GEOMOD Modeling

By Robert Gilmore Pontius Jr. (rpontius@clarku.edu) and Hao Chen. Clark University. 2006.

Land-Use & Cover Change Modeling

Land-Use & Cover Change (LUCC) modeling is a rapidly growing scientific field because land-use change is one of the most important ways that humans influence the environment. The issue is so important that scientists have formed an international organization, called "LUCC," which is connected with the International Human Dimensions of Global Change Program and the International Geosphere Biosphere Program. The LUCC organization holds regular meetings and provides a communication infrastructure for professional scientists interested in land transformation (LUCC 2002). Their publications articulate priorities for land-use change research, which include modeling as particularly important (Turner II et al. 1995; Lambin et al. 1999). The last year of the formal LUCC program is 2005; it is likely that a Global Land Program will pick up where LUCC leaves off.

Additional literature on land-use change modeling is growing rapidly. It is increasingly common for professional journals to dedicate special issues to LUCC (Veldkamp and Lambin 2001; Bian and Walsh 2002). Agarwal et al. (2002) provides a summary of many LUCC models. Additional reviews include Baker (1989) and Lambin (1994).

It is difficult to compare the performance of the numerous models because LUCC models can be fundamentally different in a variety of ways. For example, some models, such as IDRISI's GEOMOD, simulate change between two land categories (Silva and Clarke 2002; Pontius et al. 2001) while others, such as IDRISI's CA_MARKOV, can simulate change among several categories (Li and Reynolds 1997; Wagner 1997; Wu and Webster 1998; Pontius and Malanson 2005). Still others simulate change in real variables as opposed to categorical variables (Veldkamp and Fresco 1996). Most models are for raster data, while some are for vector data. Even if all researchers were to use the same model, comparison among model performance would still be difficult because researchers usually focus on one specific study region. Therefore, it is difficult to separate the quality of the model from the complexity of the landscape and of the data. For example, if a model performs poorly, it is difficult to know whether the conceptual foundation of the model is weak, or the phenomenon of land change at the study site is particularly complex, or the data is particularly detailed. Alternatively, if a model performs well, it is difficult to know whether the conceptual foundation of the model is strong, or the phenomenon of land change at the study site is particularly simple, or the data is very simplified. Perhaps most importantly, there is not yet an agreed upon method to measure the performance of LUCC models, so two modelers who use the same model on the same landscape with the same data might evaluate one simulation run differently depending on the criteria used for evaluation. Land-use change modelers seem to agree that the intellectual foundation of validation for land use change models is insufficient. The literature on rigorous validation of LUCC models is scant (Kok et al. 2001; Pontius 2002; Pontius and Schneider 2001; Pontius et al. 2004a). This paper attempts to address these issues. It presents a land-use change model called GEOMOD. In addition, it presents an intellectual foundation and statistical methods to validate land-use change models from the GEOMOD approach. All the concepts are illustrated with an application to simulate the land-use change in Central Massachusetts, USA. One advantage of using IDRISI for land-use change modeling is that in one software package you have: 1) the data, 2) the simulation model, and 3) the statistical techniques of goodness-of-fit.

Overview of the GEOMOD Approach

Pontius et al. (2001) give the most complete peer-reviewed description and application of GEOMOD. The version of GEOMOD currently in IDRISI is essentially the same model published under the names “GEOMOD1” and “GEOMOD2” (Hall et al. 1995a, 1995b, 1995c; Huffaker and Pontius 2002; Menon et al. 2001). It is the model that has been used frequently to analyze baseline scenarios of deforestation for carbon offset projects, as called for by the international agreements on climate change, such as the Kyoto Protocol. It has been used to predict land change: 1) at the continental scale for Africa, Asia and Latin America, 2) at the country scale for Costa Rica and India, and 3) at the local scale within India, Egypt, United States and several countries in Latin America. There are other examples that we learn about regularly since GEOMOD was incorporated into IDRISI.

GEOMOD is a grid-based land-use and land-cover change model, which simulates the spatial pattern of land change forwards or backwards in time. GEOMOD simulates the change between exactly two land categories denoted as 1 and 2. For simplicity of explanation, we refer to categories 1 and 2 as “non-developed” and “developed” respectively, but 1 and 2 could represent any two categories for any particular application. The user must supply a map of a beginning time and information concerning the number of grid cells of each category at an ending time, and then GEOMOD selects the location of the grid cells to classify as one of the two categories for the ending time. If there is a net increase in the developed category as the simulation proceeds from a beginning time to the ending time, then GEOMOD will search among the non-developed grid cells in order to select the cells that are most likely to be converted to become developed during the time interval. Conversely, if there is a net increase in the non-developed category as the simulation proceeds from a beginning time to a ending time, then GEOMOD will search among the developed grid cells in order to select the cells that are most likely to be converted to non-developed during the time interval.

The minimum input requirements are: the beginning time, the ending time, an image of the beginning time for two land cover types that must be denoted by 1 and 2, and an estimate of the number of cells of each of the two categories at the ending time. Most users also include either a suitability map that has already been created or driver maps that GEOMOD uses to create its own suitability map. Some users have a map of the true landscape at the ending time, which can be used for validation as described below. GEOMOD’s most important output is a map of the simulated landscape of developed versus non-developed cells at the ending time. If any of the maps have a mask, then the modeler should make certain that all maps have the exact same mask. GEOMOD has been designed such that it can take maximum advantage of data that can vary highly in availability, completeness, precision, currency, and accuracy. For example, GEOMOD requires only one beginning land-use map for calibration, while some algorithms for other popular models require maps from four times for calibration (Silva and Clarke 2002).

We illustrate the concepts of this chapter with an application of GEOMOD to the landscape of Central Massachusetts, USA. Figure 1 shows the study area’s 10 towns, which are Worcester and the nine towns that surround Worcester, Massachusetts, USA. Figures 2-4 show a process of land transformation from 1971 to 1985 to 1999 for two categories, developed and non-developed (MassGIS 2002). We use information from 1971 and 1985 to calibrate GEOMOD, then GEOMOD simulates the increase in development from 1985 to 1999, at which point we validate GEOMOD’s prediction with the image of the real landscape of 1999. Figure 5 is a crosstabulation of Figures 3 and 4, thus Figure 5 shows the phenomenon of land-use change that GEOMOD attempts to simulate.

Figure 6 shows an extrapolation of the quantity of non-developed area for each town. The symbols in the figure show the correct data according to the images for 1971, 1985 and 1999. For each town, the line is interpolated exactly through the points for 1971 and 1985, then the line is extrapolated to 1999, so that the line estimates the real quantity of non-disturbed area for 1999. For many towns, such as Worcester, the extrapolation comes extremely close to the real quantity in 1999, therefore the line intercepts the symbol for 1999. For a few towns, such as Shrewsbury and West Boylston, the line fails to predict the real 1999 quantity precisely. Thus Figure 6 shows how to use information from 1971 and 1985 to calibrate the prediction of the quantity of remaining non-developed land in 1999, with a known level of accuracy. We then use GEOMOD to simulate the locations of this estimated quantity of land-use change, as described below.

GEOMOD's Decision Rules to Specify the Location of Change

GEOMOD selects the locations of land to be converted according to four decision rules. The first decision rule is mandatory so there are no options for it in the user interface. Any of the later three decision rules can be either included or excluded at the discretion of the user.

The first decision rule concerns persistence on the landscape. GEOMOD simulates within each stratum a one way change, either from non-developed to developed or from developed to non-developed. For example, if the user-specifies a net increase of cells in the developed category, then GEOMOD simulates change from non-developed to developed for that number of additional cells. All other cells within that stratum will persist and no cells will show change from developed to non-developed. GEOMOD does not allow for conversion from non-developed to developed in some cells and from developed to non-developed in other cells simultaneously within the same stratum. The next rule concerns the strata.

The second decision rule allows for regional stratification. GEOMOD can simulate land change within a series of regions called strata, which are typically political regions for which tabular data commonly exist. If stratification is desired, the user must specify a stratum map (figure 1). For each stratum, the user specifies the quantity of each category at the ending time. GEOMOD can predict conversion from non-developed to developed in some strata and from developed to non-developed in other strata. Figure 7 shows the GEOMOD prediction of the 1999 landscape based on the first and second decision rules only, with a time step of 14 years. The quantity of non-developed area specified in Figure 6 is distributed among each of the 10 towns. Within each town, the new development is spread pseudo-randomly, since GEOMOD is not using any rules to specify where within the town the new development is likely to occur. In order to specify where the new development is likely to occur within the towns, GEOMOD must use the third and/or the fourth decision rules.

The third decision rule concerns the neighborhood constraint. It is based on a nearest neighbor principle, whereby GEOMOD restricts land change within any one time step to cells that are on the edge between developed and non-developed. This rule simulates the manner in which new development can grow out of previous development. For example, if GEOMOD is converting from non-developed to developed, then the neighborhood constraint mode will restrict the search to only those cells within a small square window around any developed cells. The width of the window, denoted by W, is called the neighborhood search width, which can be set by the user. The smallest possible search width is 3, which implies that any conversion from non-developed to developed must occur within one cell of an existing developed cell. The simulation updates the definition of edge at every time step. If GEOMOD must convert more cells than are available within a search width of W, then GEOMOD converts all available cells within the search width of W and begins to convert cells within a search width that is just sufficiently wide to contain the necessary number of cells that need converting. Figure 8 shows the GEOMOD result when the decision rules are persistent and the neighborhood constraint only, with a time step of 14 years. GEOMOD predicts the amount of new development based on the prediction of quantity of total non-developed area in each town shown in Figure 6. Notice that all of the predicted new development is located on the 1985 edge of developed and non-developed.

The fourth decision rule concerns a suitability map, which shows the suitability for land cover category 2, which is the developed category in our example. With this rule, GEOMOD simulates additional development by searching the landscape for the location of the non-developed cells that have the highest suitability value. If GEOMOD is simulating from developed to non-developed, then GEOMOD searches for the developed cells with the smallest suitability.

The user can create a customized suitability map by using Image Calculator in IDRISI or an IDRISI module such as MCE (Eastman et al. 1995). Alternatively, GEOMOD can create the suitability map based on driver maps, which are usually maps of biogeophysical attributes. Each driver map must show a categorical variable, not a real variable. If the user would like to use a real variable, such as slope, then the user must reclassify the real numbers to categorical bins, such as: category 1 = [0 degrees, 1 degree), category 2 = [1 degree, 2 degrees), etc.

GEOMOD creates the suitability map empirically, by using several driver maps and the land-cover map from the beginning time. GEOMOD's suitability map has relatively high values at locations that have biogeophysical attributes similar to those of the developed land of the beginning time, and has relatively low values at locations that have biogeophysical attributes similar to those of non-developed land of the beginning time.

GEOMOD's suitability map is created in two steps. First, GEOMOD reclassifies each driver map such that the grid cells of each category of the driver map are assigned a percent-developed real number, obtained by comparing the driver map to the beginning time land-cover map. The percent-developed real number of each category in the driver map is computed as the ratio of the quantity of developed grid cells of that category to the quantity of all grid cells of that category.

After each attribute map is reclassified, GEOMOD creates the suitability map by computing for each grid cell a weighted sum of all the reclassified attribute maps. Hence, the suitability in each cell is calculated according to the following:

$$R(i) = \left[\sum_{a=1}^{A} \left\{ W_a \times P_a(i) \right\} \right] \Big/ \left[\sum_{a=1}^{A} W_a \right] \qquad \text{Equation 1}$$

Where:
R(i) = suitability value in cell(i),
a = a particular driver map,
A = the number of driver maps,
W_a = the weight of driver map a, and
$P_a(i)$ = percent-developed in category a_k of attribute map a, where cell(i) is a member of category a_k.

We sometimes use the term "lubrication values" to refer to the percent-developed real numbers that are denoted $P_a(i)$. The larger the lubrication, the easier it is for land category 2 to overtake the grid cell. All of the lubrication values for each category of each driver map are given in the text file that has the extension LUB. There are a set of such numbers for each driver for each stratum. If a specific category is missing from the driver map, then its lubrication value is flagged by the value -1. The weights have no influence on the lubrication values. The weights influence only the process to aggregate the lubrication values from various driver maps.

Figure 9 shows a popular driver map, slope. This slope map shows categories of bins of slope as described above. GEOMOD overlays the slope map with the beginning time image of 1985, to produce the suitability map shown in Figure 10. GEOMOD reclassifies each category of slope to exactly one real number, which is the percent of the category that is developed according to the beginning time image.

GEOMOD performs this same type of reclassification with a map of land-use of 1971, shown in Figure 11. Figure 11 shows 21 categories of land-cover. GEOMOD reclassifies each category to a real number, which is the percent of the category that is developed in 1985. Figure 12 is the resulting suitability image.

Figure 13 is the average of Figures 10 and 12. Thus, Figure 13 is a suitability map that has been calibrated by comparing the beginning time image of 1985 with two driver maps, slope and historic land cover of 1971.

Figure 14 shows the 1999 landscape that GEOMOD predicts when GEOMOD uses only the persistence rule and the suitability map of Figure 13. Thus, Figure 14 is equivalent to a reclassification of Figure 13, whereby the largest suitability values show the locations of predicted new development. Just as in the previous predictions, GEOMOD predicts the amount of new development based on the prediction of quantity of total non-developed area.

Figure 15 is a prediction of 1999 that is based on the rules of persistence, regional stratification and a suitability map. GEOMOD creates the suitability map by town, based on the slope and historic land cover of 1971. Therefore, each town has its own lubrication values. Also, the quantity of new development is specified by town according to Figure 6. We consider Figure 15 to be our best simulation run, so we will use it in the remainder of this paper to discuss techniques of validation.

A model is said to be “dynamic” when the conditions at one time influence the decisions at other times. For example, GEOMOD is not dynamic in the sense that the suitability map does not change over time. On the other hand, when the neighborhood constraint rule is used with a small time step, then GEOMOD is dynamic in the sense that GEOMOD re-computes for each year the cells as candidates for change by re-computing which grid cells are on the edge between developed and non-developed.

GEOMOD Environmental Impact Analysis

GEOMOD can analyze the environmental impact of any particular simulation run. The environmental impact occurs in the cells where GEOMOD simulates land change. GEOMOD computes the impact as the product of three images as specified in Equation 2.

$$\mathrm{Impact(i)} = \mathrm{Resource(i)} \times \mathrm{Ratio1(i)} \times \mathrm{Ratio2(i)} \qquad \text{Equation 2}$$

Where:
Impact(i) is the actual environmental impact in cell i,
Resource(i) is the environmental resource in cell i,
Ratio1(i) is the ratio of potential impact / environmental resource in cell i,
Ratio2(i) is the ratio of actual environmental impact / potential impact in cell i.

Ratio1 and Ratio 2 can be either an image or a constant. The value of each cell or the constant must be in the inclusive interval [0,1].

A common example is the application of predicting carbon dioxide emission attributable to land-use change. In this case, the resource is the biomass of vegetation in each cell at the beginning time. Ratio1 is the ratio of mass of carbon / mass of biomass. Usually this ratio is approximately 0.5, since carbon accounts for about half of the mass in vegetation. Ratio2 is the proportion of the carbon that is released to the atmosphere when the grid cell is converted from non-developed to developed. Ratio2 can vary depending on the method of land change. If the method of development is to burn the biomass, then the ratio is much higher than if the method is to cut the biomass.

If the user selects to perform the environmental impact option, then GEOMOD can produce additional images that show the environmental impact. GEOMOD can produce an image of the cumulative environmental impact over the duration of the simulation, where each cell of simulated change shows the real number value of the environmental impact. Cells that do not show simulated change have an environmental impact of zero. If GEOMOD uses several time steps, then GEOMOD can produce images of environmental impact at each time step.

As a hypothetical example, Figure 16 shows the biomass existing at the beginning time image of 1985 for the Central Massachusetts example. We assume that half of this biomass is carbon and that 80% of the carbon is removed from the landscape when a grid cell becomes disturbed. Therefore, Figure 17 shows the environmental impact in terms of decrease in carbon stock as a result of the simulated new development from 1985 to 1999, shown in Figure 15.

Guidelines for the use of GEOMOD

The GEOMOD philosophy calls for a clear distinction between the information used for calibration and the information used for validation. This means that the only legitimate inputs for GEOMOD are inputs that would have been available to a scientist at the beginning time. For example, suppose GEOMOD simulates progressive development between the years 1985 and 1999, and the user would like to validate the output from GEOMOD for the year 1999. The only legitimate inputs would be the information that would have been available in the year 1985, such as the road network of 1985. It would not be legitimate to use the road network of any date after 1985. If a driver map were the distance to roads of 1999, then GEOMOD's output would likely be accurate when compared with the map of 1999, however the output would not be a legitimate prediction, because the roads of 1999 would not have been unknown in 1985. GEOMOD will generate a suitability map for whatever combination of driver maps that the user gives it, but that does not mean that any driver is scientifically legitimate, especially when the user wants to validate the model rigorously.

The majority of literature that we have encountered fails to distinguish clearly between which information is used for calibration and which is used for validation. Some authors do distinguish clearly which information was used for calibration and which was used for validation. Excellent examples include Brown et al. (2002), Geoghegan et al. (2001), Mertens and Lambin (2000), and Schneider and Pontius (2001). We can assess properly the predictive capability of models in only those cases where there is a clear distinction between calibration information and validation information.

In some cases, the user will want to use some information of the ending time to calibrate the model. For example, if the user is interested in GEOMOD's ability to specify only the location of grid cells, then the user can force the simulated map to have the correct quantity of non-developed and developed cells in each stratum. This will allow for a legitimate validation, as long as the measure of the validation takes into consideration that GEOMOD used the information of the quantity of each category in each stratum from the ending time. The section on validation explains how to account for this type of practice when measuring the validation. The motivation for using the correct quantity is to measure GEOMOD's ability to specify the location of cells for each category, unconstrained by any errors in the specification of the quantity of cells for each category. This validation procedure is recommended when the user lacks an independent prediction of the quantity of cells of each category.

Validation

Importance of Validation with Three Maps

After GEOMOD or any other model creates a simulated map of the ending time, it is desirable to validate the accuracy of the prediction. It is extremely important to validate the model output in an intelligent manner because a naïve interpretation of the accuracy can give extremely misleading results. For example, if only 10% of the landscape actually changes between the beginning time and the ending time, then the agreement between the beginning time image and the ending time image is 90%. So a naïve model that predicts no change between the beginning time and the ending time would have an accuracy of 90%. In other words, it is possible to obtain very high agreement in terms of percent correct due to simply a large signal of persistence on the landscape. It is important to know how the simulation model performs with respect to an alternative naïve model that predicts persistence only. Therefore, it is essential to assess the validity of the simulation output with respect to the reference maps of both the ending time and the beginning time. For example, the Atlanta Metropolitan Area is renown for its land use change due to urban sprawl, however 75% of the landscape persisted between 1973 and 1999, which are the years of rapid change (Yang 2002; Yang and Lo 2002; Lo and Yang 2002). Therefore a naïve model that predicts persistence only over 26 years would have an accuracy of 75%, which is a success rate that some modelers consider good. For most landscapes, the percent of persistence is much higher.

After soliciting many sophisticated land change modelers, we have been able to find only a few examples where the agreement between the simulated map and the ending time map is greater than the agreement between the beginning time map and the ending time map at the resolution of the raw data. If you know of a case, please contact the author of this chapter. There are many published papers where the agreement between the beginning time map and the ending time map appears to be better than the agreement between the simulated map and the ending time map, although this characteristic is usually not mentioned in these papers (Brown et al. 2002; Chen et al. 2002; Lo and Yang 2002; Manson 2002; Schneider and Pontius 2001; Geoghegan et al. 2001; Wear and Bolstad 1998). This illustrates the importance of thinking carefully about the techniques of validation and underscores the importance of using good validation tools, such as those that are available in IDRISI. Below we offer alternative approaches to validation, especially techniques that consider information from the beginning time map.

The authors endorse two types of approaches to validate the simulation. First is a visual approach and second is a statistical approach. The first approach is important because a visual examination is the quickest way to reveal spatial patterns, which a particular statistical method may fail to detect. The second approach is essential because the visual approach is subjective and can be misleading for a variety of reasons. For example, different palettes can cause the same image to appear very differently. In both the visual approach and the statistical approach, a naïve comparison usually reveals a large agreement, while a more sophisticated comparison usually shows that only a small portion of the agreement is attributable to the predictive ability of the model. Therefore, we recommend the following procedures for each of the approaches.

Under both the visual and statistical approach, the modeler should consider the entire study area, which is every grid cell that is not in the mask. Most importantly, if all the analyses are performed for the entire study area, then the presentation of the results is clear and the comparison of several model runs is uniform. This seems like an obvious suggested procedure, however many modelers are tempted to restrict the area of validation to only a portion of the landscape where change occurs, or to only an area that is initially non-developed. Their motivation is to focus on the ability of the model to predict a certain type of change. The danger of such a restricted focus is that it can hide various weaknesses or strengths of the model. For example, when GEOMOD simulates conversion from non-developed to developed, it can not simultaneously simulate conversion from developed to non-developed within the same stratum. Some scientists might see this as a weakness of GEOMOD. If the validation area is the entire study area, then we can see this weakness clearly. On the other hand, if the validation of GEOMOD is restricted to only those regions that were non-developed at the beginning time, then we will not see that GEOMOD generates an error for every cell that truly converts from developed to non-developed. This situation is especially important in the case where land cover category 1 means forest and category 2 means non-forest. GEOMOD is unable to predict accurately a landscape that experiences both substantial deforestation and reforestation simultaneously. But, if the cells that are initially non-forest were masked, then the mask would hide the fact that GEOMOD predicts all reforestation incorrectly. Alternatively, if there were no reforestation, the mask would hide the fact that GEOMOD predicts the persistence of non-forest correctly.

Visual Approach

For the visual approach, the modeler should first make a map of the actual change that the model attempts to predict, such as Figure 5. The CROSSTAB module should be used where the first image is the beginning time map and the second image is the ending time map. The resulting crosstabulation map should have at most five categories. The categories could be: persistence of category 1, conversion from category 1 to 2, conversion from category 2 to 1, persistence of category 2 and the mask, if any. Predictive simulation models usually attempt to identify the locations of the cells that change.

After the map of real change is created, the CROSSTAB module can be used again to compare the map of real change to the simulated map of the ending time. In this case, the first map should be the map of real change (Figure 5) and the second map should be the simulated map of the ending time (Figure 15). Depending on the landscape and the simulation model used, the result should show at most nine categories as follows, where “Beginning” means the category in the beginning time map, “Ending” means the category in the ending time map, “Simulated” means the category simulated by the model for the ending time, 1 means non-developed and 2 means developed:

A. Beginning = 1, Ending = 1, Simulated = 1
B. Beginning = 1, Ending = 1, Simulated = 2
C. Beginning = 1, Ending = 2, Simulated = 1
D. Beginning = 1, Ending = 2, Simulated = 2
E. Beginning = 2, Ending = 1, Simulated = 1
F. Beginning = 2, Ending = 1, Simulated = 2
G. Beginning = 2, Ending = 2, Simulated = 1
H. Beginning = 2, Ending = 2, Simulated = 2
I. Beginning = 0, Ending = 0, Simulated = 0

Cases A, D, E and H are correct and cases B, C, F and G are errors. However, simple persistence between the beginning time and the ending time account for cases A and H. Furthermore, the cases B and F show situations for which the model simulated an error at a location where a naïve model that predicts persistence only would have predicted correctly. In the strata where GEOMOD predicts conversion from 1 to 2, only case D shows agreement that is attributable to the model. In the strata where GEOMOD predicts conversion from 2 to 1, only case E shows agreement that is attributable to the model. Recall that GEOMOD does not predict conversion from 1 to 2 at some cells while simultaneously predicting conversion from 2 to 1 at other cells within the same stratum. Therefore, if GEOMOD simulates conversion from 1 to 2 within a stratum, then GEOMOD can not generate cases E or G in that stratum and GEOMOD is guaranteed to fail at predicting any true conversion from 2 to 1 in that stratum. Conversely, if GEOMOD simulates conversion from 2 to 1 within a stratum, then GEOMOD can not generate cases B or D in that stratum and GEOMOD is guaranteed to fail at predicting any true conversion from 1 to 2 in that stratum. Other simulation models, such as IDRISI’s CA_MARKOV model, could generate all cases A-H. We recommend a visual examination of all cases as a first step in validation. We recommend also that the modeler examine the number of grid cells in each of the cases, which can be computed easily with the module HISTO. Figure 17 shows the validation of the simulation shown in Figure 18.

Statistical Approach

The second method of validation is statistical. There are an infinite number of different ways to compare maps statistically (Turner et al. 1989; Congalton and Green 1999). The challenge is to find a technique that: 1) is easy to interpret, 2) offers information that is helpful to improve the model, and 3) answers the fundamental question, “How similar are two maps?” Some researchers consider complex measures of pattern such as fractal dimension, patch density, etc. (Messina and Walsh 2001; Crews-Meyer 2002). Before considering these relatively complex measures, we think that the modeler should consider two simpler measures. We prefer to address two elementary questions: 1) “How well do the pair of maps agree in terms of the quantity of cells in each category?”, and 2) “How well do the pair of maps agree in terms of the location of cells in each category?” In fact, it is necessary to know the answers to these question in order to interpret the more complex measures of pattern (Gustafson and Parker 1992).

There are two modules, VALIDATE and ROC, in IDRISI that are useful to address these questions. The VALIDATE module examines the agreement between two maps that show the same categorical variable. The ROC module examines the agreement between a Boolean map of one category and a suitability map for that category.

Both VALIDATE and ROC compare exactly one pair of maps. But in any analysis of validation, it is necessary to consider also the beginning time map. Therefore, a comprehensive validation should examine the results from two pairs of maps: 1) the true ending time map versus the simulated ending time map, and 2) the true ending time map versus the beginning time map. In both of these comparisons, we will refer to the true ending time map as the “reference map” and the other map as the “comparison map”. In both comparisons, the statistical methods answer the question “How well does the comparison map explain the reference map?”, where the statistical agreement between the maps indicates the level of explanation.

Under either of these techniques of statistical measurement it is essential to understand clearly the difference between error due to location versus error due to quantity. Therefore, the reader should examine closely Figure 19, which illustrates the fundamental concepts. Figure 19 shows a pair of maps containing two categories: light and dark. At the simplest level of analysis, we compute the proportion of cells that agree between the two maps. The agreement is 12/16 and the disagreement is 4/16. At a more sophisticated level, we can compute the disagreement in terms of two components: A) disagreement due to quantity, and B) disagreement due to location. A disagreement of quantity is defined as a disagreement between the maps in terms of the quantity of a category. For example, the proportion of cells in the dark category in the comparison map is 10/16 and in the reference map is 12/16, therefore there is a disagreement due to quantity of 2/16. A disagreement of location is defined as a disagreement such that a swap of the location of a pair of cells within the comparison map increases overall agreement with the reference map. The disagreement of location is determined by the amount of spatial rearrangement possible in the comparison map, so that its agreement with the reference map is maximized. In the example, it would be possible to swap the #9 cell with the #3, #10, or #13 cell within the comparison map to increase its agreement with the reference map. Any of these is the only swap we can make to improve the agreement, given the quantity of each category in the comparison map. Therefore the disagreement of location is 2/16. The distinction between error of quantity and error of location is the foundation of our philosophy of map comparison. The user must keep these concepts in mind when using the VALIDATE and ROC modules.

The VALIDATE Module Technique

The VALIDATE module offers one comprehensive statistical analysis that answers simultaneously the two important questions: 1) “How well do a pair of maps agree in terms of the quantity of cells in each category?”, and 2) “How well do a pair of maps agree in terms of the location of cells in each category?” The VALIDATE module can perform this comparison when the pair of maps show any categorical variable, which can have any number of categories. For applications to validation of GEOMOD, there are two categories, denoted as 1 and 2.

The methods of VALIDATE have been created recently at Clark University, so the reader will not find explanations in common statistical books. To learn more about these techniques, see Pontius and Suedmeyer (2004), Pontius (2002), and Pontius (2000). For applications of these techniques, see Pontius et al. (2004a), Pontius et al. (2004b), Pontius et al. (2001), Schneider and Pontius (2001).

The modeler should run the VALIDATE module twice. The first run should have the true ending time map as the reference map and the beginning time map as the comparison map. The second run should have the true ending time map as the reference map and the simulated ending time map as the comparison map. With the output from these two runs of VALIDATE, we can examine how the agreement between the beginning time map and the true ending time map compares to the agreement between the simulated ending time map and the true ending time map.

Using our Central Massachusetts example, Figures 20 and 21 are the output from VALIDATE when reference image is the true 1999 landscape and the comparison image is the true 1985 landscape. Figure 22 and 23 are the output from VALIDATE when the reference image of the true 1999 landscape and the comparison image is the 1999 landscape that is simulated by GEOMOD. In Figures 20-23, VALIDATE uses the image of the 10 towns as a strata image.

In all four figures, the vertical axis of the graphs show the proportion of the landscape, not including the mask. The vertical axis is separated into components of agreement and disagreement, based on the similarities and differences between the two maps. Each component is associated with either the first question concerning the quantity of cells in each category or the second question concerning the location of cells in each category. The graphic is a quick way to see how the pair of maps agree and disagree in the two ways asked by the two fundamental questions. We discuss these components in more detail in the next subsection below.

In Figures 21 and 23, VALIDATE plots the vertical axis perpendicular to a horizontal axis that shows resolution. The fine resolution of the raw data is on the left of the horizontal axis and the resolution becomes coarser as one moves toward the right on the horizontal axis. Figures 21 and 23 show a geometric sequence of resolutions: 1, 2, 4, 8, 16, … , 1024. The final coarsest resolution is 1116 because the entire study area is in one grid cell at that resolution. A resolution of n means that the raw grid cells of the study area are aggregated into coarse cells, each consisting of a square of n-by-n raw grid cells. The aggregation process is identical to the procedure used by IDRISI's CONTRACT module with the aggregation option, which is a simplified version of the moving window technique of Costanza (1989). At each resolution, the statistical analysis is performed on the coarse grid cells. We use 1 to denote the resolution of the raw data, for which there are 1008 columns and 1116 rows in the Central Massachusetts example. Therefore, there are 504 columns and 558 rows at a resolution of 2, and there are 252 columns and 279 rows at a resolution of 4, and so on, until there is one row and one column at a resolution of 1116, at which point the entire image is in one grid cell.

All of the components of agreement and disagreement are shown at multiple resolutions. It is important to examine the components at multiple resolutions in order to see how the pair of maps agree and disagree in terms of general location of the categories. For example, at the finest resolution of 1, the analysis is performed on a cell-by-cell basis, using the cells of the raw data. It is common that many cells disagree on a cell-by-cell basis, especially if the comparison map is slightly misregistered. If the analysis is performed at a slightly coarser resolution, then the cell-by-cell disagreement converts to agreement as discussed below.

Components of Agreement and Disagreement

Each of the curves in Figures 21 and 23 correspond to one of seven statistical calculations denoted as N(**n**), N(**m**), H(**m**), M(**m**), K(**m**), P(**m**), and P(**p**). The other eight calculations in Figures 21, 22, and 24 influence neither the components of agreement nor the components of disagreement. VALIDATE computes these statistics for each resolution. This section interprets each of the seven statistical expressions in terms of the two fundamental questions: "How well do the maps agree in quantity?", and "How well do the two maps agree in location?" In the notation of each expression, the argument in bold represents three possible levels of information of quantity:

1: **n** = no information

2: **m** = medium information
3: **p** = perfect information

Similarly, the functions denoted by the capital letters indicate three possible levels of information of location:

1: N = no information
2: H = medium stratum-level information but no grid cell level information
3: M = medium stratum-level information and medium grid cell-level information
4: K = medium stratum-level information and perfect grid cell-level information
5: P = perfect stratum-level information and perfect grid cell-level information

These levels of information are the possible levels of information that could be available to a scientist who creates the comparison map. Mathematical definitions for each expression are shown in Figure 24, which is taken from Pontius and Suedmeyer (2004), who give a detailed description of each expression. The remainder of this section interprets each expression in terms of intuitive concepts. All of these expressions are useful to compare maps of any number of categories, where the number of categories is denoted as J. For the GEOMOD example, J=2.

Expression N(**n**) is the agreement between the reference and a map that has a membership of 1/J to each category in every grid cell. If each cell of the reference map shows complete membership in exactly one category, then N(**n**)=1/J, which is usually the case at the finest resolution. If each cell of the reference map shows partial membership in many categories, then typically N(**n**)>1/J, which is usually the case at coarser resolutions. N(**n**) is the expected agreement between the reference map and a map in which every raw fine-resolution grid cell is assigned to one of the J categories with a probability of 1/J.

Four of the five points that are functions of **m** give the agreement between the reference map and a modified comparison map. N(**m**) is the agreement between the reference map and a modified comparison map, where the modification is to randomize the locations of the raw cells within the comparison map. Expression H(**m**) is the agreement between the reference map and a modified comparison map, where the modification is to randomize the locations of the cells within each stratum of the comparison map. When the modification randomizes the location of the grid cells, each cell remains within its stratum (i.e., no grid cells move across stratum boundaries).

Expression M(**m**) is the agreement between the reference map and the unmodified comparison map. M(**m**) is the proportion of grid cells classified correctly, which is the most commonly used measure of agreement between maps.

Expression K(**m**) is the agreement between the reference map and a modified comparison map, where the modification is to rearrange as perfectly as possible the locations of cells within each stratum of the comparison map in order to maximize the agreement between the modified comparison map and the reference map. The method of modification swaps the location of the grid cells, with the constraint that each cell remains within its stratum (i.e., no grid cells move across stratum boundaries).

Expression P(**m**) is the agreement between the reference map and a modified comparison map, where the modification is to rearrange as perfectly as possible the locations of cells within the entire comparison map in order to maximize the agreement between the modified comparison map and the reference map. The method of modification swaps the location of the grid cells anywhere within the comparison map (i.e., rearrangement of location of grid cells is allowed to occur across stratum boundaries). Therefore P(**m**) = 1 if and only if the proportion for each category in the comparison map is the same as the corresponding proportion in the reference map.

Expression P(**p**) is the agreement between the reference map and a map that has perfect information of both quantity and location. Therefore, P(**p**) is always 1.

When considered as a set, the seven mathematical expressions N(**n**), N(**m**), H(**m**), M(**m**), K(**m**), P(**m**), and P(**p**) constitute a sequence of measures of agreement between the reference map and other maps that have increasingly accurate information. Therefore, usually 0 < N(**n**) < N(**m**) < H(**m**) < M(**m**) < K(**m**) < P(**m**) < P(**p**) = 1. This sequence partitions the interval [0,1] into components of the agreement between the reference map and the comparison map. M(**m**) is the total proportion correct, and 1-M(**m**) is the total proportion error between the reference map and the comparison map. Hence, the sequence of N(**n**), N(**m**), H(**m**) and M(**m**) defines components of agreement, and the sequence of M(**m**), K(**m**), P(**m**) and P(**p**) defines components of disagreement.

Table 1 defines these components mathematically. Beginning at the bottom of the table and working up, the first component is agreement due to chance, which is usually N(**n**). However, if the agreement between the reference map and the comparison map is worse than would be expected by chance, then the component of agreement due to chance may be less than N(**n**). Therefore, Table 1 defines the component of agreement due to chance as the minimum of N(**n**), N(**m**), H(**m**), and M(**m**). The component of agreement due to quantity is usually N(**m**)-N(**n**); Table 1 gives a more general definition to account for the possibility that the comparison map's information of quantity can be worse than no information of quantity. The component of agreement at the stratum level is usually H(**m**) – N(**m**); Table 1 gives a more general definition to restrict this component of agreement to be non-negative. Similarly, the component of agreement at the grid cell level is usually M(**m**) – H(**m**); Table 1 restricts this component of agreement to be non-negative. Table 1 also defines the components of disagreement. It is a mathematical fact that M(**m**) ≤ K(**m**) ≤ P(**m**) ≤ P(**p**), therefore the components of disagreement are the simple definitions of Table 1.

Interpretation of Components

It is usually most helpful to interpret the components of disagreement, because usually the purpose of the assessment is to find ways to improve the method of creating the comparison map so that it agrees more closely with the reference map. Therefore, we want to budget the error so we can see the largest sources of error in the comparison map. The components of disagreement separate the error into specific interpretable sources.

The disagreement at the grid cell level, K(**m**)-M(**m**), is the amount of error associated with the fact that the comparison map fails to specify perfectly the correct locations of categories at grid cells within strata. In other words, if we were to rearrange the grid cells within each stratum of the comparison map, then we could improve the agreement between the reference map and the comparison map from M(**m**) to K(**m**). K(**m**) = 1 if and only if the quantities of each category in each stratum in the comparison map are the same as the quantities in the corresponding strata in the reference map.

The disagreement at the stratum level, P(**m**)-K(**m**), is the amount of error associated with the fact that the comparison map fails to specify perfectly the correct quantity of each category within each stratum. In other words, if we were to rearrange the grid cells of the comparison map, and to allow rearrangement across stratum boundaries, then we could improve the agreement between the reference map and the comparison map from M(**m**) to P(**m**).

The disagreement due to quantity, P(**p**)-P(**m**), is the amount of error associated with the fact that the comparison map fails to specify perfectly the correct quantity of each category according to the reference map. Thus 1-P(**m**) is the error of pure quantity, and indicates nothing about the spatial arrangement of the cells.

If the largest component of disagreement is the grid cell level, then the scientist would probably want to work to improve the specification of the location of each category at the grid cell level in the comparison map. If the largest component of disagreement is the stratum level, then the scientist would probably want to improve the specification of the quantity of each category at the stratum level in the comparison map. If the largest component of disagreement is quantity, then the scientist would probably want to improve the specification of the quantity of each category in the comparison map.

After the scientist interprets the components of disagreement, then it can be also helpful to interpret the components of agreement. The components of agreement specify distinct characteristics in which the comparison map agrees with the reference map.

The agreement due to chance, N(**n**), is the agreement that a scientist could achieve with no information of location and no information of quantity. Therefore, N(**n**) can be a good baseline upon which to compare the actual agreement. The agreement due to quantity, N(**m**) – N(**n**), is the additional agreement when the comparison map is somewhat accurate in terms of its specification of quantity of each category. The agreement due to stratification, H(**m**) – N(**m**), is the additional agreement when the comparison map is somewhat accurate in terms of its specification of quantity of each category within each stratum. The agreement due to grid cell level location, M(**m**) – H(**m**), is the additional agreement when the comparison map is somewhat accurate in terms of its specification of the grid cell level location of each category within each stratum.

Interpretation of Multiple Resolutions

The points N(**n**), N(**m**), H(**m**), M(**m**) increase as the resolution becomes more coarse. It can be helpful to examine the resolution at which each point begins to rise suddenly.

The rate at which N(**n**) increases is a function of the reference map only. If the various categories in the reference map are fairly evenly spread, then N(**n**) will rise rapidly at relatively fine resolutions. If the various categories of the reference map are arranged in large clusters, N(**n**) will remain flat until the resolution surpasses the size of the clusters, at which point N(**n**) will rise rapidly. N(**m**) and H(**m**) behave in a similar manner.

The behavior of M(**m**) at multiple resolutions offers substantial insight into the similarity between the comparison map and the reference map. As resolution becomes more coarse, M(**m**) will remain flat until it reaches a resolution at which the maps agree. For example, suppose many cells disagree in terms of cell-by-cell location at the resolution of the raw data, denoted as a resolution of 1. If the correct category in a cell of the reference map is no where near the corresponding cell in the comparison map, then M(**m**) will remain flat at fine resolutions. If the correct category in a cell of the reference map is near the corresponding cell in the comparison map, then M(**m**) will rise at fine resolutions. In other words, M(**m**) rises quickly at the resolution that corresponds to the distance at which the correct categories are found, which is the distance at which errors of location occur.

Both the agreement due to grid cell location and the disagreement due to grid cell location tend to decrease as resolution becomes coarser. This is because the grid cell level location becomes less important as the cells become coarser. Ultimately, grid cell-level location loses meaning at the final resolution where the entire study area is in one grid cell. At that resolution, the component of agreement due to grid cell location and the component of disagreement due to grid cell location is zero.

An interesting question is “At what resolutions is the agreement between the simulated map and the ending time map larger than the agreement between the beginning time map and the ending time map?” For most cases, the agreement between the beginning time map and the ending time map is greater at fine resolutions, while the agreement between the simulated time map and the ending time map is greater at coarser resolutions. The resolution at which the agreement is the same represents the distance at which the simulation model begins to perform better than a null model that predicts persistence only. Figure 25 shows that for our example this resolution is approximately 32, which corresponds to grid cells of approximately one square kilometer. VALIDATE can save text files, so crucial statistics can be imported easily into graphics software. Pontius et al. (2004) have produced a VBA program to make such graphs, and the reader can write to rpontius@clarku.edu to request the program.

The points K(**m**), P(**m**) and P(**p**) are independent of the resolution of the grid cells because they are functions of only the quantity of each category in the maps. Therefore, those three points remain constant as resolution changes.

The ROC Module Technique

The ROC module offers a statistical analysis that answers one important question: “How well is the category of interest concentrated at the locations of relatively high suitability for that category?” The answer to this question allows the scientist to answer the general question, “How well do the pair of maps agree in terms of the location of cells in a category?”, while not being forced to answer the question “How well do the pair of maps agree in terms of the quantity of cells in each category?”. Thus the ROC analysis is useful for cases in which the scientist wants to see how well the suitability map portrays the location of a particular category but does not have an estimate of the quantity of the category for the simulation. For the application to the validation of GEOMOD, typically the reference map would show the “developed” category, and the comparison map would be the suitability map that GEOMOD generates from the driver maps. In some cases, the scientist does not want to commit to any particular estimate of the quantity of simulated development at the ending time, but the scientist still would like to measure how well the suitability map agrees with the location of development (Mertens and Lambin 2000). In this sense, the ROC measures agreement of pure location, because its measure is based on many thresholds to specify the quantity of the category; it is not associated with a particular single threshold.

Pontius and Schneider (2001) explain how to use the ROC technique to examine how well a suitability map portrays the likely locations of the developed category. Schneider and Pontius (2001) apply the ROC method to compare several modeling techniques. The ROC statistical method has been used in many applications (Egan 1975; Metz 1978; Swets 1986a, 1986b, 1988). It is just starting to be used in GIS, but has been used widely in ecology (Miller and Franklin 2002). Pontius and Pacheco (2005) have recently developed a multiple-resolution ROC.

When the question is: “How well does the suitability map predict the locations of new additional development beyond the beginning time image?” then it makes sense to mask the grid cells that are already developed at the beginning time image. The ROC module in IDRISI allows for such a mask. In our Central Massachusetts example, the mask has a 1 in cells that are non-developed in 1985 and a 0 elsewhere, derived from Figure 3. The reference image is the Boolean image of new development, which is a 1 if there is real new development between 1985 and 1999, and 0 elsewhere, derived from figure 5. The comparison image is the suitability map that GEOMOD creates, shown in figure 13. The ROC considers many possible thresholds, each of which results in a prediction of the quantity of new development. For each threshold, the ROC analysis reclassifies the suitability map so that the suitability values that are greater than the threshold are reclassified as predicted new development and all other cells are classified as predicted persistence of non-development. At each possible threshold, the ROC analysis measures the agreement between the reference map and the reclassified suitability map. In this manner, the ROC analysis considers the goodness-of-fit at many possible values of quantity of the predicted category, without committing to any single predicted quantity of the predicted category. The ROC curve aggregates the measure of agreement at all thresholds into the ROC curve. The area under the curve indicates the overall agreement between the Boolean reference map and the suitability map. ROC of 0.5 indicates agreement equivalent to random chance, when the grid cells are hard classified. ROC of 1 indicates perfect agreement. Perfect agreement means that all of the suitability values that are in the locations of new development are larger than the other suitability values that are in the locations of persistence of non-development. Figure 26 shows the ROC curve where each subsequent threshold classifies 1% more of the landscape as new development. The area under the curve in Figure 26 is 0.599, meaning that the suitability map is slightly better than an alternative suitability map that has the suitability values placed at random locations.

The previous paragraph describes the most common type of ROC analysis. It answers the question, "How well does the suitability map specify the locations of true new development among those locations that are initially non-developed?". This approach is inconsistent with our validation principals mentioned earlier in previous sections of this chapter. First, the approach of the previous paragraph masks part of the study area in order to eliminate the cells that are already developed. This type of masking can make the suitability map appear better than it really is when there is substantial conversion from developed to non-developed on the real landscape. Second, the approach of the previous paragraph encourages the observed ROC statistic to be compared to a baseline of randomness. Recall that if the suitability values are in randomly selected locations, then ROC = 0.5. However, according to the GEOMOD validation philosophy, we should compare any statistic to a baseline of a null model that predicts persistence only.

Therefore, we recommend the following two runs of the ROC module. In both runs, the only mask is the mask of the entire study area, which has a 1 if the cell is in the study area, and 0 else. Also, the suitability map of each run shows a suitability of 1 for cells that are developed in 1985, which indicates the highest likelihood for the developed cells of 1985 to remain developed in 1999.

Consider a ROC analysis of a null model that predicts persistence only. Each cell of the suitability map of the null model can have one of exactly two possible unique suitability values. If the cell is developed in 1985, then its suitability value for development in 1999 is 1. If the cell is non-developed in 1985, then its suitability value for development in 1999 is 0. It is easy to create this suitability map using the RECLASS module in IDRISI and GEOMOD's 1985 beginning time image (Figure 3). For such a suitability map, it makes sense to specify one specific threshold, which is the threshold that separates the suitability values of 1 from the suitability values of 0. This threshold will separate the map into two parts, where 34% of the cells are above the threshold (hence classified as developed) and 66% of the cells are below the threshold (hence classified as non-developed). The ROC module allows the user to specify this particular threshold. By default, IDRISI's ROC analysis will automatically consider thresholds of 0% and 100%. Figure 27 shows the resulting ROC curve for the GEOMOD example. The ROC curve shows three threshold levels as the green squares on the green line. The area under the two green segments is 0.925. This green ROC curve should be compared to an ROC curve for the suitability map that GEOMOD creates, as described in the next paragraph.

Image Calculator in IDRISI can be used to make the Boolean suitability map mentioned in the previous paragraph to cover GEOMOD's suitability map using the COVER function. The result is a suitability map that has a value of 1 in any cell that is developed in 1985 and various smaller suitability values in cells that are non-developed in 1985, as shown in Figure 28. Figure 28 distinguishes among the cells that are non-developed in 1985, therefore it is more sophisticated than a suitability map of the null model. Therefore, the thresholds are set to distinguish among the cells in Figure 28 that are non-developed in 1985. Figure 27 shows the ROC curve when there are 9 thresholds, set at 0%, 34%, 40%, 50%, 60%, 70%, 80%, 90% and 100%. Notice that the ROC curve for the suitability map of the null model is 0.925, while the area under the ROC curve for GEOMOD's suitability map is 0.937. This result is consistent with Figure 26, which shows that the suitability map is slightly better than random at specifying the location of new development. Hence, GEOMOD's suitability map is slightly better than a null suitability map.

Acknowledgements

The National Science Foundation (NSF) funded this research through a grant subcontracted through the Center for Integrated Study of the Human Dimensions of Global Change at Carnegie Mellon University (CMU). We thank the CMU program with which this work is increasingly tied intellectually and programmatically through the George Perkins Marsh Institute of Clark University. We thank also the NSF-funded HERO program, which supplied the data.

Literature

Agrawal, C., G. Green, J. Grove, T. Evans, and C. Schweik. 2002. A review and assessment of land-use change models: dynamics of space, time, and human choice. USDA Forest Service, Delaware OH.

Baker, W.L., 1989. A review of models in landscape change. *Landscape Ecology*. 2(2): 111-133.

Bian, L. and S. Walsh. 2002. Characterizing and Modeling Landscape Dynamics: An Introduction. *Photogrammetric Engineering & Remote Sensing*, 68(10): 999-1000.

Brown, D., P. Goovaerts, A. Burnicki, and M-Y. Li. 2002. Stochastic Simulation of Land-Cover Change using Geostatistics and Generalized Additive Models. *Photogrammetric Engineering & Remote Sensing*, 68(10): 1051-1061.

Chen, J., P. Gong, C. He, W. Luo, M. Tamura, and P. Shi. 2002. Assessment of the Urban Development Plan of Beijing by Using a CA-based Urban Growth Model. *Photogrammetric Engineering & Remote Sensing*, 68(10): 1063-1071.

Congalton, R. and Green, K., 1999. Assessing the accuracy of remotely sensed data: principles and practices. Lewis Publishers, New York137.

Costanza, R., 1989. Model goodness of fit: a multiple resolution procedure. *Ecological Modelling*. 47: 199-215.

Crews-Meyer, K. 2002. Characterizing landscape dynamism using paneled-pattern metrics. *Photogrammetric Engineering & Remote Sensing*, 68(10): 1031-1040.

Eastman, J.R., Jin, W., Kyem, P.A.K. and Toledano, J., 1995. Raster procedures for Multi-Criteria/Multi-Objective decisions. *Photogrammetric Engineering and Remote Sensing*. 61(6): 539-547.

Egan, J.P., 1975. Signal Detection Theory and ROC analysis. Academic Press, New York.

Geoghegan, J., S. C. Villar, P. Klepeis, P. M. Mendoza, Y. Ogneva-Himmelberger, R Roy Chowdhury, B. L. Turner, and C. Vance. 2001. Modeling tropical deforestation in the southern Yucatan peninsular region: comparing survey and satellite data. *Agriculture, Ecosystems & Environment* 85(1-3) p. 25-46.

Gustafson, E. J. and G. R. Parker. 1992. Relationships between landcover proportion and indices of landscape spatial pattern. *Landscape Ecology,* 7(2): 101-110.

Hall, C., H Tian, Y Qi, R G Pontius Jr, and J Cornell. 1995a. Modelling spatial and temporal patterns of tropical land-use change. *Journal of Biogeography* 22, p. 753-757.

Hall, C., H Tian, Y Qi, G Pontius, J Cornell, and J Uhlig. 1995b. Spatially-explicit models of land-use change and their application to the tropics. DOE Research Summary 31, February. Carbon Dioxide Information and Analysis Center: Oak Ridge National Laboratory.

Hall, C., H Tian, Y Qi, G Pontius, J Cornell, and J Uhlig. 1995c. Modeling land-use change. CDIAC Communications 21, February. CDIAC: Oak Ridge National Laboratory.

Huffaker, D. and R G Pontius Jr. 2002. Reconstruction of Historical Land Cover in the Ipswich Watershed. *Biological Bulletin* 203.

Kok, K., Farrow, A., Veldkamp, T. and Verburg, P.H., 2001, A method and application of multi-scale validation in spatial land use models. *Agriculture, Ecosystems & Environment,* 85: 223-238.

Lambin, E.F., 1994. Modelling Deforestation Processes: A review. European Commission, Luxemburg.

Lambin, E.F., Baulies, X., Bockstael, N., Fischer, G., Krug, T., Leemans, R., Moran, E.F., Rindfuss, R.R., Sato, Y., Skole, D., Turner II, B.L. and Vogel, C., 1999. Land-use and

land-cover change implementation strategy. Royal Swedish Academy of Sciences, Stockholm, Sweden.
Land-Use and Land-Cover Change (LUCC). 2002. www.geo.ucl.ac.be/LUCC/lucc.html
Li, H. and J. F. Reynolds. 1997. Modeling effects of spatial pattern, drought, and grazing on rates of rangeland degradation: a combined Markov and Cellular Automaton approach. In *Scale in Remote Sensing and GIS*. D. A. Quattrochi and M. F. Goodchild, eds. (New York: Lewis Publishers).
Lo, C.P. and X. Yang. 2002. Drivers of Land-Use/Land-Cover Changes and Dynamic Modeling for the Atlanta, Georgia Metropolitan Area. *Photogrammetric Engineering & Remote Sensing,* 68(10): 1073-1082.
Manson, S. M. 2002. Integrated Assessment and Projection of Land-Use and Land-Cover Change in the Southern Yucatán Peninsular Region of Mexico. Doctoral Dissertation. Graduate School of Geography. Worcester, Massachusetts, Clark University.
MassGIS. 2002. Land Use Map. Executive Office of Environmental Affairs, Boston MA. www.state.ma.us/mgis/.
Menon, S., R G Pontius Jr, J Rose, M Kahn, and K Bawa. 2001. Identifying conservation priority areas in the tropics: a land-use change modeling approach. *Conservation Biology* 15(2), p. 501-512.
Mertens, B. and E. Lambin. 2000. Land-Cover-Change Trajectories in Southern Cameroon. *Annals of the Association of American Geographers*, 90(3): 467-494.
Messina, J. and S. Walsh. 2001. 2.5D Morphogenesis: modeling landuse and landcover dynamics in the Ecuadorian Amazon. *Plant Ecology* 156. pp. 75-88.
Metz, C.E., 1978. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*. 8: 283-298.
Miller, J. and J. Franklin. 2002. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modeling* 157. pp. 227-247.
Pontius Jr, R G and J Malanson. 2005. Comparison of the structure and accuracy of two land change models. International Journal of Geographical Information Science 19(2) p. 243-265.
Pontius Jr, R G and P Pacheco. 2005. Calibration and validation of a model of forest disturbance in the Western Ghats, India 1920 - 1990. GeoJournal 60(4) p.325-334.
Pontius Jr, R G and J Spencer. 2005. Uncertainty in extrapolations of predictive land change models. Environment and Planning B. 32 p. 211-230.
Pontius Jr, R G and M L Cheuk. 2006. A generalized cross-tabulation matrix to compare soft-classified maps at multiple resolutions. International Journal of Geographical Information Science. 20(1) p.1-30.
Pontius Jr, R G and N R Malizia. 2004. Effect of category aggregation on map comparison. Lecture Notes in Computer Science 3234 p.251-268. in M J Egenhofer, C Freksa, and H J Miller (eds): GIScience2004.
Pontius Jr, R G, D Huffaker and K Denman. 2004a. Useful techniques of validation for spatially explicit land-change models. Ecological Modelling 179(4) p.445-461.
Pontius Jr, R G, E Shusas and M McEachern. 2004b. Detecting important categorical land changes while accounting for persistence. Agriculture, Ecosystems & Environment 101(2-3) p.251-268.
Pontius Jr, R G and B Suedmeyer. 2004. “Components of agreement in categorical maps at multiple resolutions”. p.233-251. Chapter 17 in R S Lunetta and J G Lyon (eds). Remote Sensing and GIS Accuracy Assessment. CRC Press, Boca Raton FL.
Pontius Jr, R G, A Agrawal and D Huffaker. 2003. Estimating the uncertainty of land-cover extrapolations while constructing a raster map from tabular data. Journal of Geographical Systems 5(3) p.253-273.
Pontius Jr, R G and K Batchu. 2003. Using the relative operating characteristic to quantify certainty in prediction of location of land cover change in India. Transactions in GIS 7(4) p.467-484.
R G Pontius Jr. 2002. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. Photogrammetric Engineering & Remote Sensing 68(10) p.1041-1049.

Pontius Jr, R G, J Cornell and C Hall. 2001. Modeling the spatial pattern of land-use change with Geomod2: application and validation for Costa Rica. Agriculture, Ecosystems & Environment 85(1-3) p.191-203.

Pontius Jr, R G and L Schneider. 2001. Land-use change model validation by a ROC method for the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems & Environment 85(1-3) p.239-248.

Pontius Jr, R G. 2000. Quantification error versus location error in comparison of categorical maps. Photogrammetric Engineering & Remote Sensing 66(8) p.1011-1016.

Schneider, L and R G Pontius Jr. 2001. Modeling land-use change in the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems & Environment 85(1-3) p.83-94.

Silva, E. and K. Clarke. 2002. Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems* 26. pp. 525-552.

Swets, J.A., 1986a. Form of empirical ROCs in discrimination and diagnostic tasks: implications for theory and measurement of performance. *Psychological Bulletin*. 99(2): 181-198.

Swets, J.A., 1986b. Indices of discrimination for diagnostic accuracy: their ROCs and implied models. *Psychological Bulletin*. 99(1): 100-117.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. *Science*. June: 1285-1293.

Turner II, B.L., Skole, D., Sanderson, G., Fischer, L., Fresco, L. and Leemans, R., 1995. Land-Use and Land-Cover Change: Science/Research Plan. IGBP/HDP, Stockholm and Geneva.

Turner, M.G., Constanza, R., Sklar, F.H., 1989. Methods to evaluate the performance of spatial simulation models. *Ecological Modelling*. 48: 1-18.

Veldkamp, A. and Lambin, E.F., 2001, Predicting land-use change. *Agriculture, Ecosystems & Environment,* 85: 1-6.

Wagner, D.F., 1997. Cellular automata and geographic information systems. *Environment and Planning B*. 24: 219-234.

Wear, D. and P. Bolstad. 1998. Land-Use Changes in Southern Appalachian Landscapes: Spatial Analysis and Forecast Evaluation. *Ecosystems* 1: 575-594.

Wu, F. and Webster, C.J., 1998. Simulation of land development through the integration of cellular automata and multicriteria evaluation. *Environment and Planning B*. 25: 103-126.

Yang, X. 2002. Satellite monitoring of urban spatial growth in the Atlanta metropolitan area. *Photogrammetric Engineering & Remote Sensing,* 68(7): 725-734.

Yang, X. and C.P. Lo. 2002. Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. *International Journal of Remote Sensing*, 23(9): 1775-1798.

Veldkamp, A. and Fresco, L.O., 1996. CLUE-CR: An integrated multi-scale model to simulate land use change scenarios in Costa Rica. *Ecological Modelling*. 91(1/3): 231.

Table 1: Definition and values of seven components of agreement derived from the mathematical expressions of Figure 24.

Name of Component	Definition
Disagreement due to quantity	P(**p**)-P(**m**)
Disagreement at stratum level	P(**m**)-K(**m**)
Disagreement at grid cell level	K(**m**)-M(**m**)
Agreement at grid cell level	MAX [M(**m**)-H(**m**), 0]
Agreement at stratum level	MAX [H(**m**)-N(**m**), 0]
Agreement due to quantity	If MIN [N(**n**),N(**m**), H(**m**), M(**m**)]=N(**n**), then MIN [N(**m**)-N(**n**), H(**m**)-N(**n**), M(**m**)-N(**n**)], else 0
Agreement due to chance	MIN [N(**n**),N(**m**), H(**m**), M(**m**)]

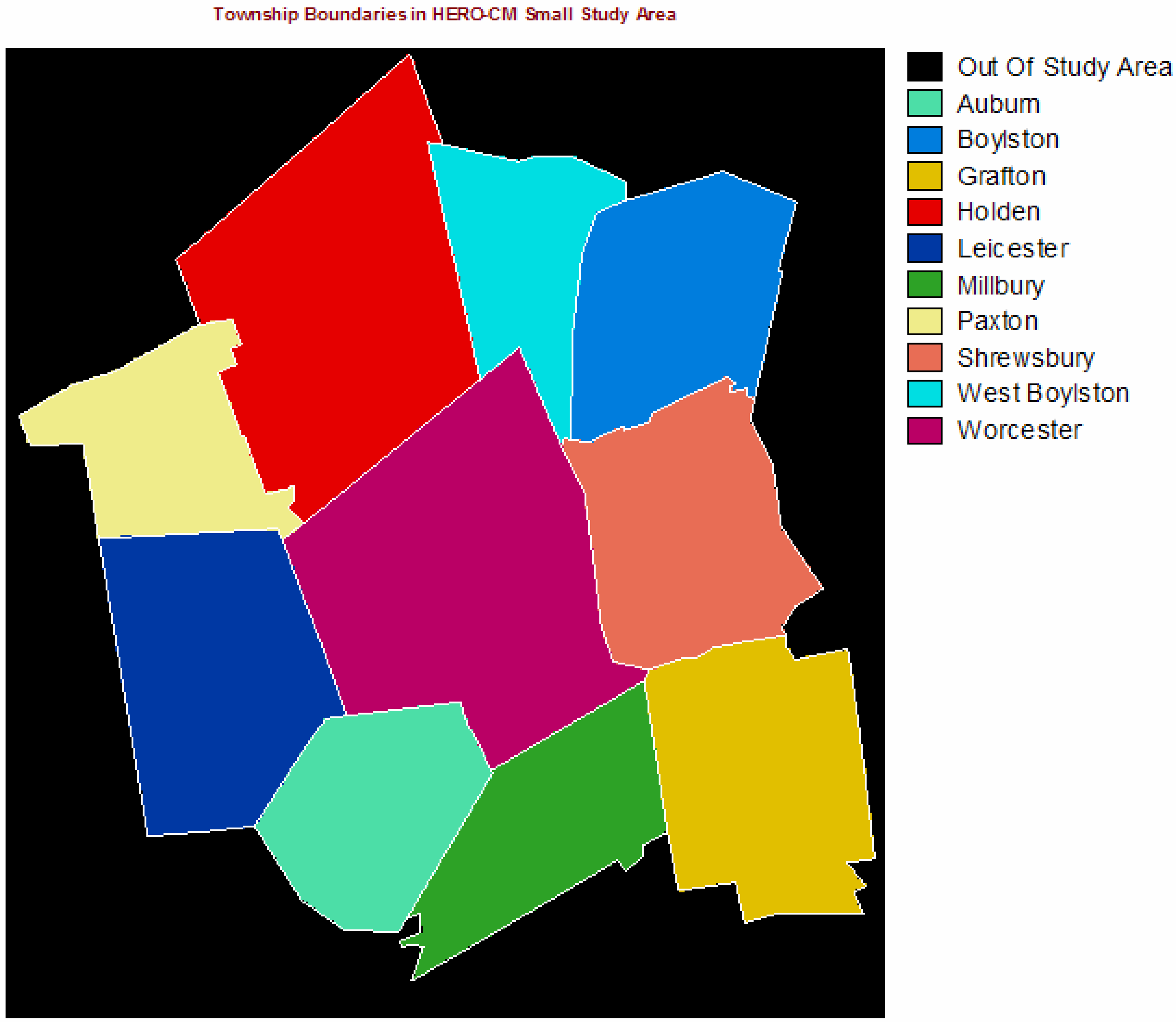

Figure 1: Study area image stratified into 10 towns of Central Massachusetts, USA.

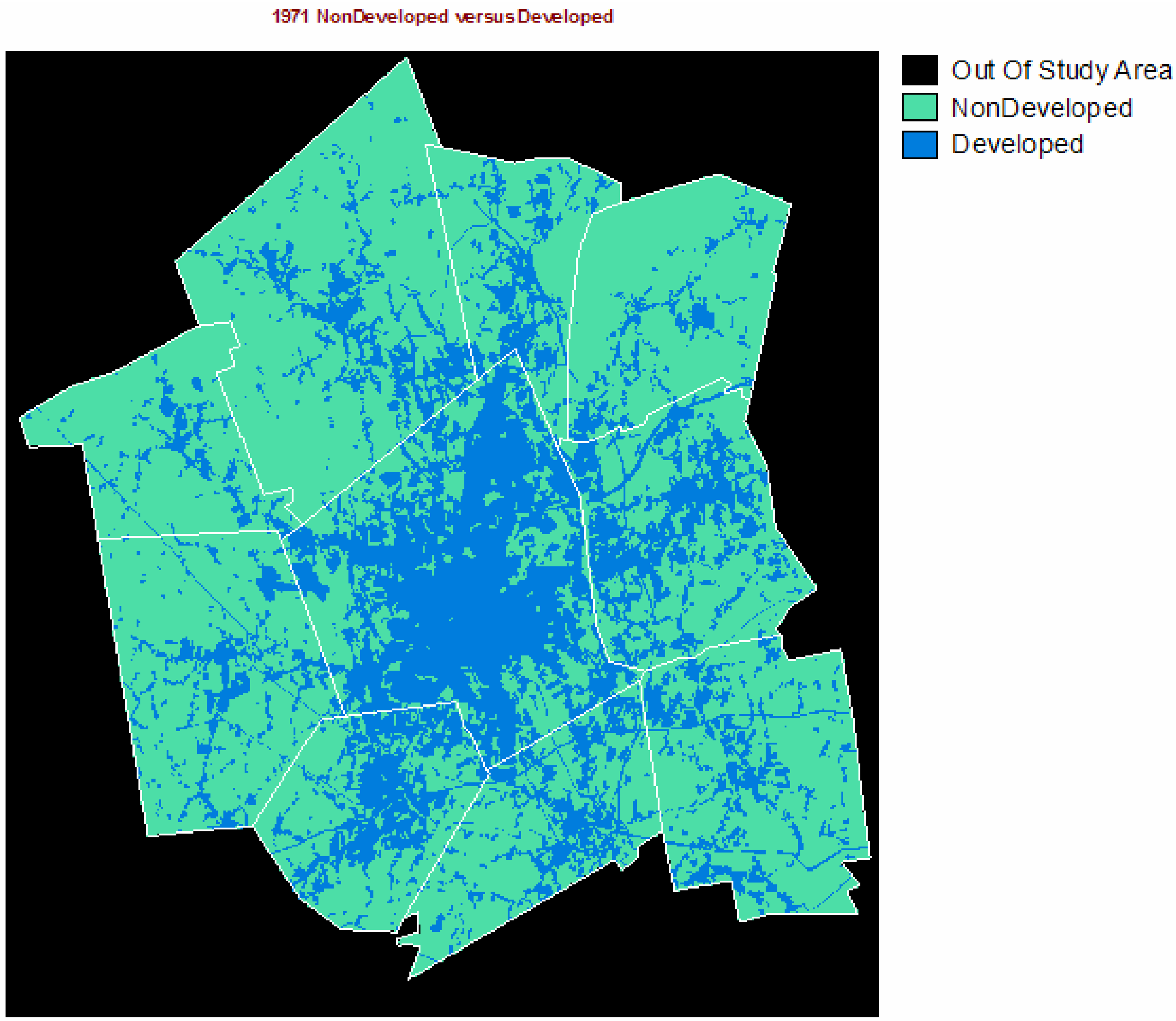

Figure 2: Image of developed versus non-developed of 1971.

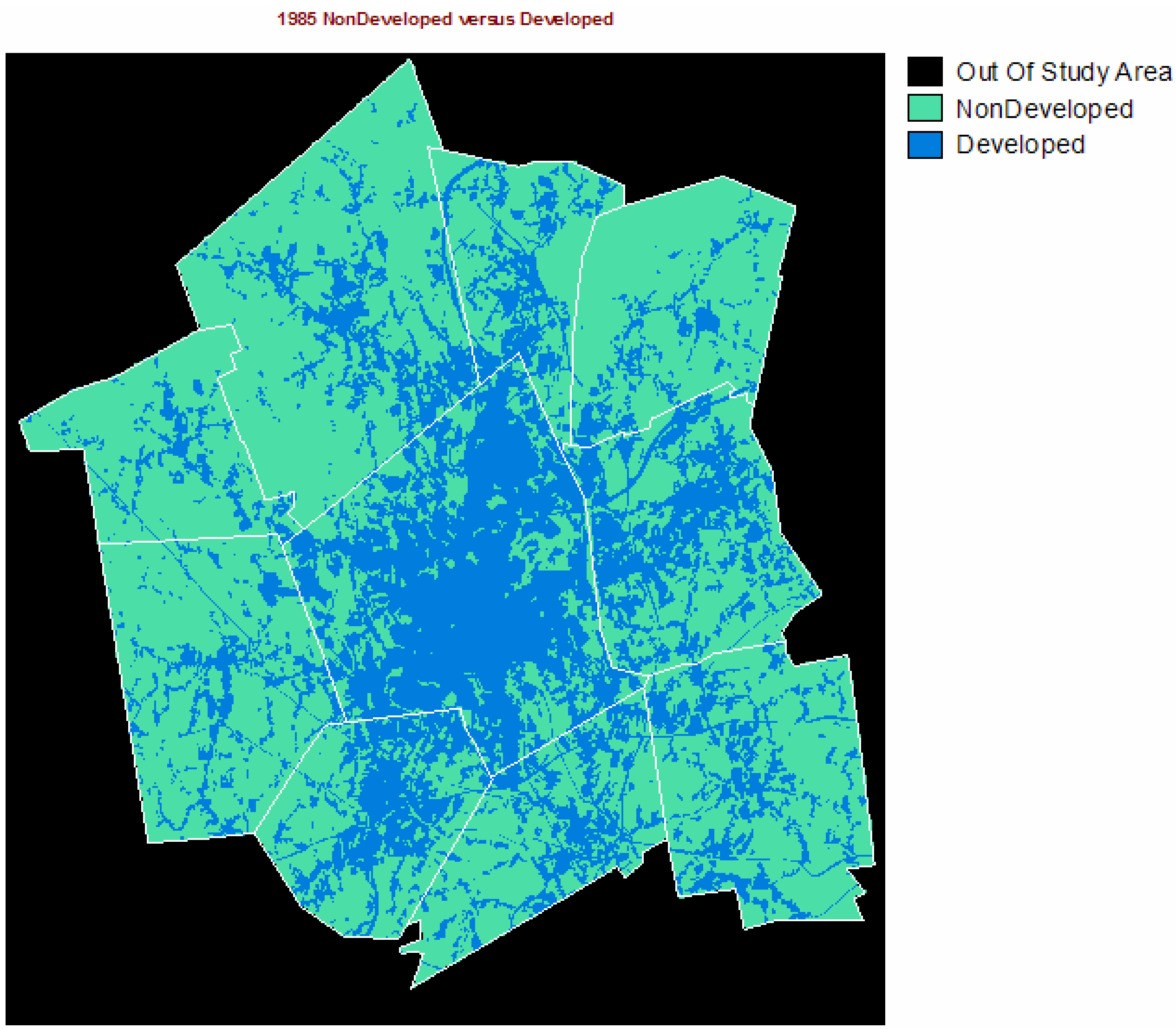

Figure 3: Beginning time image of developed versus non-developed of 1985.

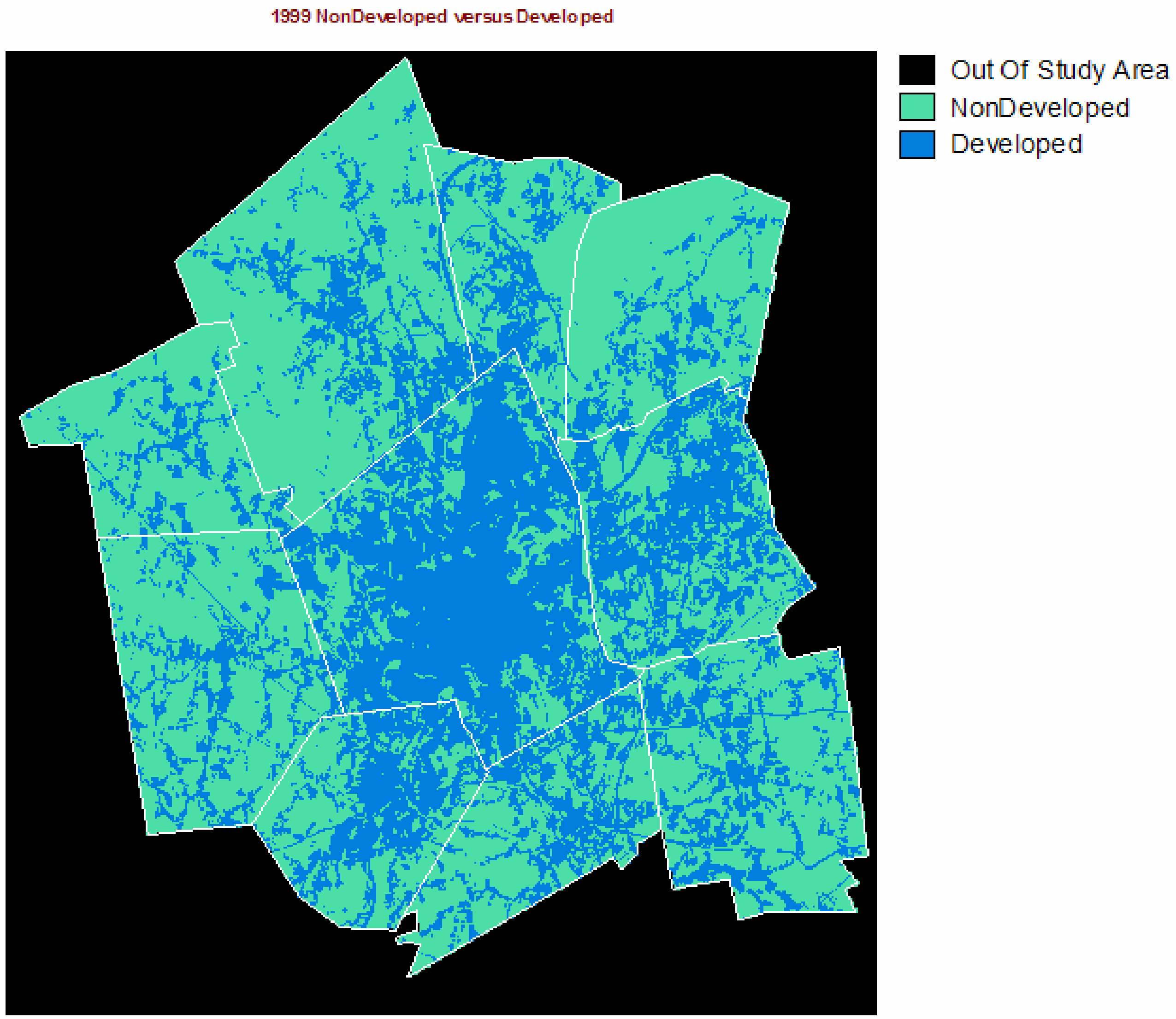

Figure 4: Validation image of developed versus non-developed of 1999.

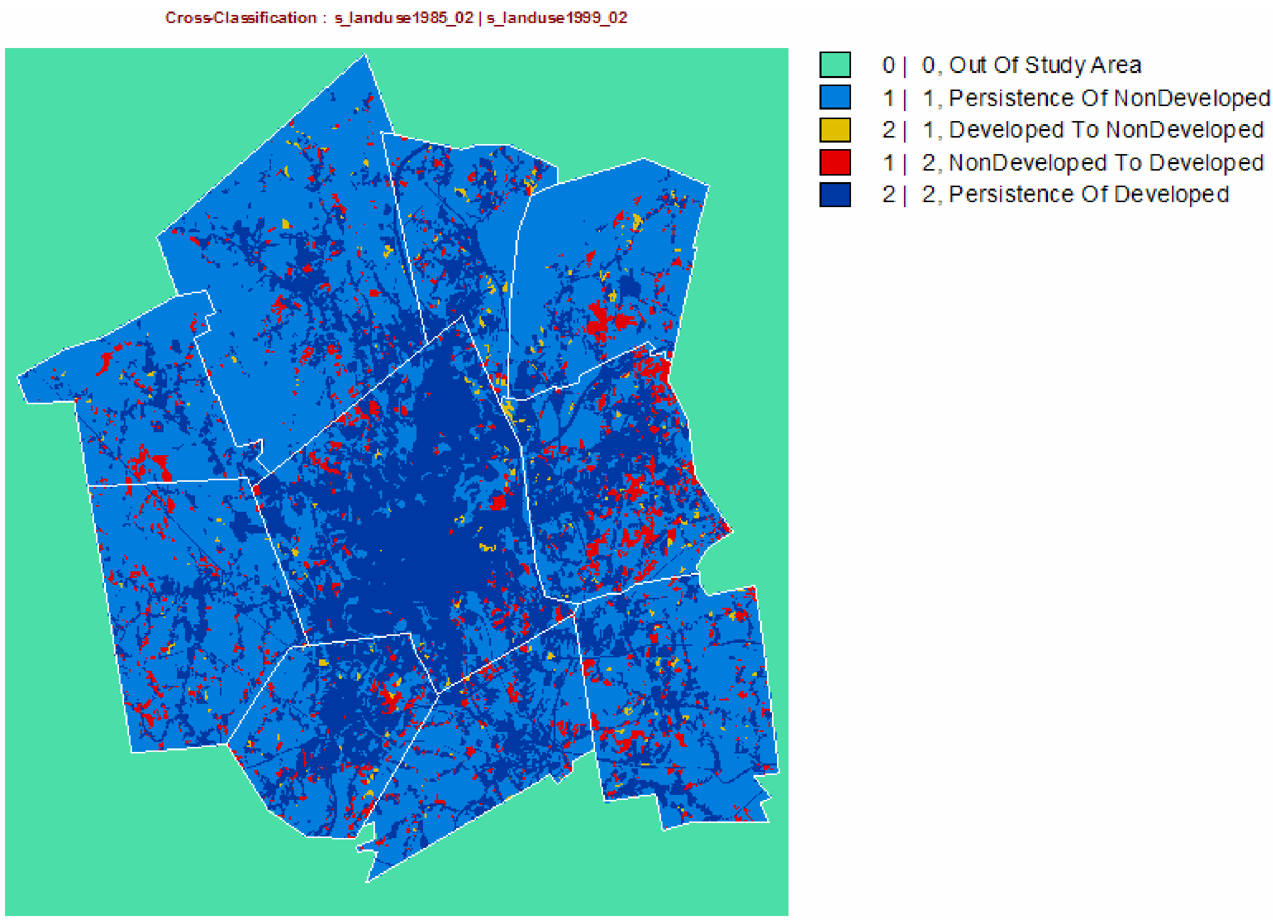

Figure 5: Image of land change between 1985 and 1999.

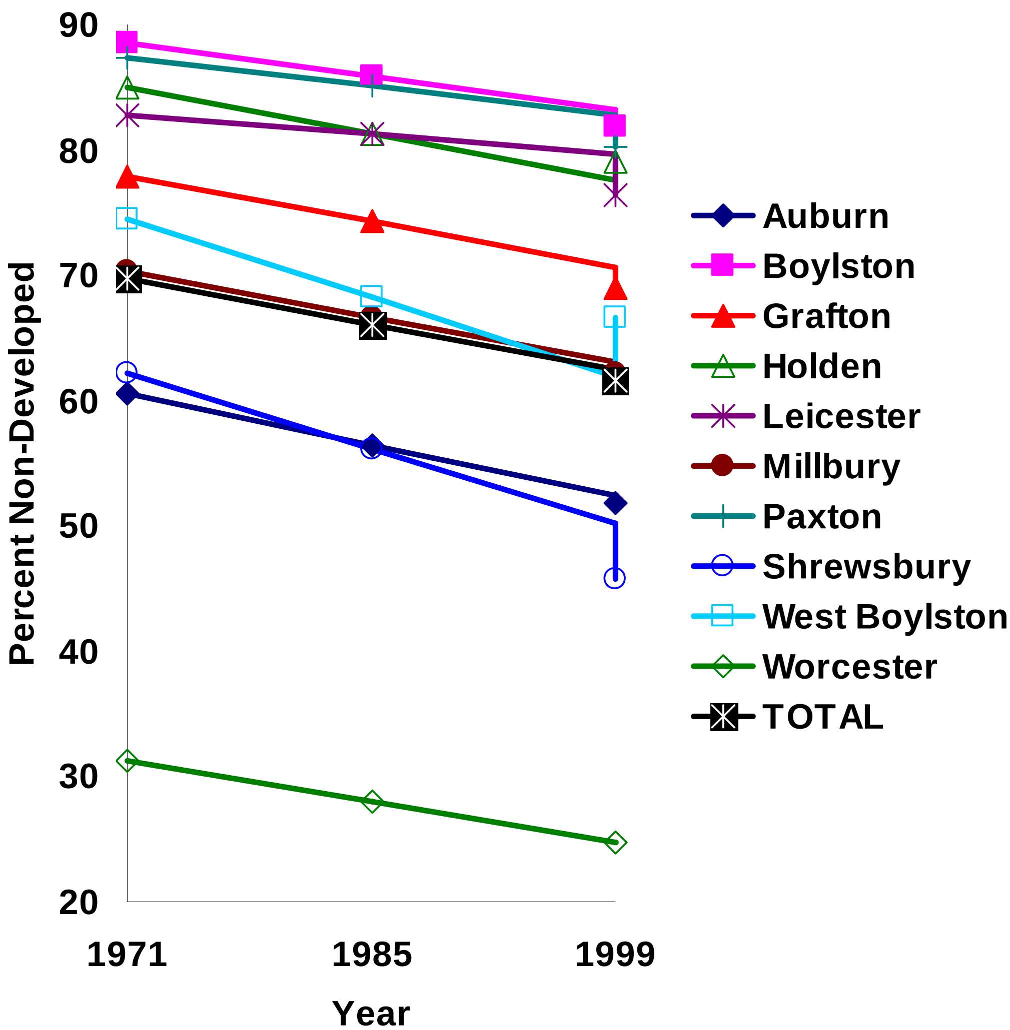

Figure 6: Percent non-developed for the 10 towns of the study area. The line interpolates the points of 1971 and 1985, then extrapolates to 1999, where it predicts with some error.

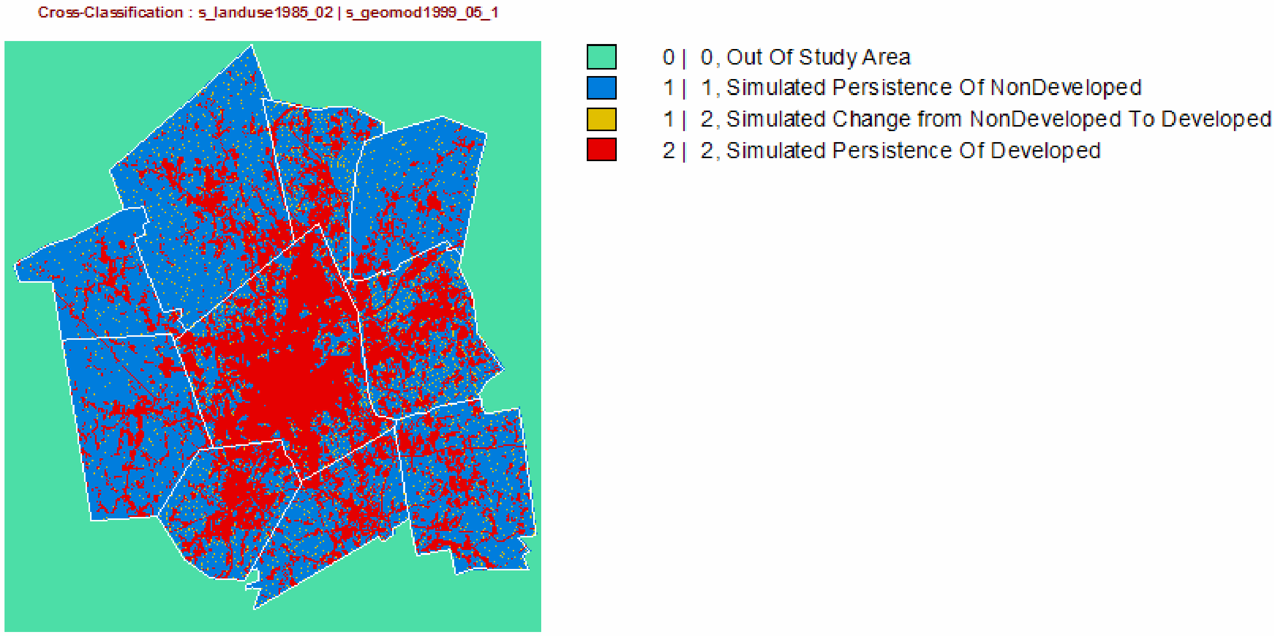

Figure 7: Predicted new development 1985-1999 image using the stratification rule only. The simulated changes are difficult to see because they are individual cells scattered among the persistence of non-development.

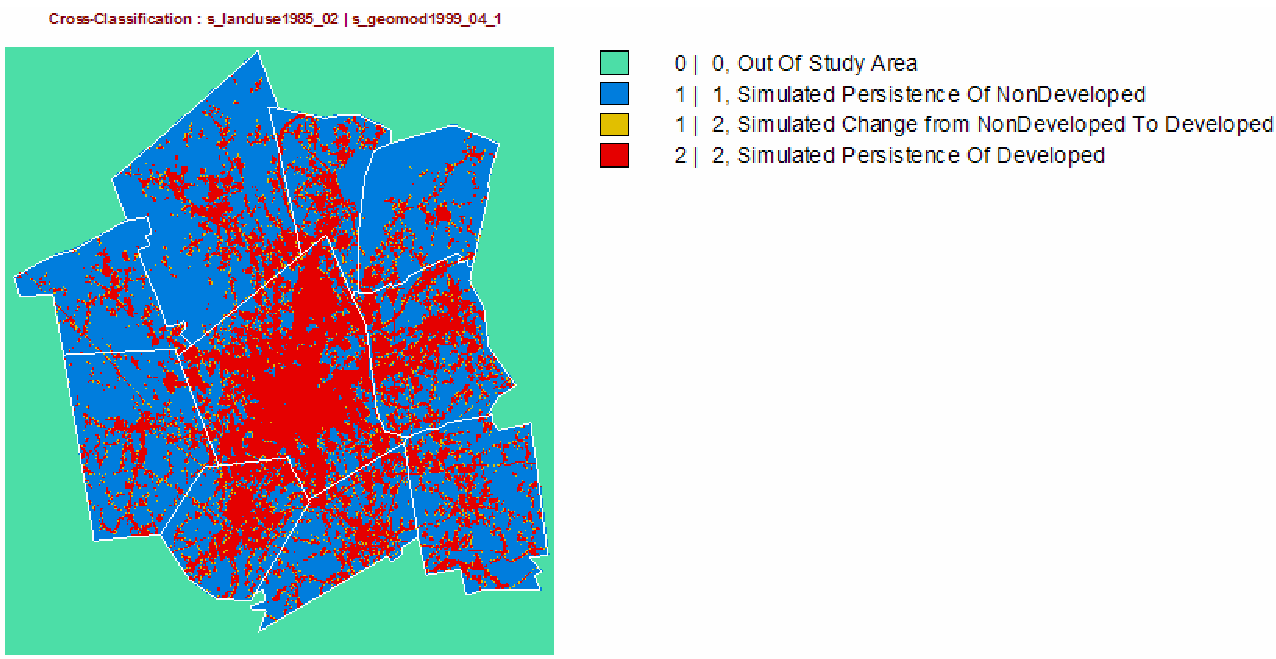

Figure 8: Predicted new development 1985-1999 image using the constraint rule only. The simulated changes are difficult to see because they are cells on the edge of developed versus non-developed of 1985.

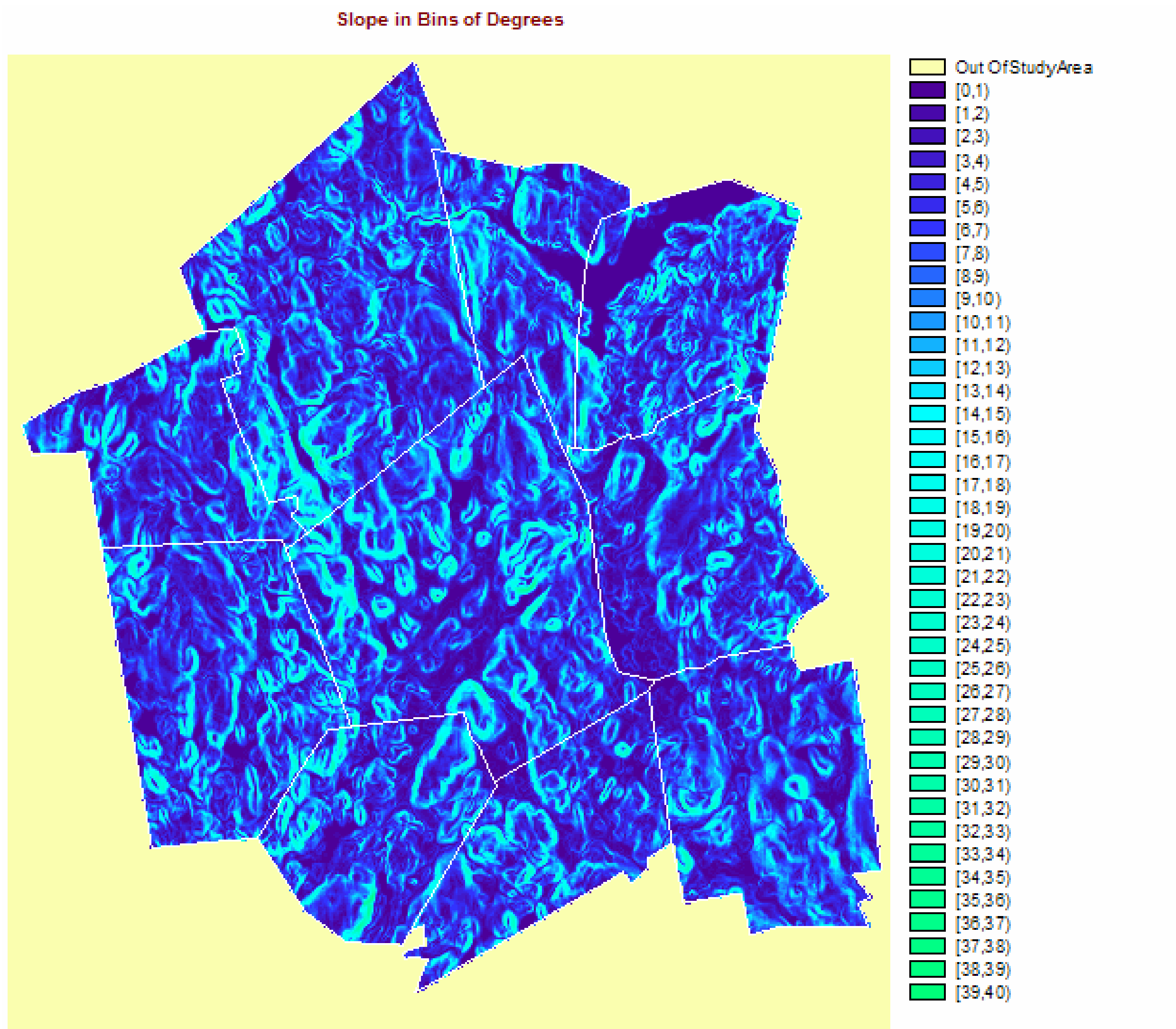

Figure 9: Image of slope, categorized into bins of one degree.

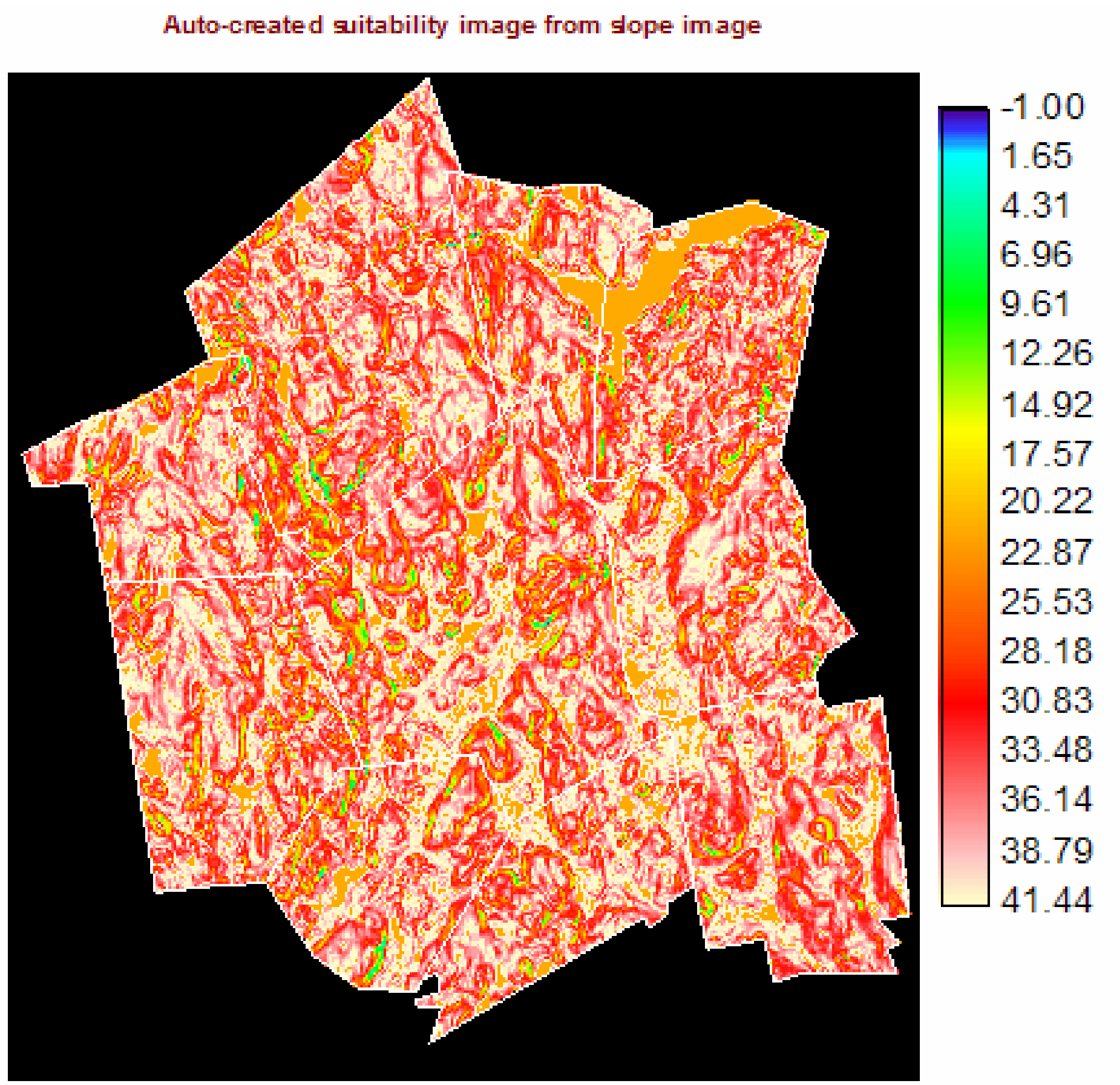

Figure 10: Suitability map created from the slope map and the 1985 beginning time image.

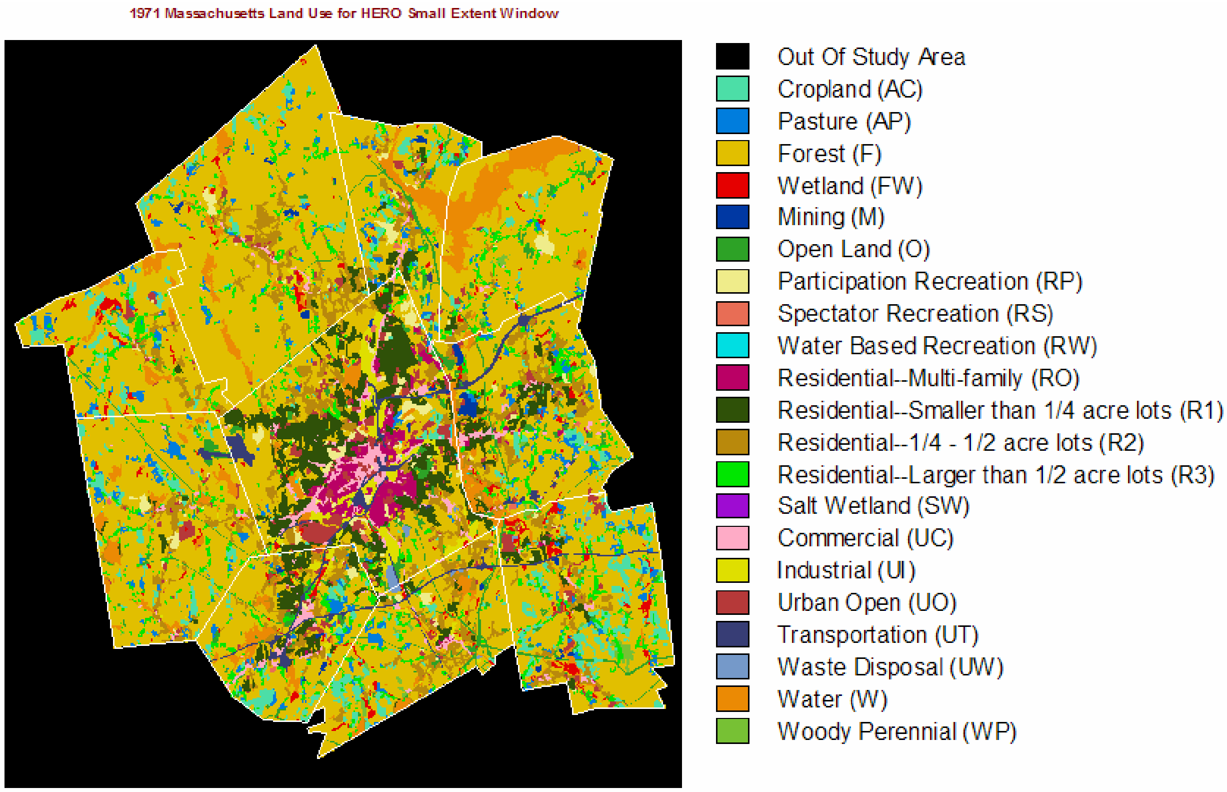

Figure 11: Image of historical land use for several categories of 1971.

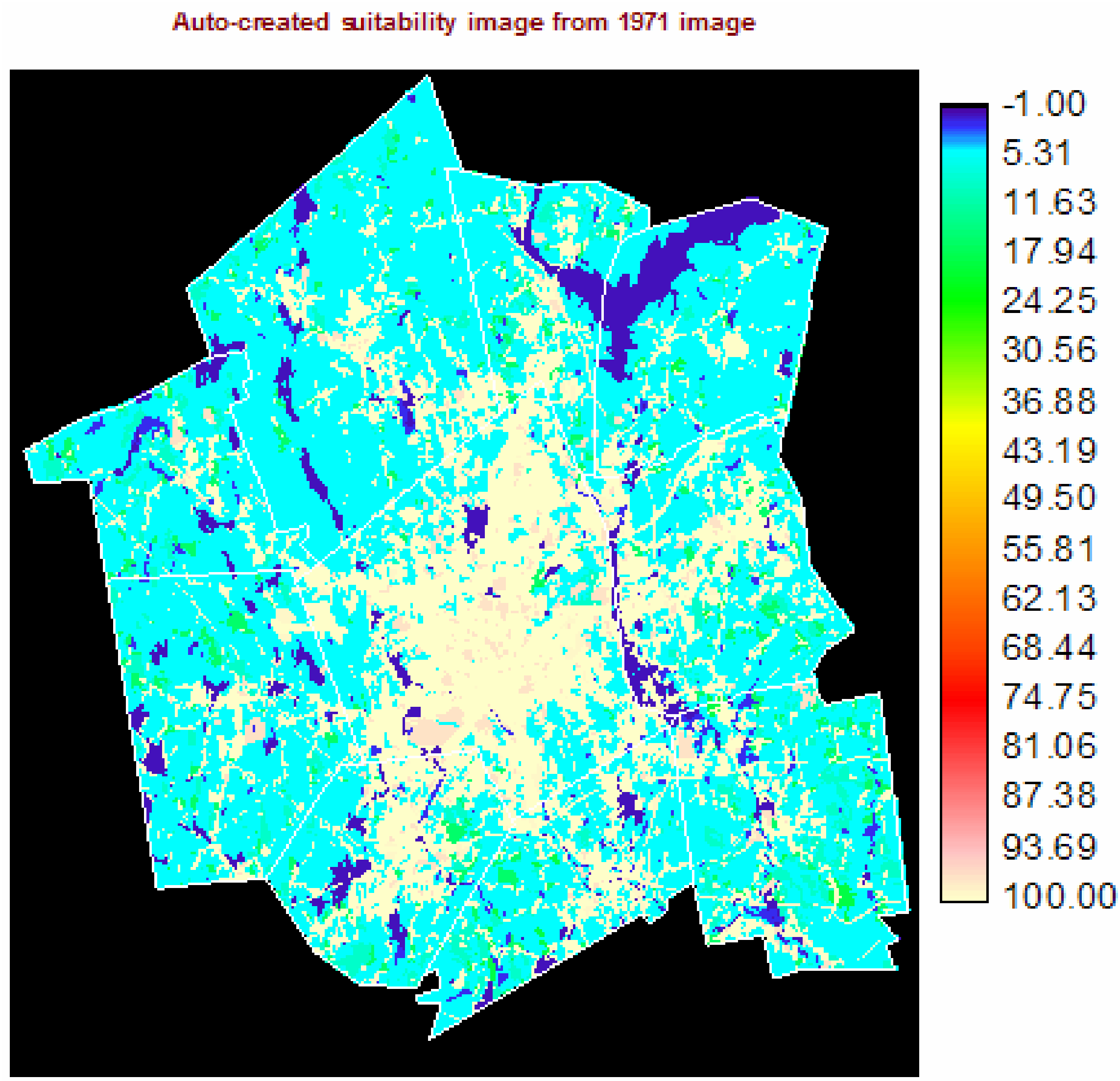

Figure 12: Suitability map created from the historic land use map of 1971 and the 1985 beginning time image.

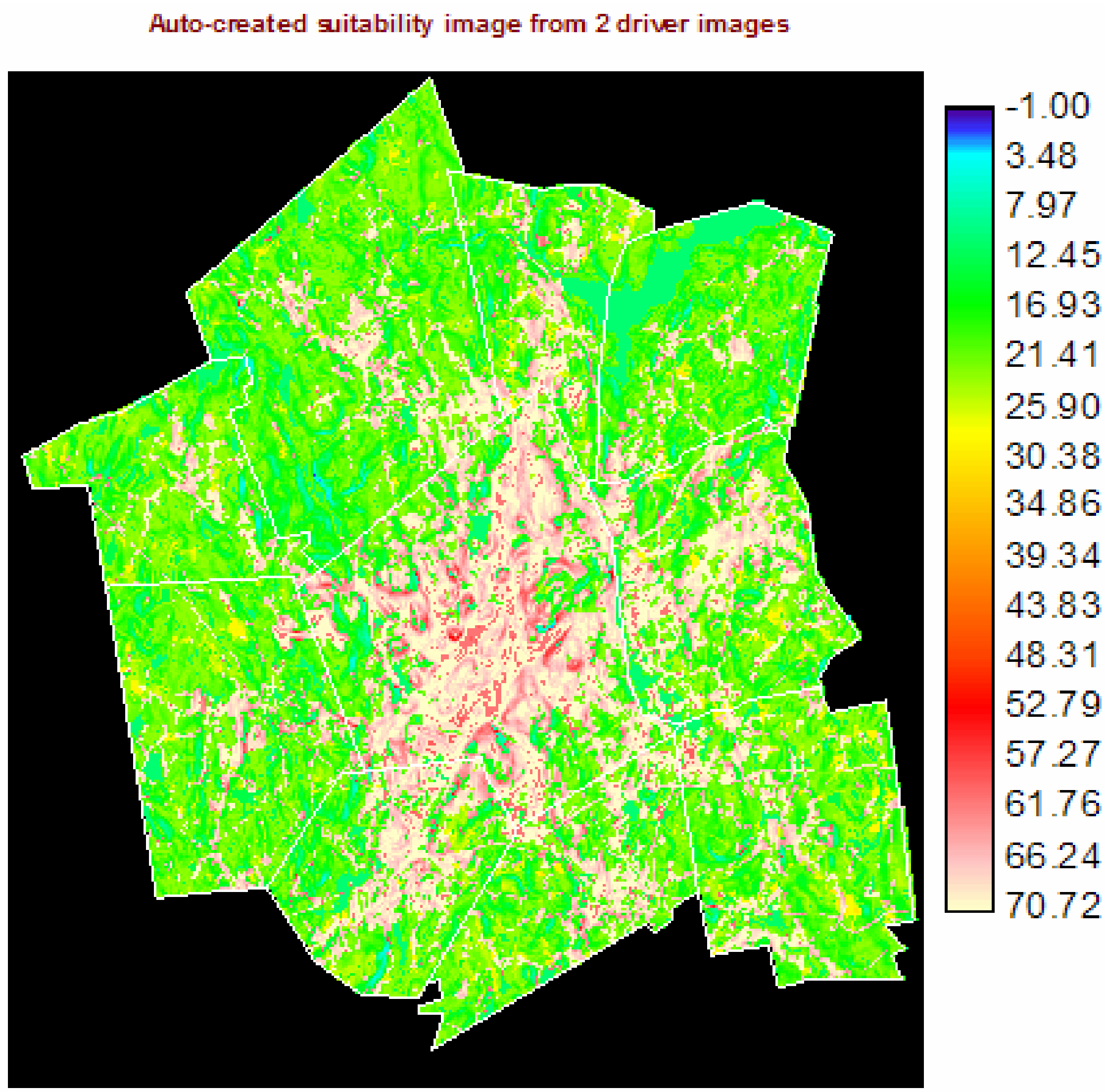

Figure 13: Suitability map created from the average of two suitability maps of Figures 10 and 12.

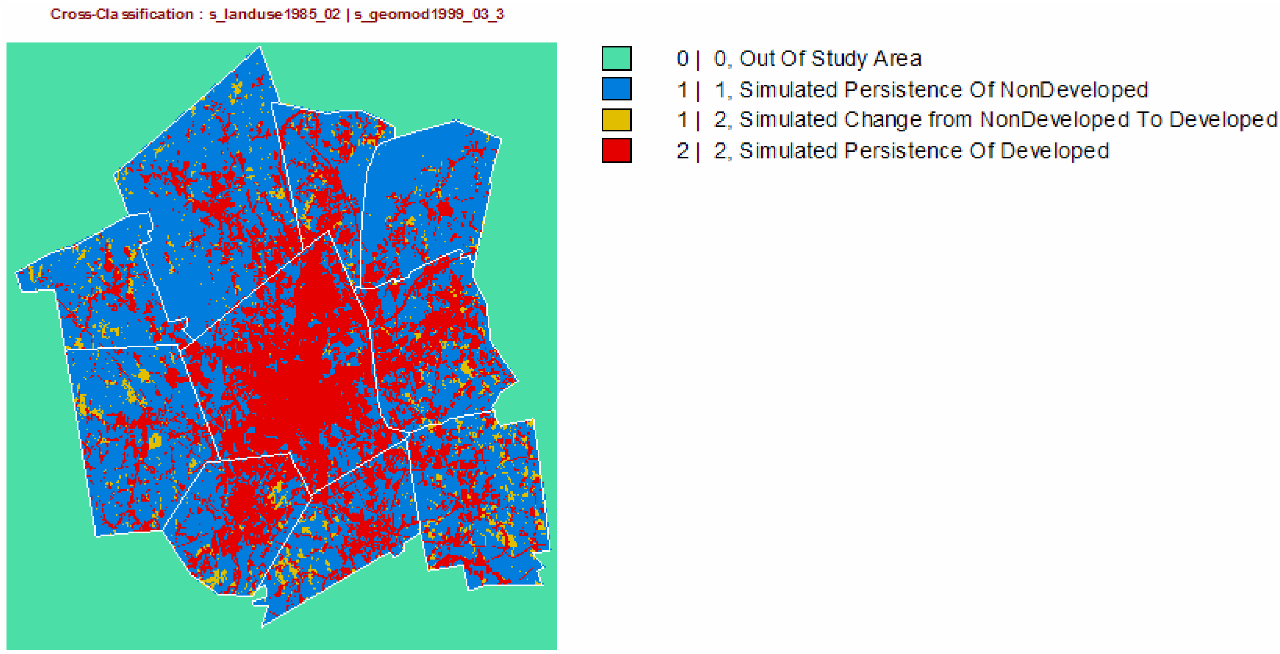

Figure 14: Predicted new development 1985-1999 image using the suitability map only.

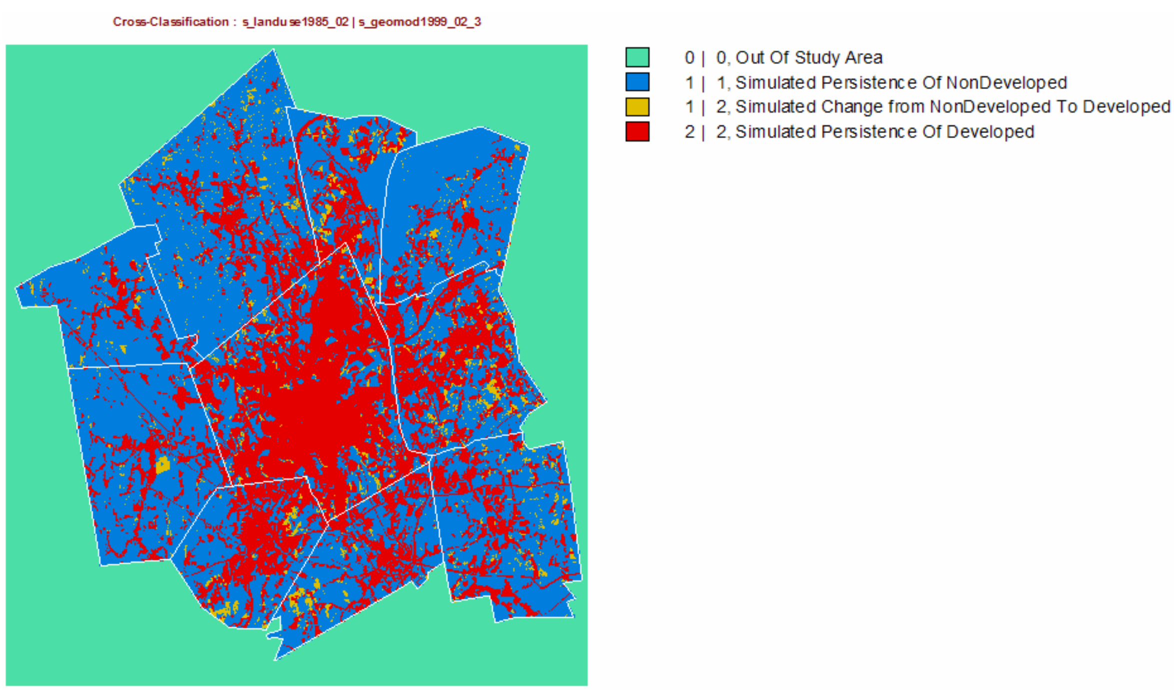

Figure 15: Predicted new development 1985-1999 image using the suitability map and regional stratification.

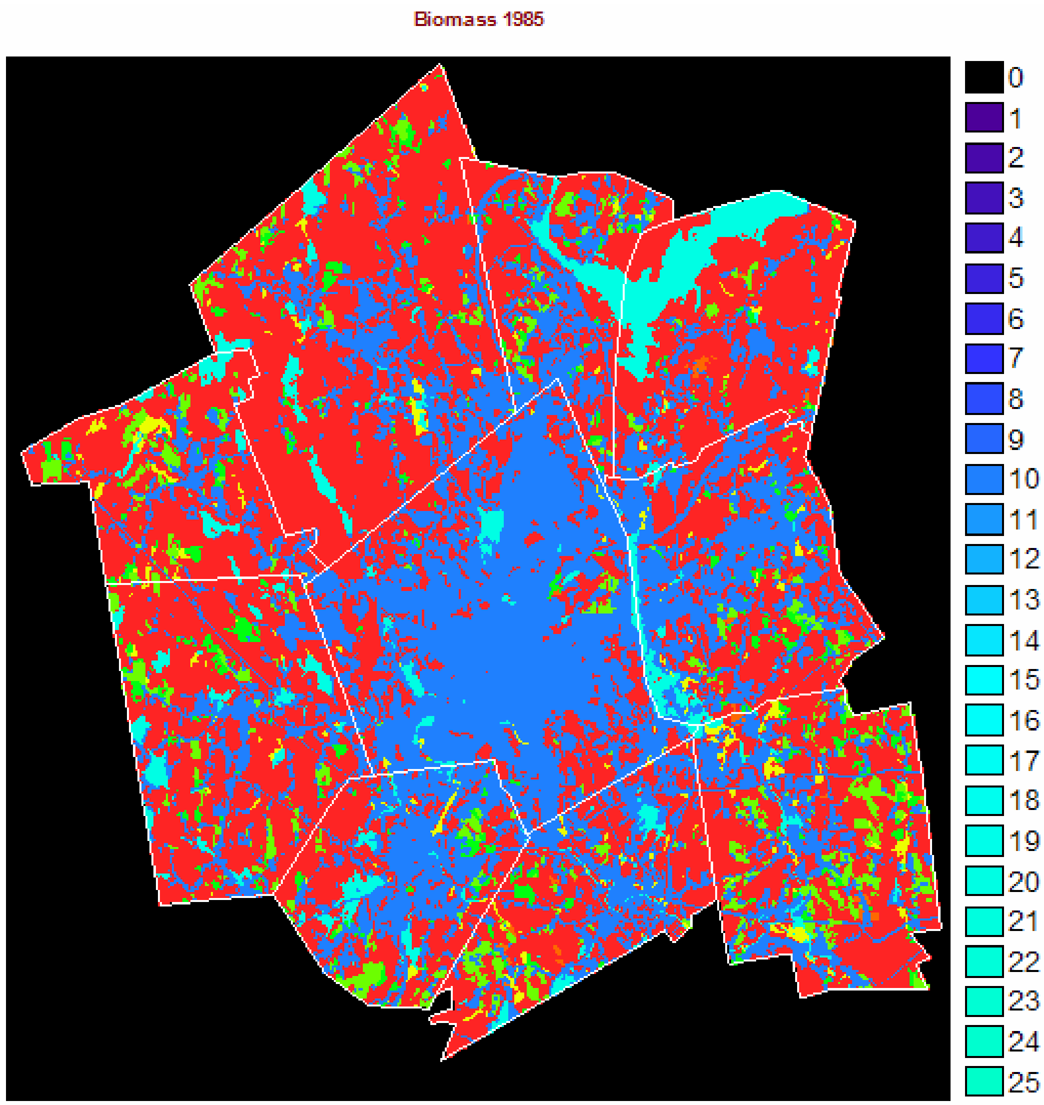

Figure 16: Environmental resource image of biomass of 1985 in tons of biomass per cell.

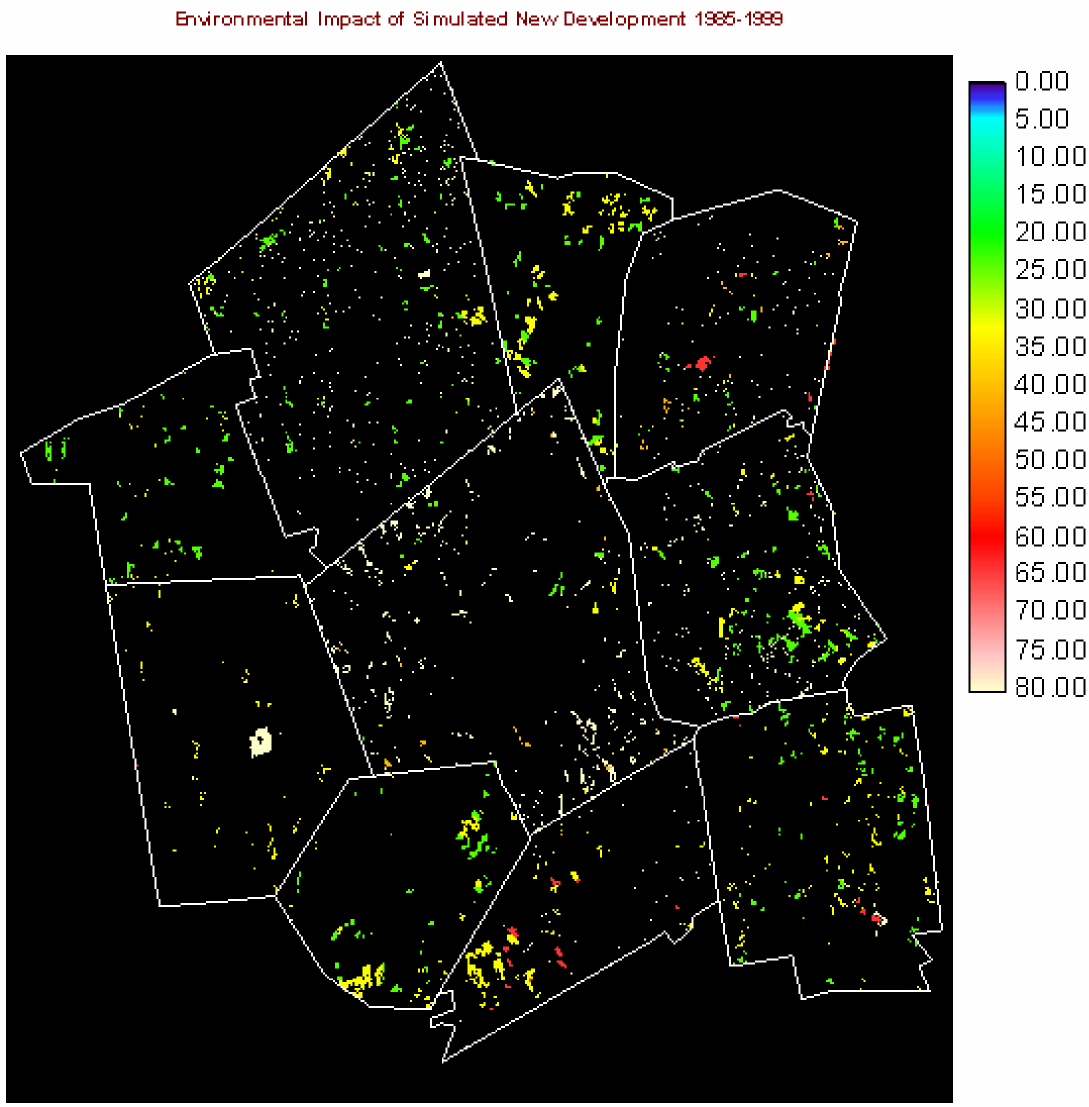

Figure 17: Environmental impact image of decrease in tons of carbon per cell as a result of simulated development from 1985 to 1999 shown in Figure 15.

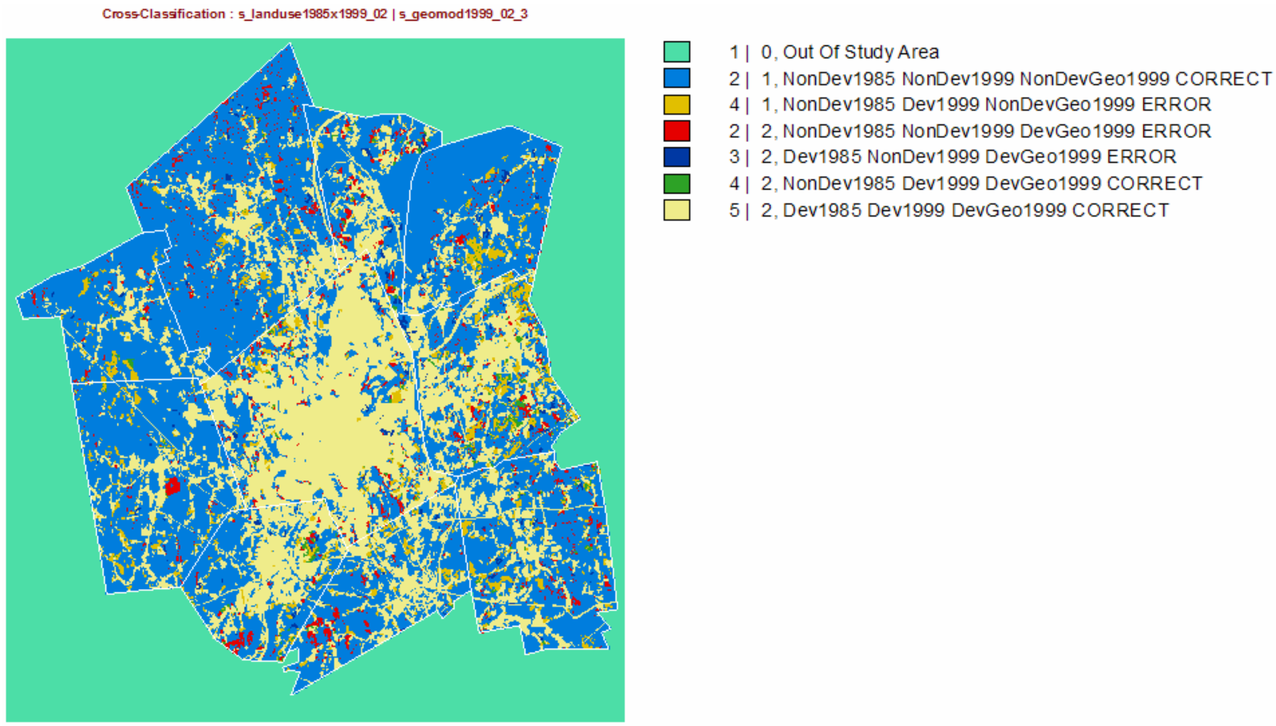

Figure 18: Visual validation of simulated development from 1985 to 1999 shown in Figure 15.

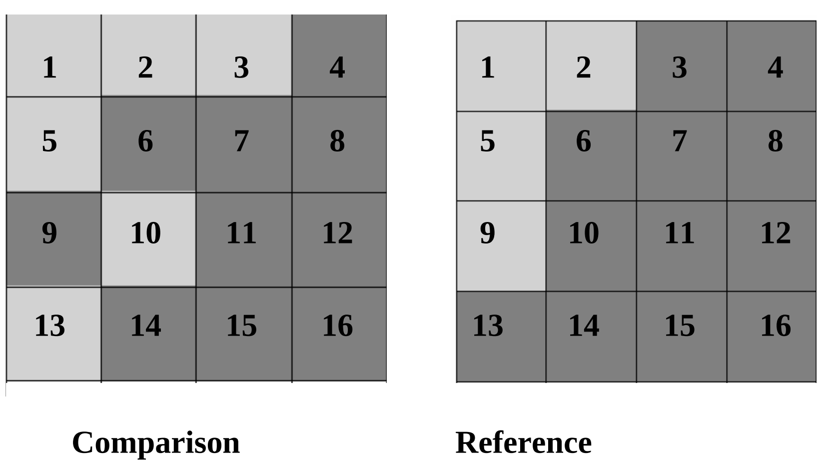

Figure 19: Demonstration puzzle to illustrate error of location versus error of quantity. Each map shows a categorical variable, with two categories: dark and light. Numbers identify the individual grid cells.

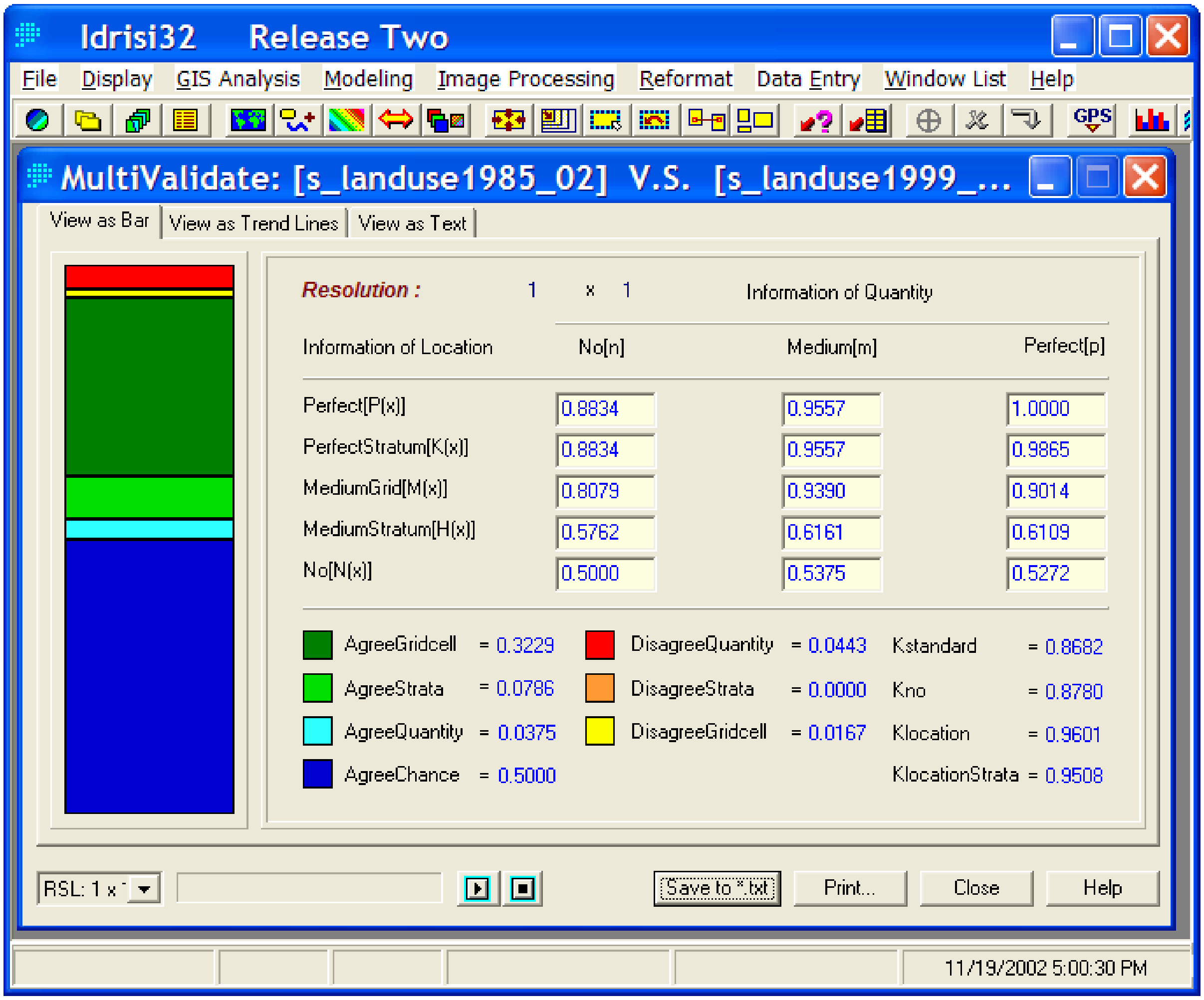

Figure 20: Components of agreement and disagreement for the comparison between the real 1999 map and the real 1985 map.

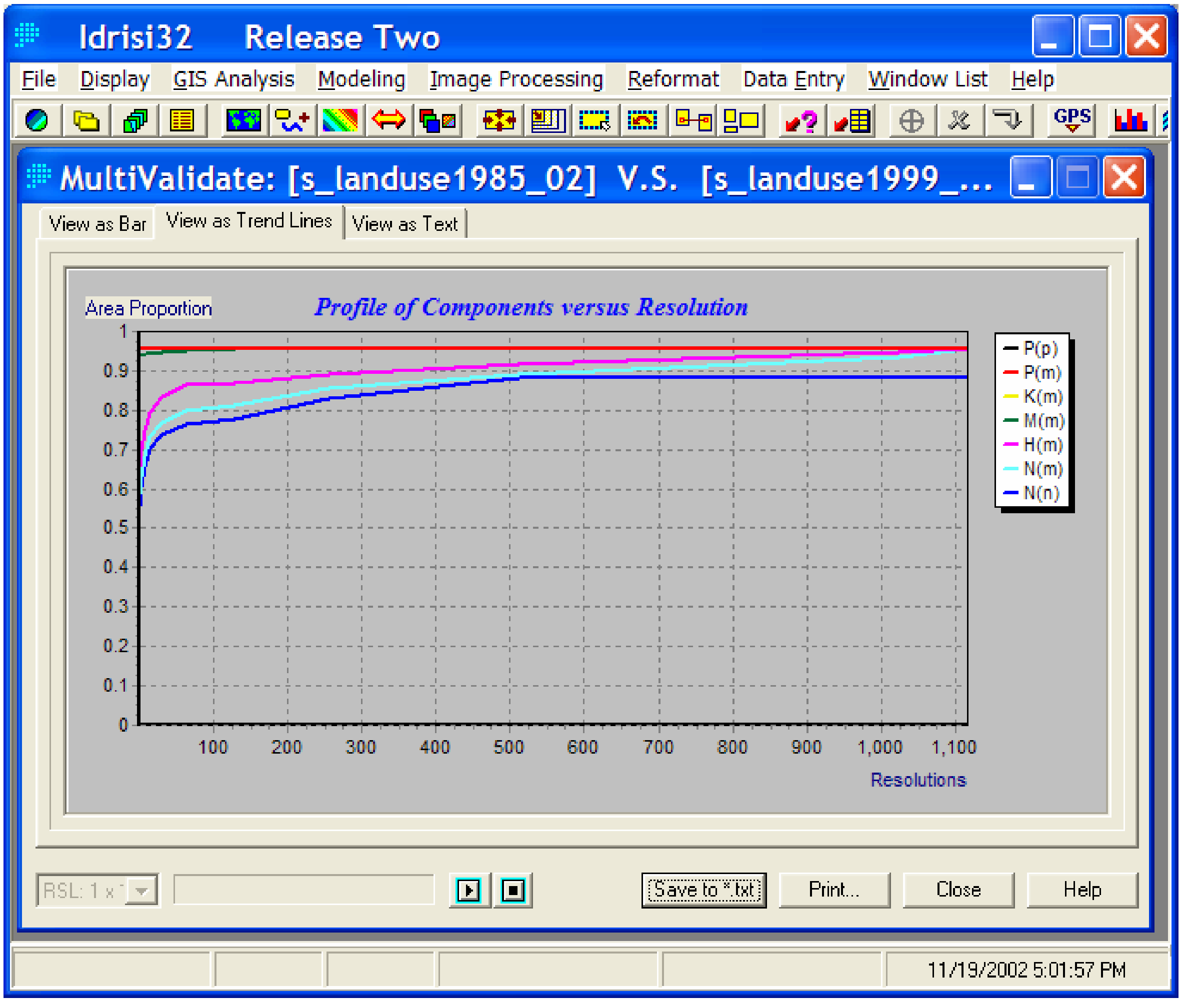

Figure 21: Components of agreement and disagreement for the comparison between the real 1999 map and the real 1985 map at multiple resolutions.

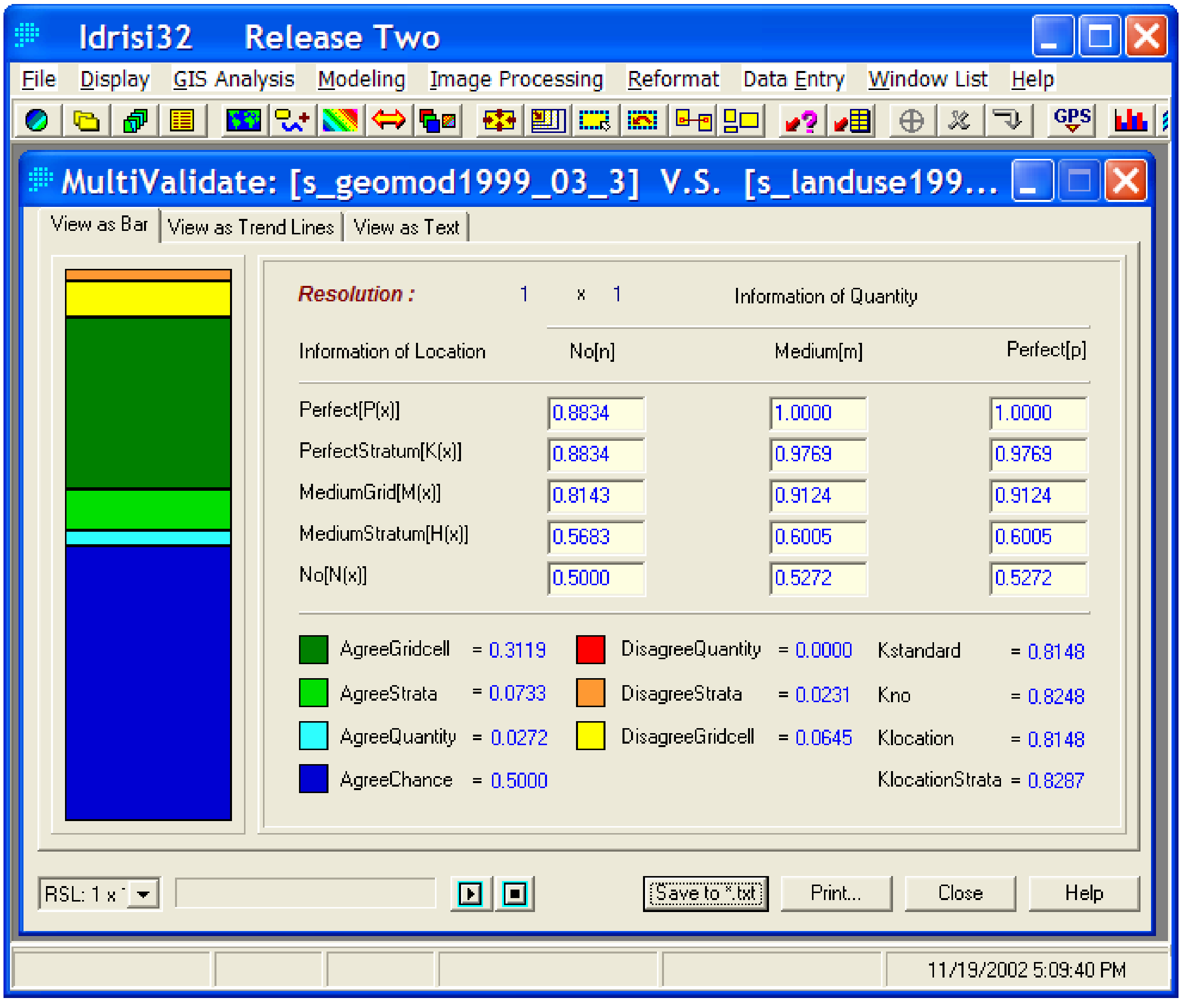

Figure 22: Components of agreement and disagreement for the comparison between the real 1999 map and the simulated 1999 map.

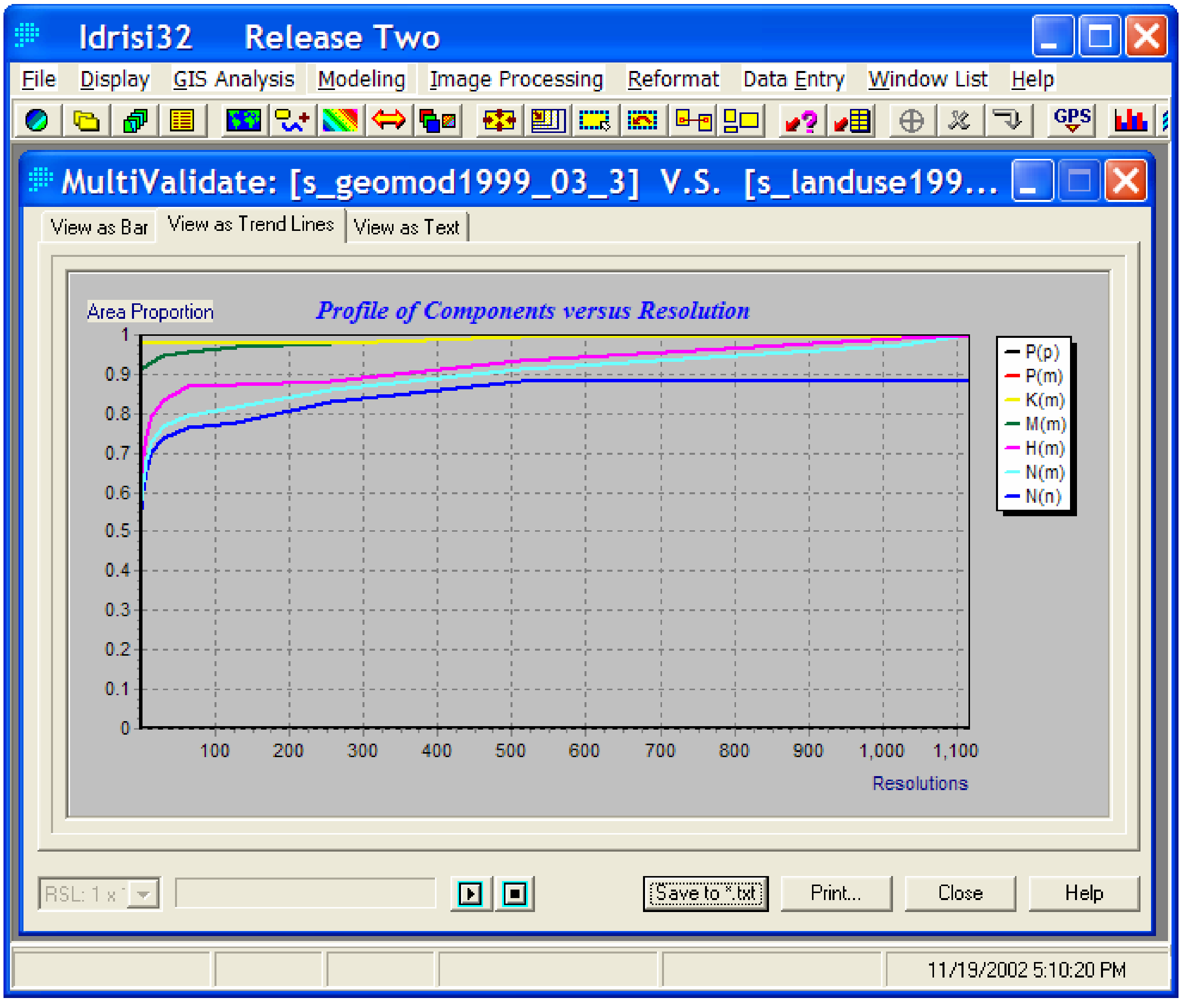

Figure 23: Components of agreement and disagreement for the comparison between the real 1999 map and the simulated 1999 map at multiple resolutions.

Information of Location	n	m	p
P(**x**)	$\sum_{j=1}^{J} MIN(R_{..j}\ ,\ \frac{1}{J})$	$\sum_{j=1}^{J} MIN(R_{..j}\ ,\ S_{..j})$	$\sum_{j=1}^{J} MIN(R_{..j}\ ,\ R_{..j})$
K(**x**)	$\frac{\sum_{d=1}^{D} W_{d\cdot}\left[\sum_{j=1}^{J} MIN(R_{d.j}\ ,\ E_{d.j})\right]}{\sum_{d=1}^{D} W_{d\cdot}}$	$\frac{\sum_{d=1}^{D} W_{d\cdot}\left[\sum_{j=1}^{J} MIN(R_{d.j}\ ,\ S_{d.j})\right]}{\sum_{d=1}^{D} W_{d\cdot}}$	$\frac{\sum_{d=1}^{D} W_{d\cdot}\left[\sum_{j=1}^{J} MIN(R_{d.j}\ ,\ F_{d.j})\right]}{\sum_{d=1}^{D} W_{d\cdot}}$
M(**x**)	$\frac{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}\left[\sum_{j=1}^{J} MIN(R_{dnj}\ ,A_{dnj})\right]}{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}}$	$\frac{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}\left[\sum_{j=1}^{J} MIN(R_{dnj}\ ,S_{dnj})\right]}{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}}$	$\frac{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}\left[\sum_{j=1}^{J} MIN(R_{dnj}\ ,B_{dnj})\right]}{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}}$
H(**x**)	$\frac{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}\left[\sum_{j=1}^{J} MIN(R_{dnj}\ ,E_{d\cdot j})\right]}{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}}$	$\frac{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}\left[\sum_{j=1}^{J} MIN(R_{dnj}\ ,S_{d\cdot j})\right]}{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}}$	$\frac{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}\left[\sum_{j=1}^{J} MIN(R_{dnj}\ ,F_{d\cdot j})\right]}{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}}$
N(**x**)	$\frac{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}\left[\sum_{j=1}^{J} MIN(R_{dnj}\ ,\frac{1}{J})\right]}{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}}$	$\frac{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}\left[\sum_{j=1}^{J} MIN(R_{dnj}\ ,S_{..j})\right]}{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}}$	$\frac{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}\left[\sum_{j=1}^{J} MIN(R_{dnj}\ ,R_{..j})\right]}{\sum_{d=1}^{D}\sum_{n=1}^{N_d} W_{dn}}$

Information of Quantity

Figure 24: Mathematical expressions of agreement between the reference map and various components of possible information in a comparison map. The vertical axis shows information of location and the horizontal axis shows information of quantity. The text defines the variables.

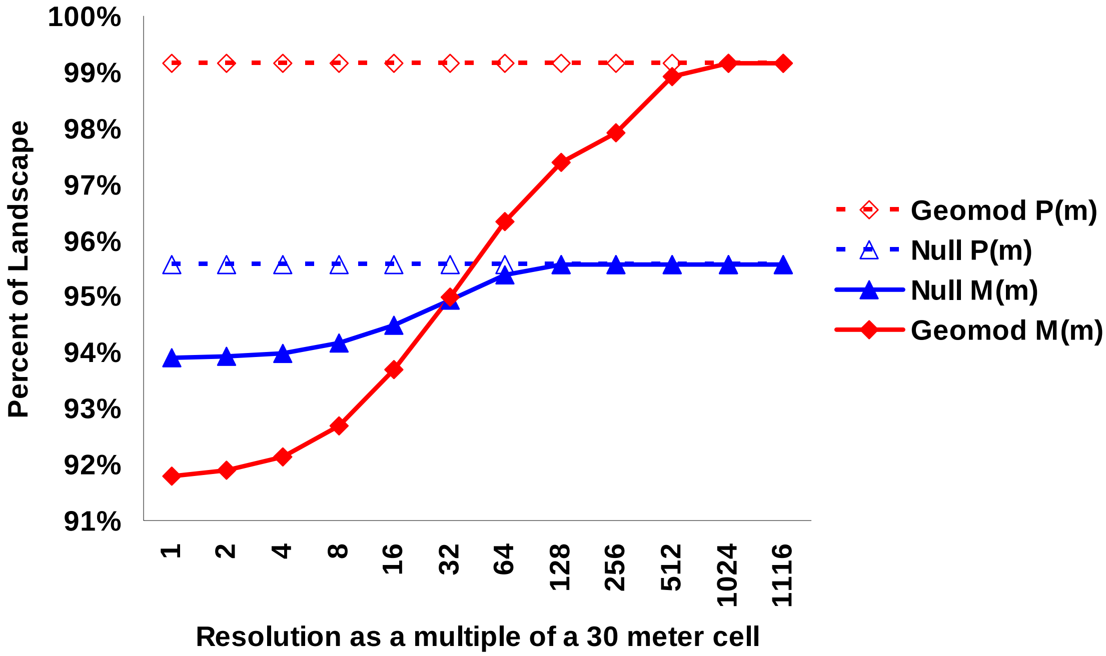

Figure 25: Multiple resolution goodness-of-fit for both the GEOMOD simulation and a null model of persistence only.

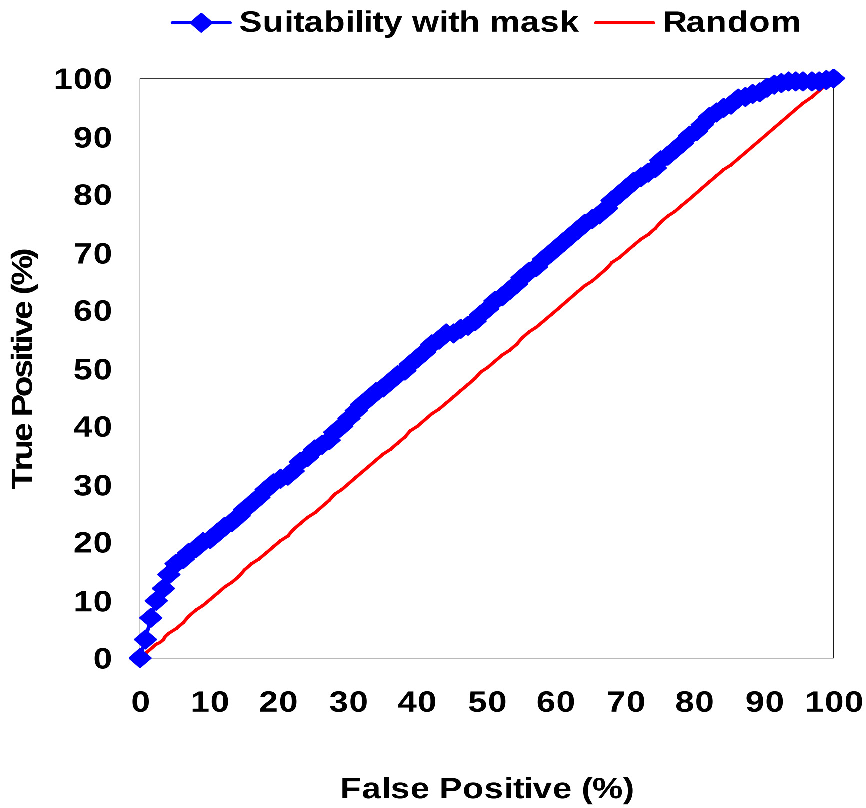

Figure 26: ROC curve for agreement between reference map of Boolean 1985-1999 development and suitability map for post-1985 development for only cells that are non-developed in 1985 for 101 thresholds. ROC=59.9%

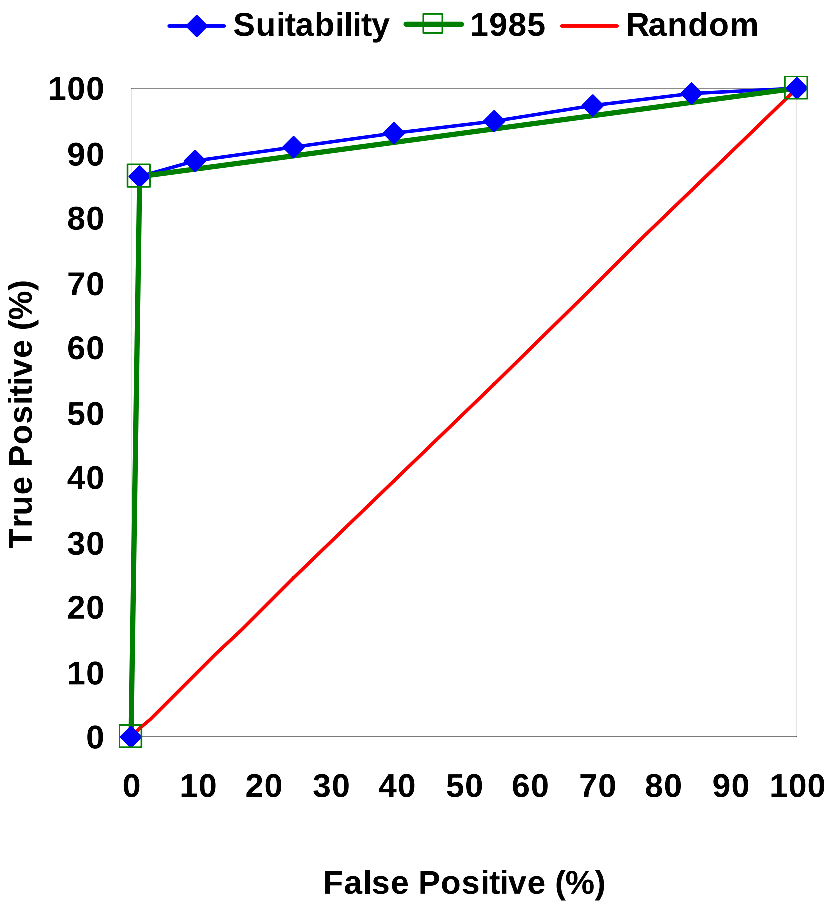

Figure 27: ROC curve for agreement between 1999 reference map of Boolean development and 1985 beginning time map of Boolean development for the entire study area. ROC = 92.5% for one threshold at 33.913% and other ROC of 93.7% for nine thresholds: 0, 33.91, 40, 50, 60, 70, 80, 90, 100.

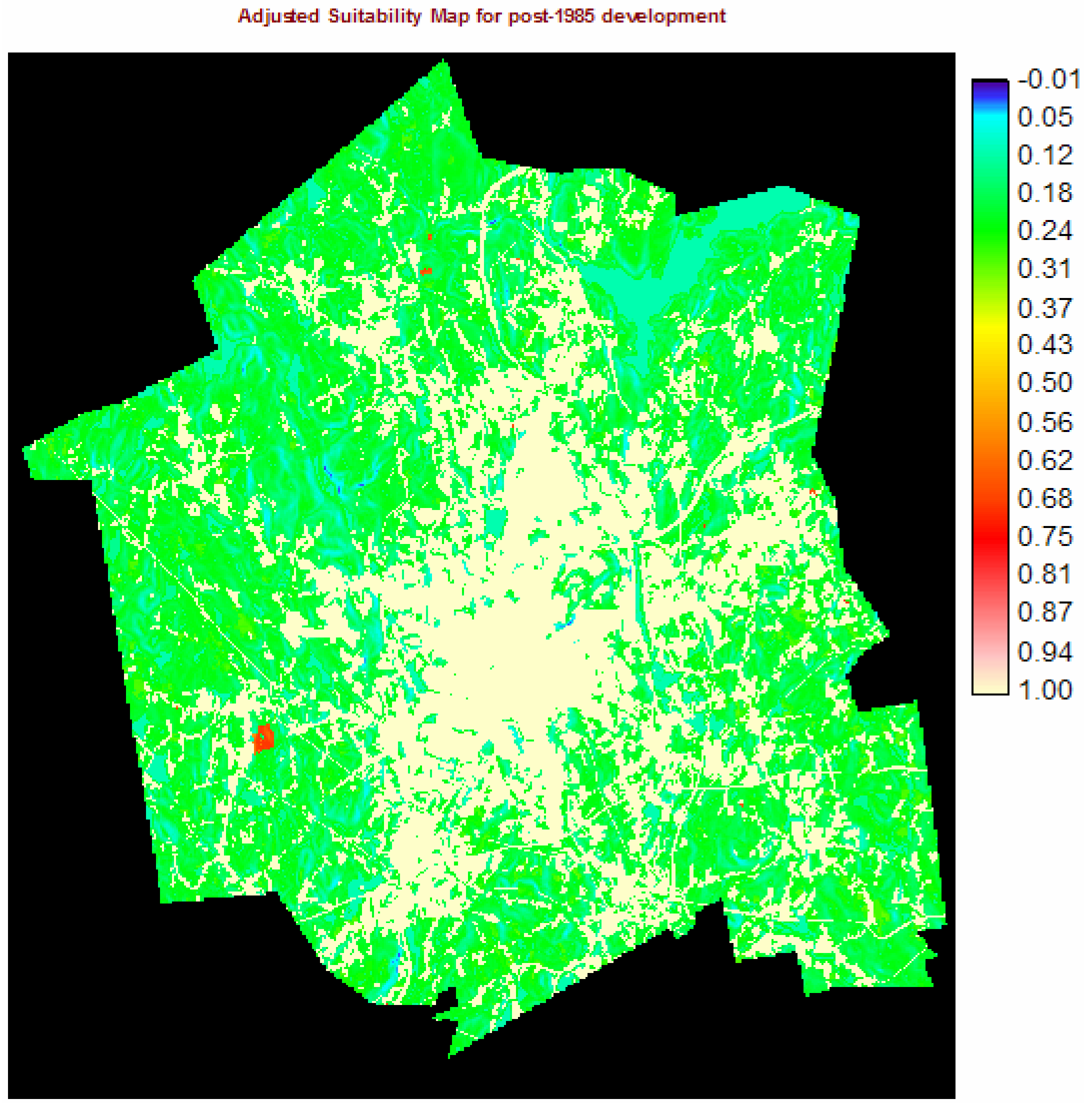

Figure 28: Adjusted suitability map for post-1985 development in which developed cells of 1985 have a suitability of 1.