

# **Stochastic spread models: a comparison between an individual-based and a lattice-based model for assessing the expansion of invasive termites over a landscape**

Francesco Tonini<sup>a, \*</sup>, Hartwig H. Hochmair<sup>a</sup>, Rudolf H. Scheffrahn<sup>a</sup>, Donald L. DeAngelis<sup>b</sup>

<sup>a</sup> *University of Florida, Fort Lauderdale Research and Education Center, 3205 College Avenue, Davie, Florida, 33314, U.S.A.*

<sup>b</sup> *Dept. of Biology, University of Miami, PO Box 249118, Coral Gables, FL 33124, U.S.A.*

*\*Corresponding author. Tel.: +1 954 577 6392; fax: +1 954 424 6851.*

*E-mail addresses: ftonini@ufl.edu (Francesco Tonini), hhhochmair@ufl.edu (Hartwig H. Hochmair), rhsc@ufl.edu (Rudolf H. Scheffrahn), Donald L. DeAngelis (ddeangelis@bio.miami.edu)*

## **Abstract**

Spatially-explicit simulation models can help state and local regulatory agencies to predict both the rate and direction of the spread of an invasive species from a set of surveyed locations. Such models can be used to develop successful early detection, quarantine, or eradication plans based on the predicted areas of infestation. Individual-based models (IBMs) are often used to replicate the dynamics of complex systems and are both able to incorporate individual differences and local interactions among organisms, as well as spatial details. In this work, we introduce a new stochastic lattice-based model for simulating the spread of invasive termites over a landscape and compare it to a recently published stochastic individual-based approach,

based on the same ecological parameters, with the goal of improving its computational efficiency. The two modeling frameworks were tested over a homogeneous landscape with randomly located sources of infestation. Further, the setting of a case-study of an invasive termite, *Nasutitermes corniger* (Motschulsky), was used to simulate the spread of the species in Dania Beach, Florida, U.S.A., and results of the proposed model were compared with an earlier application of the IBM over the same area. Results show that the extent of the infested areas predicted by the new lattice-based model are similar, thus comparable, to the individual-based model while improving computation time significantly. The simulation presented in this work could be used by regulatory authorities to draw one or more areas of intervention instead of wasting resources by randomly surveying unknown perimeters.

*Keywords:* Spatio-temporal modeling, cellular automata, invasive species, individual-based model, simulation methods

## **1. Introduction**

Termites are serious pests of the urban environment and are responsible for severe damage to structural lumber (Rust and Su, 2012). It has been estimated that every year, in the continental United States alone, termites cause property damage up to billions of dollars (Edwards and Mill, 1986). Although human-assisted (also called anthropogenic) dispersal can play a significant role in the rate of expansion of an established exotic species, its nature is complex and unpredictable. Termites are not "hitchhiking" insect pests, i.e. they cannot easily be transported aboard commercial vehicles such as cars or boats. More specifically, the nesting core of a termite colony must be moved intact and both a water and food source must be available to the core throughout

the movement (Hochmair and Scheffrahn, 2010). The natural spread of termites is expected to proceed fairly slowly for two main reasons: (i) they are weak fliers and their reproductives (winged individuals) can fly only a few hundred meters from the parent colony on average each year (Nutting, 1969); and (ii) a termite colony takes at least four years to mature and release the first reproductives from the nest (Evans, 2011). Given the unpredictability of human-assisted movements, it will be more useful here to restrict our study on termite dispersal by natural means and attempt to anticipate the rate and direction of termite spread.

The recent integration of Geographical Information Systems (GIS) with simulation models, coupled with advances in computing power, has allowed the development of spatially-explicit and dynamic simulation models, such as individual-based models (IBMs, hereafter), and lattice-based models, such as cellular automata (CA, hereafter) (Brown et al., 2005). Both types of models have been introduced in ecological modeling to capture the inherent complexity of various real-world problems (Steyaert, 1993) as alternative approaches to solving mathematical sets of equations such as partial differential equations (PDEs) (Alexanderian et al., 2011; Holmes et al., 1994). Moreover, the intricacy of a physical environment limits the applicability of mathematical models for modeling realistic dispersal of invasive species (Pitt, 2008). GIS allows for spatial complexity and simulation modeling takes GIS visualizations into the domain of temporal dynamics. Early examples of GIS-based simulation models include land-use change based on the Markov process (Burnham, 1973), and discrete state models of flow and transport (e.g. transport of pollutants) (Maidment, 1996). Other more recent computer-intensive approaches have been able to incorporate ecological information at a population level with landscape complexity (Wiegand et al., 2004).

While IBMs incorporate rules to describe the interactions between individual units, such as organisms, in which each individual can have a different set of behavioral, physiological, and other properties (Huston et al., 1988), in CA modeling the basic units of the simulation are discrete spatial cells, which can transition among different states (e.g. empty, infested, quantified occupation, etc.). CAs are the most basic form of lattice-based model and are able to model complex dynamics from biological, social, or physical processes by using simple rules (Toffoli and Margolus, 1987). Several improvements have been made to basic CAs in the past decade to include underlying landscape differences, more complex rules for changes of state, and stochastic parameters. Some examples of lattice-based models include wildfire spread models (Clarke et al., 1994), spread models of invasive ants (Pitt, 2009), spatial dynamics of urban and regional systems (Clarke and Gaydos, 1998), and epidemic propagation (Mikler et al., 2005; Morley and Chang, 2004). In the past years, different computer simulation approaches have been either compared (Ajelli et al., 2010; Lett et al., 1999) or integrated (Sudshira, 2010).

In this work, we develop a new stochastic lattice-based model in order to simulate the spatiotemporal spread of invasive termites. The model developed herein includes a set of relevant ecological parameters that can be modified according to the species of interest. Our model differs from more traditional CA approaches in that it is not based on rules for state transition depending on the state of the focal cell as well as the states of a defined set of neighboring cells. A recently published stochastic IBM for the spread of invasive termites (Tonini et al., 2013) was used as a benchmark for comparison with the new model in terms of predictions, computational runtime, and code efficiency. The main purpose of the new model is to improve computational speed over the IBM model, which is costly in terms of computational time. We believe the model proposed herein represents a substantial improvement to the IBM

and could give regulatory agencies a better tool for targeting specific areas for survey, eradication, or quarantine effort.

The format of the paper is as follows. The new model, its framework and main parameters are described in Section 2. Results are presented in Section 3, while their interpretation is discussed in Section 4. Finally, conclusions on the advantages and limitations of the models tested herein is provided in Section 5.

## **2. Materials and methods**

A side-by-side comparison between the proposed lattice-based model and the IBM by Tonini et al. (2013) was done by establishing a common ground in order to discount unwanted effects due to their different frameworks. Both models were used in two different settings: (i) a homogeneous landscape, where different sets of  $N$  termite colonies were randomly located and considered as sources of invasion, and (ii) a case-study of an invasive termite, *Nasutitermes corniger* (Motschulsky), in Dania Beach, Florida, U.S.A. (Scheffrahn et al., 2002). The average area occupied at each time step was computed over all model replications (Monte Carlo estimation) and its value compared between the two modeling approaches. Moreover, the computational time was considered to assess the speed of the proposed model against the established IBM.

### *2.1. Model design*

The lattice-based model proposed herein is spatially-explicit and stochastic with the purpose of simulating the spatiotemporal spread of a termite invasion by natural means over a realistic landscape (e.g. urban environment), starting from a set of initial surveyed colonies. The basic units of the model are cells of a grid with a pre-defined spatial resolution. Each cell may assume one or more states expressed by the number of termite colonies contained in it. The temporal scale chosen for the simulations was discrete with one time step representing a year. In order to keep the model comparable to the IBM by Tonini et al. (2013), the resolution of the spatial grid was set to 100 x 100 meters (= 1 hectare).

The main ecological parameters, shown in Table 1, were kept identical to the IBM. A general description of each one of them follows here.

**Table 1.** Model ecological parameters used in both modeling approaches and their values.

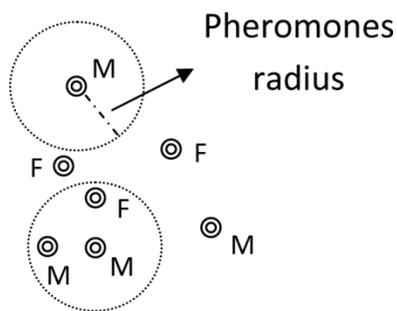
Parameter	Definition	Values
AFP	Colony age at first production of reproductives	4 yr
PHR	Maximum pheromone attraction distance	3 m
DEN	Maximum density of colonies per grid cell	1 (homogeneous landscape setting) 7 (Dania Beach case study)
SURV	Overall survival rate of reproductives *	0.01 (1%)
MAR	Prevalence of male reproductives in the colony	0.5 (50%)
SCR	Scenario of amount of reproductives generated by a colony	1,000 for colony age $4 \leq \text{yr} < 9$ 10,000 for colony age $9 \leq \text{yr} < 14$ 100,000 for colony age $\geq 14$ yr
DIST	Mean dispersal flight distance	200 m

\* Overall percentage of alates surviving all phases of a dispersal flight

*Colony age at first production of reproductives.* The age at which a termite colony starts generating the first crop of winged reproductives (AFP, hereafter) contributes to the rate of

spread of an invasion. The earlier the first reproductives are generated, the faster the invasion will proceed.

*Pheromone attraction distance.* Termite reproductives find potential mates of the opposite sex by sensing and responding to pheromones after concluding the dispersal flight (Bordereau and Pasteels, 2011). A new colony begins with a male-female (i.e. king and queen) couple of unwinged reproductives starting the nest in a proper substrate, such as soil or wood. The maximum pheromone attraction distance (PHR, hereafter) affects the chance that two heterosexual individuals find each other after the dispersal flight (see Fig. 1). The smaller this maximum distance is set, the smaller the number of new termite colonies will be.



**Figure 1 Schematic example of termite reproductives and the pheromone attraction radius within which a heterosexual individual must be found to mate after a dispersal flight. The male at the top is beyond detection range of a female while the males at the bottom left are in female detection range.**

*Colony density.* The maximum density of termite colonies over an area (DEN, hereafter) avoids overpopulation of grid cells in the simulation model. The higher this parameter is, the higher the chance for a nearby cell to become infested.

*Survival rate.* The overall survival rate of reproductives (SURV, hereafter) in a termite colony determines the number of reproductives that survive a dispersal flight and can thus potentially mate. Survival is expressed as an overall rate that considers both predation and injuries that

typically occur as reproductives start leaving the nest (i.e., preflight), volent predators in flight (capture by bats and birds), and nonvolent predators (e.g. ants and herps) as soon as they alight on the ground or on a tree (i.e., postflight) to search for a mate.

*Male prevalence.* The prevalence of male reproductives in a colony (MAR, hereafter) describes the percentage of males among the reproductives and affects the chance that a heterosexual pair unites within the pheromone attraction distance after the dispersal flight. That is, the chance decreases the further the value deviates from 50 percent of males.

*Amount of reproductives generated by a colony.* The number of reproductives generated by a colony (SCR, hereafter) after reaching maturity (determined by the aforementioned AFP parameter) increases with age. Different scenarios representing the total number of reproductives by colony age can be chosen. A higher number of reproductives increases the chance of new colonies being formed.

*Dispersal distance.* The mean dispersal flight distance (DIST, hereafter) determines how far termite reproductives are able to fly, on average, on their own. The higher this parameter value, the faster the termite spatial expansion will proceed.

The age of each colony is tracked along with the number of colonies within each grid cell. Depending on the number of colonies, their age, and the SURV parameter, each cell can be expressed as the total number of reproductives available to either stay within it or fly toward neighboring cells. The main process involved in the spread of termite colonies is the dispersal of reproductives. Typically, dispersal flights are initiated in the beginning of the wet season and are triggered by environmental factors (Jones et al., 1988; Martius, 2003; Nutting, 1969). Termites can only spread across a landscape and form new colonies beyond the territory occupied by the mother colony by using dispersal flight (excluding occasional and unpredictable human-assisted

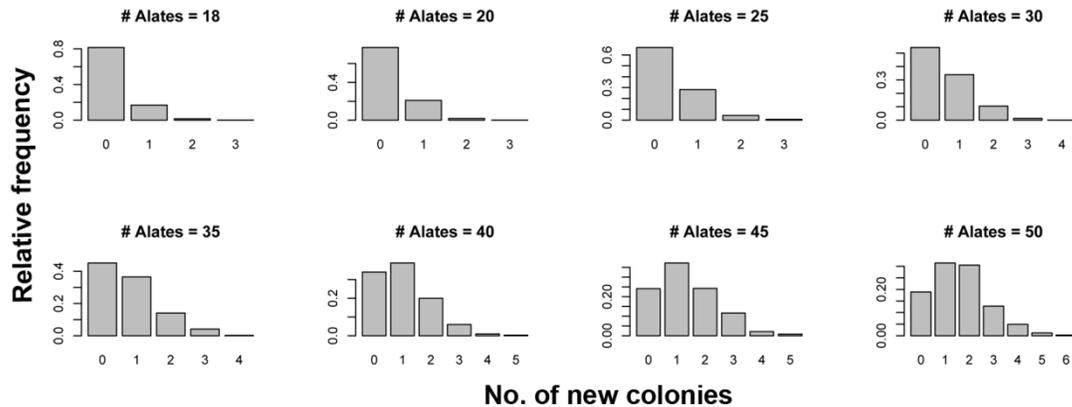
introductions). In this study, as it was assumed for the IBM study, the dispersal and the subsequent creation of new colonies was simulated as a single annual event. This simplification is necessary because of the extreme uncertainty in the frequency and magnitude of dispersal flights during each flight season. A two-parameter dispersal kernel (Clark, 1998) was used to express the proportion of reproductives that fly up to a certain distance from the mother colony:

$$f(x) = \frac{c}{2\alpha\Gamma(1/c)} \exp\left(-\left|\frac{x}{\alpha}\right|^c\right)$$

where  $\Gamma()$  is the gamma function,  $\alpha$  is the average distance parameter,  $c$  the shape parameter, and  $x$  the distance from the centroid of the central cell. The value of  $\alpha$  was set to have the same value of the baseline distance parameter in the IBM equivalent, i.e. 200 m. The kernel is exponential when  $c = 1$ , "exponentially-bounded" when  $c \geq 1$ , "fat-tailed" when  $c \leq 1$ , and Gaussian when  $c = 2$ . The probability kernel used in this study was chosen to be exponential to be consistent with the IBM model but can be changed to match a desired shape for the dispersal function. Probability densities were generated from a grid expressing values as Euclidean distances (in meters) from the central cell. This was accomplished first by replacing  $x$  in the equation with all the distances and then by normalizing the resulting  $f(x)$  values with their sum in order to obtain a true probability density kernel to use as weight matrix in the focal operation.

Termite reproductives find potential mates of the opposite sex by sensing and responding to pheromones after concluding the dispersal flight. A new colony begins with a male-female (i.e. king and queen) couple of unwinged reproductives starting the nest in a proper substrate, such as soil or wood. In the IBM model, the final location of each individual reproductive was followed and mating was assumed when two individuals of the opposite sex came within a sufficiently short distance. Thus reproduction was conceptually simple in the IBM, but very costly in terms of computational time, which the lattice-based model attempts to avoid. Therefore, a new

approach to estimating reproduction was used in which the number of new colonies formed by  $N$  reproductives, given different values of the MAR parameter, is again stochastic. This number was derived from a separate simulation study of individually moving reproductives, from which a permanent lookup table could subsequently be used for all future simulations of the lattice-based model. Using a lookup table rather than performing a detailed individual-based simulation for each reproductive event saves a great deal of time in simulations. In order to account for stochasticity in the formation of new colonies, we used a sequence of values of  $N$  randomly placed reproductives and ran 1000 replications for each of them. Fig. 2 shows some of the empirical histograms extracted from the aforementioned simulation for some values of  $N$ .



**Figure 2 A sampling of empirical histograms for the number of new colonies formed by  $N$  reproductives (alates). Histograms are derived from 1000 replications.**

A more extensive set of histograms can be found in Appendix A (online version). A numerical procedure was necessary because a formal mathematical derivation would be highly complex. Specifically, this is a combinatorial problem, in which any reproductive individual can be by chance a close neighbor of multiple other individuals. Moreover, once a heterosexual pair is found within a pre-defined pheromone attraction radius, those individuals that have mated will

not be available any more for others to mate with. Due to this complexity, simulations to develop a look-up table for future use was the most practical solution.

Because the lattice-based model is stochastic, it was run 100 times in order to account for the uncertainty associated with the outcome of a stochastic simulation. The number of replications was chosen to allow comparison with the IBM and was large enough to produce tight upper and lower simulated “envelopes” around the mean predicted area of infestation.

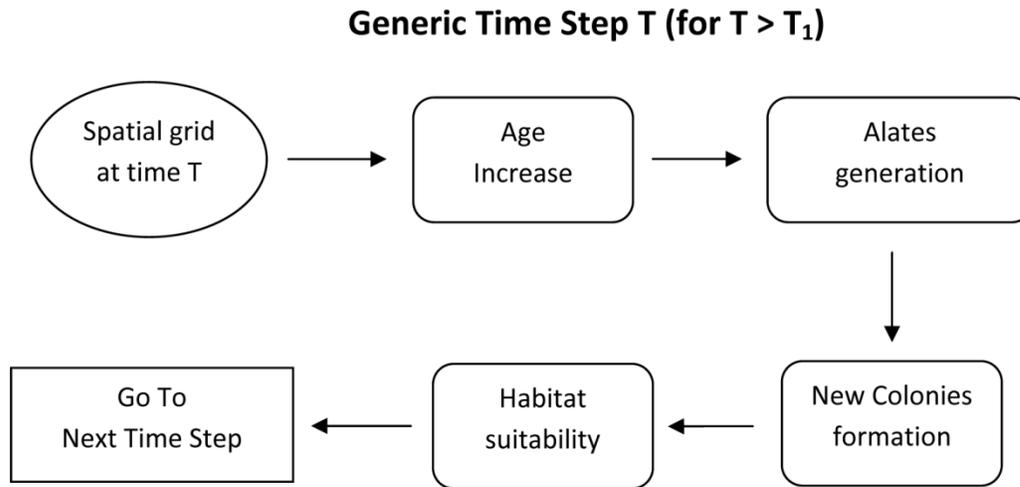
## *2.2. Framework and modules*

The simulation algorithm was implemented using a set of R functions (R Development Core Team, 2013) and the source code is provided freely<sup>1</sup>. The simulation is composed of several modules ordered in a sequential manner. The main script calls external modules that follow the steps involved in the spread of a termite invasion from the creation of reproductives to the final creation of new colonies, which are identical to those of the IBM study by Tonini et al. (2013): (1) the ecological parameters are set; (2) in case the simulation is meant to be run over a realistic landscape, a suitability habitat GIS layer is read; (3) a set of initial colonies (sources of invasion) are either read from an external file with spatial coordinates associated to each record or randomly located over the study area; (4) a spatial grid with pre-defined resolution is created and each cell is assigned the number of colonies falling within it; and (5) if the DEN parameter was set at (1), a random elimination of colonies exceeding the limit on colony carrying capacity within each cell is performed. During step 3, in the absence of specific information, each colony is assigned a random age between 0 and a pre-defined maximum lifespan.

---

<sup>1</sup><https://github.com/f-tonini/Termite-Dispersal-Simulation>

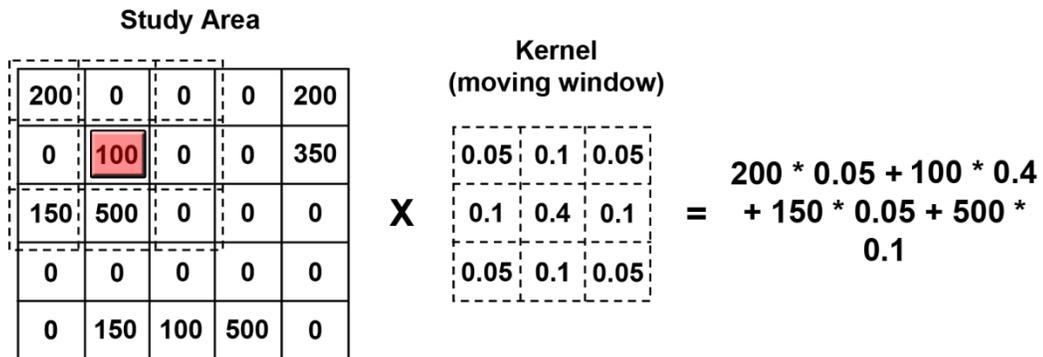
Fig. 3 shows a conceptual representation of the main modules (four boxes to the right) involved in the simulation algorithm at any generic time step.



**Figure 3 Conceptual diagram of the main modules used in the simulation algorithm at any generic time step**

At each time step, the age of each colony is increased by one (age increase module) and colonies exceeding the maximum lifespan age are removed. After each cell has been updated with the number of colonies within it, termite reproductives are generated depending on the colonies' ages (alate generation module) and their total number summed over the grid. At this stage, a dispersal kernel is used as a matrix of weights (probabilities) in combination with a summation function to calculate numbers of reproductive arriving at a focal cell from both itself and from neighboring cells. Therefore, a focal ("moving window") operation, shown in Fig. 4, can be used to return new values for the spatial grid representing the final number of potential reproductives available after the dispersal flight that may form new colonies. The 'moving window' is placed such that the focal cell is in the center. Estimates of the probabilities of reproductives from each of the eight surrounding cells and the focal cell itself arriving at the

focal cell constitute the dispersal kernel. These probabilities are then used as multipliers of the number of reproductives in the focal and adjacent cells to calculate an estimate of the number of new reproductives that arrive in the focal cell.



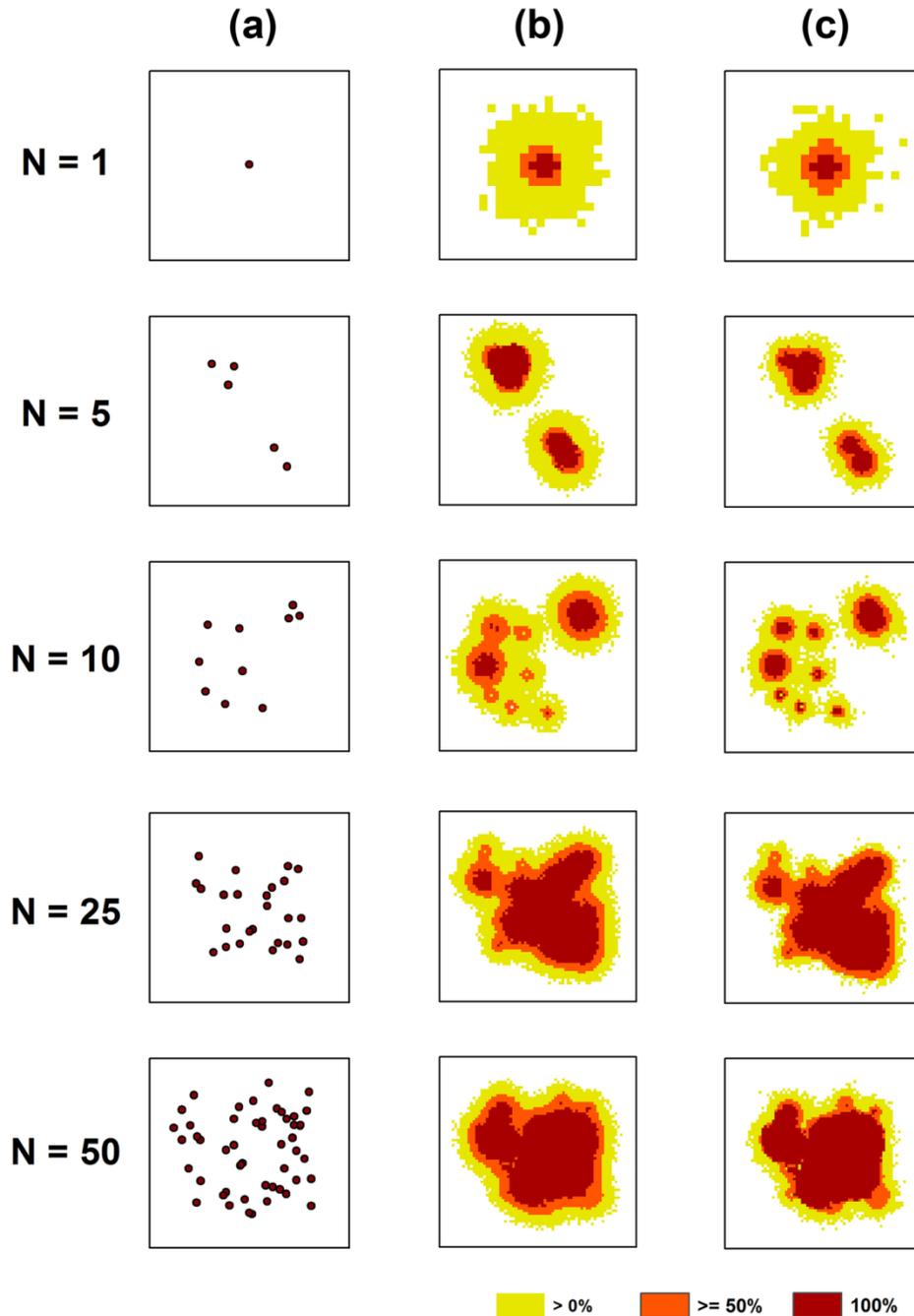
**Figure 4 Schematic representation of the focal ("moving window") operation. Focal values are calculated for the neighborhood (dashed line) of focal cells (red in this example) using the dispersal kernel as matrix of weights, in combination with a summation function. Available in color online.**

Next, the number of new colonies formed in each cell (new colonies generation module) is determined by randomly sampling from the look-up table (see Section 2.1). If the model is run over a realistic landscape, all cells occupied by colonies falling within a patch of habitat classified as unsuitable are removed (habitat suitability module). Before moving to the next time step, the spatial grid is saved as a geo-referenced raster grid to be overlaid at the end of the simulation with all other replications for a given year. This operation is the same used by Tonini et al. (2013) to compute a final occupancy envelope.

### 3. Results

### *3.1. Homogeneous Landscape*

Several initial scenarios were tested by randomly placing termite colonies across a spatially homogeneous study area, without any habitat limitations for the species. These initial scenarios were used as the first time step to run both the new lattice-based model and the IBM by Tonini et al. (2013) under the same conditions. Fig. 5 shows the results obtained from the stochastic outcomes of 100 model replications after 20 time steps for both the IBM and lattice-based models, considering different values of  $N$  randomly located colonies. The outcomes of the 100 replications are grouped such that they can be visualized by three color-coded occupancy envelopes.

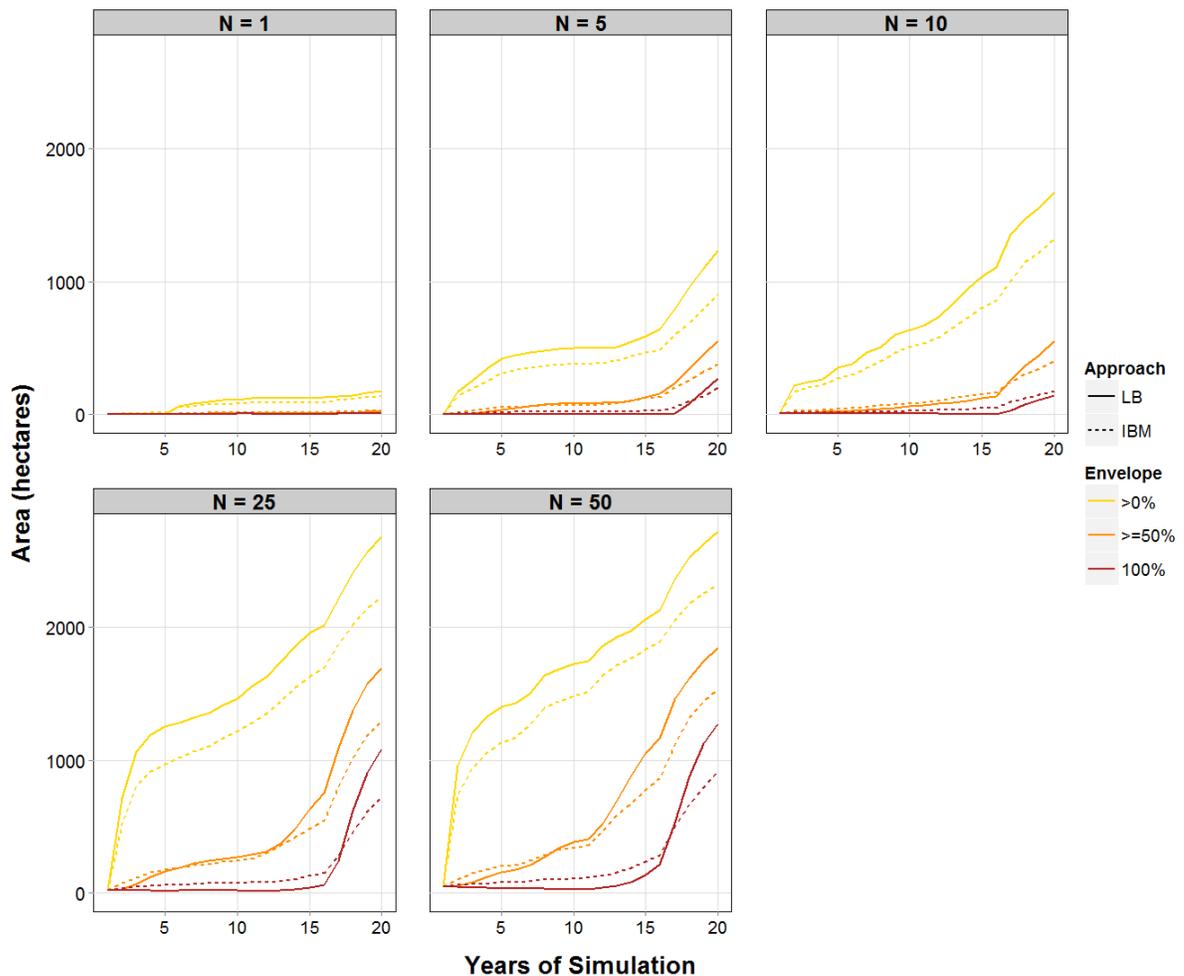


**Figure 5 Final outcome of (b) the lattice-based model and (c) the individual-based model. Initial scenarios of  $N$  randomly located colonies are shown in (a). Available in color online.**

The “ $>0\%$ ” (yellow) occupancy envelope represents all areas predicted to be occupied by the model in one or more simulation runs. Similarly, a “ $\geq 50\%$ ” (orange) occupancy represents all

predicted areas that were occupied by at least half of all simulation runs. Finally, the “100%” (red) occupancy envelope represents areas that are predicted as occupied by all runs. Although the shapes of the envelopes computed by the two approaches are similar, there are a few differences, especially in the broadest (yellow) envelope.

Fig. 6 shows a comparison between both simulation approaches in terms of occupied areas over 20 years of simulation and grouped by the same aforementioned occupancy envelopes.



**Figure 6 Predicted areas (in hectares) of infestation for the lattice-based model (LB, solid line) and the individual-based model (IBM, dashed line) for different values of N randomly distributed initial colonies. Occupancy envelopes are color coded: “>0%” (yellow), “>=50%” (orange), and “100%” (red). Available in color online.**

The broadest envelope (yellow, “>0%”) shows the largest differences between the two approaches, with the lattice-based model predicting a larger infested area. While there is less clear pattern for the “≥50%” and “100%” envelopes, the lattice-based model predicts in most cases a larger area than the IBM with increasing simulation years. Table B.1 (Appendix B, online version) shows detailed results of the multi-run average occupied area at each time step by both modeling approaches.

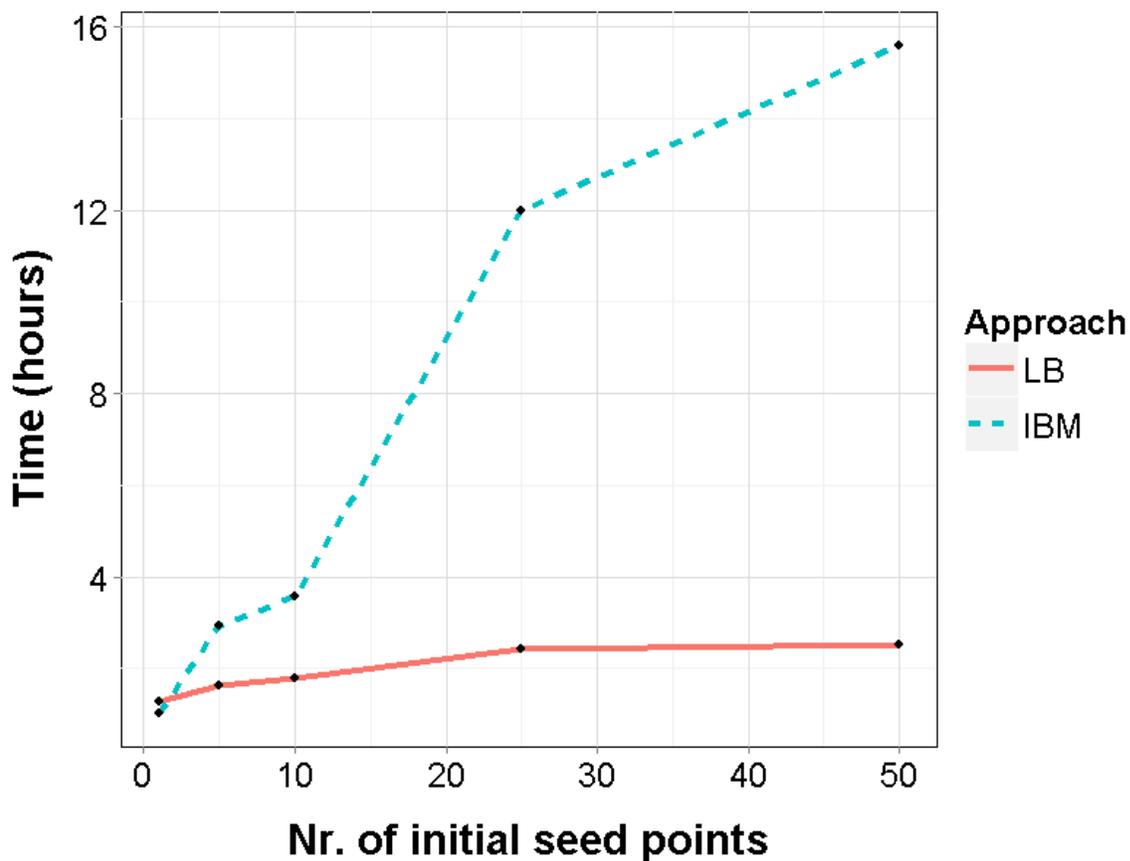
Differences between the lattice-based model and the IBM in terms of predicted areas of infestation for each simulated year were summed up across the entire timeline (Table 2). The “>0%” envelope shows the largest positive differences with an increasing magnitude for a larger number of N initial colonies. Again, there is no clear pattern for either the “≥50%” or the “100%” envelope. In the first case, positive and negative differences alternate for different values of N. In the second case, although the magnitude of differences is much smaller compared to the previous case, negative differences increase or decrease for different values of N.

**Table 2.** Sum of differences between predicted areas (in hectares) of infestation by the lattice-based model and the individual-based model over all simulation years for different values of N randomly distributed initial colonies (columns). Values are reported by occupancy envelope type (rows). Positive values indicate that the lattice-based model predicts a larger infested area compared to the IBM. Negative values indicate the opposite.

	N=1	N=5	N=10	N=25	N=50
> 0%	407	2669	3308	5660	5719
≥ 50%	0	446	-22	1867	2018
100%	0	-222	-485	-41	-70

The computational time of the two modeling approaches needed to complete the simulation and to save all results to external geo-referenced raster grids was also compared. Fig. 7 shows a

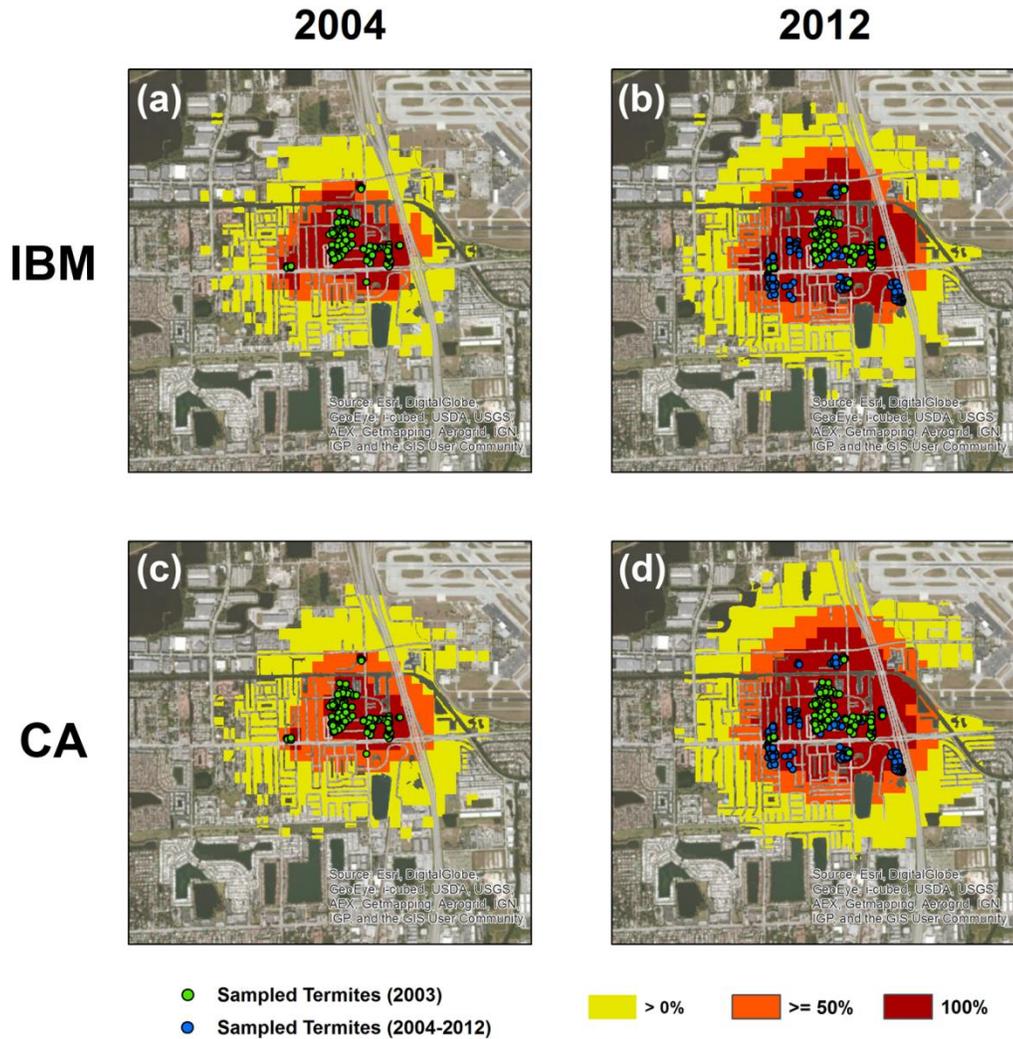
comparison of the speed (in hours) given the different values of N randomly distributed initial colonies. The runtime needed to build the look-up table (see Section 2.1) was around 48 hours. However, this was not factored in for Fig. 7 because it is a one-time operation that does not need to be repeated every time the lattice-based model is run. All simulations were run on a 3.30 GHz Intel(R) Core(TM) i7-3960X CPU with 32 GB RAM.



**Figure 7** Computational time (in hours) of the lattice-based model (LB, solid line) and the individual-based model (IBM, dashed line) for different values of N randomly distributed initial colonies. Available in color online.

### 3.2. Case Study: Dania Beach, Florida

The proposed lattice-based model was also applied to the case study presented in Tonini et al. (2013) in order to compare its stochastic outcome to the IBM. Both models used surveyed *N. corniger* colonies (Scheffrahn et al., 2002) as initial sources of dispersion and the simulated spread after 9 years was compared to the observed infested sites discovered in 2012. Values used for the Dania Beach case study were assumed identical to the ones used for the IBM. Fig. 8 shows a snapshot of two years extracted from the simulation in both approaches and the three types of occupancy envelopes (see Section 3.1). The infested locations by *N. corniger* observed in different years are also shown.



**Figure 8 Snapshot of the areas predicted as infested by the simulation models with sampled termite locations in 2003 and the period 2004-2012. (a-b) Individual-based model. (c-d) Lattice-based model. Yellow, orange, and red cells indicate the > 0%, ≥ 50%, and 100% occupancy envelopes, respectively. Available in color online.**

#### **4. Comparison of Model Results and Discussion**

Fig. 5 provides a comparison of the results between the lattice-based model and the IBM after 20 years of simulation. Although some differences can be seen across all envelopes, the number of initial colonies does not seem to have a great impact on the differences between the

spatial extents of the areas predicted by the two approaches. Visual inspection suggests that the broadest envelope (" $>0\%$ ") is the one with consistently larger differences. To assess envelope dissimilarities throughout all simulation time steps, one can refer to results in Fig. 6. An increasing trend in the dissimilarity between envelopes mirrors with an increasing number of initial colonies. However, while the " $\geq 50\%$ " and the " $100\%$ " envelopes look more similar for the two modeling approaches and do not show a consistent over/under estimation over time, the " $>0\%$ " envelope confirms the pattern observed in Fig. 5. Specifically, the lattice-based model seems to always estimate a slightly larger infested area, and this difference starts to increase approximately after the first 2-3 years of simulation. This can be explained by the larger standard deviation found in the estimated occupied area over the 100 model replications for the cell based model for the " $>0\%$ " envelope (see Table B.1, Appendix B, online version). Since the broadest (yellow) envelope is created to show spots that are predicted by one or more model replications, it is subject to high stochasticity, so that it is natural that a higher standard deviation in the occupied area is reflected more for this envelope than for the others.

Table 2 partially confirms what already seen in both Fig. 5 and Fig. 6. The lattice-based model predicts a larger overall infested area (positive difference) compared to the IBM when considering the broadest envelope (" $>0\%$ "). Both the " $\geq 50\%$ " and the " $100\%$ " envelopes cover exactly the same infested area when starting the simulation with a single colony ( $N = 1$ ). Differences in area over all simulated years are also small for  $N=5$  and  $N=10$ . However, the two modeling approaches start diverging for  $N \geq 25$  when looking at the " $\geq 50\%$ " envelope. Differences between the two modeling approaches are less marked for the most probable envelope (" $100\%$ "), with an absolute magnitude decreasing for  $N > 10$ .

As pointed out by Tonini et al. (2013), areas predicted by both the ">0%" and "≥50%" envelopes are conservative estimates that are likely to overestimate the true expansion and should not necessarily be looked at as a serious threat. As opposed to this, areas predicted by the "100%" envelope should instead be carefully inspected and monitored by local authorities as they are consistently predicted to be infested by all model replications. Higher values were also tested for the DEN parameter, in the range from 5 to 10. All models had a similar outcome to that presented herein.

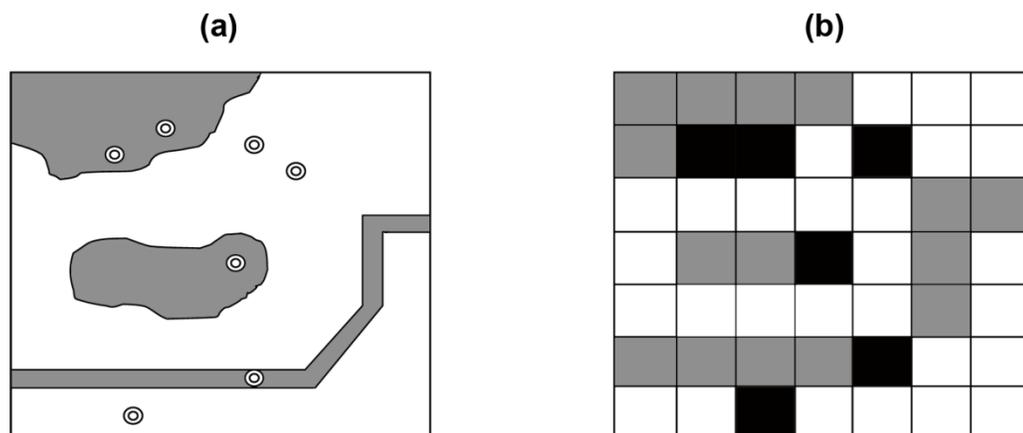
The biggest advantage of the new lattice-based model compared to the IBM can be seen clearly in Fig. 7, where computational speed is compared for several values of initial seed points (termite colonies). The larger the number of initial points, the longer it takes for the IBM to complete the simulation, as opposed to the lattice-based model which seems to tend asymptotically to a fairly stable value. Specifically, with 10 initial colonies the computational time of the IBM model already doubles that of the lattice-based one (~3.5 hours vs. ~1.8 hours). This difference increases further up to a factor six with 50 initial colonies (~15.6 hours vs. ~2.5 hours). This result should encourage the use of the proposed lattice-based model over the IBM, unless the simulation is being used to simulate the spread of invading termites from a small set of initial surveyed colonies, in which case the IBM is not much more costly and may be more accurate. Moreover, the aforementioned result was obtained from multiple simulations over a homogeneous landscape. Models allowing a higher maximum density of colonies per grid cell tended to slow down the simulation in both modeling approaches. However, while the lattice-based model only had a slight variation in speed, the IBM suffered more extreme effects, which almost halved its speed.

The runtime speed improvements with an increasing number of initial colonies (Fig. 7) suggest that the proposed lattice-based model has a lower runtime complexity compared to the IBM model. In Computer Science, the performance or complexity of an algorithm is typically described by the “Big O” notation (Reed, 2004). This can be used to describe the execution time required by an algorithm and it is useful to understand how complexity scales with the size of a problem. By ignoring all the built-in R functions used throughout the code both programs would be  $O(N^2)$ , where  $N$  represents the size of the initial number of termite colonies, because they iterate over multiple model replications and years of simulation. However, the two modeling approaches use a different number of nested loops within the modules generating new termite colonies. When considering the nested loops within each model, the lattice-based approach becomes  $O(N^3)$ , and the IBM complexity becomes  $O(N^4)$ . In other words, their performance is directly proportional to the cube and the fourth of the size of the initial number of termite colonies, respectively. This explains the divergent runtimes between the two approaches. Although a general estimate of the complexity of each algorithm can be derived, it should be noted that both programs still call a number of built-in R functions within their code. Therefore, it is hard to exactly quantify how the overall complexity is affected by these. The lattice-based model makes an extensive use of functions belonging to the “apply” family, which pre-allocate vector sizes and make the code run faster. However, these cannot be used to replace loops where each iteration is dependent on the result of previous iterations. We acknowledge that all the results presented herein should be interpreted as a comparison of the particular programs. It is possible that better algorithms would improve the timing.

The results obtained with the lattice-based model for the Dania Beach case study are also encouraging in the sense that both approaches have a similar outcome. Not only do all three

occupancy envelopes approximately cover the same areas, but the infested sites found between 2004 and 2012 all fall within the most probable envelope. Although this result does not represent a true model evaluation, given the limitations of the empirical data pointed out by Tonini et al. (2013), it corroborates the lattice-based simulation model in so far as its results enclose the surveyed locations. The "100%" envelope is smaller in the lattice-based model than in the IBM because the higher standard deviation increases the chance that at least a few grid cells will not be occupied in some of the model replications.

A major difference between the IBM and the new lattice-based model is represented by the way the underlying suitable habitat is represented, as shown in Fig. 9.



**Figure 9 Habitat representation in the IBM (left) and the lattice-based model (right). (a) Termite colonies (dots) falling within patches of unsuitable habitat (in grey) are removed. (b) Cells occupied by one or more termite colonies (black) overlaying cells of unsuitable habitat (grey) are removed.**

On one hand, the IBM can use vector-type GIS layers to consider patches of landscape that are unsuitable (e.g. roads, marshes, rivers) and all colonies that are falling within it (using the point-in-polygon operation) are removed from the simulation. On the other hand, the lattice-

based model can only use raster-type GIS layers, hence it strongly depends on the spatial resolution chosen for the simulation grid. If the grid has a coarse resolution, the underlying habitat will also be coarse and all colonies within a cell may be removed as a whole, hence losing some important detail necessary at a small spatial scale. The IBM approach is heavily influenced by the chosen size and detail of underlying suitable habitat layer. The larger the study area and the higher its detail (i.e. the combined number polygons within the habitat layer), the more the runtime will increase because of the aforementioned point-in-polygon operation. This problem is solved more efficiently in the lattice-based approach, where the simulation grid is pre-aligned with the underlying habitat raster and a simple Boolean operation is performed to remove termite colonies falling in unsuitable areas.

## **5. Conclusions**

The new lattice-based model presented in this study is a generic model for termites and can be applied to any species upon modification of the values of its ecological parameters. Although individual-based models are able to provide a very detailed scenario and incorporate both individual differences and local interactions among organisms, their computational cost is often higher compared to lattice-based models (Ajelli et al., 2010). Therefore, whenever the spatial scale is large (e.g. state, country level) or the temporal window is long, it is recommended to use a different simulation approach.

Given the findings of this study, we feel confident in concluding that the proposed lattice-based approach not only gives similar results to the IBM equivalent, but has a significant advantage in terms of computational speed. This has been shown to be particularly true when

there is a large number of initial termite colonies at the start of the simulation. However, given the higher variability found in the results of the proposed model, it is recommended to use more replications whenever possible to have a greater precision in the perimeters delineated by the occupancy envelopes. Dispersal is a stochastic process in which differences between any two simulations can grow from one time step to the next, and sometimes even lead to snowballing divergences.

One potential method to improve the computational performance of both models would be through “parallelization” of the code. Parallel processing breaks up a given task, splits it among multiple processors, and puts the components back together. Thus, parallelization can be extremely useful if one has a task that can be split up without the need of having all the different parts communicate with each other. On a standard laptop or desktop computer, the implementation of parallel processing in R may be able to speed up a program by a factor of 2 or 4. The lattice-based model presented herein and the IBM by Tonini et al. (2013) could potentially be re-written and their code be parallelized in different ways. Although there is empirical evidence that the lattice-based approach is faster than the IBM, we acknowledge that this does not imply that both algorithms were coded in the most efficient way possible. However, a parallelization of both simulation approaches goes beyond the scope of the study presented herein.

The degree of precision of the estimates used to parameterize the model represents a limitation of the proposed approach. As already pointed out by Tonini et al. (2013) for the IBM model, the lattice-based model will also greatly benefit from newer and better empirical estimations for a given species. It is often hard to find a precise estimate for cryptic species like termites and a consultation with a termite expert is always encouraged to calibrate model

parameters according to the desired species. Newer and better empirical estimations of the ecological parameters for a given species will be of great benefit to the simulation model.

Finally, we remark that the least probable envelopes should only be used for delineating a general perimeter of intervention, close to a "worst-case" scenario, instead of randomly inspecting unknown areas. A greater monitoring, quarantine, or eradication effort could be used in those areas encompassed by the most probable envelope.

### **Acknowledgements**

The authors would like to thank both anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. We would also like to thank Dr. Wendell Cropper, professor at the University of Florida, for his valuable suggestions and comments on the present work.

## References

- Ajelli, M., Goncalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J.J., et al., 2010. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infect. Dis.* 10, 190.
- Alexanderian, A., Gobbert, M.K., Fister, K.R., Gaff, H., Lenhart, S., Schaefer, E., 2011. An age-structured model for the spread of epidemic cholera: analysis and simulation. *Nonlinear Anal. Real World Appl.* 12, 3483-3498.
- Bordereau, C., Pasteels, J.M., 2011. Pheromones and chemical ecology of dispersal and foraging in termites. In: Bignell, D.E., Roisin, Y., Lo, N. (eds.), *Biology of termites: a modern synthesis*. Springer, pp. 279-320.
- Brown, D.G., Riolo, R., Robinson, D.T., North, M., Rand, W., 2005. Spatial process and data models: toward integration of agent-based models and GIS. *J. Geogr. Syst.* 7, 25-47.
- Burnham, B.O., 1973. Markov intertemporal land use simulation model. *South. J. Agric. Econ.* July, 253-258.
- Clark, J.S., 1998. Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *Am. Nat.* 152, 204-224.
- Clarke, K.C., Brass, J.A., Riggan, P.J., 1994. A cellular automaton model of wildfire propagation and extinction. *Photogramm. Eng. Remote Sens.*
- Clarke, K.C., Gaydos, L.J., 1998. Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *Int. J. Geogr. Inf. Sci.* 12, 699-714.
- Edwards, R., Mill, A.E., 1986. *Termites in buildings: their biology and control*. The Rentokil Library, 261 pp.
- Evans, T.A., 2011. Invasive termites. In: Bignell, D.E., Roisin, Y., Lo, N. (eds.), *Biology of termites: a modern synthesis*. Springer, pp. 519-562.
- Hochmair, H.H., Scheffrahn, R.H., 2010. Spatial association of marine dockage with land-borne infestations of invasive termites (Isoptera: Rhinotermitidae: *Coptotermes*) in urban South Florida. *J. Econ. Entomol.* 103, 1338-1346.
- Holmes, E.E., Lewis, M.A., Banks, J.E., Veit, R.R., 1994. Partial differential equations in ecology: spatial interactions and population dynamics. *Ecol.* 75, 17-29.
- Huston, M., DeAngelis, D.L., Post, W.M., 1988. New computer models unify ecological theory. *Biosci.* 38, 682-691.
- Jones, S.C., La Fage, J.P., Howard, R.W., 1988. Isopteran sex ratios: phylogenetic trends. *Sociobiology* 14, 89-156.
- Lett, C., Silber, C., Barret, N., 1999. Comparison of a cellular automata network and an individual-based model for the simulation of forest dynamics. *Ecol. Model.* 121, 277-293.
- Maidment, D.R., 1996. Environmental modeling within GIS. In: Goodchild, M.F., Steyaert, L.T., Parks, B.O., Johnston, C., Maidment, D., Crane, M., Glendinning, S. (eds.), *GIS and environmental modelling: progress and research issues*, Fort Collins, CO, pp. 315-324.
- Martius, C., 2003. Rainfall and air humidity: non-linear relationships with termite swarming in Amazonia. *Amazon.* 17, 387-397.
- Mikler, A.R., Venkatachalam, S., Abbas, K., 2005. Modeling infectious diseases using global stochastic cellular automata. *J. Biol. Syst.* 13, 421-439.

- Morley, P.D., Chang, J., 2004. Critical behavior in cellular automata animal disease transmission model. *Int. J. Mod. Phys. C* 15, 149-162.
- Nutting, W.L., 1969. Flight and colony foundation. In: Krishna, K., Weesner, F.M. (eds.), *Biology of Termites*, vol. 1. Academic Press, New York, pp. 233-282.
- Pitt, J.P.W., 2008. Modelling the spread of invasive species across heterogeneous landscapes. Lincoln University, p. 232.
- Pitt, J.P.W., 2009. Predicting Argentine ant spread over the heterogeneous landscape using a spatially explicit stochastic model. *Ecol. Appl.* 19, 1176-1186.
- R Development Core Team, 2013. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Reed, D., 2004. A balanced introduction to computer science. Prentice-Hall, Inc. Upper Saddle River, NJ, USA.
- Rust, M.K., Su, N.-Y., 2012. Managing Social Insects of Urban Importance. *Annu. Rev. Entomol.* 57, 355-375.
- Scheffrahn, R.H., Cabrera, B.J., Kern Jr, W.H., Su, N.-Y., 2002. *Nasutitermes costalis* (Isoptera: Termitidae) in Florida: first record of a non-endemic establishment by a higher termite. *Fla. Entomol.* 85, 273-275.
- Steyaert, L.T., 1993. A perspective for studying of environmental simulation. In: Goodchild, M.F., Parks, B.O., Steyaert, L.T. (eds.), *Environmental modelling with GIS*. Oxford University Press, New York, pp. 16–30.
- Sudshira, H.S., 2010. Integration of agent-based and cellular automata models for simulating urban sprawl. International Institute for Geo-Information Science and Earth Observation, p. 78.
- Toffoli, T., Margolus, N., 1987. Cellular automata machines: a new environment for modelling. MIT Press, Cambridge, MA.
- Tonini, F., Hochmair, H.H., Scheffrahn, R.H., DeAngelis, D., 2013. Simulating the spread of an invasive termite in an urban environment using a stochastic individual-based model. *Environ. Entomol.* 42, 412-423.
- Wiegand, T., Revilla, E., Knauer, F., 2004. Dealing with uncertainty in spatially explicit population models. *Biodivers. Conserv.* 13, 53-78.