Agriculture, Ecosystems and Environment 85 (2001) 83–94

Agriculture Ecosystems & Environment

www.elsevier.com/locate/agee

Modeling land-use change in the Ipswich watershed, Massachusetts, USA

Laura C. Schneider*, R. Gil Pontius Jr.

Graduate School of Geography, Clark University, 950 Main Street, Worcester, MA 01610-1477, USA

Abstract

The Ipswich watershed in northeastern Massachusetts, USA, is experiencing important land-use changes, which are contributing to severe environmental problems such as eutrophication, ground water depletion and loss of wildlife. The objective of this paper is to model deforestation between 1971, 1985 and 1991 in the watershed of the Ipswich River in Massachusetts, USA, where most of the forest loss is attributable to new residential development. The maps of suitability for deforestation are calibrated with maps of real change between 1971 and 1985 by using logistic regression, multi-criteria analysis and spatial filters. The maps of 1971 and 1985 serve also as the basis to extrapolate the quantity of predicted future deforestation. Then, the calibrated suitability maps and extrapolated quantities predict the location of deforestation between 1985 and 1991. The predicted deforestation maps are validated with the map of real forest loss of 1985–1991. relative operating characteristic (ROC) and variations of the Kappa index of agreement (Kno, Klocation and Kquantity) measure the validation. For most simulation runs, Kno = 93%, Klocation = 8% and Kquantity = 100%. The best predictor of quantity of deforestation from 1985 to 1991 is linear extrapolation forward in time of the deforestation that occurred from 1971 to 1985. It is difficult to predict the exact locations of deforestation in the watershed because only 2% of the watershed is deforested from 1971 to 1991, the patches of deforestation are scattered evenly across the landscape, and the some of the most important variables are not readily available in digital form. Nevertheless, the best predictor of location of deforestation (ROC = 70%) is a suitability map that uses a spatial filter and multi-criteria evaluation of elevation, slope, and proximity to existing residential areas. The locations that are most threatened are those that are unprotected, near existing residential development and in towns where the demand for new residential development is high.

Keywords: GIS; Kappa; Land-use change; Massachusetts; Model; ROC; Validation

1. Introduction

A major element of environmental change is the modification of natural land-cover due to human land uses. Human activities are altering the land at unprecedented rates, magnitudes and spatial scales (Turner, 1994; Vitousek et al., 1997). Modeling is an important tool for studying land-use change due to its ability to integrate measurements of changes in land-cover and the associated drivers (Lambin et al., 1999). Lambin (1997) points out that models of land-use/land-cover processes can help scientists generate hypotheses and in some cases answer three main questions: (1) What biophysical and socioeconomic variables explain land-cover changes? (2) Where are the locations affected by changes? (3) At what rate do land-cover changes advance? Models then help to explain and/or predict land-use/land-cover processes (Pontius, 1994; Hall et al., 1995; Veldkamp and Fresco, 1996; Lambin, 1997).

* Corresponding author. Tel.: +1-508-421-3706;
fax: +1-508-793-8881.
E-mail address: lschneider@clarku.edu (L.C. Schneider).

0167-8809/01/$ – see front matter
PII: S0167-8809(01)00189-X

One important reason for answering these questions is that land-cover change affects dramatically the structure and functioning of ecosystems (Vitousek et al., 1997). For example in watersheds, land conversion is the most important factor influencing water quality and runoff (Hopkinson and Vallino, 1995). Deforestation disrupts original patterns of water and material output from watersheds to rivers, thus affects the metabolism and productivity of important ecosystems such as estuaries (Hopkinson and Vallino, 1995).

Drivers of land transformation such as deforestation and urban sprawl are complex and regionally dependent (Kasperson et al., 1997). In industrialized countries like US, conversion of forest into residential areas is usually caused by regional economic growth, which generates jobs, attracts workers, increases per capita income and creates demand for larger residential plots (Bradshaw and Muller, 1998).

Descriptive models such as those based on regression provide an exploratory tool to test the existence of links between land-cover change and candidate driving forces or their proxies. Future land-cover change locations can be predicted by combining spatial statistical models with spatially explicit data in a geographical information system (GIS) environment. Multi-criteria evaluation can be also used to predict land-cover change. Spatial variables are defined as criteria and their information is combined to create a single index of evaluation similar to an estimation of a probability. Criteria then can be combined and trade off can be assessed (Eastman et al., 1995). To estimate the quantity of land-use change, it is important to know the extent to which socioeconomic and physical variables explain and predict land change such as deforestation.

The goal of this paper is to develop and validate a model to predict the location and quantity of deforestation in the Ipswich watershed, using logistic regression and multi-criteria evaluation. Empirical analysis calibrates maps of suitability for deforestation, and then the maps of suitability are used to predict deforestation. The methods are validated using ROC and variations on the kappa index of agreement (Pontius, 2000).

2. Methods

2.1. Study area and digital maps

The Ipswich River watershed is in Northeastern Massachusetts, 30 miles north of the city of Boston (Fig. 1). The watershed covers 404 km^2 and includes parts of 21 towns. Human's use of the watershed's resources is so intensive that the river runs dry during hot summers. When it has water, the Ipswich River flows into Plum Island Sound, which is a Long Term Ecological Research site of the Untied States National

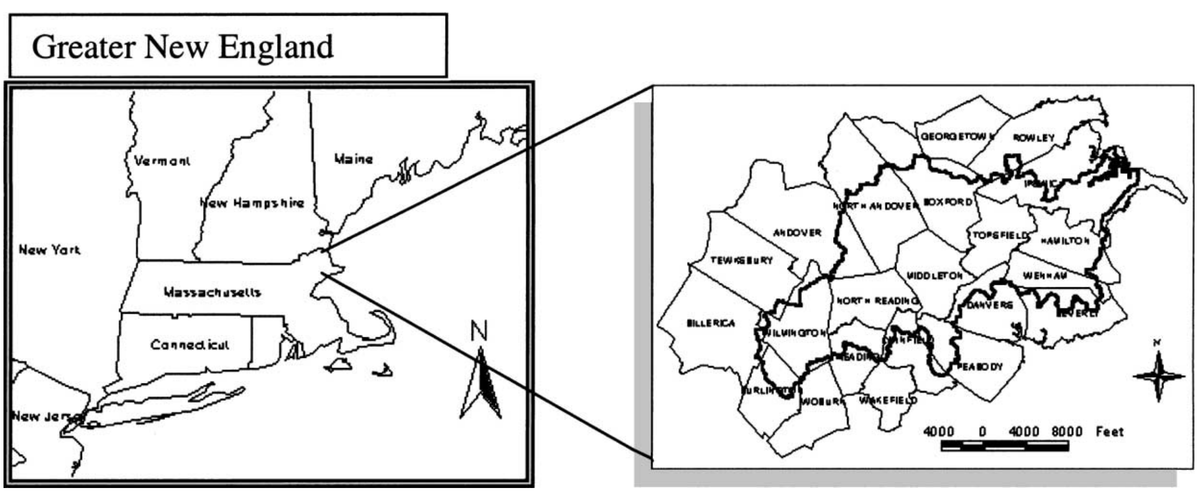

Fig. 1. The Ipswich watershed and town boundaries.

Science Foundation. In order to understand the nutrient dynamics and hence economic productivity of the sound, it is necessary to understand the land-use change in the watershed.

The watershed's human population is approximately 130,000 with median family income of $50,000, whereas the US median family income is $39,000. The area is attractive to upper-middle class people who wish to commute to the city of Boston. Urbanization occurred along the southern portion of the watershed following a rapid urban expansion of the city of Boston. In the last 20 years, the most important change in the Ipswich watershed is the replacement of forest by large residential plots. Almost a quarter of the Ipswich watershed has been set aside for land conservation and the remaining land is used primarily for suburban-residential areas.

Although deforestation is usually referred to as the transformation of old-growth forest, the current forest cover in the watershed is a product of recent succession. Primary forest in New England underwent constant cutting and burning due to agricultural practices during the last 400 years, however, the current forest is approximating the pre-settlement landscape (Foster, 1995). After decades of forest recovery following intense agricultural practices, the recent trend has been a rapid conversion of forest into residential areas. Therefore, the models developed here focus on deforestation.

The Executive Office of Environmental Affairs (EOEA) of Massachusetts supplies the maps through the MassGIS program. Land-use maps were available for three points in time: 1971, 1985 and 1991 (MassGIS, 1999). To highlight the major processes of land-cover change in the watershed, the original classification of 21 land categories was aggregated to 4: forest, residential/commercial, open land and wetlands. The original classification includes classes such cropland, pasture, open land, mining and different levels of residential based on lot areas. An additional map of wetlands, which is from 1:12,000 scale color-infrared stereo photography, was overlaid on the original land-use maps to revise the information of wetlands in the watershed.

As of 1991, the distribution of major land-cover types in the watershed is 44% forest, 31% residential/commercial, 17% wetlands and 8% open land. Residential/commercial areas include multi-family units, small and large residential plots, shopping centers and industry. Open lands include cropland, pastures, vacant undeveloped land, and open urban areas such institutional green space. Maps of land-cover for 1971, 1985 and 1991 show that a conversion of forest into residential areas is the predominant land-use change (Fig. 2).

A digital elevation model was created from a map of 1:250,000 scale contour lines using the triangulated irregular network (TIN) procedure. Slope was

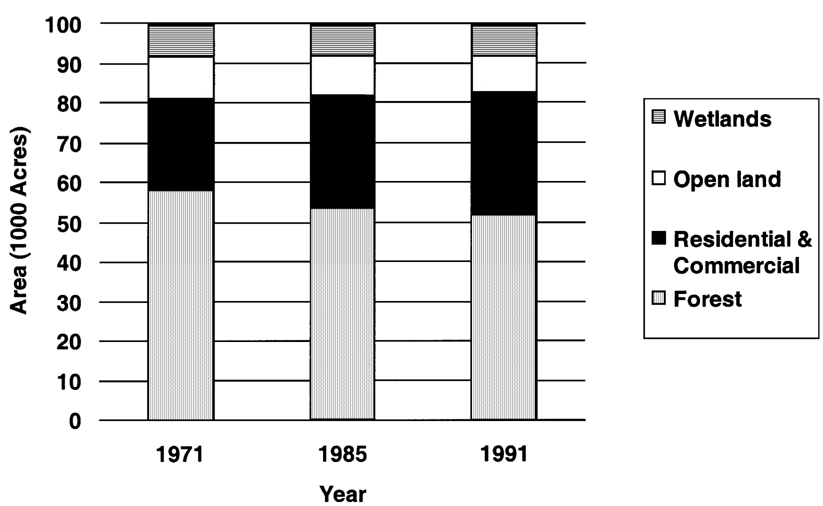

Fig. 2. Distribution of land types in 1971, 1985 and 1991.

calculated from the elevation map using the rook's case procedure. Two other important data layers are distance in meters from residential/commercial areas derived from the land-use maps of 1971 and 1985. All the maps are geographic data layers stored in the raster-based GIS IDRISI (Eastman, 1999). The resolution of all map layers is 30 m × 30 m grid cells.

2.2. Calibration of location of deforestation

Models of land-use change usually have three components: (1) maps of land-cover from more than one point in time, (2) a function of change that modifies the values and spatial arrangement of an initial land-cover map, and (3) the resulting prediction map (Lambin, 1994). Indeed, the model discussed in this paper has these three components. The change function can be created either by empirical analyses of historical patterns or by mathematical functions that describe hypothesized processes of future change (Lambin, 1994, 1997; Pontius, 2000). A variety of empirical methods to create change functions are compared. Specifically, logistic regression is compared to two types of multi-criteria evaluation (MCE) in order to predict location of change.

Each of these methods expresses its change function as a map of suitability for future deforestation at the grid cell scale. The suitability value gives the sequence in which grid cells are selected for change, where larger values of suitability are selected first. The value of the suitability indicates only the relative priority of land change, not necessarily the absolute likelihood of land change. Suitability values are not probabilities. The theorems of probability theory do not necessarily apply to suitability values. The suitability map displays all of the watershed's grid cells, where the value of each cell is a real number that indicates the predicted sequence of change from forest to non-forest. In a perfect map of suitability, the order of the suitability values would match the order in which the landscape changes, with the largest suitability values changing first and the smallest last. A map of suitability for forest loss guides the sequence in which the model selects grid cells for predicted deforestation.

The suitability maps developed for the watershed are calibrated with data from 1971 to 1985. The model predicts deforestation that occurred from 1985 to 1991. Actual deforestation from 1985 to 1991 is used for validation. The theoretical basis and application of logistic regression and MCE to model deforestation in the watershed are as follows.

2.2.1. Logistic regression

In the application of logistic regression, each "observation" is a grid cell. The dependent variable is a binary presence or absence event, where 1 = forest loss and 0 = other, for the period 1971–1985. The logistic function gives the probability of forest loss as a function of the explanatory variables. The function is a monotonic curvilinear response bounded between 0 and 1, given by a logistic function of the form:

$$p = E(Y) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_1 X_2 + \beta_1 X_3)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)} \tag{1}$$

where p is the probability of forest loss in the cell, $E(Y)$ the expected value of the binary dependent variable Y, β_0 a constant to be estimated, β_i a coefficient to be estimated for each independent variable X_i. The logistic function can be transformed into a linear response with the transformation:

$$p' = \log_e\left(\frac{p}{1-p}\right) \tag{2}$$

hence

$$p' = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_1 X_3 \tag{3}$$

The transformation (Eq. (2)) from the curvilinear response (Eq. (1)) to a linear function (Eq. (3)) is called a logit or logistic transformation. The transformed function allows linear regression to estimate each β_i. Since each of the observations is a cell, the final result is a probability score (p) for each cell.

For the Ipswich watershed, the independent variables are elevation, slope and distance from residential areas of 1971. Elevation should be important in this landscape that is prone to flooding. Slope is important to developers who want to minimize landscaping costs. Distance from residential areas should show how residential development tends to be clustered.

These three variables are the only ones available in digital form that show the landscape of 1971. Other factors such as roads and land prices are clearly important, but those data are not available for 1971.

Slope and elevation are selected as explanatory variables because they are stable through time, hence it is valid to use elevation, slope and residential areas of 1971 as variables to explain and to predict deforestation after 1971.

Distance from residential areas in 1971 is used to estimate the regression parameters, however to predict the suitability for deforestation between 1985 and 1991, the estimated parameters are applied to distance from residential areas in 1985. The model assumes the relation between deforestation and distance from residential areas is the same from 1971 to 1991. Fig. 3 shows the estimated logistic relationship between suitability for deforestation and distance from existing residential areas. The final output is a map of suitability for deforestation for the entire watershed.

The model uses the unbiased estimates from the logistic regression, because one of the major purposes is to compare the logistic regression approach to the MCE approach, where each approach uses the same variables. Spatial auto-correlation in the error terms of the logistic regression causes bias in the standard errors of the parameter estimates. Therefore it is potentially hazardous to make strong conclusions concerning the statistical significance of the parameters.

2.2.2. MCE with two sizes of bins

MCE is a decision support tool used in GIS. In this paper, the model must decide the location of future deforestation. Similar to logistic regression, MCE produces a suitability map to show likelihood of deforestation. The suitability map in MCE is created based on one or more independent variables. The independent variables are again elevation, slope and distance from residential areas of 1971. Each independent variable map is reclassified as a real number that indicates the relative suitability at a value of the independent variable. The original values (e.g., elevation, slope, and distance) are reclassified to suitability values using fuzzy membership functions (Eastman and Jiang, 1996). Fig. 3 shows how the fuzzy set membership functions are calibrated empirically.

The first step is to calibrate the fuzzy membership function is to express each factor as a map of categories. The three factors are expressed as real numbers, therefore each real number is placed into a bin in the same manner that a histogram places real numbers into bins. Then each bin is assigned a suitability value equal to the proportion of cells in the bin that experienced deforestation from 1971 to 1985. In this manner, each factor map is transformed

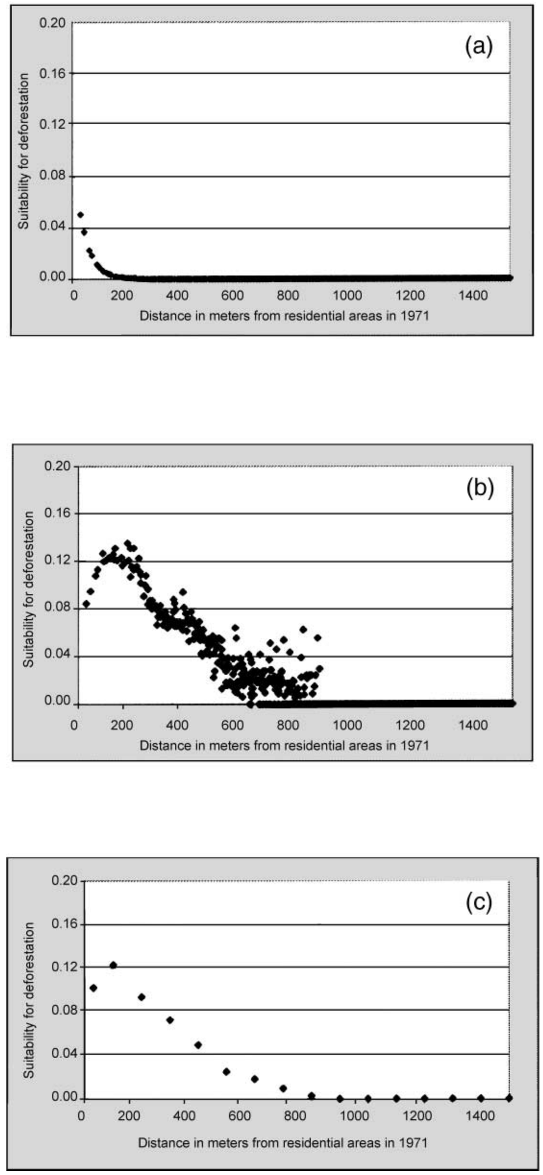

Fig. 3. Three methods to show the relationship between the percent of area that underwent deforestation from 1971 to 1985 versus the distance from residential areas of 1971. The fuzzy membership functions are based on logistic regression in the first plot (a), MCE with 30 m bins in the second plot (b), and MCE with 100 m bins in the third plot (c).

into a suitability map. The suitability maps of each of the three factors are combined by overlaying and averaging the three suitability values on a cell-by-cell basis using a weighted linear combination procedure in which each explanatory variable is weighted equally.

Two bin sizes, small and large, are compared. The small bin method places distance to existing residential areas in bins of 30 m increments, and the large bin method uses 100 m increments. For elevation, the small bins are 1 m increments and the large bins are 50 m increments. For slope, both small and large bins are 1 degree increments. Larger bins tend to make smoother fuzzy relationships (Fig. 3 (c)).

2.2.3. Constraints

A constraint is a physical or legal characteristic of a cell that prevents the cell from being deforested. Therefore, a constraint variable is a Boolean variable that indicates that the cell either is or is not available for deforestation. There are two constraints for prediction of deforestation. First, cells that are non-forest in 1985 are obviously not candidates for deforestation between 1985 and 1991. Second, cells that are protected legally from deforestation are assigned the absolute lowest suitability value in the final suitability maps. These constraints are applied to the suitability maps created by logistic regression and the two MCE methods. The legal constraints derive from a variety of laws concerning zoning, proximity to wetlands, proximity to rivers and other factors. Either the town or the state can change these laws for environmental or cultural reasons. The constraints are those that are available in digital form.

2.2.4. Spatial dependence

Spatial dependence is the phenomenon where the value at a location is a function of the value of neighboring locations. Spatial dependence is created explicitly in the suitability maps by using a spatial filter with kernel size of 35 acres, which is the average patch size of the areas deforested from 1971 to 1985. Thus, the suitability value of each cell is changed to a weighted average of the value of itself and the values of the cells in the surrounding 35 acres. This additional step is necessary because the logistic and MCE methods treat cells as independent observations without considering the characteristics of neighbor cells.

2.3. Calibration of quantity of deforestation

Independent of the suitability maps, the model predicts the quantity of cells to convert from forest to non-forest. Quantity of deforestation in each town is estimated using various methods: multivariate regression using socioeconomic factors, bivariate regression using total amount of forest of 1971, and linear extrapolation over time.

For the multivariate linear regression by town, the dependent variable is the annual amount of area deforested from 1971 to 1985, and the independent variables of 1971 are population density, median household income, percentage of the population with college degree, and poverty level.

In the bivariate linear regression by town, again the dependent variable is the annual amount of area deforested from 1971 to 1985 by town. The regression uses as its explanatory variable the amount of standing forest in 1971. Then the estimated equations with the information of 1985 are used to predict the quantity of deforestation from 1985 to 1991.

The final method to predict quantity of deforestation from 1985 to 1991 is to extrapolate linearly over time for each town. Hence the model assumes that annual quantity of deforestation from 1971 to 1985 stays constant through 1985 to 1991.

2.4. Prediction and validation

A simulation run combines a suitability map and a predicted quantity of deforestation to predict a map of deforestation between 1985 and 1991. The model simulates land-use conversion for the predicted quantity of cells that have highest relative suitability. The model can allocate deforestation to the largest suitability values within each town, or the model can allocate deforestation to the largest suitability values within the entire watershed.

For validation, the model's output is compared to a map of forest areas that changed from 1985 to 1991. There is a two-stage process of validation. The ROC is an index of discrimination accuracy that can validate suitability maps independently of any specified quantity of deforestation (Swets, 1988; Pontius and Schneider, 2001). The suitability map with the highest ROC is used for final simulations.

In the final simulations, the model combines the best suitability map with the various predicted quantities of deforestation. Kno, which is a variation of the standard kappa index of agreement, gives the overall accuracy of a simulation run. Also, two indices, Klocation and Kquantity, validate the simulation's ability to predict location and quantity, respectively. Kno, Klocation and Kquantity are equal to 1 when the simulation's success rate is perfect, and are equal to 0 when the simulation's success rate is equivalent to that due to chance (Pontius, 2000).

3. Results

3.1. Location of deforestation

Six maps of suitability of deforestation were produced. The best suitability map uses the MCE method with large bins and a spatial filter (Fig. 4). The different methodologies used to create the suitability maps yield minor differences in term of ROC, which range from 0.65 to 0.70 (Table 1). In all cases, the spatial filter improves the suitability maps. Among methodologies, the MCE methods are consistently superior to the logistic method. In logistic regression, all the coefficients are significant due to the large number of degrees of freedom. The odds ratio is largest in magnitude for slope, followed by distance from residential areas and finally elevation (Table 2). Recall that the observations are the cells in the watershed map.

Table 1
ROC to validate six maps of suitability for deforestation. Maps are calibrated by logistic regression, MCE with small bins, or MCE with large bins. Each map either uses or does not use a spatial filter

	No spatial filter		Spatial filter	
	Small bins	Large bins	Small bins	Large bins
MCE	0.6745	0.6743	0.6770	0.7044
	No spatial filter		Spatial filter	
Logistic	0.6506		0.6679	

Table 2
Logistic regression parameters estimated from actual deforestation between 1971 and 1985, hence used to predict deforestation from 1985 to 1991. Each observation is one of 151,071 grid cells. Variables are distance from developed areas, elevation and the logarithm of slope

Variable	Coefficient	t-Value	p-Value	Odds ratio
Log-slope	−0.246	−7.008	0.000	−0.218
Distance	−0.027	−2.383	0.028	−0.026
Elevation	0.004	22.947	0.000	0.004

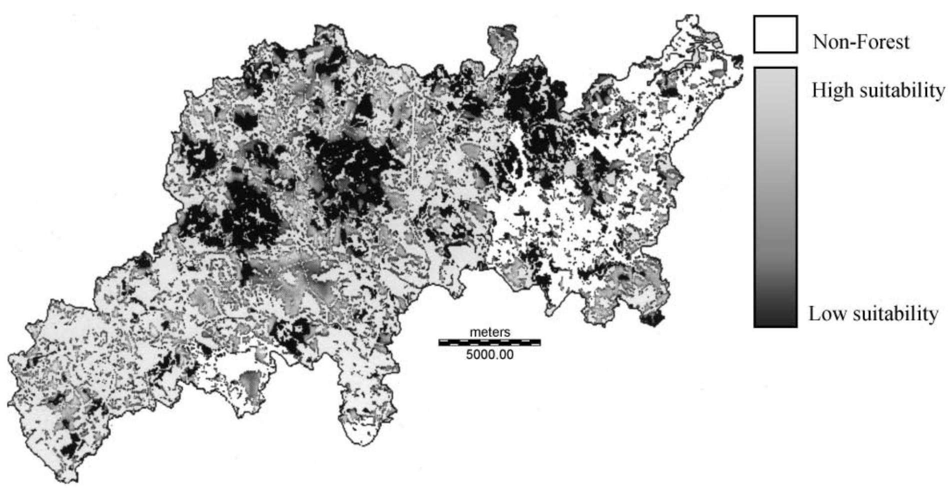

Fig. 4. Suitability map from MCE with large bins and a spatial filter.

Table 3
Variations of kappa index of agreement to validate deforestation from 1985 to 1991. Each of the six cases relates to a different specification of quantity as described and shown in Fig. 5

Kappa variation	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Kno	93.1	93.1	92.8	92.9	92.5	92.8
Klocation	8.1	8.2	7.8	9.6	7.8	11.6
Kquantity	100.7	100.7	100.4	100.4	100.0	100.0

3.2. Quantity of deforestation

Regression with town level data has poor to fair ability to explain or predict quantity of deforestation. Multivariate regression with socioeconomic variables explains only 40% of the variation in forest loss from 1971 to 1985, and none of the independent variables are significant. Therefore the multivariate regression equation is not used to predict the quantity of deforestation from 1985 to 1991. The amount of forest area in 1971 explains 81% of the variation, with more forest area being associated with significantly more deforestation (p-value $= 0.00$). When this bivariate regression equation predicts the deforestation from 1985 to 1991, the quantity predicted is 1.7% of the forested area of 1985. However, simple linear extrapolation over time predicts that 1.9% of the forested area of 1985 becomes deforested. In reality, 2.1% of the forested area of 1985 became deforested by 1991.

Table 3 shows three variations of the kappa index of agreement for each of the simulations of deforestation from 1985 to 1991. Four of the simulations are based on a slightly inaccurate prediction of the quantity of deforestation. For comparison, Table 3 shows also simulation cases 5 and 6 that have the correct quantity of deforestation. Each run uses the suitability map with the best ROC, which is the one produced by the MCE method with large bins and a spatial filter. The results from the simulated maps show that the model is more accurate at predicting the quantity than the location of deforestation. In general, the maps show high values of Kno, high values of Kquantity and low but positive values of Klocation (Table 3).

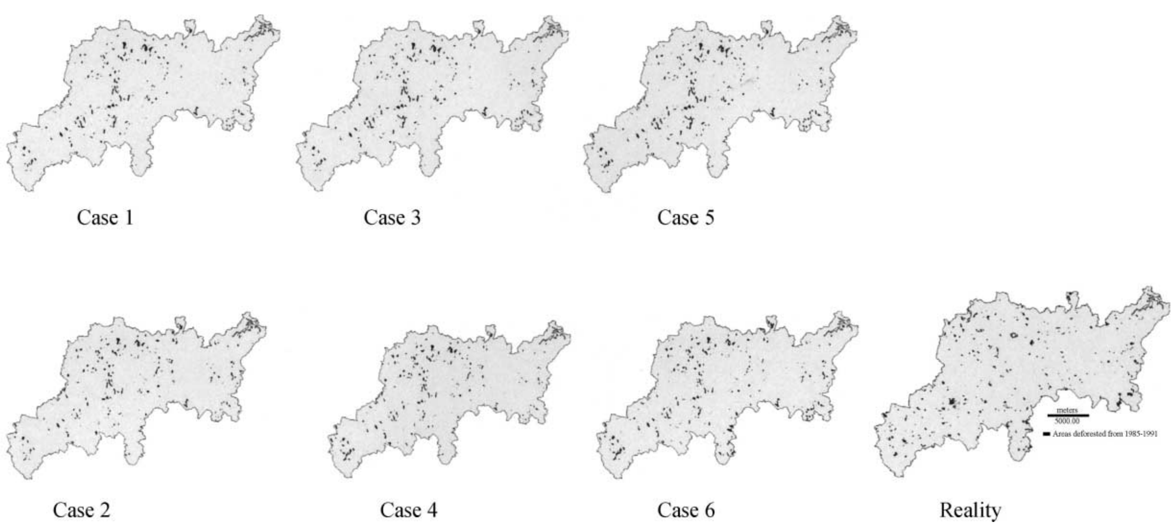

Fig. 5. Predicted deforestation from 1985 to 1991 for six cases of quantity specification. For cases 1 and 2, quantity is based on a regression equation and the amount of forest area in 1985. For cases 3 and 4, quantity is based on a linear extrapolation over time of deforestation from 1971 to 1985. For cases 5 and 6, quantity set equal to reality. In cases 1, 3 and 5, deforestation is allocated by watershed. In cases 2, 4 and 6, deforestation is allocated by town. The map of reality is the actual deforestation from 1985 to 1991. In all maps, deforested areas are in black.

However, it is important to validate also by visual inspection of the spatial pattern of deforestation in the predicted maps (Fig. 5). Variations in the spatial pattern are observed when quantity of deforestation is allocated at different scales. Some models allocate quantities of land-cover change to the study area as a whole, however in the case of the Ipswich watershed, the study area composed of several towns, each one with a particular set of regulations that affect deforestation. Allocating quantities of change to each town provides a better representation of the current spatial pattern of deforestation (Fig. 5). When the quantity is allocated by the watershed, most of the predicted deforestation occurs in the west. When the quantities are allocated by towns, the predicted deforestation is more accurately dispersed (Fig. 5).

4. Discussion

4.1. Quantity versus location

Quantitative measurements of validation, such as the variations of the kappa index of agreement, are very useful in accounting independently for the accuracy in quantity and location of simulated deforestation. This insight can help scientists to decide whether to dedicate energy to improve a simulation's ability to specify quantity versus location (Pontius, 2000).

For quantities, simple extrapolation through time gives better predictions than more complicated explanatory models. A set of variables that are related to complex social processes should be able to explain quantity of deforestation. However, proxies for these variables such population and income were not significant at explaining deforestation at the town level. Even the amount of existing forest is better at explaining and predicting the amount of deforestation than are general socioeconomic factors. If forest area change over the last half century is a guide, then the linear extrapolation should be accurate over a decade or two. In the last 50 years, the rate of forest loss has been fairly stable, with only a slight increase in the rate of forest loss recently. However, during the first half of the 1900s, forest area was increasing as a result of abandonment of agricultural land, so over periods of several decades the rate can change sign.

Clearly, the challenge in the Ipswich watershed is to improve the model's ability to predict location, i.e., to improve the suitability map. From a pure mathematical point of view, it is difficult use empirical data to capture and validate a signal of deforestation when the spatial pattern is disperse (non-clustered) and when only a small proportion of deforestation exists in the watershed. Fortunately, the ROC reveals the quality of the suitability map over a wide range of amounts of land-use change, so the ROC can validate the suitability map independent of a predicted quantity of change. ROC results show that large suitability values are more closely associated with change than are small suitability values.

The MCE bin methods yield larger ROCs than the logistic regression methods because the empirical relationship between deforestation and elevation, slope and distance from residential areas is not necessarily smooth or monotonic, as is assumed by the logistic approach (Fig. 2). Under the empirical MCE bin method, the fuzzy membership function reflects the relationship found in the data used to calibrate the model. The smoothing effect of logistic regression can lead to loss of important information that the regression treats as noise. At the other extreme, a method that uses extremely small bins can pay too much attention to the noise, hence it can miss the signal. The ROC results show that a choice of large bins is best for the application to the Ipswich watershed. This is an important result because many researchers assume that logistic regression is more robust because it is more often taught as the standard method. However, it is just as empirical as MCE analysis because both procedures are driven by the data.

There is a link between scale and the specification quantity and location. The grid cell is the finest scale, which is analyzed in terms of location. The model is slightly better than random at modeling at this finest scale. The town is a medium scale and has components of both location and quantity. The model specifies the quantity in each town, but when it allocates each quantity to a town, it specifies the general location within the watershed. The watershed is the coarsest scale, which is analyzed in terms of quantity only. The model performs well at this coarsest scale. Other models have also shown increasing performance as scale becomes coarser (Kok et al., 2001). This phenomenon can be understood in terms of quantity becoming relatively more important at coarser scales.

4.2. Explanatory variables

Nevertheless, all the simulations are only slightly better than random at predicting the location of deforestation. A reason could lie in the explanatory variables. Slope and elevation can be important biophysical variables when one tries to predict deforestation in a rural context (Lambin, 1997). In the Ipswich watershed, there is more land change at higher elevations perhaps as a result of concern for flooding, and there is more land change on flat land as developers avoid expensive landscaping on steeper slopes. However, elevation and slope do not have strong explanatory power due to both the small variation in topography and socioeconomic complexity.

Residential development in areas like the Northeastern US could be explained in terms of economic development attracting new corporations and employees, increasing the average income in the area and increasing the demand for larger residential plots (Cadwallader, 1992). The greatest proportion of land by far in the watershed is zoned for residential use. Minimum lot sizes for new homes, regulated by each town, have also increased in recent years. Developers used to be able to build before on $\frac{1}{4}$ to $\frac{1}{2}$ acre lots, but now they must now purchase between 1 and 3 acre lots on which to build a house. With land values rising sharply, lot size increases favor the more affluent classes and lead to a highly dispersed pattern of land-cover change (Burgess, 2000). These changes in town level land-use policy influence the relationship between new development and proximity to existing development, therefore the relationship between those two variables should be updated when new laws are passed.

Predicting actual location of urban sprawl in the Ipswich watershed is challenging because important social processes are specific to each of the 21 towns in the study area. In general, local zoning laws that regulate lot sizes have promoted expansion of suburban development (Leo et al., 1998). These codes prohibit both small lot sizes and mixed use development, however these laws can change quickly and some development occurs despite the laws because developers can obtain special permits from the towns. There were not sufficient resources to collect, compile and digitize this information for each of the 21 towns. Other data are highly relevant but difficult to access such as data that relate to the behavior of local developers, who are the key agents of sprawl (Jackson, 1985). Burgess (2000) interviewed several developers in the watershed. Some of their responses confirm that the model has incorporated some of the important variables. Other responses suggest additional variables that either do not exist in digital form or do not have clearly quantifiable definitions, for example proximity to reputable schools and proximity to cultural events. Additionally, there are other factors that are even less quantifiable with empirical data, such as a cultural predilection for the pastoral ideal promoted by suburban communities, white flight from inner city racial tension and a car-dependent culture (Bollier, 1988). A potential variable of interest is the age of the owner. Both developers and environmental groups know that land becomes available when the owner dies. In Massachusetts, both developers and environmental groups tend to contact elderly owners before they die in order to plan for future land use. A map of the age of owners is not available for privacy reasons.

4.3. Comparison with other approaches

This paper's model does not have complicated dynamics, in contrast to other models such as GEOMOD2 (Pontius et al., 2001) and CLUE (Veldkamp and Fresco, 1996). GEOMOD2 updates every year the candidates for deforestation, in order to force new deforestation to grow from existing non-forest. The spatial filter and the proximity to existing residential areas factor in the Ipswich example captures some of this effect, and the Ipswich model is nearly equivalent to a non-dynamic run of GEOMOD2. GEOMOD2 performs just as well when it runs in a non-dynamic mode as when it runs in a dynamic mode. So it is not clear how the model's lack of dynamics influences accuracy.

The CLUE model links the quantity of change to the location of change. This approach would be potentially fruitful in the Ipswich watershed because the quantity of deforestation at the town level is related to the quantity of forest area. Therefore, location of change within the watershed could be related to the quantity of future change within the towns over long periods of time.

Other dynamic models are based primarily on economic theory. Irwin and Geoghegan (2001) have created a model that is both theoretical and dynamic, in that it updates probabilities of conversion every year, based in part on land and housing prices. These models require a different database for calibration because theoretically based economic models assume that human behavior is driven primarily by prices. Clearly, developers are out to make money, however local politics can cause variations from this general principal. Developers in the Ipswich watershed say that they will avoid even potentially profitable ventures in towns where they have poor relations with the town planning boards because boards can cause complications in permitting. From the point of view of economics, this complication in permitting can be viewed as a cost. However, to model these costs would require data on the personal relationships between developers and town planners. Such data are probably what are required to predict accurately the precise locations of land change at the watershed scale over the period of a few years.

5. Conclusions

In this attempt to predict land-use change in the Ipswich watershed, the most successful component of model is its prediction of the quantity of deforestation between 1985 and 1991. For this component, the most accurate approach proved to be the simplest approach, i.e., linear extrapolation over time by town. The model's prediction of the exact locations of deforestation in the watershed was slightly better than the success rate expected by random chance. Both the high success in predicting the quantity and the lower success in predicting location are related to the fact that only 2% of the watershed became deforested from 1971 to 1991. In addition, the difficulty in predicting location is related to the fact that the patches of deforestation are scattered evenly across the landscape, and some of the most important variables are not readily available in digital form. Nevertheless, the best predictor of location of deforestation (ROC = 70%) is a suitability map that uses a spatial filter and MCE. The locations that are most threatened are those that are unprotected, near existing residential development and in towns where the demand for new residential development is high.

Acknowledgements

We thank the editors of this special issue for inviting this paper. For comments on a draft of this paper, we thank two anonymous reviewers. For financial support we thank the National Science Foundation (NSF) through the Water and Watersheds program grant DEB-9726862. Additional contributors include the Jesse B. Cox Charitable Trust, the Sweetwater Trust, and NSF's Long Term Ecological Research program OCE-9726921. Also, we thank Stephen Menard, our colleagues at the George Perkins Marsh Institute of Clark University and at the Center for Integrated Studies of Global Environmental Change of Carnegie Mellon University, with which this work is increasingly tied intellectually and programmatically.

References

Bollier, D., 1988. How Smart Growth Can Stop Sprawl: A Fledgling Citizen Movement Expands. Essential Books, Washington, DC.

Bradshaw, T., Muller, B., 1998. Impacts of rapid urban growth on Farmland conversion: application of new regional land use policy models and geographical information systems. Rural Sociol. 63 (1), 1–25.

Burgess, P., 2000. Conservation and development in conflict: using GIS to improve land management strategies in the Ipswich Watershed. Master's Project in International Development. Clark University, Worcester, MA.

Cadwallader, M., 1992. Migration and Residential Mobility: Macro- and Micro-approaches. University of Wisconsin Press, Madison, WI.

Eastman, J.R., 1999. Idrisi 32: User's Guide. Clark University, Worcester, MA.

Eastman, R., Jiang, H., 1996. Fuzzy measures in multi-criteria evaluation. In: Proceedings of the Second International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Fort Collins, CO.

Eastman, J.R., Jin, W., Kyem, P.A.K., Toledano, J., 1995. Raster procedures for multi-criteria/multi-objective decisions. Photogrammetric Eng. Remote Sensing 61 (6), 539–547.

Foster, D.R., 1995. Land-use history and four hundred years of vegetation change in New England. In: Turner, B.L. (Ed.), Principles, Patterns and Processes of Land Use Change: Some Legacies of the Columbian Encounter. SCOPE Publication. Wiley, New York, pp. 253–319.

Hall, C.A.S., Tian, H., Qi, Y., Pontius, G., Cornell, J., 1995. Modelling spatial and temporal patterns of tropical land use change. J. Biogeogr. 22 (4/5), 753–757.

Hopkinson, C.S., Vallino, J.J., 1995. The relationship among man's activities in watersheds and estuaries: a model of runoff effects on patterns of estuarine community metabolism. Estuaries 18, 598–621.

Irwin, E.G., Geoghegan, J., 2001. Theory, data, methods: developing spatially explicit economic models of land use change. Agric. Ecosyst. Environ. 85, 7–24.

Jackson, K.T., 1985. Crabgrass Frontier: The Suburbanization of the United States. Oxford University Press, Oxford.

Kasperson, J.E., Kasperson, R.E., Turner II, B.L., 1997. Regions at risk. Exploring environmental criticality. Environment 39, 4–15.

Kok, K., Farrow, A., Veldkamp, A., Verburg, P., 2001. A method and application of multi-scale validation in spatial land use models. Agric. Ecosyst. Environ. 85, 223–238.

Lambin, E.F., 1994. Modelling deforestation processes: a review. European Commission, Luxemburg.

Lambin, E.F., 1997. Modelling and monitoring land-cover change processes in tropical regions. Prog. Phys. Geogr. 21 (3), 375–393.

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., Vogel, C., 1999. Land-use and land-cover change (LUCC): implementation strategy. IGBP Report 48, IHDP Report 10.

Leo, C., Beavis, M.A., Turner, R., 1998. Is urban sprawl back on the political agenda? Local growth, control, regional growth management and politics. Urban Affairs Rev. 24 (2), 179–212.

MassGIS, 1999. Land use map. Executive Office of Environmental Affairs, Boston, MA.

Pontius Jr., R.G., 1994. Modeling tropical land use change and assessing policies to reduce carbon dioxide release from Africa. Graduate Program in Environmental Science. SUNY-ESF, Syracuse, 177 pp.

Pontius Jr, R.G., 2000. Quantification error versus location error in comparison of categorical maps. Photogrammetric Eng. Remote Sensing 66, 1011–1016.

Pontius Jr., R.G., Schneider, L., 2001. Land-use change model validation by an ROC method. Agric. Ecosyst. Environ. 85, 239–248.

Pontius Jr., R.G., Cornell, J.D., Hall, C.A.S., 2001. Modeling the spatial pattern of land-use change with GEOMOD2: application and validation. Agric. Ecosyst. Environ. 85, 191–203.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293.

Turner II, B.L., 1994. Global land-use/land-cover change: towards an integrated study. Ambiology 23 (1), 91–95.

Veldkamp, A., Fresco, L.O., 1996. CLUE-CR: an integrated multi-scale model to simulate land use change scenarios in Costa Rica. Ecol. Model. 91, 231–248.

Vitousek, P.M., Mooney, H.A., Lubchenco, J., Melillo, J.M., 1997. Human domination of earth's ecosystems. Science 277, 494–500.