2017

Integrating remote sensing, GIS and dynamic models for landscape-level simulation of forest insect disturbance

Lu Liang
University of Arkansas, liang@uamont.edu

Xuecao Li
Iowa State University

Yanbo Huang
USDA-ARS

Yuchu Qin
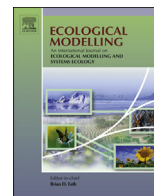*Beijing Normal University*

Huabing Huang
Beijing Normal University

Integrating remote sensing, GIS and dynamic models for landscape-level simulation of forest insect disturbance

Lu Liang [a,*], Xuecao Li [b], Yanbo Huang [c], Yuchu Qin [d], Huabing Huang [d]

[a] Arkansas Forest Resources Center, University of Arkansas, Division of Agriculture, School of Forestry and Natural Resources, University of Arkansas at Monticello, Monticello, AR 71655, USA
[b] Department of Geological & Atmospheric Science, Iowa State University, Ames, IA 50011, USA
[c] USDA-ARS Crop Production Systems Research Unit, Stoneville, MS 38776, USA
[d] State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing Normal University, Beijing 100101, China

ARTICLE INFO

ABSTRACT

Cellular automata (CA) is a powerful tool for modeling the evolution of macroscopic scale phenomena as it couples time, space, and variables together while remaining in a simplified form. However, such application has remained challenging in forest insect epidemics due to the highly dynamic nature of insect behavior. Recent advances in temporal trajectory-based image analysis offer an alternative way to obtain high-frequency model calibration data. In this study, we propose an insect-CA modeling framework that integrates cellular automata, remote sensing, and Geographic Information System to understand the insect ecological processes, and tested it with measured data of mountain pine beetle (MPB) in the Rocky Mountains. The overall accuracy of the predicted MPB mortality pattern in the test years ranged from 88% to 94%, which illuminates its effectiveness in modeling forest insect dynamics. We further conducted sensitivity analysis to examine responses of model performance to various parameter settings. In our case, the ensemble random forest algorithm outperforms the traditional linear regression in constructing the suitability surface. Small neighborhood size is more effective in simulating the MPB movement behavior, indicating that short-distance is the dominating dispersal mode of MPB. The introduction of a stochastic perturbation component did not improve the model performance after testing a broad range of randomness degree, reflecting a relative compact dispersal pattern rather than isolated outbreaks. We conclude that CA with remote sensing observation is useful for landscape insect movement analyses; however, consideration of several key parameters is critical in the modeling process and should be more thoroughly investigated in future work.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Forests worldwide increasingly cope with a variety of disturbances, such as drought, windthrow, insect and disease outbreaks, and fire (Trumbore et al., 2015), among which, insects are the agent of great concern. Through catastrophic epidemics or less apparent chronic infestation, insects exert strong impacts on the forest structure, function, productivity, and diversity (Franklin et al., 1987; Attiwill 1994; Turner 2010; Edburg et al., 2012). There is strong evidence that changing environment is a driver of insect outbreaks (Dale et al., 2001; Kurz et al., 2008). Altered temperature, precipitation, solar radiation, and atmospheric CO_2 concentrations

affect insect survival, development, and redistribution (Ayres and Lombardero 2001; Hicke et al., 2006; Bentz et al., 2010). For instance, in North American forests, unusually hot and dry weather has resulted in wide beetle spread (Logan et al., 2003). Moreover, human-intensified habitat fragmentation and pollution can modify the insect dispersal behavior and their geographic distribution (Raffa et al., 2008; Bone et al., 2013; Liang et al., 2014a). The climatic and anthropogenic stressors can further act simultaneously and lead to the increase in the frequency and severity of further insect outbreaks.

To gain insights into the past, present, and future states of the forest ecosystem, as well as the far-reaching ecological consequences, it is crucial to monitor, simulate, and predict insect disturbance processes. Those processes can be temporally uneven and spatially heterogeneous due to the highly dynamic interactions among insects and their environments. Thus, it is essential

for models that support spatial-temporal interactions and can be easily combined with geospatial data. One of such techniques is cellular automata (CA) modeling (Li and Gong 2016). Emerged from complex systems theory, CA models are mathematical representations of complex systems with a regular grid of cells, each in one of a finite number of states (White and Engelen, 1993). The cell state is governed by a set of environmental rules, neighborhood influences, and stochastic processes. The model evolves in discrete time steps, with the cell states updated synchronously. The discrete and dynamic nature of CA models allows broad applications in a variety of natural phenomena and human activities, such as urban planning (Li et al., 2014), resource management (Bone and Dragićević, 2010), forest fire spread prediction (Alexandridis et al., 2008; Yassemi et al., 2008), plant invasion (Cannas et al., 2003), insect propagation (Bone et al., 2006; Pérez and Dragićević, 2011), animal migration patterns (Bennett and Tang, 2006), and landscape ecology (Parker et al., 2008).

Despite the success of CA in studying the relationship between micro-level individual actions and emergent macro-level phenomena (Gimblett 2002), few efforts have been devoted to simulating forest insect infestation patterns, mainly because of the lack of enough reliable disturbance information for model calibration. For instance, most previous studies were built upon empirical suitability surface (e.g., Pukkala et al., 2014; Anderson and Dragicevic 2016) or few sparse observations (e.g., Bone et al., 2005; Perez and Dragicevic 2012). Considering the complexity of beetle activity, the intermediate outputs of the model should be validated by dense observations. Traditionally, forest insect and disease have long been confirmed, validated, and attributed by field and aerial sketch mapping surveys (Man, 2010). While in-situ surveys are often restricted in their geographic and temporal coverage, Forest Health Monitoring Aerial Detection Surveys were criticized for spatial imprecision as well as a temporal shift from actual attack onset (McConnell et al., 2000; Meigs et al., 2011). The lack of spatial specificity makes sketch maps of forest defoliation and mortality less suitable for fine-scale modeling (Johnson and Ross 2008). Multi-temporal moderate-resolution remote sensing imagery offers an alternative to the field or aerial surveying data due to its vast area coverage, affordable cost, frequent revisit, and historical archive (Coops et al., 2006; White et al., 2004). Most conventional change detection techniques are capable of discriminating abrupt land cover changes but are less efficiently in discerning subtle and chronic changes (Gong et al., 2016). An emerging new method is temporal trajectory analysis, which models temporal signatures of schemed land cover types from dense image stacks as a proxy for their phenological variations (Gong et al., 2016). This technique is particularly useful in assessing the extent and severity of subtle ecosystem changes (Meddens et al., 2013; Liang et al., 2014b; Meddens and Hicke, 2014; Meigs et al., 2015).

Besides the reliability of the source data used for model calibration, the uncertainties of CA models also come from the inappropriate selection of model parameter settings, of which different combinations may result in large variations in model outcomes. Typically, a CA model consists of four components: suitability surfaces, neighborhood configuration, simulation time, and stochastic perturbation (Yeh and Li 2006; Santé et al., 2010). They work in a collective way to form decisions on cell states. An arbitrary decision on their values could introduce errors and uncertainties to the model, which will further propagate through the simulation process (García et al., 2011). The whole procedure could become meaningless if the behavior of the automation was completely unstable and unrepeatable in practice. Hence, a systematic sensitivity analysis of the effects of varying the elements of the CA model design is essential to understand model behavior and sources of errors.

The overall goal of this study is to integrate satellite observation results with a Geographic Information System (GIS)-based CA model for developing a more realistic landscape-level simulation of chronic forest insect attack. The specific objectives are: 1) to test the developed insect-CA model with measured data of mountain pine beetle (MPB)—one of the most devastating disturbance agents in the pine forests of western North America for the past decade; and 2) to gain deeper insights into the complexity of the spatial process and the behavior of the proposed CA forest insect model using sensitivity analysis.

2. Methodology

### 2.1. Mountain pine beetle and study site

Mountain pine beetle (*Dendroctonus ponderosa*) has existed as a major regulator of pine forest ecosystem in the western North America for centuries. However, in recent decades it grew to epidemic levels with the area infested exceeding the extent and impacts of outbreaks documented in the past 125 years. By establishing their breeding galleries in the phloem layer of the host trees, MPBs prefer infesting and killing older and stressed trees with larger diameters. The effects of MPB are especially evident in the lodgepole pine (*Pinus contorta*) forests of Rocky Mountain ecoregion with a severe epidemic occurring in Grant County, north-central Colorado. Grand County approximately occupies an area of 4,830 square kilometers and lies entirely within one Landsat scene (path 34 row 32). The MPB outbreak started around 2002 and reached the peak of tree mortality in 2005. The majority of host trees were dead around 2008 (Klutsch et al., 2009).

2.2. Remote sensing derived time-series forest disturbance map

For each year, we acquired one cloud-free or mildly contaminated 30 m Landsat image during the most active growing season from 2000 to 2009. If no single good quality image existed, several partly cloudy images with a similar phenology were used as replacements. The effect of using combined images on the classification is very minimal as 98% of the study region contained more than nine cloud-free observations.

The MPB-induced tree mortality patterns in Grand County were generated by an automated classification workflow that integrates temporal segmentation techniques and a decision tree modeling procedure (Liang et al., 2014b, 2016). Temporal segmentation was introduced in the Landsat-based Detection of Trends in Disturbance and Recovery algorithm (LandTrendr, Kennedy et al., 2010) to capture the various disturbance processes operating at different rates. As a forest undergoes either abrupt or chronic disturbance events, its associated time series trajectory displays distinct characteristics that correspond to the duration and magnitude of the disturbances. Thus, the trajectory constructed using the Normalized Burn Ratio can be decomposed into a sequence of straight-line segments, with each segment representing one salient event. The performance of temporal segmentation is controlled by a set of parameters to minimize the omission of disturbance detection while maximizing signal-to-noise ratio in the time series. The model parameters were calibrated with 106 field samples, among which, 53 were sites with MPB mortality, 37 were clearcuts, and 16 were healthy. With the derived segment features, we then applied a set of decision tree rules to classify the disturbance types based on their distinct segment characteristics (e.g., line slope and duration) (Fig. 1).

The final output of this approach includes a series of disturbance maps depicting healthy forests, forests displaying MPB mortality, and clearcut forests at an annual time step. Validated against an independent set of points sampled across a gradient of MPB mor-

1. Trajectory-based classification

2. Insect-CA Modeling

3. Sensitivity Analysis
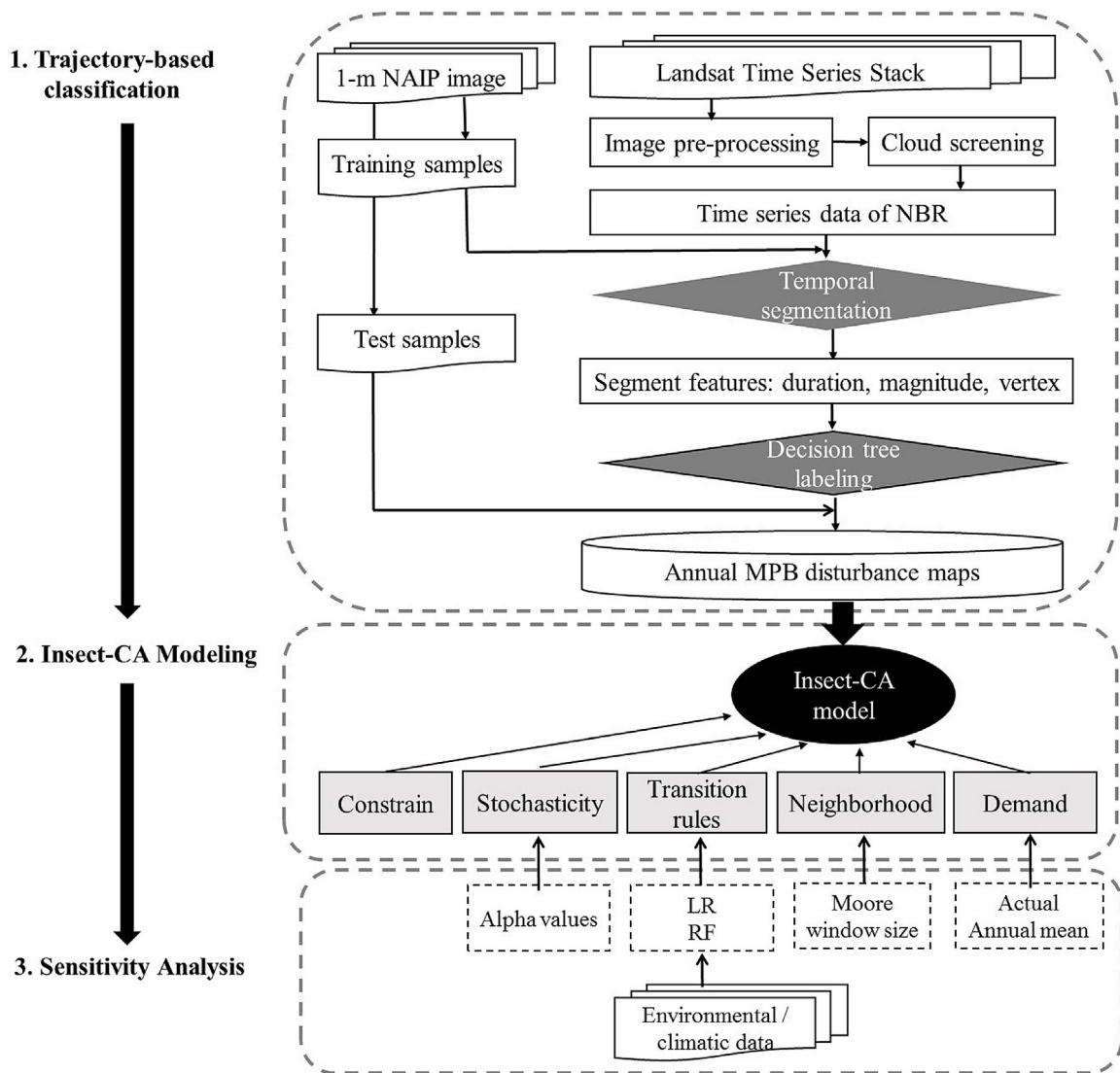
Fig. 1. Overall working flow of insect- cellular automata (Insect-CA) model and sensitivity analysis. NAIP – National Agriculture Imagery Program; NBR – Normalized Burn Ratio; LR – logistic regression; RF – random forest.

tality, the overall accuracy of the produced annual disturbance and recovery map ranged from 86.74% to 94.00% across years. The mean producer's accuracy of MPB mortality was 95.39%, and the mean user's accuracy was 84.81% (Liang et al., 2014b). As compared with traditional single-date classifications, this trajectory-based mapping approach achieved consistently higher accuracy, and significantly reduced illogical annual changes (Liang et al., 2016).

2.3. CA modeling framework overview

Cellular automata is a mathematical representation of complex systems that consist of a lattice of cells representing the landscape at the macroscopic scale (Couclelis 1985). Each cell is described by several variables whose states, being usually discrete, evolve in discrete time, according to a set of rules and the states of the neighboring cells. The cells update their state synchronously after the suitability surfaces are applied. CA models are composed of four elements: suitability surface S, neighborhood Ω, constraints C, and stochastic perturbation R (Li et al., 2014), which are combined in the

following equation form to generate the insect dispersal probability P_{ij}^t:

$$P_{ij}^t = S_{ij}^t \times \Omega_{ij}^t \times C_{ij}^t \times R_{ij}^t \tag{1}$$

where ij is the target cell, and t is the time step used to change cell states. P_{ij}^t is the potential for the insect to move to the target cell ij at time t. S_{ij}^t is the suitability surface that is a GIS-based representation of the susceptibility of each site to insect attack. Ω_{ij}^t indicates the impact of neighborhood configuration. C_{ij}^t denotes the constraint of non-host forests in insect movement. Because spatial synchrony is prevalent among insect population during epidemic years (Aukema et al., 2006), more insect presence in the immediate neighborhood increases the likelihood of a mass attack on the adjacent healthy forest. Stochastic perturbation R_{ij}^t reflects the effect of unconsidered factors, such as random noises.

Once the CA model has been calculated at time t, all cells are sorted in a descending sequence based on the dispersal probability. Pixels with higher probabilities have higher priority to be covered to infested class. At the model calibration stage, the number of converted pixels during each iteration is usually determined by empirical, surveyed or observed data (e.g. Li et al., 2014).
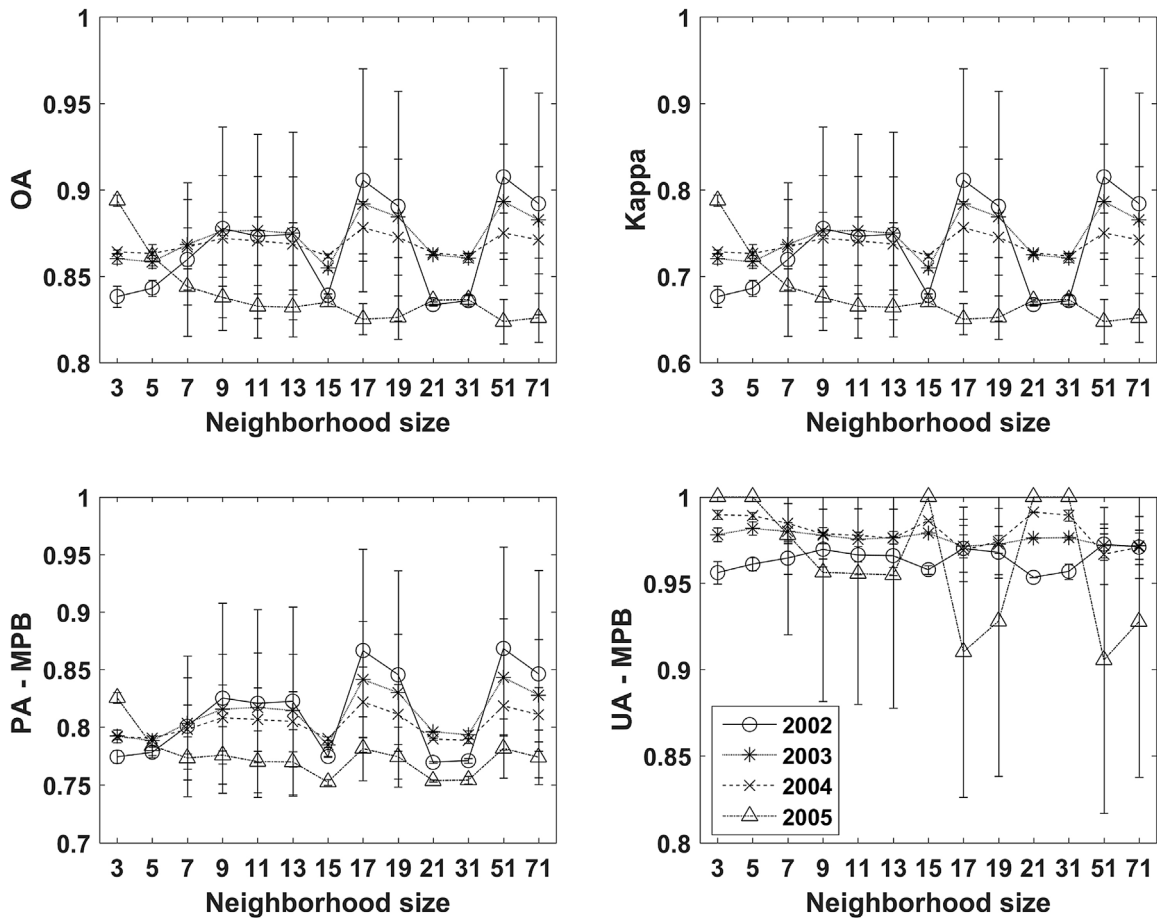
Fig. 2. Variations in overall accuracy (OA), Kappa, producer's accuracy (PA) and user's accuracy (UA) of mountain pine beetle with different neighborhood window sizes. Error bar indicates one standard deviation of the confusion estimates from the mean.

2.4. Sensitivity analysis of CA model behavior

Sensitivity analysis is a powerful tool in identifying the sources of variation influencing model performance by exploring the model's input-output relationship (Saltelli et al., 2000). In this case study, we implemented sensitivity analysis by measuring the responses of model output to different input parameter values of suitability surface, neighborhood configuration, and degree of stochasticity. Finally, we compared the model outputs with the actual observation data.

2.4.1. Development of suitability surfaces

As a crucial component in CA, the suitability surface is traditionally retrieved by empirical data without calibration. Nonetheless, their performance is closely related to the methods used, input explanatory layers, and the quantity and quality of training samples. Considering the MPB population distributions are governed by a variety of biotic and abiotic factors that are dynamically changing (Raffa et al., 2008), we updated training samples and environmental layers for each year and simultaneously construct the suitability surface at an annual time step.

(1) Response variable

We extracted the case and control samples for response variables from the remote sensing derived disturbance maps. Cases are defined as the recently died forests where healthy trees can be found in the previous year. And 100 samples were randomly placed in the annual newly emerged areas of MPB mortality. Con-

trol samples are a collection of pixels without MPB presence, and the sample size is equivalent to that of the case.

(2) Explanatory variables

Twenty-four variables that are considered important in determining the distance and direction of beetle movement were employed. Their selection justifications are explained below. Detailed data sources and calculation can be found in Table S1.

- Forest characteristics, including homogenous, even-aged, high-density, and large-diameter stands, are favorable for MPB mass attack (Raffa and Berryman 1983). Tree cover prior to the MPB outbreak was used as a proxy for the pre-disturbance abundance of MPB host species.
- Topographic metrics, including elevation, aspect, southwestness, topographic wetness index, and distance to water canal network, were employed to characterize the topographic-moisture gradient that controls the growing environment of lodgepole pine forests as an integrated function of soil water holding capacity, evapotranspiration, and surface runoff (Schmidt and Persson 2003).
- Residential housing density and distance to the nearest road were utilized as proxies for the impacts of human activities on MPB dispersal. For instance, habitat fragmentation due to human land use can cause dispersal barriers, and silvicultural treatments (e.g., thinning, logging) can significantly reduce the beetle population as well as host density.
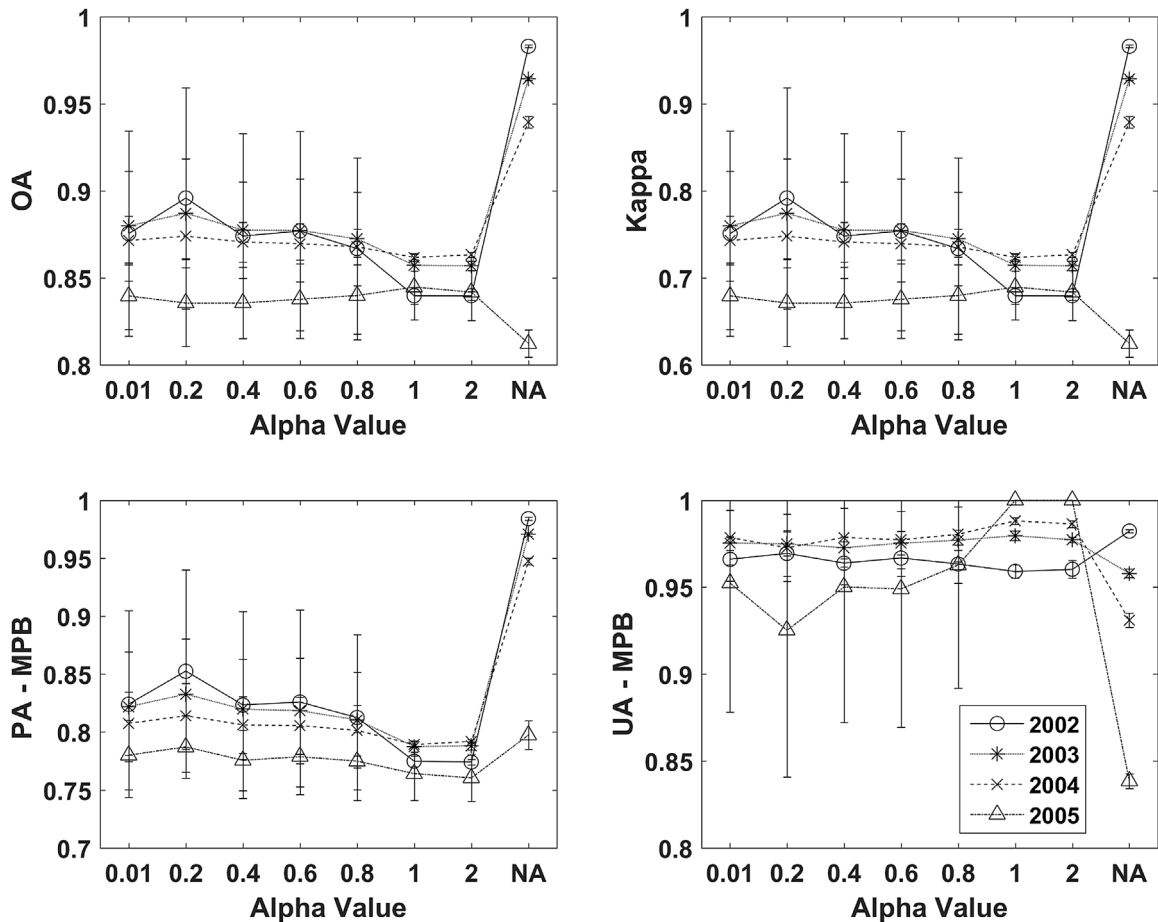
Fig. 3. Variations in overall accuracy (OA), Kappa, producer's accuracy (PA) and user's accuracy (UA) of mountain pine beetle (MPB) with different alpha values of stochastic perturbation. Error bar indicates one standard deviation of the confusion estimates from the mean. NA – no stochastic perturbation.

- Thermal and precipitation regimes impact beetles' developmental timing, cold-induced mortality, and the associated fungal community (Bentz et al., 2010; Preisler et al., 2012). A warmer climate reduces the cold-induced mortality in the adult and larval stages and accelerates the developmental timing within one generation. Longer drought periods may lead to increased host susceptibility and thus result in a higher probability of outbreak intensification. We used annual mean temperature and precipitation information from 4 km climate datasets, as well as six derived variables that are controlling factors in the beetles' life cycle (Kaufmann et al., 2008), separately summer mean temperature, winter mean temperature, summer mean precipitation, warmest summer temperature, coldest winter temperature, and summer mean precipitation in the previous year. Since climate change has been indicated to have both direct and indirect impact on MPB outbreaks (Kurz et al., 2008), we also computed eight climate anomalies by taking the differences between each climate variable and their 30-year averages from 1981 to 2010.

(3) Model selection

We used a traditional logistic regression (LR) model and a random forest (RF) model to construct the suitability layer and compared their effects on the model outcomes. LR has been more frequently applied in CA-related studies because of its interpretable results. In contrast, RF is a non-linear machine learning technique that combines the results of many weak classifiers, such as classification and regression trees, and is often praised for its robust performance (Dietterich 2000; Breiman 2001). The same sets of response and explanatory variables were used in both models. The agreement between the probability maps of suitability surface and the binary observed maps was measured by the receiver operating characteristic value (ROC, Hu and Lo 2007), which calculates the proportion of true positives and false positives for a range of specified threshold values and relates them to each other in a graph.

2.4.2. Neighborhood configuration

The scale of the neighborhood depends on the beetles' dispersal ability that varies by topography, wind speed, preflight weight, flight duration, lipid content, and host availability (Evenden et al., 2014). Two modes have been summarized for MPB from field observations and laboratory flight mill bioassay: short-distance and long-distance dispersal (Robertson et al., 2008, 2009; Evenden et al., 2014). Short-distance dispersal happens within stands typically ranging from 30 to 50 m (Safranyik et al., 1989, 1992; Robertson et al., 2007). The long-distance dispersal usually occurs when beetles are transported above the forest canopy by wind and ranges from several to tens of kilometers (Robertson et al., 2007; Jackson et al., 2008; Robertson et al., 2009; de la Giroday et al., 2011). We used the 3×3 Moore neighborhood as the configuration setting, which is formulated as:

$$\Omega_{ij}^t = \frac{\sum_{n^2} Con(L_{ij}^t = \text{MPB mortality})}{n \times n - 1} \qquad (2)$$

where $Con()$ is the conditional function that returns 1 if the state of cell L_{ij}^t is beetle mortality at time t. n is the window size and we
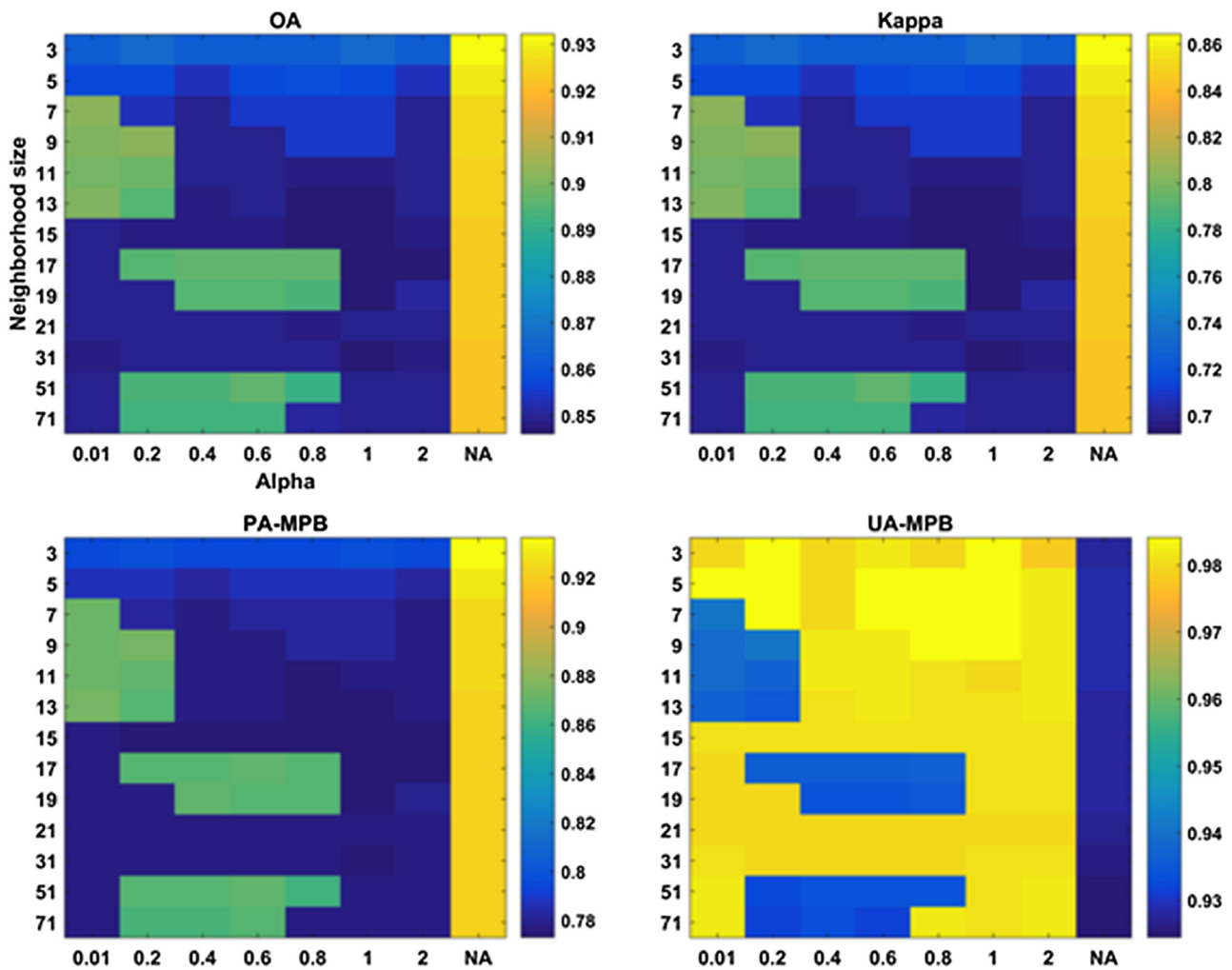
Fig. 4. Variation in the averaged accuracy metrics with different alpha values and different neighborhood configurations over the training phase. OA – overall accuracy; PA – producer's accuracy; UA – user's accuracy; MPB – mountain pine beetle.

tested different sizes for the assessment on neighborhood effects (n = 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 31, 51, 71).

2.4.3. Stochastic perturbation

Stochastic perturbation is quantified according to the formula developed by White and Engelen (1993) since it has control over the randomness size contrary to some other approaches.

$$R_{ij}^t = 1 + \left(-ln\left(rand_{ij}^t\right)\right)^\alpha \tag{3}$$

where R_{ij}^t is the scalable random perturbation term of the target cell at a specific time. $rand_{ij}^t$ is a uniform random variable varying spatially and temporally and the values always stay between 0 and 1. α is a parameter that controls the size of the perturbation. A higher value of α means that extreme values of rand are given more weight and thus correspond to a more disperse landscape (Li et al., 2014; Mustafa et al., 2014). The different runs of the model performance without stochastic perturbation and with different values of α in stochastic perturbation (α = 0.01, 0.2, 0.4, 0.6, 0.8, 1, and 2) were compared to analyze the effect of introducing the stochastic perturbation in the traditional CA model.

2.5. Model evaluation and future scenario simulation

For each set of input parameter values, we ran the associated model for ten times to estimate the model uncertainties. In total,

910 runs were executed during the training phase (13 window size × 7 alpha values × 10 iterations). By assuming the satellite-derived disturbance patterns were an accurate representation of landscape-level MPB mortality, we compared them against the result of each model run. An independent set of validation points was selected for each year by extracting 300 healthy sites and 300 mortality sites from the annual Landsat observed maps. Overall accuracy (OA), producer's accuracy (PA), and user's accuracy (UA) were calculated for each run. Producer's accuracy corresponds to the error of omission whereas user's accuracy measures the error of commission.

The set of model parameters that retained the highest accuracy during the training stage was used to simulate MPB dispersal patterns in the testing years. To validate the outcomes, we compared the predicted patterns with the satellite observed patterns in both quantitative and qualitative manners. In the quantitative assessment, besides the pixel-based accuracy measures mentioned above, we also included three landscape metrics, separately percentage agreement, omission percentage, and commission percentage. Percentage agreement calculates the ratio of pixels that have been labeled as the same class (either MPB mortality or healthy) in both the predicted map and the satellite-derived map to the total number of pixels. Omission percentage is the proportion of pixels that have been predicted as healthy but observed as MPB mortality. Moreover, commission percentage is the pro-
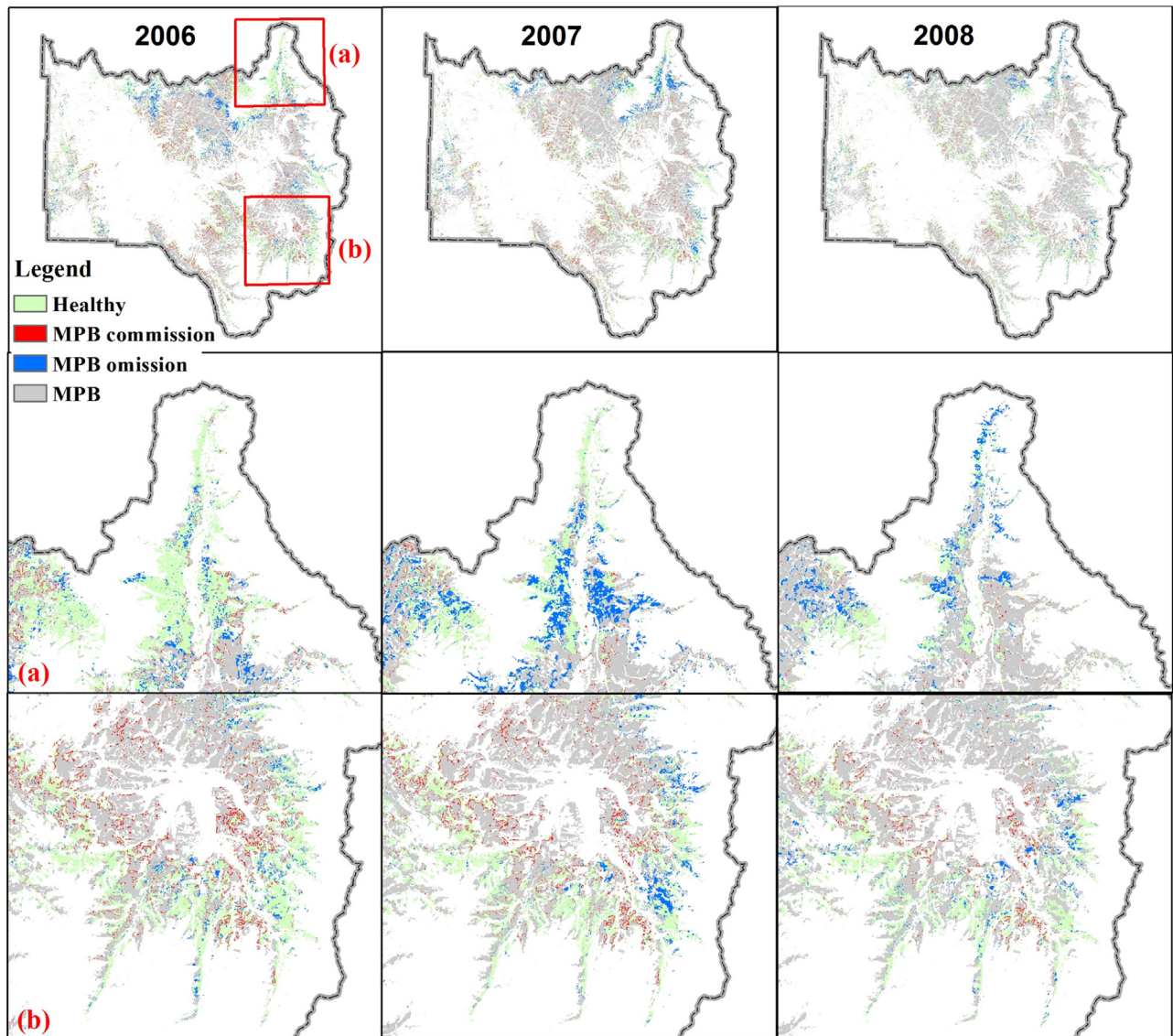
Fig. 5. Comparison of disagreements between model predicted and Landsat-observed disturbance pattern in Grand County. Panel (a) and (b) are zoomed-in views of two sites. In the legend, healthy refers to the matched healthy pixels between predicted and observed; mountain pine beetle (MPB) is the matched predicted and observed MPB mortality; MPB commission indicates the pixels with predicted MPB mortality and observed healthy label; MPB omission refers to pixels with predicted healthy and observed MPB mortality.

portion of pixels that have been predicted as MPB mortality but observed as healthy. Regarding the qualitative assessment, we visually inspected the beetle mortality pattern between the model predicted results and the satellite observations, to understand the predictive capacity of the model in a spatially explicit context.

3. Results

3.1. Responses of simulation accuracy to suitability surface, neighborhood size, and stochastic perturbation

We built the maps of suitability surface with LR and RF and tested their performance via the means of ROC. The ROC statistic measures the area under the curve (AUC) in the graph, and a value of 1 represents a perfect fit while an area of 0.5 suggests a random fit. The average AUC value of RF model was 0.701, with the range from 0.659 to 0.727. LR has an overall lower average AUC (0.595) but with less inter-annual variation (Table 1).

Table 1
The area under the curve (AUC) values for random forest and logistic regression models in constructing suitability surface.

Year	AUC	
	Random forest	Logistic regression
2002	0.707	0.590
2003	0.712	0.594
2004	0.727	0.599
2005	0.659	0.595
MEAN	0.701	0.595

We also tested the effects of neighborhood configuration by changing the window size from 3 pixels (i.e., 30 m dispersal distance) to 71 pixels (i.e., 1050 m dispersal distance). The overall classification accuracy was found to be sensitive to changes in neighborhood size (Fig. 2), and several features of this response were notable. First, two accuracy peaks occurred at the scales 17 (i.e., 240 m dispersal distance) and 51 (i.e., 750 m dispersal distance) for the years 2002–2004 but happened at the scale 3 for the

year 2005. Second, the first years of the training phase responded most dramatically to changes in neighborhood configuration, while decreased gradually in the subsequent years.

The model accuracy compared with those without the component of randomization got decreased (Fig. 3). With the alpha values increased, the average OA, Kappa, and MPB-PA decreased. Similar to the neighborhood size effect, the variations across different alpha values was most noticeable for the first few years and became stable for the later years.

The combined effects of neighborhood configuration and stochastic perturbation on the model performance were assessed by calculating the overall accuracies for each pair of window scale and alpha value. The result is illustrated using a 2-D mesh plot (Fig. 4). Warm colors in the mesh plot indicated high accuracy. The most significant accuracy peak involved a neighborhood size of three and no stochastic perturbation.

3.2. Simulated future MPB dispersal scenario

MPB mortality patterns in the test years were projected using the insect-CA model with the most optimal set of model parameters retrieved from the sensitivity analysis. The quantitative evaluations with independent dataset suggested a good model performance in future scenario simulation. Overall accuracy ranged from 88% to 94%, and the average was 90%. Producer's accuracies for MPB mortality in all years were about 10% lower than user's accuracies. Regarding the pattern-based assessment, pixels showing consistent labels in both the model prediction and satellite observation count about 89.60–94.58% of the host forest area in Grand County. The mean commission percentage and omission percentage was 4.00% and 5.00% with small variations among years (Table 2).

Through visual inspection, we observed that the predicted areas of MPB mortality matched well with the Landsat-based observations. Although the landscape-level assessment suggested a mean of 92% agreement, some problematic areas were identified. For instance, the northeast corner of Grand County has more progressive MPB outbreak during the test years than the training period, but the insect-CA model did not capture the MPB's rapid dispersal very well (Fig. 5a). Some minor issues were raised in the peripheral areas of the epicenters, where erroneously classified pixels were concentrated at the border between healthy forests and new mortality zones (Fig. 5b).

4. Discussion

4.1. Impacts of different suitability surface construction methods on model outcomes

Comparison of results with two distinct suitability surface construction methods revealed a 6.4%-12.8% more variation in the AUC values of LR than RF, suggesting that the ensemble learning algorithm outperforms the traditional parameter-based LR in our study case and future similar simulations should give more considerations to those new techniques. Most suitability potential models are defined in a linear form because of the easy interpretation merits (Yang et al., 2008). However, most complex processes are often characterized by nonlinearity, complexity, and self-organization (Torrens 2000; Liu et al., 2008). In contrast, the ensemble algorithm such as RF relies on all outputs produced by each tree that is built upon the random selection of input layers. By grouping many individual classifiers, RF combines the strengths and minimized the weaknesses of each (Ghimire et al., 2010). Another advantage of RF is that they are free of normal distribution assumptions and robust in dealing with outliers and noise. Moreover, the ability of RF in

evaluating the variable contribution to variance can partly compensate the lack of model interpretability in non-linear methods.

4.2. Effects of neighborhood size

Mountain pine beetles may move over short or long distances (Safranyik et al., 1989; Robertson et al., 2007; Jackson et al., 2008; de la Giroday et al., 2011). MPB is usually a relatively poor disperser (Simard et al., 2012), however, under suitable wind conditions or if no suitable host is around, beetles can fly over greater distances for a longer period of time (Robertson et al., 2009).

Results revealed that the highest overall and class-specific accuracies occurred at the coarser scales for the first few years and happened at the three-pixel scale for the last year of the training period. With regard to MPB, three-pixel scale corresponds to short-distance dispersal model, while 17 and 51 pixels correspond to long-distance dispersal mode correspond to long-distance dispersal mode. This study implies that long-distance dispersal is the dominant mode of MPB expansion in Grand County for the early years and then replaced by short-distance dispersal mode. This finding is also in agreement with some previous studies. Chen and Walton (2011) implied that while long-distance movement promoted the initiation of infestations at early stages, short-distance dispersal became crucial when infestations got intensified, and populations reach epidemic levels. In this case, for the first few years of the study period, the MPB epidemic has already been established at the northern and southern part of Grand County, and thus long-distance dispersal mode became less prevalent in the subsequent beetle progression.

4.3. The ineffectiveness of stochastic perturbation randomness

For most complex systems, the forecast of their future states cannot exclusively rely on deterministic rules, while neglecting the measurement of uncertainties. For instance, Mustafa et al. (2014) found that the model accuracy increased by introducing a stochastic perturbation component with a limited degree of randomness in the cellular automata urban growth model. However, unlike most urban models, the inclusion of stochastic components did not improve the insect-CA model performance after testing a broad range of degrees of randomness.

The lack of improvement in classification accuracy with added stochastic perturbation in our study could be attributed to three explanations. First, the dominance of short-distance dispersal mode during our study period reduces the random occurrence of MPB mortality in this region. This inference is further supported by the satellite-derived MPB mortality pattern, where MPB progression was observed most often in continuously expanding patterns instead of isolated outbreaks. Second, model uncertainty could be a result of insufficient input variables. The lack of certain geophysical or socio-economic factors can affect the overall output of the system (Li et al., 2014). In this study, we have included a wide spectrum of exogenous factors whose combined effects in predicting the movement direction and strength have minimized the potential model uncertainties. Finally, the forecast of land use in urban systems is not only influenced by global factors like population and transportation network but also heavily manipulated by political decisions and societal factors such as population life cycle, migration movements, societal values and standards (Mustafa et al., 2014). The difficulties in predicting those factors make the urban modeling easily subject to uncertainties. In contrast, beetle behavior in a natural forest ecosystem is less affected by those factors.

On the other hand, we also noticed that the introduction of stochastic perturbation at various degrees did not cause substantial variations in the model accuracy. Within each year, the variation is around 5% or less. This implies that its property of stability is

Table 2
Accuracy metrics in the test years.

Year	OA	Kappa	PA	UA	Perc Agreement	Commission Perc	Omission Perc
2006	0.89	0.78	0.86	0.92	89.60%	4.66%	5.73%
2007	0.88	0.75	0.81	0.95	91.15%	3.96%	4.89%
2008	0.94	0.87	0.90	0.97	94.58%	2.10%	3.32%
mean	0.90	0.80	0.86	0.95	92.00%	4.00%	5.00%

Note: OA – overall accuracy; PA – producer's accuracy; UA – user's accuracy; Perc – percentage.

only mildly affected by perturbations. If we use robustness as a measure of assessment to which extent behavior or performance is violated by system perturbations, this model can be considered robust against noises.

4.4. Limitations

Despite the success of our modeling approach in projecting the majority of the disturbance pattern, there is an observed failure in capturing the fast expanding MPB mortality areas and the random noises around the edges of epidemic zones. It is possibly due to the different behavior of MPB in various site-specific microenvironments. The suitability layer can explain a certain portion of the effects of exterior physical or anthropogenic conditions on MPB ecology. However, constrained by the 1-km coarse resolution of the climatic factors, microclimatic environments can hardly be fully characterized. This is so far the biggest limitation for our model to achieve higher values in accuracy metrics and better representation of dispersal pattern in a spatial context. To further improve the model performance, finer and accurate downscaled climatic datasets should be used.

5. Conclusion

The last decades have witnessed growing increases in the frequency, duration, extent, and severity of insect disturbances (Kurz et al., 2008). There is a strong need for geosimulation methodology to simulate spatiotemporal dynamics of insect infestation and understand how a pattern of spread may react to events such as climate change, abrupt disturbance (e.g., fire), and management strategies. This study presents a prototype that combines the advantages of the state-of-art earth observation technique in documenting the evolving pattern of insect outbreak over an extended period of time and the merits of cellular automata models in dealing explicitly with the processes underlying spatial patterns. Meanwhile, the success of the insect-CA model in projecting the MPB movement in Grand County demonstrates the potential for this framework to understand the mechanisms resulting in population patterns, to evaluate the role of habitat and other factors, and to characterize and understand other emerging insect issues in various ecosystems, such as emerald ash borer and Asian longhorned beetle. Our study also illuminates the importance of sensitivity analysis as an essential stage of modeling practice since it determines the model reliability, aiding in assessing model uncertainties as well as the relation of input parameters and outcomes.

References

Alexandridis, A., Vakalis, D., Siettos, C.I., Bafas, G.V., 2008. A cellular automata model for forest fire spread prediction: the case of the wildfire that swept through Spetses Island in 1990. Appl. Math. Comput. 204 (1), 191–201.

Anderson, T., Dragicevic, S., 2016. A geosimulation approach for data scarce environments: modeling dynamics of forest insect infestation across different landscapes. ISPRS Int. J. Geo-Inf. 5 (2), 9, http://dx.doi.org/10.3390/ijgi5020009.

Attiwill, P.M., 1994. The disturbance of forest ecosystems: the ecological basis for conservative management. For. Ecol. Manage. 63 (2), 247–300.

Aukema, B.H., Carroll, A.L., Zhu, J., Raffa, K.F., Sickley, T.A., Taylor, S.W., 2006. Landscape level analysis of mountain pine beetle in British Columbia, Canada: spatiotemporal development and spatial synchrony within the present outbreak. Ecography 29, 427e441.

Ayres, M.P., Lombardero, M.J., 2001. Assessing the consequences of global change for forest disturbance from herbivores and pathogens. Sci. Total Environ. 262 (3), 263–286.

Bennett, D.A., Tang, W., 2006. Modelling adaptive, spatially aware, and mobile agents: Elk migration in Yellowstone. Int. J. Geogr. Inf. Sci. 20 (9), 1039–1066.

Bentz, B.J., Régnière, J., Fettig, C.J., Hansen, E.M., Hayes, J.L., Hicke, J.A., Kelsey, R.G., Negron, J.F., Seybold, S.J., 2010. Climate change and bark beetles of the western United States and Canada: direct and indirect effects. Bioscience 60 (8), 602–613.

Bone, C., Dragićević, S., 2010. Incorporating spatio-temporal knowledge in an Intelligent Agent Model for natural resource management. Landscape Urban Plann. 96 (2), 123–133.

Bone, C., Dragicevic, S., Roberts, A., 2005. Integrating high resolution remote sensing, GIS and fuzzy set theory for identifying susceptibility areas of forest insect infestations. Int. J. Remote Sens. 26 (21), 4809–4828.

Bone, C., Dragicevic, S., Roberts, A., 2006. A fuzzy-constrained cellular automata model of forest insect infestations. Ecol. Model. 192, 107–125.

Bone, C., White, J.C., Wulder, M.A., Robertson, C., Nelson, T.A., 2013. Impact of forest fragmentation on patterns of mountain pine beetle-caused tree mortality. Forests 4, 279e295.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Cannas, S.A., Marco, D.E., Paez, S.A., 2003. Modelling biological invasions: species traits, species interactions, and habitat heterogeneity. Math. Biosci. 183, 93–110.

Chen, H.P., Walton, A., 2011. Mountain pine beetle dispersal: spatiotemporal patterns and role in the spread and expansion of the present outbreak. Ecosphere 2, art66.

Coops, N.C., Wulder, M.A., White, J.C., 2006. Integrating remotely sensed and ancillary data sources to characterize a mountain pine beetle infestation. Remote Sens. Environ. 105, 83e97.

Coucrelis, H., 1985. Cellular worlds: a framework for modeling micro—macro dynamics. Environ. Plann. A 17 (5), 585–596.

Dale, V.H., Joyce, L.A., McNulty, S., Neilson, R.P., Ayres, M.P., Flannigan, M.D., Hanson, P.J., Irland, L.C., Lugo, A.E., Peterson, C.J., Simberloff, D., 2001. Climate change and forest disturbances. BioScience 51 (9), 723–734.

de la Giroday, H.M.C., Carroll, A.L., Lindgren, B.S., Aukema, B.H., 2011. Incoming! Association of landscape features with dispersing mountain pine beetle populations during a range expansion event in western Canada. Landscape Ecol. 26, 10971110.

Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Mach. Learn. 40, 139–157.

Edburg, S.L., Hicke, J.A., Brooks, P.D., Pendall, E.G., Ewers, B.E., Norton, U., Gochis, D., Gutmann, E.D., Meddens, A.J., 2012. Cascading impacts of bark beetle-caused tree mortality on coupled biogeophysical and biogeochemical processes. Front. Ecol. Environ. 10, 416–424.

Evenden, M.L., Whitehouse, C.M., Sykes, J., 2014. Factors influencing flight capacity of the mountain pine beetle (Coleoptera: Curculionidae: Scolytinae). Environ. Entomol. 43, 187e196.

Franklin, J.F., Shugart, H.H., Harmon, M.E., 1987. Tree death as an ecological process. BioScience 37 (8), 550–556.

García, A.M., Santé, I., Crecente, R., Miranda, D., 2011. An analysis of the effect of the stochastic component of urban cellular automata models Computers. Environ. Urban Syst. 35 (4), 289–296.

Ghimire, B., Rogan, J., Miller, J., 2010. Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the getis statistic. Remote Sens. Lett. 1, 45–54.

Gimblett, H.R. (Ed.), 2002. Integrating Geographic Information Systems and Agent-based Modeling Techniques for Simulating Social and Ecological Processes. Oxford University Press.

Gong, P., Yu, L., Li, C.C., Wang, J., Liang, L., Li, X.C., Ji, L.Y., Bai, Y.Q., Cheng, Y.Q., Zhu, Z.L., 2016. A new research paradigm for global land cover mapping. Ann. Gis 22 (2), 87–102.

Hicke, J.A., Logan, J.A., Powell, J., Ojima, D.S., 2006. Changing temperatures influence suitability of modeled mountain pine beetle (Dendroctonus ponderosae) outbreaks in the western United States. J. Geophys. Res. 111, G02019, http://dx.doi.org/10.1029/2005JG000101.

Hu, Z.Y., Lo, C.P., 2007. Modeling urban growth in Atlanta using logistic regression. Computers. Environ. Urban Syst. 31 (6), 667–688.

Jackson, P.L., Straussfogel Lindgren Mitchell, D.B.S.S., Murphy, B., 2008. Radar observation and aerial capture of mountain pine beetle, Dendroctonus ponderosae Hopk. (Coleoptera: Scolytidae) in flight above the forest canopy. Can. J. For. Res. 38, 23132327.

Johnson, E.W., Ross, J., 2008. Quantifying error in aerial survey data. Aust. For. 71, 216e222.

Kaufmann, M.R., Aplet, G.H., Babler, M.G., Baker, W.L., Bentz, B., Harrington, M., et al., 2008. The Status of Our Scientific Understanding of Lodgepole Pine and Mountain Pine Beetles: A Focus on Forest Ecology and Fire Behavior. Nature Conservancy, Arlington, VA.

Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1: LandTrendr—temporal segmentation algorithms. Rem. Sens. Environ. 114 (12), 2897–2910.

Klutsch, J.G., Negron, J.F., Costello, S.L., Rhoades, C.C., West, D.R., Popp, J., Caissie, R., 2009. Stand characteristics and downed woody debris accumulations associated with a mountain pine beetle (dendroctonus ponderosae hopkins) outbreak in Colorado. For. Ecol. Manag. 258, 641–649.

Kurz, W.A., Dymond, C.C., Stinson, G., Rampley, G.J., Neilson, E.T., Carroll, A.L., . . . Safranyik, L., 2008. Mountain pine beetle and forest carbon feedback to climate change. Nature 452 (7190), 987–990.

Li, X., Gong, P., 2016. Urban growth models: progress and perspective. Science, Bulletin1–Bulleti14.

Li, X., Liu, X., Yu, L., 2014. A systematic sensitivity analysis of constrained cellular automata model for urban growth simulation based on different transition rules. Int. J. Geogr. Inf. Sci. 28 (7), 1317–1335.

Liang, L., Hawbaker, T.J., Chen, Y., Zhu, Z., Gong, P., 2014a. Characterizing recent and projecting future potential patterns of mountain pine beetle outbreaks in the Southern Rocky Mountains. Appl. Geogr. 55, 165–175.

Liang, L., Chen, Y., Hawbaker, T.J., Zhu, Z., Gong, P., 2014b. Mapping mountain pine beetle mortality through growth trend analysis of time-series Landsat data. Remote Sens. 6 (6), 5696–5716.

Liang, L., Hawbaker, T.J., Zhu, Z., Li, X., Gong, P., 2016. Forest disturbance interactions and successional pathways in the Southern Rocky Mountains. For. Ecol. Manage. 375, 35–45.

Liu, X., Li, X., Shi, X., Wu, S., Liu, T., 2008. Simulating complex urban development using kernel-based non-linear cellular automata. Ecol. Model. 211, 169–181.

Logan, J.A., Regniere, J., Powell, J.A., 2003. Assessing the impacts of global warming on forest pest dynamics. Front. Ecol. Environ. 1 (3), 130–137.

Man, G., 2010. Major Forest Insect and Disease Conditions in the United States: 2009 Update. US Department of Agriculture, Forest Service, Forest Health Protection, Washington.

McConnell, T., Johnson, E.W., Burns, B., 2000. A Guide to Conducting Aerial Sketchmap Surveys. USDA Forest Service, Forest Health Technology Enterprise Team, Fort Collins, Colo, FHTET 00-01 88pp.

Meddens, A.J., Hicke, J.A., 2014. Spatial and temporal patterns of Landsat-based detection of tree mortality caused by a mountain pine beetle outbreak in Colorado, USA. For. Ecol. Manage. 322, 78–88.

Meddens, A.J., Hicke, J.A., Vierling, L.A., Hudak, A.T., 2013. Evaluating methods to detect bark beetle-caused tree mortality using single-date and multi-date Landsat imagery. Remote Sens. Environ. 132, 49e58.

Meigs, G.W., Kennedy, R.E., Cohen, W.B., 2011. A Landsat time series approach to characterize bark beetle and defoliator impacts on tree mortality and surface fuels in conifer forests. Remote Sens. Environ. 115, 37073718.

Meigs, G.W., Kennedy, R.E., Gray, A.N., Gregory, M.J., 2015. Spatiotemporal dynamics of recent mountain pine beetle and western spruce budworm outbreaks across the Pacific Northwest Region, USA. For. Ecol. Manage 339, 71–86.

Mustafa, A., Saadi, I., Cools, M., Teller, J., 2014. Measuring the effect of stochastic perturbation component in cellular automata urban growth model. Procedia Environ. Sci. 22, 156–168.

Pérez, L., Dragićević, S., 2011. ForestSimMPB: a swarming intelligence and agent-based modeling approach for mountain pine beetle outbreaks. Ecol. Inf. 6 (January (1)), 62–72.

Parker, D.C., Hessl, A., Davis, S.C., 2008. Complexity, land-use modeling, and the human dimension: fundamental challenges for mapping unknown outcome spaces. Geoforum 39 (2), 789–804.

Preisler, H.K., Hicke, J.A., Ager, A.A., Hayes, J.L., 2012. Climate and weather influences on spatial temporal patterns of mountain pine beetle populations in Washington and Oregon. Ecology 93, 24212434.

Pukkala, T., Möykkynen, T., Robinet, C., 2014. Comparison of the potential spread of pinewood nematode (Bursaphelenchus xylophilus) in Finland and Iberia simulated with a cellular automaton model. For. Pathol. 44 (5), 341–352.

Raffa, K.F., Berryman, A.A., 1983. The role of host plant resistance in the colonization behavior and ecology of bark beetles (Coleoptera: Scolytidae). Ecol. Monogr. 53, 27e49.

Raffa, K.F., Aukema, B.H., Bentz, B.J., Carroll, A.L., Hicke, J.A., Turner, M.G., Romme, W.H., 2008. Cross-scale drivers of natural disturbances prone to anthropogenic amplification: the dynamics of bark beetle eruptions. Bioscience 58, 501e517.

Robertson, C., Nelson, T.A., Boots, B., 2007. Mountain pine beetle dispersal: the spatialetemporal interaction of infestations. For. Sci. 53, 395e405.

Robertson, C., Wulder, M.A., Nelson, T.A., White, J.C., 2008. Risk rating for mountain pine beetle infestation of lodgepole pine forests over large areas with ordinal regression modelling. For. Ecol. Manage. 256, 900e912.

Robertson, C., Nelson, T.A., Jelinski, D.E., Wulder, M.A., Boot, B., 2009. Spatialetemporal analysis of species range expansion: the case of the mountain pine beetle, Dendroctonus ponderosae. J. Biogeogr. 36, 14461458.

Safranyik, L., Silversides, R., McMullen, L.H., Linton, D.A., 1989. An empirical approach to modeling the local dispersal of the mountain pine beetle (Dendroctonus ponderosae Hopk.) (Col., Scolytidae) in relation to sources of attraction, wind direction and speed. J. Appl. Entomol. 108, 498e511.

Safranyik, L., Linton, D.A., Silversides, R., McMullen, L.H., 1992. Dispersal of released mountain pine beetles under the canopy of a mature lodgepole pine stand. J. Appl. Entomol. 113, 441e450.

Saltelli, A., Chan, K., Scott, E.M. (Eds.), 2000. Sensitivity Analysis, vol. 1. Wiley, New York.

Santé, I., García, A.M., Miranda, D., Crecente, R., 2010. Cellular automata models for the simulation of real-world urban processes: a review and analysis. Landscape Urban Plann. 96 (2), 108–122.

Schmidt, F., Persson, A., 2003. Comparison of DEM data capture and topographic wetness indices. Precis. Agric. 4, 179–192.

Simard, M., Powell, E.N., Raffa, K.F., Turner, M.G., 2012. What explains landscape patterns of tree mortality caused by bark beetle outbreaks in Greater Yellowstone? Global Ecol. Biogeogr. 21, 556e567.

Torrens, P.M., 2000. How Cellular Models of Urban Systems Work (1. Theory); CASA Working Paper Series (28). Centre for Advanced Spatial Analysis (UCL), London, UK, Available online: http://www.casa.ucl.ac.uk/working_papers/paper28.pdf (Accessed 17 August 2009).

Trumbore, S., Brando, P., Hartmann, H., 2015. Forest health and global change. Science 349 (6250), 814–818.

Turner, M.G., 2010. Disturbance and landscape dynamics in a changing world. Ecology 91, 2833–2849.

White, R., Engelen, G., 1993. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. Environ. Plann. 25 (8), 1175–1199.

White, J.C., Wulder, M.A., Brooks, D., Reich, R., Wheate, R.D., 2004. Mapping mountain pine beetle infestation with high spatial resolution satellite imagery. For. Chron. 80, 743e745.

Yang, Q., Li, X., Shi, X., 2008. Cellular automata for simulating land use changes based on support vector machines. Comput. Geosci. 34, 592–602.

Yassemi, S., Dragićević, S., Schmidt, M., 2008. Design and implementation of an integrated GIS-based cellular automata model to characterize forest fire behaviour. Ecol. Modell. 210 (1), 71–84.

Yeh, A.G., Li, X., 2006. Errors and uncertainties in urban cellular automata computers. Environ. Urban Syst. 30 (1), 10–28.