Assessing Multi-Agent Parcelization Performance in the MABEL Simulation Model using Monte Carlo Replication Experiments

Konstantinos Alexandridis¹, Bryan C. Pijanowski², and Zhen Lei³

Abstract: In this paper we present and test the functionality of a parcelization algorithm implemented in our spatially-explicit, agent-based, land-use change model which we call the Multi-Agent Behavioral Economic Landscape (MABEL) model. In order to test the best possible spatial configuration of the algorithm and its efficiency compared to historically observed land use changes, we employed a Monte Carlo simulation approach using a series of replication experiments across time, comparing observed changes between 1970 and 1990, and across two different landscapes in Michigan USA. We compare the simulated parcel shapes to historically observed land use changes using the landscape ecology metric program, FRAGSTATS.

Keywords: agent-based; land use change; Monte Carlo; dynamic simulation; replication; shape metrics; accuracy assessment.

1. Introduction

Understanding the nature of the proximate and underlying driving forces of change in land use is challenging and often requires analysis of decisions at the individual, farm or community levels, and of how they are influenced by policy and other factors at local, state, national and international scales. Previous efforts to simulate land use change have focused on a grid-cell modeling environment approach (e.g., Parker and Meretsky, 2004; Pijanowski et al., 2002; Pontius, 2002). In such a modeling environment, cells within a grid are assigned values related to the probability of them transitioning which are often times derived from spatial relationships of a cell’s location in the grid compared to the location of “driver cells” (e.g., its distance to transportation). Few attempts (e.g., Brown et al., 2001), however, have been made to simulate land use change of ownership parcels over space and time. Challenges of modeling in this environment include the need to determine how parcels subdivide after land transactions and accounting of cells contained within a parcel during a simulation and determining boundaries for new ownership parcels occurring during land transactions.

The main research question that we are addressing in this paper is whether we can represent the parcelization nature of land use change using Geographic Information Systems (GIS) and the Swarm agent-based modeling (Minar et al., 1996; Swarm Intelligence Group, 2000) environment.

¹ Human-Environment Modeling & Analysis Facility (HEMA), Purdue University, Department of Forestry and Natural Resources. Corresponding Author: 195 Marsteller Street, West Lafayette IN 47907, USA. E-mail: kostas@fnr.purdue.edu.
² Human-Environment Modeling & Analysis Facility (HEMA), Purdue University, Department of Forestry and Natural Resources. E-mail: bpijanow@purdue.edu
³ Krannert Business School, Purdue University. E-mail: leizhen@msu.edu
We report here the results of our Multi-Agent based Behavioral Economic Landscape (MABEL) model simulations of agents who are selected to participate in a simple market model that represents the bidding process in land transactions. We use a historical parcel database developed for Michigan, USA, to initialize and compare the results of a series of Monte Carlo simulations. We compare simulated parcelization with observed parcelization patterns quantifying parcel shape differences using the landscape ecology metric software, FRAGSTATS (McGarigal et al., 2002).

The structure of the paper is as follows. First, we provide a brief overview of the MABEL model describing the composition of agents and the process of introducing the agents in our simple market model. Second, we present the parcelization algorithm that we have developed. Third, we then describe the simulations that we conducted that tests the applicability of the parcelization algorithm. Fourth, a description of the results of our simulations is presented comparing the results of the simulation with historical parcel data. Finally, we discuss the results of our study and highlight some of the important aspects of the results.

2. Overview of the MABEL Model

A more complete description of the MABEL model can be found in Lei et al. (in review) and Alexandridis et al. (in revision), where technical and mathematical aspects of the model are described, respectively. We briefly describe the MABEL model components that are relevant to this study.

We recognize two different types of agents in MABEL. Base agents in MABEL are agents that own land, and non-base agents that do not necessarily own land, but represent policy-makers, local and regional planners, among others. Land-use based attributes are the main drivers of the simulation, and land-use driven acquisition of land in a market model represents the basic framework for determining base agents’ actions. Currently, base agents in MABEL can be formed from any land use category contained in the study, frequently including agents whose land use is agriculture, residential, or forest.

Each agent also contains a variety of attributes that are either assigned to individual agents (socioeconomic), are geographic (distance to a road) or are biophysical (soil type, elevation/slope). Agents can also interact within the Swarm development environment as they are able to draw (on the screen) and update themselves in the simulated world, and respond to users’ inquiries to view an agent’s attributes (see, Lei et al., in press). Currently, an agent’s behavioral model is configured using a Bayesian Belief Network (BBN) model (see Alexandridis et al., in revision, Lei et al., in press). For the purposes of this paper, the BBN component is replaced with a stochastic simulator (see section 4.3) in order to examine the performance of the presented parcelization routine. Data to drive the MABEL simulation are input from either a GIS or from a tabular database using an assigned parcel ID as an item to join spatial and tabular data.

One of the important modules of MABEL is the market model. Here, agents whose intention is to buy or sell are sent to the area’s market model where they “bid” for portions of existing parcels. Agents attempt to determine which land transaction represents the most profitable deal. The principle of a market model is to help a potential agent who intends to buy (for convenience sake we call these buyer agents, but they have not changed agent types but only announced their intentions to buy) make the most profitable deal with any corresponding agent who desires to sell. The degree of the profit in a transaction depends on how close the transaction quantity of the buyer agent meets that of the seller agent.

The market model first classifies all participating agents in different transaction groups by the action type (i.e., buy or sell), and sorts the agents in each group by the transaction quantity. Then, a buyer agent, which we call the premium buyer agent possessing the largest transaction quantity in the buyer agent list, checks the seller agent list for a matched premium seller agent. If successful, a transaction is made between the premium buyer agent and the premium seller agent.
If the premium seller agent does not sell all its land to the premium buyer agent, a new agent is created within the transaction area. The new agent inherits the attributes from the original seller agent, and adopts the land use type of the buyer agent. The original premium buyer agent and premium seller agent update themselves, and are inserted back to the seller/buyer agent list, so long as they still satisfy the policy/rule requirement, such as minimum parcel size, of the transaction. The process of finding a matched premium buyer agent and a premium seller agent for transaction continues until the buyer or seller lists are empty.

3. The Parcelization Algorithm in MABEL

3.1 Agent Partitioning Algorithm Details

In the process of creating the new agent, a seller agent has to determine how to partition the original area into two blocks with a corresponding shape and position for the new agent (i.e., the buyer agent) and seller agent after the land transaction. There are many different ways to partition land that achieve multiple objectives of maintaining accessibility to a road, achieving width/length ratios that are allowable by local government, etc. To address this challenge, we first define several possible partition algorithms. For illustrative purposes, assume that the following square area of sixteen parcels (Figure 1) is occupied by a seller agent, and each parcel is indexed by a number as shown. Furthermore, let five alternative shapes (labeled as A-E in Figure 1) indicate the configuration of land owned by the seller agent to be examined. These configurations are meant to span different levels of complexity in parcel shapes; although these five are not exhaustive.

![Algorithm Scanning Grid](image)

**Figure 1.** The grid-cell numbering scheme and the five different shape configurations (locations of cells in gray) ranging from simple (A) to complex (E).
Scan and Allocate Pathways
1. Vertical Priority Top-Left Order Scan Algorithm (VTL)
2. Horizontal Priority Top-Left Order Scan Algorithm (HTL)
3. Vertical Priority Top-Right Order Scan Algorithm (VTR)
4. Horizontal Priority Top-Right Order Scan Algorithm (HTR)
5. Vertical Priority Bottom-Left Order Scan Algorithm (VBL)
6. Horizontal Priority Bottom-Left Order Scan Algorithm (HBL)
7. Vertical Priority Bottom-Right Order Scan Algorithm (VBR)
8. Horizontal Priority Bottom-Right Order Scan Algorithm (HBR)
9. Square Priority Top-Left Order Scan Algorithm (STL)
10. Square Priority Top-Right Order Scan Algorithm (STR)
11. Square Priority Right-Top Order Scan Algorithm (SRT)
12. Square Priority Bottom-Left Order Scan Algorithm (SBL)
13. Square Priority Bottom-Right Order Scan Algorithm (SBR)
14. Square Priority Right-Right Order Scan Algorithm (SRR)

Figure 2. The sixteen scanning algorithms for shapes A-E in the MABEL parcelization routine.

When the seller agent begins to sell its land by the unit of parcel, it scans and allocates its land of parcels in a specific order (or pathway), and "gives" (or allocates) the cells to the new agent until the required area of the new agent is reached. We consider sixteen different 'scan and allocate pathways' (see Figure 2) for each parcel configuration. The scanning algorithms listed in the upper left corner of Figure 2 are defined by the search pattern describing two scanning directions, namely the left-right direction, and the top-down scanning direction. The scanning sequence of the algorithms always starts from one of the four corners of the closest confined rectangle that encapsulates the agent’s parcel shape. In all cases shown in Figure 2, the buyer is attempting to purchase five cells (shown in dark gray). The original parcel from the seller not being considered in the purchase is in light gray and areas outside the original parcel are white.

Since any scan-and-allocate order may induce different effects in different agent shapes, seller and buyer agents must intelligently choose the partition algorithm that yields the best possible shape for the ownership parcel after the partition. We currently use three metrics for our assessment of the best possible parcel shape: Occupancy Area Ratio (O), Width/Height Ratio (W), and Landscape Shape Index (L). O describes how fully an agent occupies the contiguous scanned rectangle compared to its land transaction amount. The later closely confined rectangle in landscape terms can be considered the smallest possible patch. It resembles the concept of patch
occupancy used often in landscape ecology studies of population and patch dynamics, with the addition of considering the simulation model’s scanning algorithm performance (Fortin et al., 2003; Rutledge, 2003). \( W \) describes how irregular we allow for the confined rectangle to be. We can allow for an application or combination of more determinant factor/measurements to be taken into consideration when a simulation-specific policy/rule assessment is used. Lastly, \( L \) accounts for the “clumpiness” properties of an agent’s parcel, taking into account both the area and the perimeter of the parcel. We quantify \( L \) using a minor modification from McGarigal and Marks’ (1994) and Patton’s (1975) inverse of the landscape shape index (LSI) metric computed by FRAGSTATS (McGarigal et al., 2002). An inverse LSI is employed here so that it is normalized consistently to the other two metrics.

### 3.2 An Example of the Parcelization Algorithm Functionality

The following example (Figure 3) illustrates how \( O \), \( W \) and \( L \) are calculated. As the figure shows, the seller agent originally occupies the dark-shaded area; areas not shaded are not owned by the seller. The original seller area (we refer to this as the original area, to distinguish this from the area of the seller after the parcel is divided following the land transaction) is mostly confined to rectangle A, which we call the smallest confined rectangle. In this example, the seller agent begins a scan starting in the upper left corner of the smallest confined rectangle and then counts, in one of sixteen possible pathways (summarized in Figure 2), the number of cells that is owned by the seller until the area to be purchased is met. All of the cells within this area scanned by the seller are then allocated into a rectangle, B, as shown in Figure 3. The unscanned area representing the remaining shaded agent area is confined in rectangle C, and is left to the seller agent.

![Figure 3. An example of quantity measurements of best possible shape for a hypothetical MABEL agent. (a) rectangle A: the smallest confined rectangle for the agent i is denoted with the thick-bordered rectangle, and contains the entire area of the agent i; (b) rectangle B: denotes the scanned area for agent i (buyer agent’s smallest confined rectangle), and (c) rectangle C: denotes the unscanned area for agent i.](image)
If we consider only the original agent, the occupancy area ratio is:

\[ O_i = \frac{A_i}{A_{\text{rectangle } A}} = \frac{\text{AREA}_{\text{shaded, rectangle } A} - \text{AREA}_{\text{non-shaded, rectangle } A}}{\text{AREA}_{\text{rectangle } A}} \]  

(1)

The Width/Height Ratio is:

\[ W_i = \min \left\{ \frac{\text{WIDTH}_{\text{rectangle } A}}{\text{HEIGHT}_{\text{rectangle } A}}, \frac{\text{HEIGHT}_{\text{rectangle } A}}{\text{WIDTH}_{\text{rectangle } A}} \right\} = \min \left\{ \frac{x_A}{y_A}, \frac{y_A}{x_A} \right\} \]  

(2)

And the Landscape Shape Index is:

\[ L_i = \frac{\min e_i}{i} \]  

(3)

The calculation of the $L$ metric for the MABEL parcelization algorithm is provided in the end of this paper, in the Appendix A.

Let us now consider that the parcel shape after selling and buying needs to be quantified simultaneously. In the same way, we can define $O$ and $W$ for the new agent of the scanned area (rectangle B) and remaining seller agent of the unscanned area (rectangle C), as follows:

\[ O_b = \frac{A_b}{A_{\text{rectangle } B}} = \frac{\text{AREA}_{\text{shaded, rectangle } B}}{\text{AREA}_{\text{rectangle } B}} \]  

\[ O_s = \frac{A_s}{A_{\text{rectangle } C}} = \frac{\text{AREA}_{\text{shaded, rectangle } C}}{\text{AREA}_{\text{rectangle } C}} \]  

\[ W_b = \left\{ \frac{x_b}{y_b}, \frac{x_b}{y_b} \right\} = \left\{ \frac{x_{\text{rectangle } B}}{y_{\text{rectangle } B}}, \frac{y_{\text{rectangle } B}}{x_{\text{rectangle } B}} \right\} \]  

\[ W_s = \left\{ \frac{x_s}{y_s}, \frac{x_s}{y_s} \right\} = \left\{ \frac{x_{\text{rectangle } C}}{y_{\text{rectangle } C}}, \frac{y_{\text{rectangle } C}}{x_{\text{rectangle } C}} \right\} \]  

\[ L_b = \frac{\min e_b}{e_b} \]  

\[ L_s = \frac{\min e_s}{e_s} \]  

To select the best partition algorithm, we consider the three parcel metrics for both the new agent and seller agent. For a certain partition algorithm $i$, the best parcel algorithm metric value, $B_i$, is used to quantify the best possible parcel shape. We use a generalized expression of:

\[ B_j = \frac{A_j}{A_i} \left( w_1 \cdot O_s + w_2 \cdot W_s + w_3 \cdot L_s \right) + \frac{A_i}{A_i} \left( w_1 \cdot O_b + w_2 \cdot W_b + w_3 \cdot L_b \right) \]  

(4)
where \( w_1 + w_2 + w_3 = 1 \), and \( w_1, w_2 \) and \( w_3 \) are defined as the weights for \( O \), \( W \) and \( L \) of the best parcel shape, respectively. Agents calculate \( B_j \) at each time step and select the partition algorithm with the largest \( B_j \).

### 4. Monte Carlo Simulation Analysis in MABEL

#### 4.1 Simulation Methodology

We used a Monte Carlo simulation approach to test the applicability of the parcelization algorithm in MABEL. Briefly, the Monte Carlo procedure we employed follows a five-step process described by Mooney (1997) where we: (a) defined the artificial pseudo-population to reflect the “real” observed population; (b) generated a pseudo-sample from that pseudo-population; (c) estimated the statistic of interest, and store it as a vector element; (d) repeated (b) and (c) \( t \) times, where \( t \) is the number of the replications, and; (e) constructed a relative frequency distribution of all the estimated statistics across all \( t \) replications.

An important consideration for all Monte Carlo experiments is the necessary size of the replications required to provide a sufficiently adequate confidence level on the observed behavior of the statistic of interest. We follow (Kleijnen, 2004) who suggested using the \( cn \) rule, where \( n \) is the number of attributes of interest, and \( c \) is a constant that is around 10 or more. Our goal is to observe and analyze the behavior and landscape attributes of the parcelization algorithm in MABEL across different landscapes \( (n=2; \) a county dominated by urban and another county dominated by agriculture use) and land use types \( (n=3; \) urban, agriculture and forest). Thus, combining the landscape level and land uses we get \( n=2 \times 3 = 6 \). Thus, a Monte Carlo simulation of 60 or more runs is necessary to analyze two landscapes and three land uses.

#### 4.2 Study Areas

Our study area focused on two counties (Grand Traverse County and Mecosta County) located in Michigan, USA, where ownership parcels where digitized from plat maps for eight 3x3 square miles blocks and land uses were assigned to classes based on aerial photography interpretation. Data represent three sequential decadal time-steps of 1970, 1980 and 1990. A more detailed description of these data can be found in Brown et al. (2001). It was our aim to examine the effects (and accuracy) of simulation results across these two diverse collections of landscapes which are commonly encountered across the Midwest region. Grand Traverse County (hereafter as GTC) blocks represent a dynamically changing urban/suburban landscape at the fringe of forest and agricultural land uses and with several scattered lakes that influence land uses. Between 1970 and 1990, the number of parcels in the 720 square miles in GTC double. In 1970, the GTC study area was 9.4% urban and the remainder nearly split between agricultural and forest use. In contrast, Mecosta County (hereafter as MC) represents a more static landscape for agriculture, with nearly two-thirds of this study area remaining agricultural through the twenty years. The amount of urban increase, however, nearly three fold, represented a greater increase in urban than GTC during the same period. Thus, parcelization was more dramatic in GTC than in MC despite a greater proportional increase in urban in MC than in GTC from 1970 to 1990.
Table 1. Historical land use parcel changes in eight 3x3 mile sampling frames in two Michigan counties (1970-1990).

<table>
<thead>
<tr>
<th>Region</th>
<th>Time Scale</th>
<th>No. of Parcels</th>
<th>No. Changed</th>
<th>Percent Changed (2-way)</th>
<th>Percent Changed (3-way)</th>
<th>Percent Urban</th>
<th>Percent Agriculture</th>
<th>Percent Forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grand Traverse</td>
<td>1970</td>
<td>711</td>
<td>9</td>
<td>9.4%</td>
<td>41.8%</td>
<td>45.6%</td>
<td>41.8%</td>
<td>35.6%</td>
</tr>
<tr>
<td></td>
<td>1980</td>
<td>1074</td>
<td>363</td>
<td>51.06%</td>
<td>12.1%</td>
<td>36.5%</td>
<td>48.1%</td>
<td>35.6%</td>
</tr>
<tr>
<td></td>
<td>1990</td>
<td>1432</td>
<td>358</td>
<td>33.33%</td>
<td>101.41%</td>
<td>15.8%</td>
<td>29.9%</td>
<td>50.3%</td>
</tr>
<tr>
<td></td>
<td>1970</td>
<td>651</td>
<td>3</td>
<td>3.6%</td>
<td>66.8%</td>
<td>26.3%</td>
<td>35.6%</td>
<td>41.8%</td>
</tr>
<tr>
<td></td>
<td>1980</td>
<td>913</td>
<td>262</td>
<td>40.25%</td>
<td>6.6%</td>
<td>66.8%</td>
<td>26.3%</td>
<td>35.6%</td>
</tr>
<tr>
<td></td>
<td>1990</td>
<td>1089</td>
<td>176</td>
<td>19.28%</td>
<td>9.1%</td>
<td>66.8%</td>
<td>21.1%</td>
<td>35.6%</td>
</tr>
</tbody>
</table>

For each of the historical time periods, an intermediate change was calculated yielding two sequential land use change datasets (1970-1980, and 1980-1990), the GIS polygon interface data was provided as input in the MABEL model. For each, we ran one hundred Monte Carlo Simulations according to the initial simulation assumptions and configurations outlined in the next section in order to account for the uncertainty of the simulation estimations, and provide a consistent range of potential simulation outcomes that are subject to the variation embedded in the dynamic nature of the agent-based modeling approach.

Thus, three temporal categories of Monte Carlo simulations were generated:

a. MABEL Monte Carlo Simulations (100), starting from 1970, to 1980.
b. MABEL Monte Carlo Simulations (100), starting from 1980, to 1990.
c. Two-step MABEL Markov Chain Monte Carlo Simulations (100), starting from 1970, intermediate at 1980 (step1), and continuing to 1990 (step2), resetting the random seed between decadal runs.

4.3 Simulation Assumptions

Time-steps in MABEL simulations were fixed so the number of the transactions in the simulation equaled the number of transactions in the historical database for each county (see Table 1). We conducted six sets of Monte Carlo replication experiments, two sets of the three temporal decadal simulations, for a total of 600 simulations. These experiments were repeated for each of the potential best algorithm weight configurations (see assumption g below), which resulted in a total of 3000 different simulations for this study.

We introduced the following stochastic conditions and rules for the parcelization transitions:

a. For each simulation run, there are fifty (50) potential buyers and fifty (50) potential sellers participating in the market bidding procedure, chosen randomly from the existing agents using a random number generator, but using a different random seed for each run. This does not mean, however, that there are fifty transactions occurring on each replication experiment, since the later is dynamically determined by the model dynamics, but rather introduces an adequate level of stochasticity in the simulation that within the entire Monte Carlo replication span, accounts for detrending the potential spatial dependencies that might be present in the initial landscapes.

b. The basic agent classes are designed to represent a level-I land use classification, i.e. urban (resident-type agents), agriculture (farmer-type agents), and forest/wetlands (forest-type agents) and a set of allowable transactions were imposed as follows:
   i. **Sell Action Function**: Forest & Wetlands → Agriculture → Urban; or Forest & Wetlands → Urban.
ii. **Buy Action Function**: Urban → Agriculture → Forest & Wetlands; or Urban → Forest & Wetlands.

In practical terms, this means that a forest-type agent can sell his or her land to any of the three agent classes (that is, he or she can sell to another forest-type agent, or to either a farmer-type agent or a resident-type agent); a farmer-type agent can sell his or her land to either another farmer-type agent or a resident-type agent; and a resident-type agent can only sell his or her land to another resident-type agent. Similarly, in terms of the buy action function (which is defined only for new agents created in the simulation), a new resident-type agent can buy land from any of the three agent classes, a new farmer-type agent can buy land from another farmer-type agent or a forest-type agent, and a new forest-type agent can only buy land from another forest-type agent.

c. There are two sub-classes of agents that purposely were restricted in the simulation: the water-type agents, and the transportation-type agents. They are still present in the simulation but their action space is set to null.

d. There are no restrictions imposed on the repeatability of agent actions throughout the Monte Carlo simulation replications. Thus, an agent already transitioned in a previous time step can freely participate in subsequent time steps of the simulation.

e. While the amount of land to be sold was fixed at 50% of the seller agent’s area, no restrictions were imposed on the spatial location of the sell action functions that the agents utilize. Certainly, one can impose such a restriction, especially in cases where the goal of the simulation is to determine and assess the simulation performance of spatially dependent types of situations (for example, when studying sprawl dynamics of land use), yet the purpose of our current Monte Carlo experiments is to assess the parcelization algorithm performance throughout the landscape, therefore keeping the number of assumptions and restrictions imposed to the model’s structure at a minimum for our analysis is essential.

f. We fixed both the minimum transaction area and excluded from consideration parcels where the width to height ratio exceeded certain values. Both of these factors are contained in MABEL because in the future we will introduce land use zoning ordinances that encompass these two factors. In our Monte Carlo simulations, we fixed the minimum transaction areas to 64 cells: the resolution of our land use maps is 30 square meters, which amounts to approximately 0.5 acre a minimum area that can be transitioned at any time. We also fixed the range of dimensions that such a parcel can have to the following:

\[
4 \leq \frac{x}{y} \leq \frac{1}{4}
\]  

(5)

where, \(x\) and \(y\) are the parcel’s vertical and horizontal length dimensions respectively, as measured by the length of the line connecting the longest path on each direction (in non-square parcels). The \(x\) and \(y\) length dimensions decompose into the length of the edges in a completely square parcel.

g. Finally, we compared five different parcelization algorithm weight configurations (Table 2) described in equation 4. We tested whether the weight configurations have significant implications on the parcelization routines, given the shape properties of the three metrics of the \(B_{ij}\) algorithm. We attempted to increase the significance of the squarness \(O\) component while we introduce increased significance of the clumpiness \(L\) component (configurations 1, 4, and 5), and compare their performance with relative balanced weight configurations (2 and 3).
Table 2. Parcelization algorithm weight configurations for MABEL Monte Carlo simulation runs.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>$w_1 (O)$</th>
<th>$w_2 (W)$</th>
<th>$w_3 (L)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3/5</td>
<td>1/5</td>
<td>1/5</td>
</tr>
<tr>
<td>2</td>
<td>2/4</td>
<td>1/4</td>
<td>1/4</td>
</tr>
<tr>
<td>3</td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td>4</td>
<td>3/6</td>
<td>1/6</td>
<td>2/6</td>
</tr>
<tr>
<td>5</td>
<td>3/7</td>
<td>1/7</td>
<td>3/7</td>
</tr>
</tbody>
</table>

4.4 Monte Carlo Analysis

Output from the simulation were introduced to FRAGSTATS (McGarigal et al., 2002) and a series of landscape ecology spatial pattern metrics were computed for each of the 3000 replications of the experiment. The use of landscape ecology spatial pattern metrics for assessing spatial accuracy has been used by Pijanowski et al. (2004) to test the ability of the artificial neural network (ANN)-based LTM model to simulate land use changes in two large metropolitan areas in the Upper Midwest area of the US. FRAGSTATS computes spatial pattern metrics for three different levels of spatial patterns: patch, class and landscape. We employ the latter two here for this study and focus on those spatial pattern metrics that quantify shape. Class pattern metrics calculate summary statistics (e.g., means, standard errors) for cells in patches (i.e., parcel) belonging to the same class; land use in our case. Landscape shape metrics calculate summary statistics for entire landscapes without regard to classes contained in the landscape. We report on class shape metrics whenever we compare land uses between locations and between temporal simulations. Landscape shape metrics are used to compare shape metrics between locations and/or years.

Analyses of shape metric data were accomplished using SPSS (SPSS Inc., 2003). Class shape metrics used are: shape index, clumpiness index and fractal dimension index and landscape division index. Landscape shape metrics used are: shape index, fractal dimension index, related circumscribing circle, landscape shape index, contiguity index, perimeter-area fractal dimension and landscape division index. The shape metric equations and their descriptions are given in Appendix B of this paper.

We also conducted partial correlation analysis to examine whether the shape metrics that we used provided us with significantly different results. This analysis, following Sokal and Rholf (1969), and Sprott (2000), controlled for the differences in location (county), simulation year and the simulation algorithm calibration weights.

We used the Kolmogorov–Smirnov test (KS) (D’Agostino and Stephens, 1986; Massey, 1951) to examine the null hypothesis that the patterns generated by the parcelization routine do not significantly differ from a random distribution of patterns. The KS test utilizes an empirical cumulative distribution function,

$$
\hat{F}(x) = \begin{cases} 
0 & \text{if } -\infty < x_i < x_1 \\
\frac{i}{N} & \text{if } x_i \leq x_i < x_{i+1} , \forall i=1,2,...,N-J \\
1 & \text{if } x_N \leq x_i < \infty
\end{cases}
$$

where,
The $i$th observation in the sample, ranging from 1 to N-1 (N=the total number of observations in the sample, in ascending order). The KS tests the underlying empirical cdf against a normal distribution, with mean $\mu$ and standard deviation $\sigma$,

$$
\mu = \frac{\sum_{i=1}^{N} x_i}{N}, \text{ and } \sigma = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}{N(N-1)}} \tag{7}
$$

by calculating the theoretical cumulative distribution function, $F_0(x_i) = F_{0.1}\left(\frac{x_i - \mu}{\sigma}\right)$, and computing the differences as such,

$$
D_i = \hat{F}(x_{i-1}) - F_0(x_i), \text{ and } \tilde{D}_i = \hat{F}(x_i) - F_0(x_i) \tag{8}
$$

The KS statistic is $Z_{KS} = \sqrt{N} \max_i(|D_i|,|\tilde{D}_i|)$, and the two-tailed probability level is calculated by

$$
\begin{cases}
  if \quad 0 \leq Z_{KS} < 0.27 & p = 1 \\
  if \quad 0.27 \leq Z_{KS} < 1 & p = 1 - \frac{2.506628(O + Q^9 + Q^{25})}{Z_{KS}} \quad [Q = e^{-1.2337Z_{KS}^2}] \\
  if \quad 1 \leq Z_{KS} < 3.1 & p = 2(Q - Q^4 + Q^9 - Q^{16}) \quad [Q = e^{-2Z_{KS}^2}] \\
  if \quad Z_{KS} \geq 3.1 & p = 0
\end{cases} \tag{9}
$$

Finally, we used the Reliability Analysis to assess how reliable our stochastic simulation methods were in terms of capturing the variability present in our real world observed changes. We used Cronbach’s alpha as a measurement of reliability. It is the lower bound of the true reliability of the sample. It is defined as the part of the variability in the sample results that can be accounted by differences in the parcels (spatial items). That is, we expect that the Monte Carlo simulation results will mainly vary across land uses due to the size, shape and individual spatial characteristics of the initial (original) land parcels, and not due to the stochastic nature of the process. The mathematical form of Cronbach’s alpha is,

$$
\alpha = \frac{k \cdot (Cov/Var)}{1 + (k-1) \cdot (Cov/Var)} \approx \frac{n \cdot \bar{r}}{1 + (n-1) \cdot \bar{r}} \tag{10}
$$

where $\alpha$ is Cronbach’s coefficient, or alpha, $k(=n)$ is the number of items in the sample, and $\bar{r}(=\overline{Cov/Var})$ is the ratio of the average inter-item covariance to the average item variance, but under the assumption of equal variances of the sample items, it denotes the inter-item correlation coefficient. When $r$ is used, the ratio is also known as standardized item alpha, or Spearman-Brown stepped-up reliability coefficient. Cronbach’s alpha values range from 0.0 to 1.0, with greater values reflecting great reliability. Nunnally (1978) has shown that an alpha value of 0.7 or greater is best although values between 0.5 and 0.7 could cautiously be considered to be reliable (Limnios and Nikulin, 2000).
5. Assessment of the Monte Carlo Results

As described in the previous section, shape metrics (class or landscape) were computed for the six sets of Monte Carlo simulations. Table 3 summarizes the results of the landscape shape metric partial correlation analysis. This analysis controlled for the differences in location, simulation year and simulation weight configuration. We can see from this table that there are four different groups of shape metrics. The first contains shape index, fractal dimensions, related circumscribing circle and landscape shape index. All are highly correlated with one another with p < 0.001; all correlation coefficients are greater than 0.800. The continuity index, on the other hand, is inversely correlated with the first four and has correlation coefficients around 0.5 with metrics in the first group. The perimeter-area fractal dimension index has small correlation coefficients with the first four shape metrics which are not significant. The landscape division index varies greatly with the other three types of metrics with correlation coefficients near 0.0 with weak significance levels for these coefficients.

Table 3. Landscape-level Shape Metrics Partial correlation coefficients. The correlation analysis controls for differences in the area (county), simulation year (by replication experiment), and simulation configuration (by algorithm calibration weights), and corresponding 2-tailed significance levels.

<table>
<thead>
<tr>
<th>Control Variables</th>
<th>Shape Index</th>
<th>Fractal Dimension Index</th>
<th>Related Circumscribing Circle</th>
<th>Landscape Shape Index</th>
<th>Contiguity Index</th>
<th>Perim-Area Fractal Dimension</th>
<th>Landscape Division Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area &amp; Year &amp; Weights Configuration</td>
<td>1.000</td>
<td>.999</td>
<td>.984</td>
<td>.810</td>
<td>-.525</td>
<td>.202</td>
<td>-.009</td>
</tr>
<tr>
<td>Fractal Dimension Index</td>
<td>.999</td>
<td>1.000</td>
<td>.987</td>
<td>.812</td>
<td>-.548</td>
<td>.180</td>
<td>-.001</td>
</tr>
<tr>
<td>Related Circumscribing Circle</td>
<td>.984</td>
<td>.987</td>
<td>1.000</td>
<td>.801</td>
<td>-.570</td>
<td>.116</td>
<td>.008</td>
</tr>
<tr>
<td>Landscape Shape Index</td>
<td>.810</td>
<td>.812</td>
<td>.801</td>
<td>1.000</td>
<td>-.422</td>
<td>.517</td>
<td>.348</td>
</tr>
<tr>
<td>Contiguity Index</td>
<td>-.525</td>
<td>-.548</td>
<td>-.570</td>
<td>-.422</td>
<td>1.000</td>
<td>.281</td>
<td>-.088</td>
</tr>
<tr>
<td>Landscape Division Index</td>
<td>-.009</td>
<td>-.001</td>
<td>.008</td>
<td>.348</td>
<td>-.088</td>
<td>.155</td>
<td>1.000</td>
</tr>
</tbody>
</table>
The KS test was computed across configuration weights for the two locations and for the three temporal simulation sets for each five landscape metrics. The results (Table 4) indicate that in the overwhelming majority of the Monte Carlo simulations, we can reject the null-hypothesis, that is, the landscape metrics of the sample replications observe a random normal distribution. The findings support the fact that the dynamic parcelization algorithm that the MABEL simulation model utilizes is able to provide a non-random distribution of parcels across the landscape.

Table 4. One sample Kolmogorov-Smirnov test for the MABEL Monte Carlo simulation replications. The significance of the KS test (p-values) is reported in italics beneath each test’s Z-score.

<table>
<thead>
<tr>
<th>Year</th>
<th>Configuration</th>
<th>Shape Index Mean</th>
<th>Fractal Dimension Index</th>
<th>Related Circumscribing Circle Index</th>
<th>Contiguity Index</th>
<th>Landscape Division Index</th>
<th>Asymp. Sig. (2-tailed)</th>
<th>Kolmogorov-Smirnov Z</th>
<th>Asymp. Sig. (2-tailed)</th>
<th>Kolmogorov-Smirnov Z</th>
<th>Asymp. Sig. (2-tailed)</th>
<th>Kolmogorov-Smirnov Z</th>
<th>Asymp. Sig. (2-tailed)</th>
<th>Kolmogorov-Smirnov Z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cfg4 (3-1-2)</td>
<td>Kolmogorov-Smirnov Z</td>
<td>.004</td>
<td>0.026</td>
<td>0.024</td>
<td>0.000</td>
<td>0.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cfg5 (3-1-3)</td>
<td>Kolmogorov-Smirnov Z</td>
<td>.000</td>
<td>0.012</td>
<td>0.014</td>
<td>0.000</td>
<td>0.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4 shows the mean perimeter-area fractal dimension index across 100 run simulations organized by location and across the three temporal simulations. The entire simulation mean and 95% confidence interval are also plotted here as solid and dash lines, respectively. Note that for the six sets of simulations nearly all runs are within the upper and lower 95% confidence limits.
When different landscape shape metrics (Figure 5) for the 1970-1980 simulations are compared for both locations, the MC metrics have large deviations from the mean with more consistency between values within simulations for the fractal dimension index. This is probably due to the fact that the MC landscape contains few, very large parcels (in agriculture), compared to the landscape size, which results in larger deviations from the mean when these parcels subdivide. Again, most of the simulation runs have values that are within the 95% confidence interval of each 100-run simulation.
Figure 5. Three patterns of mean landscape metrics: Fractal Dimension Index, Contiguity Index, and Landscape Division Index for the Monte Carlo Replication Experiments. (a) Grand Traverse County by simulation run year; (b) Mecosta County by simulation run year. In all graphs, solid and dotted lines denote replication run means and 95% confidence interval (CI) limits, respectively.

Mean shape index values, by land use, location and simulation year, graphed across simulation runs, are shown in Figure 6. The values for shape index are greater than or equal to 1.0; larger values reflect patches that depart in shape from a perfect square of the same size. In all six simulation sets, forest use has the largest mean shape index, followed by agriculture and then urban. The differences between mean shape index values between the two decal simulations (1970-1980 and 1980-1990) are not great. However, of significance is the 1970-1990 simulation in MC; here, the mean shape index fluctuated wildly between simulations and contained values greater than 1.50 in approximately 20% of the entire population of simulations that were not observed in the two decal simulations or any of the three temporal simulations in GTC. Given
that there are few, large patches of agriculture in this landscape, some simulations may have resulted in highly irregular shapes, especially during the early portion of the simulation.

**Figure 6.** Mean shape index for all patches in each location composed of urban (solid, grey line), agriculture (solid black line) and forest (dotted line). Simulations are grouped according to simulation year.

The land use class shape metrics are shown for the sequential decades 1970-1980, 1980-1990 and 1970-1990 across the two county groups, namely the Grand Traverse County group (sub-figures a.1-3), and the Mecosta County Group (sub-figures b.1-3) for the Monte Carlo simulation replications. Of the results, worth noticing that there seem to be a general, although weak, pattern in place indicating the different degree of variability and thus, fragmentation dynamics across the three level-I land use classes: urban, agriculture and forest. From the observed real changes maps (Figure 4), one can notice that there is a general trend of the urban parcels to be relatively smaller than the forest/wetlands, and these in turns, smaller than the agricultural ones. In addition, in Grand Traverse County, the dynamics of urban sprawl are noticeable and present in the landscape configuration, while in Mecosta County, the landscape configuration is still dominated by large, mainly agricultural-type parcels.
Figure 7. Means and error bars for 95% CI of the mean shape index in the MABEL Monte Carlo simulations. (a) Grand Traverse County land use categories by simulation temporal runs; (b) Mecosta County land use categories by simulation temporal runs. In all graphs, the solid and dotted lines denote the mean shape index of the real landscapes in the 1970-80 and 1980-90/1970-90 simulation runs respectively.
Figure 8. Means and error bars for 95% CI of the mean fractal dimension index in the MABEL Monte Carlo simulations. (a) Grand Traverse County land use categories by simulation temporal runs; (b) Mecosta County land use categories by simulation temporal runs. In all graphs, the solid and dotted lines denote the mean fractal dimension index of the real landscapes in the 1970-80 and 1980-90/1970-90 simulation runs respectively.
Figure 9. Means and error bars for 95% CI of the clumpiness index in the MABEL Monte Carlo Simulations. (a) Grand Traverse County land use categories by simulation temporal runs; (b) Mecosta County land use categories by simulation temporal runs. In all graphs, the solid and dotted lines denote the clumpiness index of the real landscapes in the 1970-80 and 1980-90/1970-90 simulation runs respectively.
Another interesting result can be seen in Figures 9 to 10, is that the transition dynamics across the two counties are apparent: as the land transitions from agriculture to forest to urban ultimately, the landscape divisional properties settle down to more stable equilibrium forms. It is
rather intuitive, but relatively easy to comprehend that the dynamics that initiate these transitions, including the deterministic simulation assumptions about the hierarchical process of parcelization included in our simulation design, are in part responsible for the level of entropy present in our landscapes. As the land changes from agriculture to forest to urban, the informational content necessary to characterize our landscapes is reduced, and eventually (in terms of urban or suburban parcels) will asymptotically reach zero: knowing that a given parcel is currently a residential or commercial parcel, one would not expect it to change further in terms of its land use class type. When we look at the side-by-side county comparisons, and considering the landscape configuration of the two counties, we can see that the transition dynamics occur in a spatial sequence as if we start from Mecosta County, and reach a level in the Grand Traverse County. Then the large variability of changes emerging in the agricultural parcels initially is reduced, in benefit of the change variability observed to the forest/wetland land use type parcels.

By comparing the different algorithm configurations across both areas and all land use types, we can see (Figures 7-10) that there is a significant amount of variation present among them, and a diversity of observed patterns of change. Given the nature of the two areas (i.e., predominately suburban landscape in GTC, versus an agriculture-dominated landscape in MC), and the patterns observed in our computed landscape indexes (i.e., high correlation between shape index and fractal dimension index, versus distinct patterns for clumpiness and landscape division indexes), we can make the following observations. In the first group (Figures 7-8) of metrics (shape index, fractal dimension index), different algorithm configurations appear to present more realistic results. In GTC, the configurations with emphasis on squareness (with higher weight on \( O_i \)) or clumpiness (higher weight on \( L_i \)) perform better compared to the mean metric value of the real changes, while in MC, more balanced configurations (with relatively similar weights among the algorithm components) display significantly better results compared to the mean metric value of the real changes. In the case of the clumpiness index (Figure 9), all configurations seem to perform relatively well, although running a continuous simulation sequence (i.e., from 1970 to 1990) seems to deviate disproportionately from the mean of running the simulations separately. In terms of the landscape division index there seem to be no noticeable difference on the performance of the five algorithm configurations. Finally, consistently over almost all metrics, the results observed for MC are significantly closer to the real observed mean values, than in GTC, which is an indication of the involved increased level of complexity of factors influencing the landscape configurations as we move from a more static (agriculture-dominated) to a more dynamic (suburban) landscape.

In our analysis, we used seven basic shape metrics, and computed Cronbach’s \( \alpha \) using the covariance matrix calculated for the simulation data. These results (Figure 11) indicate clearly that the majority of the simulation runs and metrics measured display a relative low level of reliability, not without exceptions however. While the Cronbach’s \( \alpha \) values are more appropriate measures of reliability in our analysis, mainly due to the use of the covariance matrix used in the analysis (thus, the assumption of equal variances cannot be easily made), we reported also the standardized version of the index, as an indicator of the relative importance of the values computed. In most cases, the simulation results cannot capture the process that has been observed in the simulation replications. It is our claim here that this inability to capture the underlying landscape processes is due to the significant impact that the human-related activities have on the landscape and parcel configuration observed in the real world. A significant finding to be noted here also is that there are obvious observable differences on the levels of reliability between the two spatial regions of the simulation, mainly due to the nature and type of land use changes present on each area. In addition, the statistics based on standardized item correlations indicate the level of diversity present in the different land use types. The relative heterogeneity of the urban landscape in terms of their landscape shape composition, and the relative homogeneity of the agricultural landscapes, respectively, is the reason for the observed differences across land use types.
6. Discussion and Conclusions

In this paper, we presented and tested a parcelization algorithm implemented in our spatially-
explicit, agent-based, land-use change model which we call the Multi-Agent Behavioral Economic
Landscape (MABEL) model. We employed a Monte Carlo simulation approach using a series of
replication experiments across time, comparing observed changes between 1970, 1980 and
1990, and across two groups of landscapes that present different kinds of land use changes. We
compare the simulated parcel shapes to historically observed land use changes using several
landscape ecology spatial pattern metrics generated by FRAGSTATS.

We examined seven different landscape-level shape metrics using a partial correlation analysis
controlling for differences in location, simulation year and simulation configuration and found that
four of these (shape index, fractal dimension, related circumscribing circle and landscape shape
index) were highly autocorrelated. Three others (congruity perimeter-area fractal dimension and
the landscape division index) provided different results from one another and from the four highly
related shape metrics. This underscores the importance of examining a variety of shape metrics
for these types of studies and to examine how they are autocorrelated. A diverse set of shape
metrics are likely to provide more a robust assessment of model performance for any parcelization
routine. We were also able to show that the parcelization routine produced shapes across
locations that departed significantly from random distribution of shapes and that nearly all shape
metrics provided us with the ability to detect these differences from normal distribution, except the
related circumscribing circle metric. A reliability analysis showed that the shape metrics that we
used were best suited for urban patterns, followed by forest and then agriculture. Standardized
Cronbach’s alpha exceed very good levels (greater than 0.4) for urban in all years and for both
locations. Standardized Cronbach’s alpha for the other years and uses were less than 0.4.

A comparison of the performance of the parcelization routine across different locations (GTC
and MC) and difference temporal periods (1970-1980; 1980-1990; and 1970-1990) suggest that
the routine performed equally well across locations and temporal periods. There was also a very
consistent pattern of shape for land uses within simulation years and locations except for the
twenty year (i.e., 1970-1990) simulation in Mecosta were over 20% of the runs produced mean
shape index values that were much larger than in the other simulation sets.

We also examined how different configuration weights for the best parcel algorithm used by
seller and buyer agents prior to splitting the original parcel. We found that the weights varied

---

Figure 11. Reliability analysis level comparisons between Grand Traverse and Mecosta Counties by
land use type and simulation year. Left graph: Cronbach’s Alpha Statistic; Right graph: Cronbach’s
Alpha based on standardized item correlations.
considerably between locations and simulation years produced shape metrics that either was significantly larger or smaller than in the observed dataset.

The parcelization routine will require additional components that we believe are important to the parcelization process in the real world. First, parcelization is greatly influenced by accessibility to roads. Parcels need to be subdivided in a way that provides access to roads for all new parcels. Second, proximity to natural resources such as rivers, lakes, and views/vantagepoints (e.g., placing homes on hilltops) as well as infrastructure amenities such as utilities, need to be considered as well. Third, parcels are likely to be split and possibly be split in large aggregations via subdivision development based on neighborhood characteristics. Fourth, many parcelization decisions are made on the basis of complex planning and management rules which are designed for economic or environmental goals at local to international scales (Münier et al., 2002; Kleijnen, 2004; Maret, 1988; Holling et al., 2001; Bousquet et al., 2001; Holling, 2001). These need to be examined and incorporated into the parcelization rules as well. Finally, all of these behaviors have an economic and psychological determinants (Parker et al., 2003; Parker and Berger, 2002; Alexandridis et al., in revision; Bockstael, 1996; Irwin and Bockstael, 2002; Roe et al., 2002), need to be incorporated into the parcelization process. Efforts to do this are underway using Bayesian Belief Networks (Alexandridis et al, in revision) and Role Playing Simulation (Campbell and Pijanowski, in prep.).

7. Acknowledgements

We wish to acknowledge funding from a grant from NASA’s Land-Cover and Land-Use Change Program (NAG5-6042) and a cooperative agreement with the USDA Forest Service North Central Forest Experiment Station (#23-95-50), grants from the National Science Foundation BE/CNH 0308420 and WCR 0233648 and a grant from the Great Lakes Fisheries Trust. We would also like to thank Dan Brown for directing the development of the parcel database used in this study. Snehal Pithadia provided technical assistance as well.

8. Bibliography


Appendix A: Calculation of the Landscape Shape Metric (LSI) for the MABEL Parcelization Algorithm Computation

In computing the MABEL LSI metric for the parcelization algorithm, we can consider the agent area as shown in the example of Figure 3. Such an area can be expressed as the product of the vertical and horizontal dimensions of the smallest confined rectangle. In the specific example, \( i = \{A, B, C\} \), where \( A, B \) and \( C \) are the closest confined rectangles of the original agent, the buyer-agent and the seller-agent respectively.

\[
a_i = x_i \cdot y_i
\]

We can also define as \( n \) the closest integral of the square root of the area \( a \), and \( m \) as the difference between the area \( a \) and the squared \( n \). (McGarigal and Marks, 1994).

\[
n_i = \text{int}\left(\sqrt{a_i}\right)
\]

\[
m_i = a_i - n_i^2
\]

By replacing (A1) into (A2), we have

\[
n_i = \text{int}\left(\sqrt{x_i \cdot y_i}\right)
\]

And by replacing (A1) and (A2) into (A3) we have,

\[
m_i = (x_i \cdot y_i) - \text{int}\left(\sqrt{x_i \cdot y_i}\right)^2
\]

The conditions for the \( e_i \) calculations are (McGarigal et al., 1994; Bogaert et al., 2000):

\[
e_i = \begin{cases} 
4n_i & \text{if } m_i = 0 \\
4n_i + 2 & \text{if } n_i^2 < a_i \leq n_i(1+n_i) \\
4n_i + 4 & \text{if } a_i > n_i(1+n_i)
\end{cases}
\]

By replacing (A1), (A2), (A3), (A4), and (A5) to (A6), we have:

\[
e_i = \begin{cases} 
4\left(\text{int}\left(\sqrt{x_i \cdot y_i}\right)\right) & \text{if } (x_i \cdot y_i) - \left(\text{int}\left(\sqrt{x_i \cdot y_i}\right)\right)^2 = 0 \\
4\left(\text{int}\left(\sqrt{x_i \cdot y_i}\right)\right) + 2 & \text{if } \left(\text{int}\left(\sqrt{x_i \cdot y_i}\right)\right)^2 < x_i \cdot y_i \leq \left(\text{int}\left(\sqrt{x_i \cdot y_i}\right)\right)\left(1+\text{int}\left(\sqrt{x_i \cdot y_i}\right)\right) \\
4\left(\text{int}\left(\sqrt{x_i \cdot y_i}\right)\right) + 4 & \text{if } x_i \cdot y_i > \left(\text{int}\left(\sqrt{x_i \cdot y_i}\right)\right)\left(1+\text{int}\left(\sqrt{x_i \cdot y_i}\right)\right)
\end{cases}
\]

Also, we can define,
\[
\min e_i = \min \{e_i^k \}_{i=1}^k \\
\text{(A8)}
\]

The LSI metric can be calculated by,

\[
LSI_i = \frac{\min e_i^k}{e_i} \\
\text{(A9)}
\]

Or, by replacing (A8) into (A9), we have,

\[
LSI_i = \frac{\min \{e_i^k \}_{i=1}^k}{e_i} \\
\text{(A10)}
\]
Appendix B: 
Landscape and Land Use Class Metrics Used in the Analysis

Note: the following list of landscape and class metrics represents a short summary of the metrics and their attributes. Additional information can be found on the Fragstats documentation by McGarigal et al. (1994). A more complete metric database is the IDEFIX Indicator Database for Scientific Exchange (Larg, 2003).

1. Computed Landscape Metrics

Mean Shape Index: Measures the complexity of a patch shape at the landscape level. It is used to compare the landscape patch shape against a simple square shape of the same size. Its functional form is (McGarigal et al., 1994):

\[
SHAPE_{\text{MN}} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \left( \frac{P_{ij}}{\min P_{ij}} \right)}{N}
\]

where,
- \( P_{ij} \): the perimeter of the parcel \( ij \) in terms of the number of cell sizes;
- \( \min P_{ij} \): the minimum perimeter of a parcel \( ij \) in terms of the cell sizes. The forms that \( \min P_{ij} \) can take are three (\( a_{ij} \) is the area of the parcel \( ij \), and \( n \) is the side of a larger integer square smaller than \( a_{ij} \)):
  i. \( \min P_{ij} = 4n \) when \( m = a_{ij} - n^2 = 0 \);
  ii. \( \min P_{ij} = 4n+2 \) when \( n^2 < a_{ij} < n(n+1) \);
  iii. \( \min P_{ij} = 4n+4 \) when \( a_{ij} > n(n+1) \);
- \( N \): the number of parcels in the landscape.

The mean shape index equals 1 when the parcel is approaching a square shape, and increases without limit as the shape of the parcel becomes more irregular.

Mean Fractal Dimension Index: Indexes the relative degree of complexity of a parcel by computing its fractal dimension \( D \), such that the perimeter of a patch is related to the area of the same patch by \( P \approx \sqrt{AD} \) (or, \( \log P \approx 0.5D \cdot \log A \)). For simple Euclidean shapes (e.g., circles and rectangles), \( D=1 \). For more complex parcels, the perimeter becomes increasingly plane-filling and \( P \approx A \) with \( D \rightarrow 2 \). (McGarigal et al., 1994). The functional form of the index is:

\[
FRAC_{\text{MN}} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \left( \frac{2 \cdot \ln(0.25 P_{ij})}{\ln a_{ij}} \right)}{N}
\]

where,
- \( P_{ij} \): perimeter \( (m) \) of parcel \( ij \).
- \( a_{ij} \): area \( (m^2) \) of parcel \( ij \).

The mean fractal dimension index range is \( 1 \leq FRAC_{\text{MN}} \leq 2 \).
**Mean Related Circumscribing Circle:** It is similar to the mean shape index, with the difference that it uses the smallest circumscribing circle instead of the smallest circumscribing square despite the raster data format because it is much simpler to implement, and it also accounts for effects that deviate from the main horizontal-vertical directions, thus providing a measure of overall patch elongation (McGarigal et al., 1994). The functional form of the metric is:

\[
CIRCLE_{\text{MN}} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[ 1 - \left( \frac{a_{ij}}{a_{ij}^c} \right) \right]
\]

where,
- \(a_{ij}\) = area (\(m^2\)) of parcel \(ij\).
- \(a_{ij}^c\) = area (\(m^2\)) of smallest circumscribing circle around parcel \(ij\).

The range of the values of the metric is \(0 \leq CIRCLE_{\text{MN}} < 1\), and approaches zero for relatively circular patches of parcels, or 1 for elongated, relative linear parcels.

**Landscape Shape Index:** Measures the class aggregation at the landscape level. Equals (McGarigal et al., 1994). The functional form of the index is:

\[
LSI = \frac{e_i}{\min e_i}
\]

where,
- \(e_i\) : total length of the perimeter of a landscape class \(i\), given the number of cell units.
- \(\min e_i\) : the minimum length of the landscape class perimeter required for a maximally aggregated class.

The LSI ranges are defined as follows:
- i. \(\min e_i = 4n\) when \(m=0\);
- ii. \(\min e_i = 4n+2\) when \(n^2 < a_i < n(n+1)\);
- \(\min e_i = 4n+4\) when \(a_i > n(n+1)\);

where, \(a_i\) is the area of the landscape class \(i\), \(n\) is the smaller integer square smaller than \(a_i\), and \(m = a_i - n^2\).

**Mean Contiguity Index:** Measures the spatial connectedness of a landscape by accounting for boundary configurations. (McGarigal et al., 1994). The functional form of the index is:

\[
CONTIG_{\text{MN}} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[ \frac{\sum_{z=1}^{z} c_{ijr}}{a_{ij}} \right] - 1
\]

where,
$c_{ijr}$: the contiguity value for a cell $r$ in a given patch $ij$.

$v$: the sum of the values in a 3x3 cell template.

$a_{ij}$: the area of the patch $ij$.

The mean contiguity index range is $0 \leq \text{CONTIG}_\text{MN} \leq 1$.

**Perimeter-Area Fractal Dimension Index**: It accounts for consideration of a variety of spatial scales in measuring shape complexity. It involves computation of the slope of the regression line between the log of patch area and the log of patch perimeter (McGarigal et al., 1994). It uses OLS regression for the equation

$$\ln(\text{area}) = b_0 + b_1 \ln(\text{perim}) + \epsilon_i$$

(B6)

The functional form of the index is:

$$\text{PAFRAC} = \frac{2}{\sqrt{\frac{N \sum_{i=1}^{m} \sum_{j=1}^{n} \left( \ln p_{ij} \cdot \ln a_{ij} \right)}{\left( \sum_{i=1}^{m} \sum_{j=1}^{n} \ln p_{ij} \right) \left( \sum_{i=1}^{m} \sum_{j=1}^{n} \ln a_{ij} \right)} - \frac{\left( \sum_{i=1}^{m} \sum_{j=1}^{n} \ln p_{ij} \right)^2}{\sum_{i=1}^{m} \sum_{j=1}^{n} \ln p_{ij}}}}$$

(B7)

where,

$p_{ij}$: perimeter (m) of parcel $ij$.

$a_{ij}$: area ($m^2$) of parcel $ij$.

$n$: number of patches in the landscape of class $i$.

The perimeter-area fractal dimension index range is $0 \leq \text{PAFRAC} \leq 1$.

**Landscape Division Index**: Is the probability that two randomly chosen raster cells in the landscape are not located in the same landscape patch (McGarigal et al., 1994). The functional form of the index is:

$$\text{DIVISION} = \left[ 1 - \sum_{i=1}^{m} \sum_{j=1}^{n} \left( \frac{a_{ij}}{A} \right)^2 \right]$$

(B8)

where,

$a_{ij}$: the area of the patch $ij$.

$A$: total landscape area

The landscape division index range is $0 \leq \text{DIVISION} \leq 1$.

**2. Computed Land Use Class Metrics**

**Mean Shape Index**: It resembles the mean shape index at the landscape level, but it is computed separately for different land use class types. Its functional form is (McGarigal et al., 1994):
SHAPE\_MN_c = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \left( \frac{p_{ij}}{\min p_{ij}} \right)}{N_c} \quad (B9)

where,

- \( p_{ij} \): the perimeter of the parcel \( ij \) in terms of the number of cell surfaces;
- \( \min p_{ij} \): the minimum perimeter of a parcel \( ij \) in terms of the cell surfaces.
- \( c \): the land use class (between 1 and C)
- \( N_c \): the number of parcels in the landscape that belongs to class \( c \).

Although computed separately, the landscape-level index and the class-level index are associated in a way that:

\[ SHAPE\_MN \approx \sum_{c=1}^{C} (SHAPE\_MN_c) \quad (B10) \]

**Mean Fractal Dimension Index**: Again, it is similar to the fractal dimension index at the landscape level, but it is computed separately for different land use class types. The functional form of the index is (McGarigal et al., 1994):

\[ FRAC\_MN_c = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \left[ \frac{2 \cdot \ln(0.25 p_{ij})}{\ln a_{ij}} \right]}{N_c} \quad (B11) \]

where,

- \( p_{ij} \): perimeter (m) of parcel \( ij \).
- \( a_{ij} \): area (m²) of parcel \( ij \).

The landscape-level metric and the class-level metric are associated in the following way:

\[ FRAC\_MN \approx \sum_{c=1}^{C} (FRAC\_MN_c) \quad (B12) \]

**Mean Related Circumscribing Circle**: Similar to the metric at the landscape level, but computed for different land use classes. The functional form of the metric is (McGarigal et al., 1994):

\[ CIRCLE\_MN_c = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} \left[ 1 - \left( \frac{a_{ij}}{a_{\_ij}} \right)^{\_j} \right]}{N_c} \quad (B13) \]

where,

- \( a_{ij} \): area (m²) of parcel \( ij \).
\( a_{ij} \) = area \((m^2)\) of smallest circumscribing circle around parcel \( i j \).

The landscape-level and class-level metrics are associated in the following way:

\[
CIRCLE\_MN \approx \sum_{c=1}^{C} (CIRCLE\_MN_c) \quad (B14)
\]

**Landscape Division Index**: Represents the probability that two randomly chosen pixels in the landscape are not located in the same class patch. (McGarigal et al., 1994). Its functional form is:

\[
DIVISION = \left[ 1 - \sum_{j=1}^{n} \left( \frac{a_{ij}}{A} \right)^2 \right] \quad (B15)
\]

where,

- \( a_{ij} \) = area \((m^2)\) of parcel \( i j \).
- \( A \) = total landscape area \((m^2)\).

The range of the index is \( 0 \leq DIVISION < 1 \). The value of the landscape division index approaches zero when the landscape consists of a unique class.