban to forest transition with 80% agreement was the most accurate land use transition. Following, agriculture to forest transition, urban to agriculture transition and forest to agriculture transition with 76, 71, and 62% agreement were the most accurate land use transition, respectively.

6. Discussion

Modeling and model calibration for multiple LUCs at a regional scale is a challenge in land change science. Previous back-cast models (Ray and Pijanowski, 2010; Ray et al., 2012) have been limited in their ability to simulate multiple land use transitions simultaneously. Rather, each land use class was trained separately and simple what-if rules were applied to determine actual land use transitions. In this study, we developed a new back-cast model that simulates multiple land use (urban, agriculture and forest) changes at large spatial scales using an advanced computer modeling technique (e.g., artificial intelligence) on a HPC platform. We also demonstrated a new calibration approach to test quantity and location errors of the back-cast model for the ORB. Previous studies showed that the accuracy of models decreased slightly when models are applied

to multiple LUCs at the same time rather than modeling separate binary transitions (Yang et al., 2008). We also found this to be the case when comparing our results to previous studies that used a similar model for binary back-casting of LUC (Ray and Pijanowski, 2010).

The sequence of LUC at a given location is important to understand. For example, Brown et al. (2005) showed that many non-urban areas undergo three or more changes over several decades, with agriculture transitioning to forest and then to urban. Future research can benefit from land use maps produced by the back-cast model to address questions related to the process of LUC. For example, Tayyebi et al. (2015) recently analyzed LUCs for the entire ORB of United States for three LUCs (forest, agriculture and urban) by comparing annual land use maps across ORB between 1980 ~ 2001. They found that three general trends emerge from back-cast simulations, an increase in urban, an increase in forests and a decrease in agriculture. In other words, agriculture to forest and agriculture to urban transitions were the most common LUCs in the ORB.

Considering past land use in establishing future mitigation measures is crucial to overcoming the legacies of human stressors in the region and the effective restoration and protection of ecosystem resources at regional scales (Stark, 2011). Thus, products resulting from large regional LUC simulations like the one presented here could support regional scale environmental programmatic assessments that are becoming more common with national government agencies (NEON) (Schimel et al., 2007). For example, Tayyebi et al. (2015) used the developed land use maps in this study to determine the location of catchments in the ORB exceeding critical land use thresholds for water quality and other aquatic resources. They found that while the proportion of catchments to exceed critical urban land use thresholds will increase from 15% in 1980 to 17.5% in 2001, the proportion of catchments to exceed critical agricultural land use thresholds will decrease from 45 to 42%.

One of the main challenges in LUC science is the number of desired outputs in the model (Tayyebi and Pijanowski, 2014). One of the limitation of this study is that it has few land use classes as output; however, this may not reflect the reality of LUCs in nature. In future, there is a need to test models for back-casting of unlimited LUCs or the modeling of land classes hierarchically (simulating urban and then the type of urban, such as residential or commercial). Second, this study only performed model calibration at 1 km resolution for entire the ORB. Pijanowski et al. (2014) and Tayyebi et al. (2013) performed model calibration at multiple spatial and temporal scales for the entire conterminous United States. There is thus an urgent need to develop an accurate back-cast model that can work at fine scale resolution (e.g., 30 m), at large spatial scales, and for multiple transitions all simultaneously. Finally, this study only used few LUC scenarios to back-cast multiple LUCs for short time interval (from 2001 to 1980). Previous studies showed that these LUC scenarios are suitable for a short time (Brown et al., 2005); however, for a longer term it is required to explore more sophisticated LUC scenarios and perform model evaluation in a longer term using aerial photo (Ray and Pijanowski, 2010).

We made several assumptions to develop the back-cast model. First, we assumed that the pattern of spatial drivers and ANN network remained constant over time. Thus, spatial rules used to determine the potential locations for land use transitions were assumed to remain constant over time. Second, we used Cubic function to determine the quantity of agriculture and urban areas in the past. Using other functions such as polynomial function can change these quantities and these quantities directly affect the back-cast LUC scenario. Third, since the quantity of forest areas in the past is not available online like urban and agriculture quantity, our back-cast LUC scenarios were limited and depended on either agriculture gain or agriculture loss. Therefore, further analysis to determine quantity of forest in the past using high resolution satellite imageries or aerial photos can lead to define better LUC scenarios for back-casting. Finally, more rules need to be explored in order to determine how they would be applied to the back-cast model and it is possible that in a large region several transition rules may need to be applied in the same simulation (Dietzel and Clarke, 2006).

7. Conclusions

This study developed a new back-cast model to simulate multiple LUCs over broad spatial scales using a HPC platform. The study represents an advancement in computational modeling as well as a successful broad-scale modeling framework that has direct applications to land use management. This study also developed a new metric that enables LUCs modelers to assess simultaneously the locational and quantity errors of multiple LUCs at large spatial scales. We demonstrated that our back-cast model can simulate historical persistence and LUCs accurately at a regional scale by adjusting land use transitions across counties. Finally, we produced new land use maps between 2001 and 1980 at new spatial and temporal scales that can be coupled with other models to assess land use and environment interactions.

Acknowledgments. Funding to complete this research was obtained through the U.S. Environmental Protection Agency, and the Department of Forestry and Natural Resources, Purdue University.

References

- Azari, M., Tayyebi, A., Helbich, M., and Reveshty, M.A. (2015). Integrating Cellular Automata, Artificial Neural Network and Fuzzy Set Theory to Simulate Threatened Orchards: Application to Maragheh, Iran. *GISci. & Remote Sens.* (accepted).
- Blanchard, S.D., Pontius Jr, R.G., and Urban, K.M. (2015). Implications of using 2 m versus 30 m spatial resolution data for suburban residential land change modeling. *J. Environ. Inf.*, 25(1), 1-13. <http://dx.doi.org/10.3808/jei.201400284>
- Brown, D.G., Pijanowski, B.C., and Duh, J.D. (2000). Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. *J. Environ. Manage.*, 59(4), 247-263. <http://dx.doi.org/10.1006/jema.2000.0369>
- Brown, D.G., Johnson, K.M., Loveland, T.R., and Theobald, D.M. (2005). Rural land-use trends in the conterminous United States, 1950-2000. *Ecol. Appl.*, 15, 1851-1863. <http://dx.doi.org/10.1890>

/03-5220

- Brown, L.R., Cuffney, T.F., Coles, J.F., Fitzpatrick, F., McMahon, G., Steuer, J., Bell, A.H., and May, J.T. (2009). Urban streams across the USA: Lessons learned from studies in 9 metropolitan areas. *J. N. Am. Benthol. Soc.*, 28(4), 1051-1069. <http://dx.doi.org/10.1899/08-153.1>
- Clarke, K.C. (2003). Geocomputation's future at the extremes: High performance computing and nanoclients. *Parallel Comput.*, 29(10), 1281-1295. <http://dx.doi.org/10.1016/j.parco.2003.03.001>
- Dietzel, C. and Clarke, K. (2006). The effect of disaggregating land use categories in cellular automata during model calibration and forecasting. *Comput. Environ. Urban Syst.*, 30(1), 78-101. <http://dx.doi.org/10.1016/j.compenvurbsys.2005.04.001>
- Feurdean, A.N., Willis, K.J., and Astalos, C. (2009). Legacy of the past land-use changes and management on the 'natural' upland forest composition in the Apuseni Natural Park, Romania. *Holocene*, 19(6), 967-981. <http://dx.doi.org/10.1177/0959683609337358>
- Foster, D., Swanson, F., Aber, J., Burke, I., Brokaw, N., Tilman, D., and Knapp, A. (2003). The importance of land-use legacies to ecology and conservation. *BioScience*, 53(1), 77-88. [http://dx.doi.org/10.1641/0006-3568\(2003\)053\[0077:TIOLUL\]2.0.CO;2](http://dx.doi.org/10.1641/0006-3568(2003)053[0077:TIOLUL]2.0.CO;2)
- Fry, J.A., Coan, M.J., Homer, C.G., Meyer, D.K., and Wickham, J.D. (2008). *Completion of the National Land Cover Database (NLCD) 1992-2001 Land Cover Change Retrofit Product*, US Geological Survey Open-file Report, 1379, 18.
- Hammer, R.B., Stewart, S.I., Winkler, R.L., Radeloff, V.C., and Voss, P.R. (2004). Characterizing dynamic spatial and temporal residential density patterns from 1940 to 1990 across the North Central United States. *Landscape Urban Plann.*, 69(2-3), 183-199. <http://dx.doi.org/10.1016/j.landurbplan.2003.08.011>
- Huang, G.H. and Chang, N.B. (2003). The perspectives of environmental informatics and systems analysis. *J. Environ. Inf.*, 1(1), 1-7. <http://dx.doi.org/10.3808/jei.200300001>
- Hill, M.J. and Aspinall, R.J. (2000). *Spatial Information for Land Use Management*, CRC Press.
- Islam, K.R. and Weil, R.R. (2000). Land use effects on soil quality in a tropical forest ecosystem of Bangladesh. *Agric. Ecosyst. Environ.*, 79(1), 9-16. [http://dx.doi.org/10.1016/S0167-8809\(99\)00145-0](http://dx.doi.org/10.1016/S0167-8809(99)00145-0)
- Jantz, C.A., Goetz, S.J., Donato, D., and Claggett, P. (2010). Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model. *Comput. Environ. Urban Syst.*, 34(1), 1-16. <http://dx.doi.org/10.1016/j.compenvurbsys.2009.08.003>
- Jia, K., Liang, S., Liu, J., Li, Q.Z., Wei, X.Q., Yuan, W.P., and Yao, Y.J. (2014). Forest cover changes in the Three-North Shelter Forest Region of China during 1990 to 2005. *J. Environ. Inf.*, 26: 112-120.
- Kumar, S., Merwade, V., Rao, P.S.C., and Pijanowski, B.C. (2013). Characterizing long-term land use/cover change in the United States from 1850 to 2000 using a nonlinear bi-analytical model. *Ambio*, 42(3), 285-297. <http://dx.doi.org/10.1007/s13280-012-0354-6>
- Lepczyk, C.A., Hammer, R.B., Stewart, S.I., and Radeloff, V.C. (2007). Spatiotemporal dynamics of housing growth hotspots in the North Central U.S. from 1940 to 2000. *Landscape Ecol.*, 22(6), 939-952. <http://dx.doi.org/10.1007/s10980-006-9066-2>
- Liu, Z. and Tong, S.T.Y. (2011). Using HSPF to model the hydrologic and water quality impacts of riparian land-use change in a small watershed. *J. Environ. Inf.*, 17(1), 15-24. <http://dx.doi.org/10.3808/jei.201100183>
- Maizel, M., White, R.D., Gage, S., Osborne, L., Root, R., Stitt, S., and Muehlbach, G. (1998). Historical interrelationships between population settlement and farmland in the conterminous United States, 1790 to 1992, pp. 5-12, in T.D. Sick (Ed.), *Perspectives on the Land-use History of North America: A Context for Understanding Our Changing Environment: USGS/BRD/BSR-1998-0003*, U.S. Geological Survey, Washington, D.C., USA.
- Nassauer, J.I., Cooper, D.A., Marshall, L.L., Currie, W.S., Hutchins, M., and Brown, D.G. (2014). Parcel size related to household behaviors affecting carbon storage in exurban residential landscapes. *Landscape Urban Plann.*, 129, 55-64. <http://dx.doi.org/10.1016/j.landurbplan.2014.05.007>
- Pauleit, S., Ennos, R., and Golding, Y. (2005). Modeling the environmental impacts of urban land use and land cover change - A study in Merseyside, UK. *Landscape Urban Plann.*, 71(2-4), 295-310. [http://dx.doi.org/10.1016/S0169-2046\(04\)00083-0](http://dx.doi.org/10.1016/S0169-2046(04)00083-0)
- Pekin, B.K. and Pijanowski, B.C. (2012). Global land use intensity and endangerment status of mammal species. *Divers. Distrib.*, 18(9), 909-918. <http://dx.doi.org/10.1111/j.1472-4642.2012.00928.x>
- Pijanowski, B.C., Brown, D.G., Shellito, B.A., and Manik, G.A. (2002). Using neural networks and GIS to forecast land use changes: A Land Transformation Model. *Comput. Environ. Urban Syst.*, 26(6), 553-575. [http://dx.doi.org/10.1016/S0198-9715\(01\)00015-1](http://dx.doi.org/10.1016/S0198-9715(01)00015-1)
- Pijanowski, B., Ray, D.K., Kendall, A.D., and Duckles, J.M. (2007). Using backcast land-use change and groundwater travel-time models to generate land-use legacy maps for watershed management. *Ecol. Soc.*, 12(2), 375-386.
- Pijanowski, B.C., Tayyebi, A., Delavar, M.R., and Yazdanpanah, M.J. (2009). Urban expansion simulation using geospatial information system and artificial neural networks. *Int. J. Environ. Res.*, 3(4), 493-502.
- Pijanowski, B.C., Tayyebi, A., Doucette, J., Pekin, B.K., Braun, D., and Plourde, J. (2014). A big data urban growth simulation at a national scale: Configuring the GIS and neural network based Land Transformation Model to run in a High Performance Computing (HPC) environment. *Environ. Model. Software*, 51, 250-268. <http://dx.doi.org/10.1016/j.envsoft.2013.09.015>
- Pontius Jr, R.G., Cornell, J.D., and Hall, C.A.S. (2001). Modeling the spatial pattern of land-use change with GEOMOD2: Application and validation for Costa Rica. *Agric. Ecosyst. Environ.*, 85(1-3), 1-3. [http://dx.doi.org/10.1016/S0167-8809\(01\)00183-9](http://dx.doi.org/10.1016/S0167-8809(01)00183-9)
- Pontius Jr, R.G. and Millones, M. (2011). Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.*, 32(15), 4407-4429. <http://dx.doi.org/10.1080/01431161.2011.552923>
- Radeloff, V.C., Hammer, R.B., Stewart, S.I., Fried, J.S., Holcomb, S.S., and McKeefry, J.F. (2005). The wildland-urban interface in the United States. *Ecol. Appl.*, 15, 799-805. <http://dx.doi.org/10.1890/04-1413>
- Ray, D.K. and Pijanowski, B.C. (2010). A backcast land use change model to generate past land use maps: Application and validation at the Muskegon river watershed of Michigan, USA. *J. Land Use Sci.*, 5(1), 1-29. <http://dx.doi.org/10.1080/17474230903150799>
- Ray, D.K., Pijanowski, B.C., Kendall, A.D., and Hyndman, D.W. (2012). Coupling land use and groundwater models to map land use legacies: Assessment of model uncertainties relevant to land use planning. *Appl. Geogr.*, 34, 356-370. <http://dx.doi.org/10.1016/j.apgeog.2012.01.002>
- Reid, W.V., Chen, D., Goldfarb, L., Hackmann, H., Lee, Y.T., Mokhele, K., Ostrom, E., Raivio, K., Rockström, J., Schellnhuber, H.J., and Whyte, A. (2010). Earth system science for global sustainability: Grand challenges. *Science*, 330(6006), 916-917. <http://dx.doi.org/10.1126/science.1196263>
- Rhemtulla, J.M., Mladenoff, D.J., and Clayton, M.K. (2007). Regional land-cover conversion in the U.S. Upper Midwest: Magnitude of change and limited recovery (1850-1935-1993). *Landscape Ecol.*, 22(1), 57-75. <http://dx.doi.org/10.1007/s10980-007-9117-3>
- Rittenhouse, C.D., Pidgeon, A.M., Albright, T.P., Culbert, P.D., Clayton, M.K., Flather, C.H., Masek, J.G., and Radeloff, V.C. (2012). Land-cover change and avian diversity in the conterminous United

- States. *Conserv. Biol.*, 26(5), 821-829. <http://dx.doi.org/10.1111/j.1523-1739.2012.01867.x>
- Schimel, D., Hargrove, W., Hoffman, F., and MacMahon, J. (2007). NEON: A hierarchically designed national ecological network. *Front. Ecol. Environ.*, 5(2), 59-59. [http://dx.doi.org/10.1890/1540-9295\(2007\)5\[59:NAHDNE\]2.0.CO;2](http://dx.doi.org/10.1890/1540-9295(2007)5[59:NAHDNE]2.0.CO;2)
- Silva, E.A. and Clarke, K.C. (2002). Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Comput., Environ. Urban Syst.*, 26(6), 525-552. [http://dx.doi.org/10.1016/S0198-9715\(01\)00014-X](http://dx.doi.org/10.1016/S0198-9715(01)00014-X)
- Tayyebi, A., Meehan, T.D., Dischler, J., Radloff, G., Ferris, M., and Gratton, C. (2016a). SmartScape™: A web-based decision support system for assessing the tradeoffs among multiple ecosystem services under crop-change scenarios. *Computers and Electronics in Agriculture*, 121, 108-121. <http://dx.doi.org/10.1016/j.compag.2015.12.003>
- Tayyebi, A., Arsanjani, J.J., Tayyebi, A.H., Omrani, H., and Moghadam, H.S. (2016b). Group-based crop change planning: Application of SmartScape™ spatial decision support system for resolving conflicts. *Ecological Modelling*, 333, 92-100. <http://dx.doi.org/10.1016/j.ecolmodel.2016.04.018>
- Tayyebi, A. and Jenerette, G.D. (2016). Increases in the climate change adaption effectiveness and availability of vegetation across a coastal to desert climate gradient in metropolitan Los Angeles, CA, USA. *Science of The Total Environment*, 548, 60-71. <http://dx.doi.org/10.1016/j.scitotenv.2016.01.049>
- Tayyebi, A., Pekin, B.K., Pijanowski, B.C., Plourde, J.D., Doucette, J.S., and Braun, D. (2013). Hierarchical modeling of urban growth across the conterminous USA: Developing meso-scale quantity drivers for the Land Transformation Model. *J. Land Use Sci.*, 8(4), 422-442. <http://dx.doi.org/10.1080/1747423X.2012.675364>
- Tayyebi, A. and Pijanowski, B.C. (2014). Modeling multiple land use changes using ANN, CART and MARS: Comparing tradeoffs in goodness of fit and explanatory power of data mining tools. *Int. J. Appl. Earth Obs. Geoinf.*, 28, 102-116. <http://dx.doi.org/10.1016/j.jag.2013.11.008>
- Tayyebi, A., Perry, P.C., and Tayyebi, A.H. (2014a). Predicting the expansion of an urban boundary using spatial logistic regression and hybrid raster-vector routines with remote sensing and GIS. *Int. J. Geogr. Inf. Sci.*, 28(4), 639-659. <http://dx.doi.org/10.1080/13658816.2013.845892>
- Tayyebi, A., Pijanowski, B.C., Linderman, M., and Gratton, C. (2014b). Comparing three global parametric and local non-parametric models to simulate land use change in diverse areas of the world. *Environ. Model. Software*, 59, 202-221. <http://dx.doi.org/10.1016/j.envsoft.2014.05.022>
- Tayyebi, A.H., Tayyebi, A., and Khanna, N. (2014). Assessing uncertainty dimensions in land-use change models: Using swap and multiplicative error models for injecting attribute and positional errors in spatial data. *Int. J. Remote Sens.*, 35(1), 149-170. <http://dx.doi.org/10.1080/01431161.2013.866293>
- Tayyebi, A., Pijanowski, B.C., and Pekin, B.K. (2015). Land use legacies of the Ohio River Basin: Using a spatially explicit land use change model to assess past and future impacts on aquatic resources. *Appl. Geogr.*, 57, 100-111. <http://dx.doi.org/10.1016/j.apgeog.2014.12.020>
- Theobald, D.M. (2001). Land-use dynamics beyond the American urban fringe. *Geogr. Rev.*, 91(3), 544-564. <http://dx.doi.org/10.2307/3594740>
- U.S. Bureau of the Census (2000). *Census of Population and Housing 1980: Summary File 1C [MRDF]*, Washington, D.C., USA.
- U.S. Department of Agriculture (1997). *Census of agriculture, geographic area series, AC97-CD-Volume 1*, U.S. Department of Agriculture, National Agricultural Statistics Service.
- Utz, R.M., Hilderbrand, R.H., and Boward, D.M. (2009). Identifying regional differences in threshold responses of aquatic invertebrates to land cover gradients. *Ecol. Indicators*, 9(3), 556-567. <http://dx.doi.org/10.1016/j.ecolind.2008.08.008>
- Veldkamp, A. and Verburg, P.H. (2004). Modelling land use change and environmental impact. *J. Environ. Manage.*, 72(1-2), 1-3. <http://dx.doi.org/10.1016/j.jenvman.2004.04.004>
- Verburg, P.H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., and Mastura, S.S.A. (2002). Modeling the spatial dynamics of regional land use: The CLUE-S model. *Environ. Manage.*, 30(3), 391-405. <http://dx.doi.org/10.1007/s00267-002-2630-x>
- Vesterby, M. and Krupa, K.S. (1997). *Major uses of land in the United States, 1997. Statistical Bulletin No. (SB-973)*, 47, Economic Research Service, U.S. Department of Agriculture, Washington, D.C., USA.
- Vogelmann, J.E., Sohl, T., and Howard, S.M. (1998). Regional characterization of land cover using multiple sources of data. *Photo gram. Eng. Remote Sensing*, 64(1), 45-57.
- Waisanen, P.J. and Bliss, N.B. (2002). Changes in population and agricultural land in conterminous United States counties, 1790 to 1997. *Global Biogeochem. Cycles*, 16(4), 84-1-84-19. <http://dx.doi.org/10.1029/2001GB001843>
- Song, W., Pijanowski, B.C., and Tayyebi, A. (2015). Urban expansion and its consumption of high-quality farmland in Beijing, China. *Ecol. Indicators*, 54, 60-70. <http://dx.doi.org/10.1016/j.ecolind.2015.02.015>
- Yang, X. and Lo, C.P. (2002). Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. *Int. J. Remote Sens.*, 23(9), 1775-1798. <http://dx.doi.org/10.1080/01431160110075802>
- Yang, J.S., Wang, Y.Q., and August, P.V. (2004). Estimation of land surface temperature using spatial interpolation and satellite-derived surface emissivity. *J. Environ. Inf.*, 4(1), 40-47. <http://dx.doi.org/10.3808/jei.200400035>
- Yang, Q., Li, X., and Shi, X. (2008). Cellular automata for simulating land use changes based on support vector machines. *Comput. Geo sci.*, 34(6), 592-602. <http://dx.doi.org/10.1016/j.cageo.2007.08.003>
- Yeh, A.G.O. and Li, X. (2006). Errors and uncertainties in urban cellular automata. *Comput. Environ. Urban Syst.*, 30(1), 10-28. <http://dx.doi.org/10.1016/j.compenurbysys.2004.05.007>