Landscape Ecology vol. 10 no. 3 pp 177-189 (1995)
SPB Academic Publishing bv, Amsterdam

Forecasting the spatial dynamics of gypsy moth outbreaks using cellular transition models

Guofa Zhou[1,2] and Andrew M. Liebhold[1]*
[1] *USDA Forest Service, Northeastern Forest Experiment Station, 180 Canfield St., Morgantown, WV 26505, USA;* [2] *Department of Applied Mathematics, Branch Campus of Beijing University, Beijing 10083, People's Republic of China*

Keywords: Markov chain, cellular automata, *Lymantria dispar*, Lepidoptera: Lymantriidae

Abstract

A series of cellular transition probability models that predict the spatial dynamics of gypsy moth (*Lymantria dispar* L.) defoliation were developed. The models consisted of four classes: Simple Markov chains, Rook's and Queen's move neighborhood models, and distance weighted neighborhood models. Historical maps of gypsy moth defoliation across Massachusetts from 1961 to 1991 were digitized into a binary raster matrix and used to estimate transition probabilities. Results indicated that the distance weighted neighborhood model performed better then the other neighborhood models and the simple Markov chain. Incorporation of interpolated counts of overwintering egg mass counts taken throughout the state and incorporation of historical defoliation frequencies increased the performance of the transition models.

Introduction

Gypsy moth, *Lymantria dispar* L., is one of the most destructive exotic organisms in North America. Since it was introduced to the northeastern U.S. in 1869, it has expanded its range over 1,000,000 km^2 (Liebhold *et al.* 1992), and it is likely that gypsy moth populations will ultimately invade most of North America. In many areas, epidemics are common and resultant defoliation can cause substantial ecological and economic effects.

As with other pests, an ability to predict the occurrence of gypsy moth outbreaks is critical to the management of these populations. Unfortunately, gypsy moth populations tend to be erratic through space and time and outbreaks are consequently very difficult to forecast. Previously developed models for predicting gypsy moth outbreaks have largely been based on the relationship between pre-season egg mass density and defoliation within individual stands (Gansner *et al.* 1985; Williams *et al.* 1991; Liebhold *et al.* 1993). These models often do not perform reliably (Liebhold *et al.* 1993).

Previously developed models for predicting gypsy moth outbreaks have largely ignored the spatial component of outbreak dynamics even though at least some research has indicated that spatial processes may be important to gypsy moth dynamics. Campbell (1973, 1976) found that outbreaks were more persistent between years when insect densities ranged widely among subpopulations in the region and conversely that when numerical variability was minimal among subpopulations, outbreaks were likely to decline. Liebhold and Elkinton (1989) showed that the existence of nearby areas of defoliation increased the probability of an outbreak developing in an area and that defoliation begins at various locations and then spreads outward in all

* Corresponding author.

directions. Valentine and Houston (1979) and Campbell (1976) hypothesized that this spread of outbreaks is caused by immigration of larvae from epicenters or foci. However Liebhold and McManus (1991) concluded that the apparent spread of outbreaks could not be caused by larval migration but was instead an artifact of the distribution of susceptible forest types.

Transition models, especially Markov chains, have been used with some success to model the dynamics of various ecological systems (Isaacson and Madsen 1976; Pielou 1962, 1977; Usher 1981; Badger *et al.* 1987). For example, Kemp (1987) used a two-state Markov chain to model the probability of rangeland grasshopper outbreaks in Montana. Most of these applications have not incorporated information about conditions in spatially adjacent areas into transition probabilities. Lippe *et al.* (1985) used Markov models to quantify ecological succession and concluded that transition probabilities are affected by conditions in adjacent areas and suggested an extension or modification of the Markov model to incorporate spatial influence. Turner (1987, 1988) developed spatially explicit transition models to simulate land use change in the Georgia piedmont that incorporated these types of modifications. Under her models, land use status in the four or eight nearest cells influenced transition probabilities.

Most of the previously developed, spatially-explicit transition models can be considered types of 'cellular automata' models (Preston and Duff 1984). In this paper, we provide a brief description of cellular automata and how we used this approach to build spatially explicit transition models to predict gypsy moth outbreaks. We present various models that differ in the way that spatial influence is simulated and we evaluate the usefulness of incorporating counts of overwintering egg mass populations. The performance of the various models are compared and discussed.

Cellular Automata

Wolfram (1984) broadly defined cellular automata (CA) as 'systems of cells interacting in a simple way but displaying complex overall behavior.' Phipps (1992) refined this broad definition by describing three characteristics common to all CA: 1) the models act on a 1-, 2-, or 3-dimensional matrix of cells; 2) at any point in time, a given cell may exist in any one of k possible discrete states; 3) at every time step, the state of each cell is updated according to a transition rule that takes into account the state of neighboring cells. CA were first conceived over 40 years ago (von Neumann 1966) but were never extensively explored because computer power was limiting (Preston and Duff 1984). Recently, there have been several applications of CA for modeling a variety of physical processes but there are still relatively few ecological applications.

In any CA, each cell can take on k possible values (states) and is updated in discrete time steps according to a rule f that depends on the value at sites in some neighborhood around it. The value, V, of a variable at position x and time step $t+1$ is described in a one-dimensional cellular automata with a rule that depends only on values of nearest neighbors (Vichniac *et al.* 1986):

$$V_{x,t+1} = f(V_{x,t}, V_{x-1,t}, V_{x+1,t}) \tag{1}$$

There are several possible lattices and neighborhood structures for two-dimensional cellular automata (Birman and Trebin 1985; Packard and Wolfram 1985; Langton 1990; Wootters and Langton 1990). For square cellular matrices, adjacency can be described according to Rook's move and Queen's move definitions (Cliff and Ord 1973). In a Rook's move neighborhood the cellular automata uses:

$$V_{x,y,t+1} = f(V_{x,y,t}, V_{x-1,y,t}, V_{x+1,y,t}, V_{x,y-1,t}, V_{x,y+1,t}) \tag{2}$$

and a Queen's move definition uses:

$$V_{x,y,t+1} = f(V_{x,y,t}, V_{x-1,y,t}, V_{x+1,y,t}, V_{x,y-1,t}, V_{x,y+1,t}, V_{x+1,y+1,t}, V_{x-1,y+1,t}, V_{x+1,y-1,t}, V_{x-1,y-1,t}) \tag{3}$$

where x,y represent the coordinates of a cell.

Many CA adopt one of many neighborhood summation rules, in which the value at a site depends only on the sum of the values in the neighborhood (Ripley 1981; Packard and Wolfram 1985).

For example in the Rook's move neighborhood definition a summation rule would be:

$$V_{x,y,t+1} = f(V_{x,y,t} + V_{x-1,y,t} + V_{x+1,y,t}, V_{x,y-1,t} + V_{x,y+1,t}) \quad (4)$$

These types of rules obviously assume that neighborhood effects are not influenced by direction. Many attempts to use CA for modeling biological growth processes, such as 'the game of life' have utilized Conway's transition rules (Gardner 1971). This summation rule consists of 3 transition paths under 0–1 state space and Rook's-move neighbor definition and mimics the essential spatio-dynamic features of life processes; colonies invade unoccupied spaces and they die as a result of crowding.

These types of CA have been extensively used to model physical processes and to a lesser extent they have been used to model biological processes. Considerable attention has been focused on the chaotic and other nonlinear behavior of these models when they are allowed to run through thousands of iterations (Langton 1990; Wolfram 1983, 1984).

The models we present here differ from most previously developed CA in several ways. First, our study focuses on the use of these models for prediction across a single time step, rather than focusing on complex behaviors over thousands of iterations. Secondly, most previous CA models have used rules that incorporate the state of a cell at time $t+1$ as a function (rule) only of its own state and the states of immediate neighbors at time t. In our study we expanded the neighborhood definition beyond adjacent cells. Finally, while most CA have been deterministic, in that the dynamics of the state at any given cell is uniquely determined by the transition function, in our model we applied a stochastic transition function.

Data description

A geographic information system (GIS), IDRISI (Eastman 1987), was employed to assemble and collate gypsy moth data. IDRISI is a raster-based (grid cell) GIS for capturing, storing, analyzing and displaying geographical data. The study area was defined using a base-map of Massachusetts county boundaries that was generated from latitude and longitude coordinates projected using the azimuthal equal-distant projection (Snyder 1987). This projection conserves true distance linearly from a designated point (Boston, Massachusetts). A 2 × 2 km grid cell size was selected as standard for all map layers in the GIS. Each map layer comprised 198 by 93 cells. The grid size was selected because it represented the minimum dependable spatial resolution of the available defoliation maps.

The Massachusetts Department of Environmental Management monitored gypsy moth defoliation annually from 1961–91 in all parts of the state using maps sketched during a series of low level reconnaissance flights in late July when defoliation is at its peak. Thirty percent (30%) defoliation is considered the lower threshold for detection from the air. In situations where there is doubt as to the cause of the defoliation, ground checks for the presence of gypsy moth life stages are made. Initially the aerial sketch maps were overlaid on standard U.S. Geological Survey (1:24,000) topographical maps. Subsequently a composite mosaic map was generated for the entire state at 1:760,000 scale. Mapping processes may vary from region to region and year to year in these sketch maps, possibly resulting in systematic and non-systematic data errors (Talerico 1981; Chrisman 1987). The likely presence of these errors dictated the coarse spatial resolution of the digital representation these maps used in this study (2 × 2 km rasters).

To create a uniform set of geographically-referenced defoliation data, the composite defoliation maps for the period 1960 to 1991 were first transferred to mylar stable-base sheets. At least four geo-reference points, on clearly recognizable intersections of county boundaries, were accurately located. The prepared maps were then scanned using a digital scanner set at 150 dots per inch resolution. Binary TIFF files from the scanner were converted to IDRISI map layers. The transformation of each map layer to a common base map resolution and projection was achieved through a 'rubber-sheeting' procedure. In transforming maps of various scales and projections, IDRISI resamples each scanned defoliation image to match the location of the four geo-reference points on the

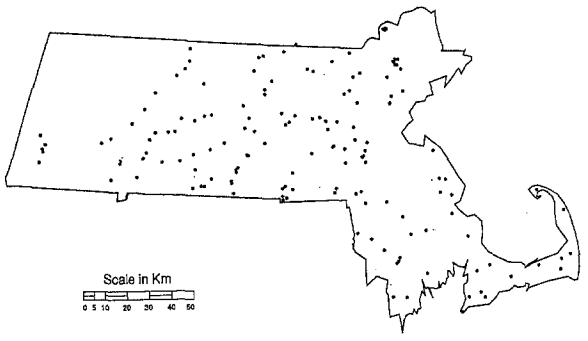

Fig. 1. Distribution of eggs mass sampling points in Massachusetts.

base map (Eastman 1987). Because defoliation data were coded as either 0 or 1, depending on whether it exceeded the threshold for visual detection (30%), the variable is an 'indicator' or binary variable.

From 1985–90 the Massachusetts Department of Environmental Management counted overwintering gypsy moth egg masses at a network of 150 plots irregularly distributed through out the state (Fig. 1). Plots were located in forest stands that were considered susceptible to gypsy moth defoliation. At each plot, burlap bands were placed around the stem of the 20 closest oaks over 6″ DBH. In the late summer, plots were visited and all egg masses under burlap bands were counted. Plot counts were calculated as mean numbers of egg masses per tree.

Model descriptions

Simple Markov chain model (Model I)

In this model, the state of neighboring cells is not incorporated in the transition function and the model can be described by

$$V_{i,t+1} = f(V_{i,t}). \quad (5)$$

We used a two-state transition function (Parzen 1962) to quantify the transition of cells from 0 (undefoliated) to 1 (defoliated) and from 1 to 1. There are four possible transitions which can occur for each cell from one year to the next (i.e., $0 \Rightarrow 0$, $0 \Rightarrow 1$, $1 \Rightarrow 0$ and $1 \Rightarrow 1$). If V_t is the state (0 or 1) of defoliation at the tth point in time, then we can let

$$P_{ij} = \text{Prob}(V_t = j \mid V_{t-1} = i) \quad i,j \in \{0,1\} \quad (6)$$

be the probability that the state of the process (0 or 1) at the tth year is j, given that it was in state i in year $t-1$. P_{ij} is known as the one-step transition probability because the transition spans a single time step (Bhat 1972). Numerous studies have reported that gypsy moth population densities are highly correlated between successive generations (Campbell 1976; Liebhold and Elkinton 1989). These observations of temporal autocorrelation in gypsy moth population levels suggest that a simple Markov chain model may be useful for modeling outbreak dynamics (Kemp 1987).

The one-step transition probabilities for each cell can be estimated by determining the proportion of times that the system moved from one state to another (*e.g.*, $0 \Rightarrow 0$, $0 \Rightarrow 1$, $1 \Rightarrow 0$ and $1 \Rightarrow 1$) (*i.e.* $P^*_{ij} = N_{ij}/N_i$). In this study, we used the transition frequencies over all cells, in all years to estimate transition probabilities. These estimates can then be represented by P^*_{00}, P^*_{01}, P^*_{10} and P^*_{11}. We estimated P^*_{01} as 7509/170666 = 0.044 and P^*_{11} as 4145/11450 = 0.362. The highest transition probability was P^*_{00}, ($= 1 - P^*_{01} = 0.956$) indicating that nondefoliated cells generally remain in that state (Liebhold and Elkinton 1989). The P^*_{10} was greater than P^*_{11} indicating that defoliation terminates at a faster rate than it persists.

Rook's move neighborhood model (Model II)

Liebhold and Elkinton (1989) and Hohn *et al.* (1993) documented the autocorrelation of gypsy moth defoliation in space and time; defoliation state in one cell is correlated with the state in the same cell and in neighboring cells in the previous year. Therefore *Model I* was modified to take into account the state of immediate neighbors under a Rook's move definition of adjacency (Cliff and Ord 1973). This is the simplest case of a two-dimensional cellular automata (equation 2).

The sum of the V_t of each neighboring cell was calculated and used to estimate transition probabilities as in *Model I* (Table 1)

$$P_{ij}(k) = \text{Prob}(V_t = j \mid V_{t-1} = i \;\&\; L_{t-1} = k) \quad i,j \in \{0,1\};\ k \in \{0,1,2,3,4\} \quad (7)$$

Table 1. Estimated transition probabilities for Rook's move neighborhood model (Model *II*). Numbers in parentheses are the number of observations that were used to estimate probabilities.

Number of defoliated neighbors	Probabilities	
	$0 \Rightarrow 1$	$1 \Rightarrow 1$
0	0.032 (144587)	0.249 (743)
1	0.275 (5632)	0.350 (998)
2	0.366 (1666)	0.386 (1439)
3	0.370 (319)	0.403 (1706)
4	0.343 (105)	0.360 (5673)

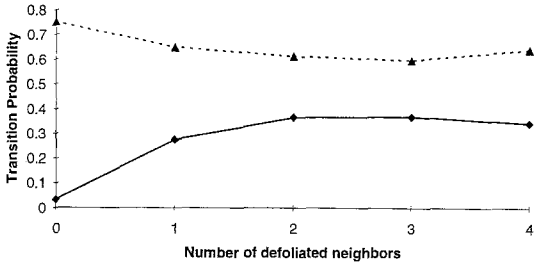

Fig. 2. Transition probabilities under the Rook's move neighborhood model. Solid line: P^*_{01}; dotted line: P^*_{10}.

Table 2. Estimated transition probabilities for Queen's move neighborhood model (Model *III*). Numbers in parentheses are the number of observations that were used to estimate probabilities.

Number of defoliated neighbors	Probabilities	
	$0 \Rightarrow 1$	$1 \Rightarrow 1$
0	0.027 (136828)	0.228 (464)
1	0.202 (6321)	0.311 (634)
2	0.300 (2798)	0.367 (482)
3	0.352 (1518)	0.405 (563)
4	0.367 (780)	0.401 (749)
5	0.406 (303)	0.414 (831)
6	0.294 (109)	0.417 (736)
7	0.395 (76)	0.348 (1330)
8	0.300 (67)	0.354 (4510)

in which $L_{t-1} = V_{x,y,t-1} + V_{x-1,y,t-1} + V_{x+1,y,t-1}$ $V_{x,y-1,t-1} + V_{x,y+1,t-1}$ is the number of defoliated neighbors. As the numbers of defoliated neighbors increased P^*_{01} increased and P^*_{10} decreased, except when all 4 neighbors were defoliated, P^*_{10} increased slightly (G test with Williams' correction [Sokal and Rohlf 1981]: $G = 10.48$; Prob. $> G = 0.001$) (Fig. 2). These results confirm the results of Liebhold and Elkinton (1989) that the presence of nearby defoliation apparently increased the probability of a cell becoming defoliated. Liebhold and McManus (1991) concluded that this apparent 'spread' of defoliation was caused by populations simultaneously increasing across regions where susceptible forest types were spatially dependent in their distribution, rather than insects physically spreading *en masse* through a region. The increase in P^*_{10} reflects the phenomenon in which there is a regional collapse of populations once outbreaks are widely present (Campbell 1976; Liebhold and Elkinton 1989; Liebhold and McManus 1991). These collapses are typically associated with epizootics of a nuclear polyhedrosis virus (Elkinton and Liebhold 1990).

Queen's move neighborhood model (Model III)

This model is identical to the Model II described above except the Queen's move (Cliff and Ord 1973) adjacency definition was used. The model incorporated the state of the 8 immediate neighbors in the transition function:

$$P_{ij}(k) = \text{Prob}(V_t = j \mid V_{t-1} = i \;\&\; L_{t-1} = k) \quad i,j \in \{0,1\};\ k \in \{0,1,2,3,4,5,6,7,8\} \qquad (8)$$

in which $L_{t-1} = V_{x,y,t-1} + V_{x-1,y,t-1} + V_{x+1,y,t-1} + V_{x,y-1,t-1} + V_{x,y+1,t-1} + V_{x+1,y+1,t-1} + V_{x-1,y+1,t-1} + V_{x+1,y-1,t-1} + V_{x-1,y-1,t-1}$ is the number of defoliated neighbors.

Transition probabilities were estimated as in Model II (Table 2). The relationship between transition probabilities and numbers of neighboring cells defoliated was similar to that of Model II (Fig. 3).

Weighted mean neighborhood model (Model IV)

Previous analyses of the autocorrelation of gypsy moth defoliation through space and time indicated that autocorrelations tend to exist beyond the 2 km

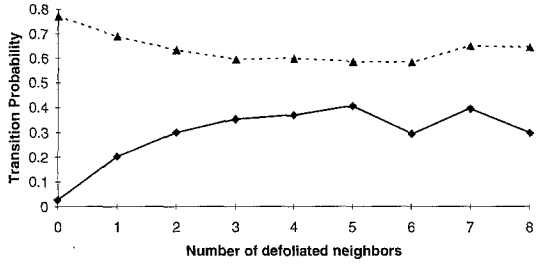

Fig. 3. Transition probabilities under the Queen's move neighborhood model. Solid line: P^*_{01}; dotted line: P^*_{10}.

Table 3. Estimated transition probabilities for weighted average neighborhood model (Model *IV*). Numbers in parentheses are the number of observations that were used to estimate probabilities.

	Probabilities	
Weighted average of neighborhood	0⇒1	1⇒1
0	0.009 (130172)	0.159 (88)
0.00–0.05	0.063 (20682)	0.147 (307)
0.05–0.10	0.153 (7009)	0.214 (373)
0.10–0.20	0.261 (6267)	0.349 (786)
0.20–0.40	0.354 (4670)	0.400 (1550)
0.40–0.60	0.353 (1393)	0.438 (1626)
0.60–1.00	0.324 (463)	0.358 (6720)

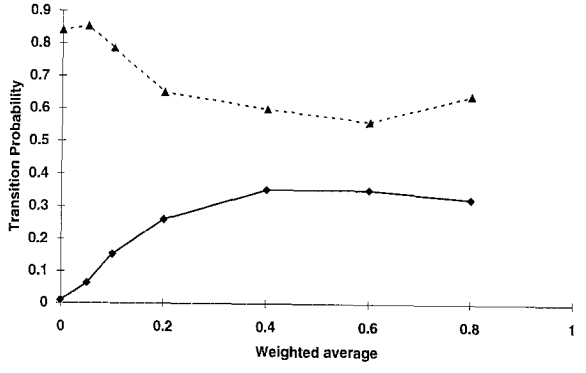

Fig. 4. Transition probabilities under the weighted average model. Solid line: P^*_{01}; dotted line: P^*_{10}.

cell distance, used in this study, but the correlation decreased with distance (Liebhold and Elkinton 1989; Hohn *et al.* 1993). To account for this declining autocorrelation, we used an inverse distance weighted mean of defoliation in the neighborhood to adjust the transition function. The transition probabilities were thus defined:

$$P_{ij}(k) = \text{Prob}(V_t = j \mid V_{t-1} = i \;\&\; L_{t-1} = k) \quad i,j \in \{0,1\};\ k \in \{0-1\} \tag{9}$$

where $L_{t-1} = \dfrac{\sum_{a=x-3}^{x+3} \sum_{b=y-3}^{y+3} V_{a,b,t-1}/D_{a,b,x,y}}{\sum_{a=x-3}^{x+3} \sum_{b=y-3}^{y+3} 1/D_{a,b,x,y}}$ and

$D_{a,b,x,y}$ represents the distance from point a,b to point x,y. The weighted mean is a continuous number between 0 and 1 but we discretized it into 7 classes (Table 3); it has a maximum value when all of the cells in the neighborhood are defoliated, and a minimum value when no defoliation exists in the neighborhood.

Estimated transition probabilities for this model are given in Table 3. The effects of neighborhood defoliation on transition probabilities (Fig. 4) were similar to those seen in the rooks and queen's move neighborhood models (Figs. 2 & 3).

Weighted mean in susceptible areas model (Model V)

Liebhold and Elkinton (1989) noted that across a region, such as the state of Massachusetts, there was considerable variation in defoliation frequency over 30 yr; many locations were never defoliated. This variation is largely explained by the spatial variation in climate and forest composition (Liebhold *et al.* 1994a; Gansner *et al.* 1993). Presumably this variation in the habitat can affect the transition probabilities in the models described above.

We attempted to adjust Model *IV* by estimating transition probabilitiesa for only those cells that had any defoliation from 1961–91. Obviously P^*_{01} will always be zero when the 30 yr defoliation frequency is zero. Also, at any cell where the 30 yr defoliation frequency is zero, P^*_{10} cannot be estimated (but it would never be used in simulations). Estimated transition probabilities are shown in Table 4. The elimination of locations where the defoliation frequency was zero caused P^*_{01} to be higher than when all cells were used (Table 3).

Table 4. Estimated transition probabilities for weighted average neighborhood model after excluding unsuitable cells (those that were never defoliated from 1961–1991) (Model *V*). Numbers in parentheses are the number of observations that were used to estimate probabilities.

	Probabilities	
Weighted average of neighborhood	0⇒1	1⇒1
0.00–0.05	0.021 (104897)	0.171 (327)
0.05–0.10	0.158 (5627)	0.205 (322)
0.10–0.20	0.295 (5002)	0.389 (599)
0.20–0.30	0.384 (2515)	0.439 (624)
0.30–0.40	0.396 (1389)	0.417 (634)
0.40–0.50	0.415 (773)	0.488 (680)
0.50–0.60	0.430 (1108)	0.462 (1407)
0.60–0.70	0.316 (247)	0.421 (729)
0.70–1.00	0.338 (311)	0.347 (5751)

Weighted mean with frequencies model (Model VI)

We modified model *IV* to incorporate geographical variation in forest susceptibility beyond the simple elimination of immune cells as done in model *V*. Instead, we calculated separate transition probabilities for three categories of cells grouped according to their defoliation frequency, f, from 1961–1991: $f = 0$; $0 < f \leq 2$; $f > 2$. Estimated transition probabilities are shown in Table 5. For cells where $f = 0$, $P^*_{01} = 0$. The transition probability, P^*_{01}, was generally greater for cells where $f > 2$ than for cells where $0 < f \leq 2$.

Weighted mean with egg mass counts model (Model VII)

As previously mentioned, egg mass counts are at least loosely correlated with defoliation at the stand level, and a distinct correlation exists at the landscape level (Liebhold et al. 1994b). Liebhold *et al.* 1994b showed that the spatial correlation between defoliation and egg mass counts is greatest when egg mass counts are transformed to an indicator variable (binary: 1,0) such that the transformed variable is zero if there are less than 1 egg mass per tree and 1 if there are 1 or more egg masses per tree. We incorporated $E_{x,y}$, the transformed egg mass count, in model *IV*:

$$P_{ij}(k) = \text{Prob}(V_t = j \mid V_{t-1} = i \;\&\; L_{t-1} = k \;\&\; E_t) \quad i,j \in \{0,1\};\ k \in \{0-1\} \qquad (10)$$

in which L_{t-1} is the same as in equation (9) and E_t is the estimated egg mass count.

Because egg mass counts were only taken in a small fraction of the cells being simulated, it was necessary to interpolate these egg mass counts. Egg mass counts were interpolated using ordinary Kriging (Isaaks and Srivistava 1989). The exact procedures, including variogram models, are described in Liebhold *et al.* (1991). Estimated transition probabilities are shown in Table 6. When egg mass counts were high (above 1 per tree), both P^*_{01} and P^*_{11} were generally greater than when egg mass counts were low. The only exception to this trend was that when neighborhood defoliation was high (weighted mean = 0.4 to 1.0) P^*_{11} was actually lower when egg mass density was high than when egg mass counts were low. This probably reflects a tendency of regional populations to collapse when they are at high population levels and outbreaks are extensive.

The kitchen sink model (Model VIII)

As the name implies, this model incorporated all of the adjustements described above. It essentially was model *VII* modified by restricting the transition probability estimation to only cells where defoliation occurred from 1961–91 as in model *V*. We did not have adequate numbers of observations to incorporate the more detailed variation in 30 yr defoliation frequencies as was done in model *VI*. Again, we assumed that P^*_{01} will always be zero when the 30 yr defoliation frequency is zero. Estimated transition probabilities are given in Table 7.

Simulation results

Data from 1961–90 were used to initialize each of models *I*–*VI* and simulations were run for one time step. This represented a total of 30 × 6 = 180

Table 5. Estimated transition probabilities for weighted average neighborhood model adjusted for 1961–1991 defoliation frequency, f(Model *VI*). Those cells that were never defoliated from 1961–1991 were excluded. Numbers in parentheses are the number of observations that were used to estimate probabilities.

	Probabilities			
	$0 < f \leq 2$		$f > 2$	
Weighted average of neighborhood	$0 \Rightarrow 1$	$1 \Rightarrow 1$	$0 \Rightarrow 1$	$1 \Rightarrow 1$
0.0–0.1	0.0151 (75540)	0.0805 (149)	0.0598 (37885)	0.2205 (508)
0.1–0.2	0.2650 (3208)	0.3439 (157)	0.3558 (2015)	0.3645 (513)
0.2–0.3	0.3827 (1581)	0.4167 (204)	0.4479 (940)	0.4309 (434)
0.3–0.4	0.4122 (922)	0.3730 (185)	0.5000 (504)	0.4325 (474)
0.4–0.5	0.4083 (529)	0.3224 (214)	0.5072 (278)	0.5391 (499)
0.5–0.6	0.3923 (339)	0.2956 (203)	0.5323 (186)	0.5026 (577)
0.6–0.8	0.4428 (271)	0.2921 (558)	0.3929 (112)	0.5213 (1105)
0.8–1.0	0.3310 (142)	0.1804 (2822)	0.5116 (43)	0.4627 (2736)

Table 6. Transition probabilities under the method of weighted mean with eggmass data (Model *VII*). Numbers in parentheses are the number of observations that were used to estimate probabilities.

	Probabilities			
	< 1 egg mass/tree		≥ 1 egg mass/tree	
Weighted average of neighborhood	$0 \Rightarrow 1$	$1 \Rightarrow 1$	$0 \Rightarrow 1$	$1 \Rightarrow 1$
0.00–0.10	0.011 (29219)	0.112 (34)	0.022 (3065)	0.121 (33)
0.10–0.40	0.103 (1745)	0.165 (399)	0.136 (154)	0.280 (25)
0.40–1.00	0.214 (393)	0.315 (1046)	0.393 (28)	0.235 (85)

Table 7. Transition probabilities under the method of weighted mean with eggmass data and excluding the unsuitable cells (Model *VIII*). Numbers in parentheses are the number of observations that were used to estimate probabilities.

	Probabilities			
	< 1 egg mass/tree		≥ 1 egg mass/tree	
Weighted average of neighborhood	$0 \Rightarrow 1$	$1 \Rightarrow 1$	$0 \Rightarrow 1$	$1 \Rightarrow 1$
0.00–0.10	0.006 (10100)	0.067 (90)	0.044 (12802)	0.163 (80)
0.10–0.40	0.039 (502)	0.132 (244)	0.167 (827)	0.245 (67)
0.40–1.00	0.129 (651)	0.177 (1401)	0.325 (126)	0.419 (210)

simulations. Egg mass data were only available from 1985–1990 and therefore models *VII* and *VIII* were only initialized from 6 years. These models were thus used in $6 \times 2 = 12$ simulations. Maps of predicted defoliation probabilities from selected models are shown in Fig. 5.

Even model *I*, which did not incorporate any spatial influence, yielded maps of simulated defoliation probabilities that resembled the spatial distribution of actual defoliation (Fig. 5). The spatial

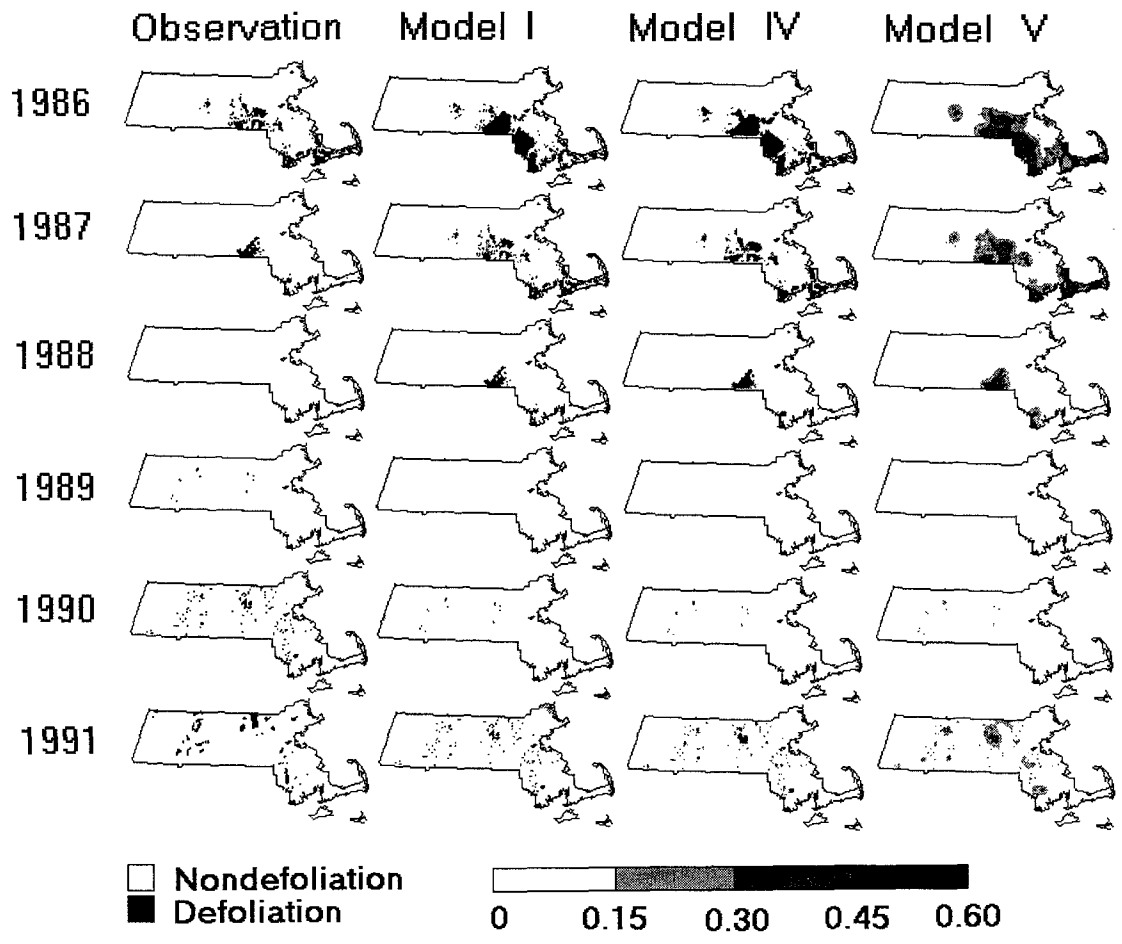

Fig. 5. Maps of observed defoliation and predicted defoliation probabilities from selected transition models from 1986–1991.

distribution of defoliation predicted by model *I* mimicked the patterns of defoliation in the previous year. Thus, the similarity of predicted patterns with observed defoliation patterns in the same year is a result of the temporal dependence or persistence of defoliation (Hohn *et al.* 1993). Incorporation of additional information (neighborhood defoliation state, defoliation frequency, and egg mass counts) increased the visual similarity of the spatial patterns of observed and predicted defoliation (Fig. 5).

The sequences of mean probabilities generated by each model are presented in Figs. 6 and 7. All models shared a similar deficiency: predicted probabilities were too high when defoliation was low and probabilities were too low when populations were high. This is probably a result of estimating transition probabilities from all data, pooled across years. The graphs also illustrate that for most models, predicted probabilities troughed or peaked 1 yr after troughs and peaks in the real data. This behavior is also probably due to the fact that a single set of transition probabilities were derived from all years. Previous research indicates that regional trends in gypsy moth dynamics may be strongly synchronous and affected by regional weather patterns, specific for that year (Miller *et al.* 1989; Liebhold and McManus 1991). Thus, by averaging transitions over all years we may have failed to capture the among year variation in population trends. Preseason egg mass counts were incorporated into models *VII* and *VIII* in an attempt to incorporate the current trend in population levels. Though egg mass data were available for only 6 years, simulated defoliation levels appeared to

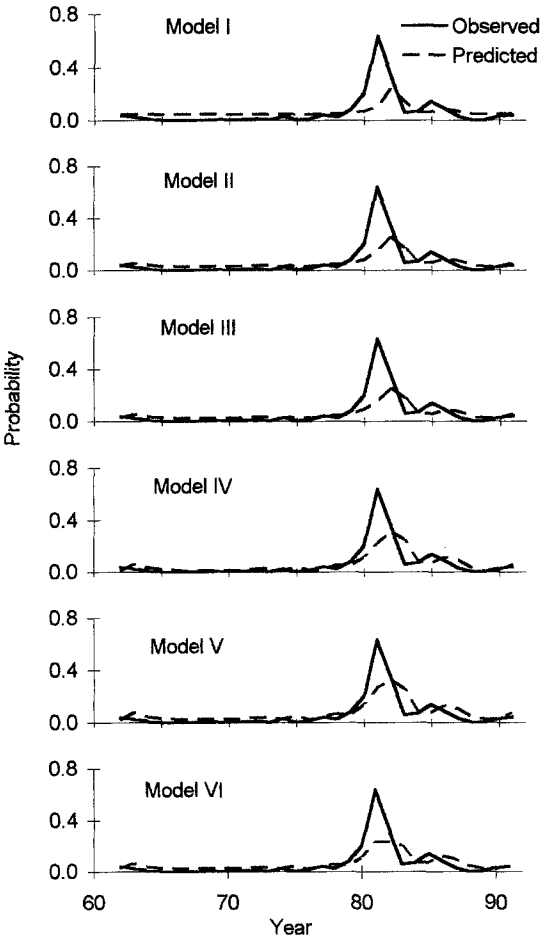

Fig. 6. Graphs of yearly average defoliation and average predicted defoliation probabilities from 1961–1991.

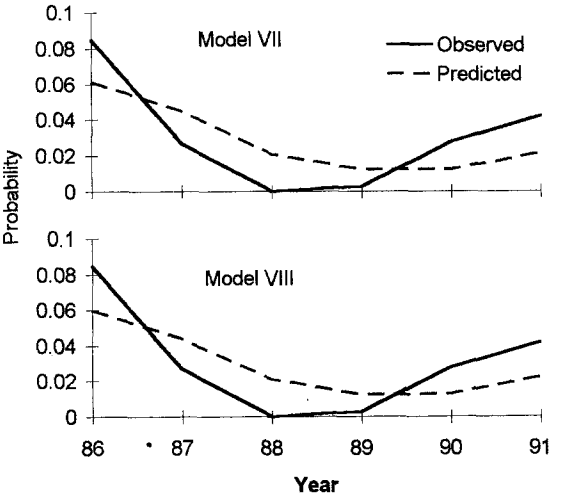

Fig. 7. Graphs of yearly average defoliation and average predicted defoliation probabilities from 1986–1991.

Table 8. Mean errors and mean squared error of the defoliation probabilities from the seven different models compared with observed defoliation frequencies.

	1961–1991		1986–1991	
Model	Mean error	Mean squared error	Mean error	Mean squared error
I	0.0009	0.0563	−0.0283	0.0287
II	0.0051	0.0549	−0.0226	0.0294
III	0.0070	0.0534	−0.0202	0.0383
IV	0.0029	0.0478	−0.0236	0.0272
V	−0.0107	0.0549	−0.0368	0.0343
VI	−0.0010	0.0451	−0.0303	0.0301
VII			0.0018	0.0245
VIII			0.0019	0.0249

track observed defoliation more closely (Fig. 7).

The mean error and mean squared error were used here to compare the performance of the different models. The mean error was simply the difference between the expected probability of defoliation, P^*, and the observed defoliation status, $V_{x,y,t}$, averaged over all points and years. The mean squared error was calculated in the same manner, except errors were squared before averaging. Errors were computed for all models from the 1986–1991 period and errors were computed from 1961–1991 data for all models except models *VII* and *VIII* since egg mass data were not available during the entire period.

Errors computed from 1961–1991 simulations indicated that the precision of predictions generally increased as additional information was included (Table 8). Model *I* used only the defoliation status in the previous year and it yielded the largest mean squared error. Progressively more information about the defoliation status at neighboring locations was added in models *II*, *III*, and *IV* and mean squared errors progressively decreased. Addition of information on total (30 yr) defoliation frequency at the cell in model *V* and *VI* did not yield consistent results. While the mean squared error of model *VI* was less than that of model *IV*, the mean squared error of model *V* was actually greater. There was no consistent pattern in the mean errors among the

different models. The lowest mean error was obtained using model *I*, indicating that this model was the most accurate, even though it was the least precise.

Errors computed from 1986–1991 simulations indicated that the incorporation of egg mass data increased the precision and accuracy of predictions (Table 8); both the mean error and mean squared error of models *VII* and *VIII* were lower than in any other model during the 1986–1991 period. The mean errors of models *I*–*V* were all negative during this period indicating that these models, which did not incorporate information about overwintering densities, tended to over-predict defoliation probabilities (Fig. 6). Model *VIII*'s errors were slightly larger than that of model *VII*, indicating that incorporating the 30 yr defoliation frequency did not improve accuracy during the 1986–1991 period. Model *VIII* used only two defoliation frequency classes because insufficient observations were available to parameterize a more complex model. We suspect that if sufficient data were available to parameterize a more complex defoliation frequency effect, such as was used in model *VI*, a model that incorporated both egg mass data and defoliation frequency would perform better than any of the other models.

Conclusions

Generally, models that incorporated the most information (*e.g.*, defoliation status of neighbors, egg mass densities, etc.) performed better than the other models tested. Models *VI*, *VII* and *VIII* performed the best (in terms of accuracy and precision) but there was still considerable error in their predictions. One component of this error was probably the error in estimating egg mass densities at unsampled locations (Liebhold *et al.* 1991). Though kriging minimizes the estimation error, that error can be considerable when sample points are widely spaced as in this study (Isaaks and Srivistava 1989). Spatial scale is probably another factor contributing to error in predictions. Within the 2 × 2 km raster cells used in this study there was likely considerable heterogeneity in actual defoliation, egg mass density, and forest composition. Our measures of both defoliation and egg mass density were essentially averaged within these cells and considerable error may have been introduced during this process. It would be valuable to test the CA methodologies developed here using data collected at a smaller spatial scale. Since most gypsy moth management decisions are made at the stand level (*ca.* 10–100 ha), higher resolution models would be more useful for management purposes.

In this study, models were parameterized and tested using the same data. For this reason, these models probably cannot be reliably applied during other time intervals and in other geographical locations without more extensive testing. Nevertheless, tests of these models indicated that these types of CA can be useful for predicting the spatial dynamics of gypsy moth outbreaks.

The CA transition probability models developed here behaved very similarly to a previously developed 3-dimensional kriging model (Hohn *et al.* 1993; Liebhold *et al.* 1994c). The three-dimensional kriging model was based upon the statistical dependence of defoliation through space and time. Kriging incorporates this dependence through a series of semivariogram models that are then used to optimally form weighted averages based upon expected covariances among samples. The models *II*, *III* and *IV*, while less complex than the 3-dimensional kriging model of Hohn *et al.* (1993), essentially simulated the same dependence of defoliation through space and time and this is probably why the two modeling techniques behaved so similarly. Liebhold *et al.* (1994c) provide a detailed comparison of these models with the 3-dimensional kriging model.

Acknowledgments

We thank Joseph Elkinton, University of Massachusetts, and the Massachusetts Department of Environmental Management for providing the defoliation and eggs mass data, and Michael Hohn, Linda Gribko and Eugene Luzader for their assistance with various aspects of this research. We thank William Hargrove, Thomas Jacob, and Alexei

Sharov for reviewing an earlier version of this manuscript. This research was funded by the USDA Forest Service Gypsy Moth Research and Development program and by grants 89-37250-4533 and 91-37302-6278 from the USDA CSRS Competitive Grants Program.

References

Badger, G.J., Vacek, P.M. and Reichman, R.C. 1987. A Markov model for a clinical episode of recurrent genital herpes. Biometrics 43: 399–408.

Birman, J.L. and Trebin, H.R. 1985. 'Statistical' symmetry with applications to phase transitions. Journal of Statistical Physics 38: 371–391.

Bhat, U.N. 1972. Elements of applied stochastic processes. Wiley, New York.

Campbell, R.W. 1973. Numerical behavior of a gypsy moth population system. Forest Science 19: 162–167.

Campbell, R.W. 1976. Comparative analysis of numerically stable and violently fluctuating gypsy moth populations. Environmental Entomology 5: 1218–1224.

Chrisman, N.R. 1987. The accuracy of map overlays: a reassessment. Landscape and Urban Planning 14: 427–439.

Cliff, A.D. and Ord, J.K. 1973. Spatial autocorrelation. Pion Press, London. 175 pp.

Eastman, J.R. 1987. Access to technology: the design of the IDRISI research system, Proc. GIS/LIS-87. pp. 166–175, ASPRS/ASCM, San Francisco.

Elkinton, J.S. and Liebhold, A.M. 1990. Population dynamics of gypsy moth in North America. Annual Review of Entomology 35: 571–596.

Gansner, D.A., Herrick, O.W. and Ticehurst, M. 1985. A method for predicting gypsy moth defoliation from egg mass counts. Northern Journal of Applied Forestry 2: 78–79.

Gansner, D.A., Drake, D.A., Arner, S.L., Hershey, R.R. and King, S.L. 1993. Defoliation potential of gypsy moth. USDA Forest Service Research Note NE-354. 4 pp.

Gardner, M. 1971. On cellular automata, self-reproduction, the Garden of Eden and the game 'life'. Scientific American 224: 112–117.

Hohn, M.E., Liebhold, A.M. and Gribko, L.S. 1993. A geostatistical model for forecasting the spatial dynamics of defoliation caused by gypsy moth, Lymantria dispar (*Lepidoptera*: *Lymantriidae*). Environmental Entomology 22: 1066–1075.

Isaacson, D.L. and Madsen, R.R. 1976. Markov chains: theory and applications. Wiley, New York.

Isaaks, E.H. and Srivistava, R.M. 1989. An Introduction to Applied Geostatistics. New York: Oxford University Press. 561 pp.

Kemp, W.P. 1987. Probability of outbreak for rangeland grasshoppers (*Orthoptera*: *Acrididae*) in Montana: Application of Markov principles. Journal of Economic Entomology 80: 1100–1105.

Langton, C.G. 1990. Computation at the edge of Chaos: Phase transitions and emergent computation. Physica D 42: 12–37.

Liebhold, A.M. and Elkinton, J.S. 1989. Characterizing spatial patterns of gypsy moth regional defoliation. Forest Science 35: 557–568.

Liebhold, A.M. and McManus, M.J. 1991. Does larval dispersal cause the spread of gypsy moth outbreaks? Northern Journal of Applied Forestry 8: 95–98.

Liebhold, A.M., Zhang, X., Hohn, M.E., Elkinton, J.S., Ticehurst, M., Benzon, G.L. and Campbell, R.W. 1991. Geostatistical analysis of gypsy moth (Lepidoptera: Lymantriidae) egg mass populations. Environmental Entomology 20: 1407–1417.

Liebhold, A.M., Halverson, J.A. and Elmes, G.A. 1992. Gypsy moth invasion in North America: a quantitative analysis. Journal of Biogeography 19: 513–520.

Liebhold, A.M., Simons, E., Sior, A. and Unger, J. 1993. Predicting gypsy moth defoliation from field measurements. Environmental Entomology 22: 26–32.

Liebhold, A.M., Elmes, G.A., Halverson, J.A. and Quimby, J. 1994a. Landscape characterization of forest susceptibility to gypsy moth defoliation. Forest Science 40: 18–29.

Liebhold, A.M., Elkinton, J.S., Zhou, G., Hohn, M.E., Rossi, R.E., Boettner, H.J., Boettner, C.W., Burnham, C. and McManus, M.L. 1994b. Regional correlation of gypsy moth defoliation (Lepidoptera: Lymantriidae) with counts of egg masses, pupae, and male moths. Environmental Entomology (in review).

Liebhold, A., Zhou, G., Gribko, L. and Hohn, M. 1994c. Models that predict the spatial dynamics of gypsy moth, *Lymantria dispar* (L.), populations. *In* Population Dynamics of Forest Insects. Edited by F. Haine and T. Payne. (in press).

Lippe, E., De Smidt, J.T. and Glen-Lewin, D.C. 1985. Markov models and succession: a test from a heathland in the Netherlands. Journal of Ecology 73: 775–791.

Miller, D.R., Mo, T.K. and Wallner, E.E. 1989. Climate influences on gypsy moth defoliation in southern New England. Environmental Entomology 18: 646–650.

Packard, N.H. and Wolfram, S. 1985. Two-dimensional cellular automata. Journal of Statistical Physics 30: 901–946.

Parzen, E. 1962. Stochastic processes. Holden-Day, Oakland, CA. 324 pp.

Phipps, M.J. 1992. From local to global: the lesson of cellular automata. *In* Individual-based models and approaches in ecology. pp. 165–187. Edited by D.L. DeAngelis and L.J. Gross. Chapman and Hall, New York.

Pielou, E.C. 1962. Runs of one species with respect to another in transects through plant populations. Biometrics 18: 579–593.

Pielou, E.C. 1977. Mathematical Ecology. John Wiley & Sons. New York.

Preston, K.J. and Duff, M.J.B. 1984. Modern Cellular Automata: Theory and applications. Plenum Press, New York. 340 pp.

Ripley, B.D. 1981. Spatial Statistics. Wiley, New York.

Snyder, J.P. 1987. Map projections – a working manual. U.S. Geol. Surv. Prof. Pap. 1395.

Sokal, R.R. and Rohlf, F.J. 1981. Biometry. W.H. Freeman and Co., New York. 859 pp.

Talerico, R.L. 1981. Defoliation as an indirect means of population assessment. *In* The Gypsy Moth: Research Toward Integrated Pest Management. pp. 38–49. Edited by C.C. Doane and M.L. McManus. USDA Tech. Bull. 1584. 757 pp.

Turner, M.G. 1987. Spatial simulation of landscape changes in Georgia: a comparison of 3 transition models. Landscape Ecology 1: 29–36.

Turner, M.G. 1988. A spatial simulation model of land use changes in a piedmont county in Georgia. Applied Mathematics and Computation 27: 39–51.

Usher, M.B. 1981. Modeling ecological succession, with particular reference to Markovian models. Vegetatio 46: 11–18.

Valentine, H.T. and Houston, D.R. 1979. A discriminant function for identifying mixed-oak stand susceptibility to gypsy moth defoliation. Forest Science 25: 468–474.

Vichniac, G.V., Tamayo, P. and Hartman, H. 1986. Annealed and quenched inhomogeneous cellular automata (INCA). Journal of Statistical Physics. 45: 875–883.

von Neumann, J. 1966. Theory of self-reproducing automata. Univ. of Illinois Press, Urbana, IL.

Williams, D.W., Fuester, W., Metterhouse, W.W., Balaam, R.J., Bullock, R.H. and Chianese, R.J. 1991. Oak defoliation and population density relationships for the gypsy moth. Journal of Economic Entomology 84: 1508–1514.

Wolfram, S. 1983. Statistical mechanics of cellular automata. Rev. Mod. Phys. 55: 601–644.

Wolfram, S. 1984. Universality and complexity in cellular automata. Physica D 10: 1–35.

Wootters, W.K. and Langton, C.G. 1990. Is there a sharp phase transition for deterministic cellular automata? Physica D 45: 95–104.