MODELING THE SPREAD OF ASIAN LONGHORNED BEETLE IN NEW YORK CITY

Jacqueline W.T. Lu[1,2] and Gareth J. Russell[1]

[1]Columbia University, Department of Ecology, Evolution and Environmental Biology, 10th Floor Schermerhorn Ext., 1200 Amsterdam Avenue, New York, NY 10027

[2]City of New York, Department of Parks & Recreation, Central Forestry & Horticulture, Olmsted Center, Flushing Meadow Corona Park, Flushing, NY 11368

Abstract

Introduction

The goal of this project was to develop a model that describes the local infestation dynamics of the Asian longhorned beetle *Anoplophora glabripennis* (ALB) in Bayside, New York City, using data on infested trees and uninfested host trees from the New York Eradication Program. In most epidemiological data there is uncertainty about the timing of events, such as the time a host was first infested, first contagious to others, or removed and no longer contagious. In the case of ALB, uncertainty exists because of our inability to precisely 'backdate' the time of first exit from the level of damage observed on the infested tree. The use of a likelihood approach, combined with integration over timing uncertainty, allowed us to build models incorporating several factors that may affect the pattern of ALB spread. Information based statistics allowed us to directly compare and select the model best supported by the observed data. The selected model may then be applied to both management and research applications.

From Program Data to Estimated Infestation Events for Modeling

The New York Asian Longhorned Beetle Eradication Program records the date each infested tree was found, the date the infested tree was removed, and a damage code based on the level of damage observed. We used a previously developed relationship between damage level and the age of an infestation, based on a Jersey City outbreak discovered in October 2002 (Sawyer et al. 2004), to 'backdate' the timings of infestation events in a Bayside, NYC outbreak. For six of the infested Jersey City trees, Sawyer et al. (2004) used the numbers of growth rings around exit holes to calculate the periods of time during which each tree must have been first infested and when it was first potentially able to infest other trees. These trees represented each damage code that was recorded, as well as the four earliest and most heavily damaged trees (the oldest exit hole found was made in 1998). Because Jersey City trees that were classified into the same damage code were found through dissection to have been infested in different years, the backdating algorithm can at best give an interval within which each tree was likely to have been infested. This means that there is also an interval within which beetles will have first begun to emerge and possibly infest other trees (Figure 1).

Uncertain Timing and the Use of Likelihood Models

In most epidemiological data there is uncertainty about the timing of events, such as when a node is first infested, first contagious, and first removed or no longer contagious (t_n, c_n and r_n). We incorporated such uncertainty into a likelihood model of the spread of an infestation based on a 'distance' matrix D and associated distance function $f(d_{ij})$. The distance function converts an element of D into a transmission rate T_{ij}. Our model considers the problem of fitting various models to data of an epidemic spread of an infection on a set of nodes $n = 1, 2, 3, \ldots, N$ with fixed spatial relationships to one

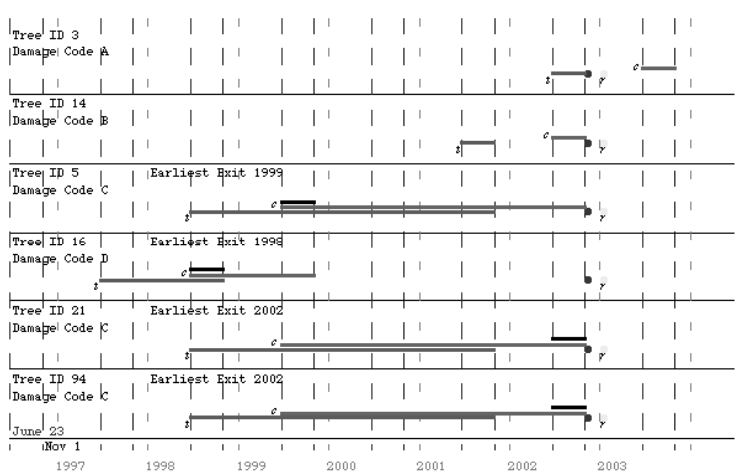


Figure 1.—Calculation of t_n and c_n intervals for Jersey City trees and comparison to earliest date of exit as determined from dissections.

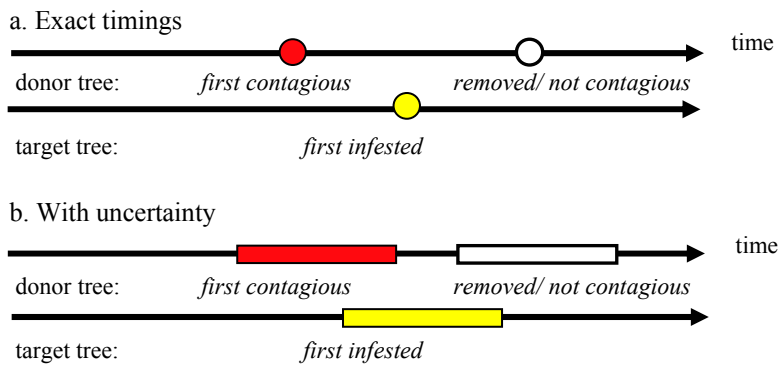


Figure 2.—Overlaps in infestation events between a donor tree and a target tree.

another; in the case of ALB, these nodes are individual trees. So, this is a spatially explicit model that considers the density of available hosts and their relationships to infested trees. The best-fit parameters for the distance function are those that maximize the likelihood of the observed ALB dataset.

If we knew the timing of infestation events exactly, we would be able to calculate the probability a target tree i was not infested by a donor tree j as $P = e^{T_{ij}(r_j - c_j)}$ (the probability of drawing zero from a Poisson distribution with rate parameter $T_{ij}(r_j - c_j)$ - Figure 2a). We can then calculate the probability of at least one infestation as 1 - P.

With uncertainty, however, first infestation occurs during an interval, and the contagious period of a donor tree is also an interval with uncertain beginning and end (Figure 2b). Our list of $\{t_n, c_n, r_n\}$ event triples for each node becomes a list of 'pair-triples' $\{\{t_{n,1}, t_{n,2}\}, \{c_{n,1}, c_{n,2}\}, \{r_{n,1}, r_{n,2}\}\}$. To calculate the probabilities we therefore have to integrate over this uncertainty, which we do under the assumption that no time within each interval is more likely than any other. Depending on how the intervals overlap (there are 15 possible combinations in the general model), calculating the probability of infestation can involve a combination of double and single integrals. The product of the probabilities of infestation for each infested tree during the contagious intervals of all the other infested trees, and the probabilities of no infection for uninfested trees, gives the likelihood of the entire dataset (for further details on the general model, see Russell and Lu in prepsaration). The general model was adapted to specifically analyze the Bayside infestation data from New York City, where there was no uncertainty about the removal date.

Model Fitting to Bayside, New York

We initially fitted four different distance functions to the backdated infestation data in Bayside, New York. Limitations in computing speed and memory prevented the calculation of all the pairwise probabilities for every tree in the data at once, so we created spatially nested subsets of the entire dataset and fit the models to each. A search function was used to calculate the maximum likelihood estimates of the fitted parameters. The MLEs were similar across data set extents for each distance function except the exponential decay ($f(d_{ij})=ae^{-bd}$), so model selection was limited to the remaining three distance functions.

Maximum likelihood models are a natural approach for dealing with probabilistic infestation data. Furthermore, information-based statistics like Akaike's Information Criterion (AIC) allow comparison of multiple models with different levels of complexity (Burnham and Anderson 2002). Table 1 shows the results of the model selection process for the three distance functions. AIC is defined as -2 log L + 2 V, where L is the maximized likelihood of a candidate model with V free parameters. Comparison of AIC values chose the radius decay distance function (lowest AIC value) as the model that was best supported by the data.

Simulations and Potential Tools

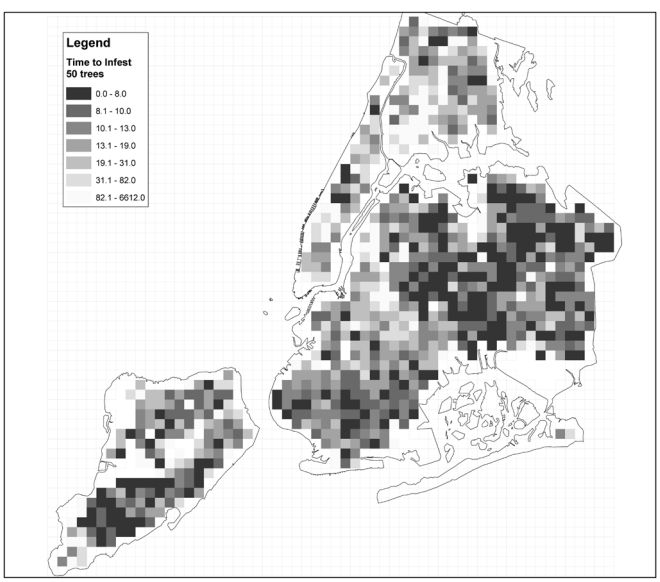
Using data on host tree locations throughout New York City, we created a risk map of ALB spread by performing simulations of spread using the selected distance model.

Table 1.—Maximum likelihood and AIC values for fitted models.

Model name	$f(d_{ij})$	Max. likelihood L	AIC value
Radius	a / d	-85.235	172.47
Radius decay	$a*e^{-bd}/d$	-69.0487	142.097
Radius power	ad^{-b}	-77.1233	158.247

The host tree population was divided into cells, into each of which a single infestation was seeded. We simulated the spread of an infestation until 50 trees in each cell were infested, recording the duration of the simulation (Map 1). The risk map is visually similar to a map of host tree density, which is not surprising given that our model is a simple distance function. Nevertheless, the risk map provides a quantitative measure that could help inform managers of the level of survey that would need to be done before an area could be declared uninfested. This is especially important in a pest with a high detection threshold such as the ALB; in an area where ALB spreads more slowly, a higher level of survey effort would be required.

Neighborhood-scale simulations were also performed on Bayside's host tree data and compared to the observed progression of the infestation. Trees that were infested in the greatest proportion of the simulations were the trees that were also infested earliest in the real data. There was more variation in the proportion of simulations in which infestation occurred for trees that were infested later, but this is expected because the timing of later events is strongly affected by the stochastic nature of the early stages of spread when just a few trees are infested. These simulations also identified areas that were only infested in the simulations and not actually found infested, and vice versa. Differences between model predictions and observations may be due to the role of other factors in dispersal, such as ALB host preference or the role of buildings as barriers to dispersal. Simulations at this scale may also be used to guide program managers to areas where more survey is required, by locating areas where simulated infestations occurred but where none have been reported to date.



Map 1.—Simulations of selected model on hosts in New York City. Dark areas have faster rates of spread.

Future Directions and Other Applications

Future work will add other factors that may affect dispersal to the basic distance functions fitted above. Because the distance function converts an element of D (in this case the distance between two trees) into a transmission rate T_{ij}, the matrix D does not have to be an actual distance, but can be an 'effective' distance representing the relative difficulty of transmission between trees due to host preference and environmental characteristics. The general model may also be applied to a variety of types of uncertain epidemiological data. One such type is 'regular census' data, where infestations occur at some unknown time in the interval between surveys (survey uncertainty). Timing uncertainty may also be due to incomplete knowledge of an organism's biology, such as the duration of an incubation period before a node is contagious, or before it is no longer contagious (biological uncertainty). This study is an example of how imperfect operational data can nevertheless contribute to research in factors affecting invasive species spread, and generate results which can be used in turn as management tools for program managers.

References

Burnham, K.P., and D.R. Anderson. 2002. **Model selection and multimodel inference : a practical information-theoretic approach, 2nd edition.** Springer, New York.

Russell, G.J., and J. W.T. Lu. 2004. **Modeling the spread of an epidemic under timing uncertainty.** In prep.

Sawyer, A.J., V.C. Mastro, B.C. Emens, and T. Denholm. 2004. **Analysis of the Asian longhorned beetle infestation in Jersey City, NJ**, poster. 2004 USDA Interagency Research Forum on Gypsy Moth and other Invasive Species, Annapolis MD.