# MODELING THE SPREAD OF *ALLIARIA PETIOLATA* ACROSS DIFFERENT LANDSCAPE DISTRIBUTIONS

A thesis presented to the faculty of the Graduate School of Western Carolina University in partial fulfillment of the requirements for the degree of Master of Science in Applied Mathematics.

By

Michael F. Lomuscio III

Director: Dr. Erin McNelis
Associate Professor of Mathematics
Mathematics and Computer Science Department

Committee Members: Dr. Beverly Collins, Biology
Dr. John Wagaman, Mathematics and Computer Science

April 2010

# ACKNOWLEDGEMENTS

I would like thank my thesis advisor, Dr. Erin McNelis, for her support and guidance throughout every stage of this research. I truly appreciate your dedication to teaching, as well as the excitement and enthusiasm that you bring the subject of mathematics.

Special thanks to my committee members, Dr. John Wagaman and Dr. Beverly Collins, who were constantly available and contributed greatly to my understanding of Ecology and Statistical Analysis.

I also wish to thank Dr. Gary Wein, without whom this research could not have happened. Dr. Wein provided and pre-processed all of the GIS landscape data for this study.

# Contents

# List of Tables

# List of Figures

# ABSTRACT

MODELING THE SPREAD OF *ALLIARIA PETIOLATA* ACROSS DIFFERENT LANDSCAPE DISTRIBUTIONS

Michael F. Lomuscio III

Western Carolina University (April 2010)

Director: Dr. Erin McNelis

Invasive plant species are believed to be the second greatest threat to biodiversity within the plant community. Among these invasives is the noxious weed *Alliaria petiolata*, otherwise known by its more common name: garlic mustard. Because garlic mustard is such an aggressive and entrenching invasive species, modeling its spread across landscapes is an invaluable tool in the battle to protect landscapes and their biodiversity. This research models the spread of garlic mustard using a stochastically driven simulation in order to analyze landscape characteristics that drive or dampen its spread.

# 1 Introduction

## 1.1 Garlic Mustard

Invasive plant species are believed to be the second greatest threat to biodiversity (Burls et al. 2008). Among these invasives is the noxious weed *Alliaria petiolata*, otherwise known by its more common name garlic mustard. A native of western Eurasia, garlic mustard can be found in Europe as far south as Italy and as far north as Sweden. It also exists in parts of North Africa and Asia Minor. It first appeared in North America during colonial times when colonists used it as a garlic substitute and for medicinal purposes (Rodgers et al. 2008). Although in its native habitats garlic mustard existed in small populations, many of its properties have given it an extreme advantage over native species in North American ecosystems, where it has spread over varying landscapes. Currently it spreads at an estimated rate of over 6400 km per year (Rodgers et al. 2008).

There are many physiological traits of garlic mustard that give it an advantage over native species. In particular, garlic mustard has advantageous chemical properties. Garlic mustard contains flavonoids, defense proteins, glycosides, and glucosinolates which make it an unfavorable plant for consumption by herbivores. This effect is so strong that some species of butterflies have been shown to preferentially lay their eggs on garlic mustard leaves (Rodgers et al. 2008). In addition to the chemicals listed above, garlic mustard is also capable of producing toxic cyanide compounds through hydrolysis of glucosinolates (Rodgers et al 2008). These compounds can be toxic to fungi, soil pests and pathogens, insect herbivores, and other plants, giving garlic mustard a competitive advantage (Rodges et al. 2008). Garlic mustard is thus able to move into regions that are initially unsuited for its survival by sterilizing the habitat of life forms that hinder its growth. The ability of garlic mustard to cleanse the soil of certain forms of life is so strong that it has been used in the past as a break crop to biofumigate soil. Garlic mustard tissue kills many soil-born diseases and has been show to increase crop yield when ground up and used as a soil fertilizer (Rodgers et al. 2008). These chemicals also have a dual advantage. Since they make garlic mustard unpalatable to herbivores, herbivores are forced to feed on the surrounding native plants. This causes a decrease in the local native species and gives garlic mustard yet another advantage over native plants (Rodgers et al. 2008).

In addition to chemical traits, there are also physical traits that give garlic mustard an advantage over native species. Most woodland herbs, the major group of native competitors with garlic mustard, are perennials, have limited seed production, have growth primarily limited by understory light availability, and have spatial distributions strongly influenced by the availability of resources (Rodgers et al. 2008). Garlic mustard, on the other hand, has a different physiology that gives it an advantage over the natives. As an obligate biennial, garlic mustard has a two year life cycle under all environments. Garlic mustard's development consists of two phases: a rosette stage and an adult stage.The plant reaches its rosette stage in early spring before most native herbs. In the rosette stage garlic mustard produces dark green leaves (Rodgers et al. 2008) that could block light from reaching the soil below it, potentially hindering the germination and growth of native species who are limited by light availability. After the rosette stage, in March or April of the following year, garlic mustard transitions into its adult stage. It grows at a rate of 1.9 centimeters per day and produces flowers from April through July and fruits from June through September (Rodgers et al. 2008). After fruits drop their seeds, all second year plants die (Rodgers et al. 2008).

The properties of garlic mustard's seeds are one of its most important physiological considerations. Garlic mustard produces seeds in its fruits, four to six millimeter siliques (Rodgers et al. 2008). Each silique contains approx 10 to 20 seeds. On average, a single plant can produce more than 3500 seeds and a population of adult plants can produce from 9500 seeds per square meter to more than 107,000 seeds per square meter (Rodgers et al. 2008). When the seeds fall from the plant they typically require at least 14 weeks of cold stratification at temperatures from 1 degree Celsius to 10 degrees Celsius for seeds to germinate in the spring (Rodgers et al. 2008). If seeds do get the required temperature exposure then about 70% of fallen seeds will germinate in the following spring (Rodgers et al. 2008). Any seeds that do not germinate can remain viable underground in the seed bank for up to ten years (Pardini et al. 2009). The length of time that garlic mustard seeds remain viable in the seed bank makes it a particularly difficult plant to get rid of. Once garlic mustard establishes itself in an area, it rapidly becomes a permanent fixture and controlling further invasion and spread becomes an extremely hard task.

Because garlic mustard is such an aggressive and entrenching invasive species, modeling its spread across landscapes and being able to predict where it will travel is an invaluable tool in the battle to fend it off and protect landscapes and their biodiversity. Biennials, in particular, can be very difficult to model due to their often unstable population dynamics (Winterer et al. 2004). Garlic mustard is no exception. There have been attempts to model the population dynamics of garlic mustard that have been met with varying levels of success (Pardini et al. 2009). Accurately modeling the population dynamics of an aggressive invasive species is difficult because aggressive invasives adapt very well to their environments and are, in general, very plastic. The plasticity of a plant refers to the level of variability within plant properties. For instance, a very plastic plant can have properties that vary greatly with geographical location, while a plant that is not very plastic can have very similar properties in all geographical locations. The plasticity of garlic mustard affects things like number of seeds per plant, plant height, and percentage of seeds that germinate per year. Therefore models that depend on factors such as these may only be accurate locally. On a large scale there is too much variability in these properties to create an accurate general model to predict spread.

## 1.2 Mathematical Modeling

Mathematical modeling of spatial and temporal systems is a well-developed discipline. There are many powerful techniques and models that have been proven to yield valuable insight into these systems. However, many of these methods are not suitable for application to the garlic mustard plant. The multi-stage dynamics of garlic mustard, along with the plants plasticity and extreme dependence on location specific factors, violates many of the fundamental assumptions of well-established spatial-temporal models including metapopulation and diffusion modeling.

Most spatial-temporal models used for modeling the spread of an invasive species across a landscape share key components and strategies. They each assume as a starting point that the landscape can be broken up into blocks or regions. These regions can vary based on the dynamics of the species being modeled. They can be distributed as blocks on a grid where each block represents a location on the landscape, they can be defined regions on the landscape that represent locations that share common characteristics or landscape type, or they can be a combination of the two. Garlic mustards ability to spread depends on extremely variable landscape characteristics such as the percentage of the landscape that is open field of forest. Its ability to spread is also extremely dependent on the distance that the susceptible landscape sites are from the nearest landscape edge. A landscape edge is defined to be any landscape boundary or landscape transitional zone. A landscape transitional zone exists wherever the landscape changes from one classification to another. For instance, the edge of a forest that is adjoined to a field would be considered a landscape transitional zone, and hence a landscape edge. Because garlic mustards ability to spread is dependent on extremely variable landscape characteristics it would be impractical to break the landscape up into regions that share common landscape characteristics. Hence we will consider a modeling technique that breaks the landscape up into a square grid over landscape locations as seen in Figure 1.

With the assumption that the landscape is broken up into a square grid over landscape locations, the next step in modeling the spread of an invasive is to create a model for how the species spreads from one block on the grid to another. This is typically done through finding a function that takes certain properties of the infected

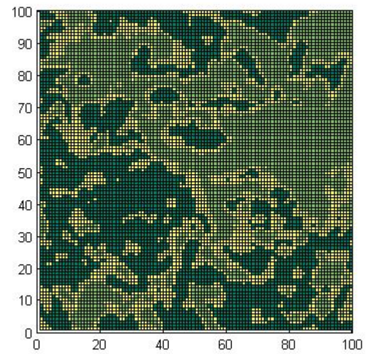



Figure 1: *Sample landscape broken up into a* $100 \times 100$ *grid of landscape locations.*

and infecting blocks as inputs, and gives as an output either the amount of plant spread, or the probability of the plant spreading. Once this function is defined, it can be used across the whole landscape grid and iterated over time to determine the long term spread of the invasive. Even if this sounds easy in principle, it is very difficult in practice to find such a function.

Since there is so much variability in the properties of the garlic mustard plant across different geographic locations, it may not be feasible to accurately model the spread of the plant with a model that is based on the plants properties. For example, a model that uses plant height and number of seeds per plant to predict spread may only be locally accurate since these properties vary by geographic location. However, it is feasible to model the spread of the plant with a model that is based on the landscape properties. Since garlic mustard isn't eaten by any herbivores its seeds are not generally spread by animals or insects. It turns out that the driving force of spread is falling seeds. Seeds that exist at the edge of a patch of garlic mustard fall into surrounding uninfected patches. Garlic mustard seeds fall within a few meters of the adult plant (Rodgers et al. 2008). Although some seeds do get dispersed further by somewhat stochastic vectors, such as strong winds, falling seeds account for most of garlic mustard's spread. Garlic mustard is also a very shade tolerant plant. Its shade tolerance effects where it can exist on a landscape. Because of these reasons, it is possible to model the spread of garlic mustard by studying how different landscape distributions affect spread via falling seeds and shade tolerance.

# 2 Methods

## 2.1 Landscapes

Since the goal of this study is to determine how different landscape factors affect the spread of garlic mustard, it is desirable to have many diverse landscapes to analyze. However, due to time constraints and availability of landscape data, only data from three landscapes were obtained. The landscapes were very large and sufficiently diverse, but three landscapes were not enough to complete this study. To solve this problem each landscape was broken up into smaller disjoint landscapes. Because the original landscapes were sufficiently diverse with regards to landscape variation, it is reasonable to view smaller sections as somewhat randomly varied independent landscapes. The following method was used to break the larger landscapes into smaller disjoint landscapes.

The three landscapes were broken up into grids over the landscapes. The grid sizes of the three landscapes are $200 \times 200$, $100 \times 100$, and $75 \times 75$. Landscapes of different sizes were chosen due to availability. Each of the landscapes were then broken into smaller disjoint landscapes of size $20 \times 20$. Since 20 divides 100 and 200, all of the landscape cells were used from the $200 \times 200$ and the $100 \times 100$ landscapes. Since 20 does not divide 75, some of the landscape cells from the $75 \times 75$ landscape were disregarded. Figure 2 shows an example of a $4 \times 4$ landscape being broken up into four disjoint $2 \times 2$ landscapes.

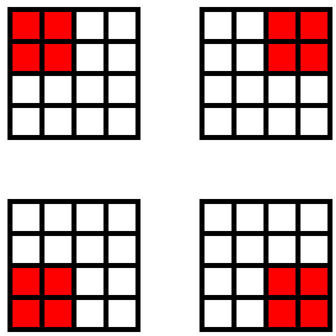


Figure 2: *Example of a* $4 \times 4$ *landscape being broken up into four disjoint* $2 \times 2$ *landscapes.*

Since $\frac{200}{20} = 10$ and $\frac{100}{20} = 5$, the $200 \times 200$ landscape was broken into $10^2 = 100$ disjoint landscapes and the $100 \times 100$ landscape was broken into $5^2 = 25$ disjoint landscapes. In order to keep landscape sizes uniform, only the first 60 rows and columns of the $75 \times 75$ landscape were considered. Since $\frac{60}{20} = 3$, the $75 \times 75$ landscape was broken into $3^2 = 9$ disjoint $20 \times 20$ landscapes and excess cells were disregarded. Therefore, this study considered a total of $100 + 25 + 9 = 134$ sufficiently varied disjoint $20 \times 20$ landscapes.

Each landscape was composed of three different landscape classifications: open field, forest, and landscape boundary edge. Landscape boundary edges are the areas where the landscape changes from open field to forest or from forest to open field.

These are also referred to as landscape transitional zones.

## 2.2 Simulating Spread

A landscape site is considered infected with garlic mustard if either a rosette or an adult plant exists on the site. Since garlic mustard is so entrenching, random death of garlic mustard at a site (disinfection) is extremely rare and is not considered in this study. Hence, once a site is infected, it is assumed that it will stay infected. In this study 1% of each of the 134 landscapes is randomly infected with garlic mustard. Since the landscapes are $20 \times 20$ landscapes there are $20^2 = 400$ cells in each landscape. Hence 4 cells are randomly chosen to be initially infected with rosettes in each landscape.

When considering the spread of an invasive plant, the first five years is a relatively small period of time. However, five years is still plenty of time for the plant to firmly establish itself within the landscape. One of the major goals of studying invasive species is to contribute knowledge that could be beneficial in preventing and controlling invasions. When it comes to controlling and preventing invasions of garlic mustard, the first five years of invasion are clearly important. Considering a time period greater than five years could be detrimental to this study because given enough time, even the least probable spreading events could occur. The effects that landscape properties have on the spread could change as the invasion progresses. This study is concerned with the way that a landscape distribution affects the initial spread of the plant. Hence considering a small time period, yet one that is long enough for the plant to establish itself, will allow for an adequate study of which landscape characteristics have the biggest impact on an initial invasion.

### 2.2.1 The spread algorithm

The algorithm is fairly simple and is presented below in Figure 4. Note, with regards to which surrounding cells an infected cell can infect, only the cells directly above, below, to the right, and to the left are considered, see Figure 3.

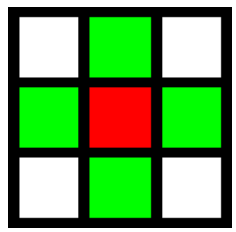


Figure 3: *The red center square is an infected cell. The infection can only spread in the cardinal directions; these are the green susceptible squares.*

Each run of this algorithm is equivalent to one year of growth and spread of garlic mustard. Hence five iterations of this algorithm is equivalent to five years of invasion.

### 2.2.2 Probability functions

Garlic mustard is a very shade tolerant plant. Its shade tolerance allows it to penetrate the forest canopy up to about 30 feet. Although it can penetrate the canopy pretty far, it grows near the forest edge with a greater probability. The probability of infection spreading to a forest landscape cell can be modeled as a linear function of distance from nearest forest edge.

The probability of spread on a landscape edge is approximately 100% while the

Check each cell for infection

- **if** a cell is infected with a rosette
  - the rosette turns into an adult plant,
- **else if** a cell is infected with an adult plant
  - check the four surrounding cells (up, down, left, and right) for susceptibility
    - **if** one of the surrounding 4 cells is an uninfected open field or forest
      - use the correct probability function to determine whether the cell will become infected with a rosette
    - **else**
      - do nothing
- **else**
  - do nothing

Figure 4: *Spread Simulation Algorithm*

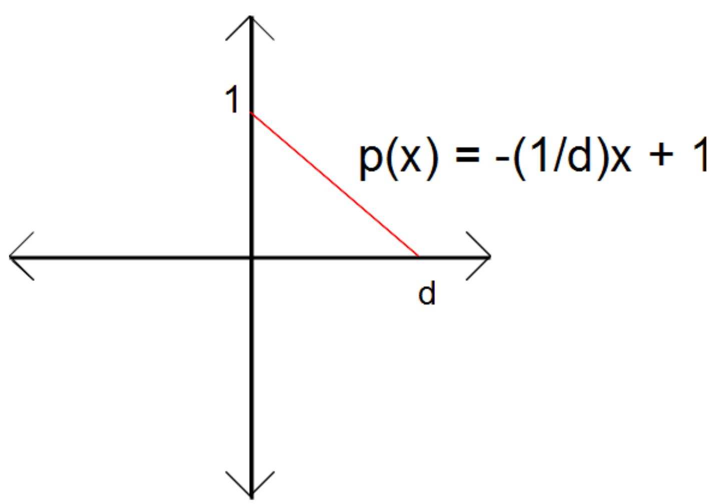



Figure 5: *The probability of spreading as a function of distance from the nearest landscape edge. The probability of spreading along a landscape edge is* 100% *while the probability of spreading at a distance greater than d is* 0%. *Therefore the probability of spreading at a distance $x$ from the nearest landscape edge is* $p(x) = -\frac{1}{d}x + 1$.

probability of spread at a distance greater than 30 feet is approximately 0%. Since the probability, $p_{\text{Forest}}(x)$, is modeled by a linear function of distance from the nearest edge the vertical intercept of $p_{\text{Forest}}(x)$ is 1 and the horizontal intercept is 30. For a linear function the horizontal intercept is equal to $-\frac{\textit{vertical intercept}}{\textit{slope}}$. Since the vertical and horizontal intercepts are known, the slope can be solved for and is equal to,

$$slope = \frac{-\textit{vertical intercept}}{\textit{horizontal intercept}} = \frac{-1}{30}$$

Therefore $p_{\text{Forest}}(x) = -\frac{1}{30}x + 1$.

Although garlic mustards shade tolerance allows it to penetrate deep into the forest canopy, the shade tolerance also keeps it from being able to spread very far into open fields. The plant cannot handle a lot of sunlight. In addition to intense sunlight, open fields tend to have an abundance of well-established plant life. Although garlic mustard is a fierce competitor, it is difficult for it to penetrate more than about five feet from the nearest landscape edge into open field. The probability, $p_{\text{Field}}(x)$, of infection spreading to an open field landscape cell is also modeled as a linear function of distance from nearest landscape edge. The probability of spread on a landscape edge is still approximately 100%, but the probability of spread at a distance greater than five feet is approximately 0%. Hence the vertical intercept of $p_{\text{Field}}(x)$ is 1 and the horizontal intercept is 5. Since the vertical and horizontal intercepts are known, the slope can be solved for and is equal to,

$$slope = \frac{-\textit{vertical intercept}}{\textit{horizontal intercept}} = \frac{-1}{5}$$

Therefore $p_{\text{Field}}(x) = -\frac{1}{5}x + 1$.

# 3 Analysis

In this study each of the 134 landscapes were selected one at a time. For each simulation 1% of the landscape cells were randomly selected to be designated as infected, and the spread algorithm was iterated to simulate a time range of five years. Such simulations were performed 50 times for each landscape. For a given landscape, the following properties of each were recorded for each of the 50 simulations and then averaged.

- Cover: The percentage of the landscape cells that are infected with garlic mustard at the end of the five year period.

- Percent Field: The percentage of the landscape cells that are open field.

- Percent Forest: The percentage of the landscape cells that are forest.

- Average Distance: The average distance from each cell to its nearest edge.

## 3.1 Correlations

The goal of this study is to determine how cover is influenced by each of the three landscape characteristics: percent field, percent forest, and average distance. Scatter plots of cover with each of the three characteristics are shown below in Figures 6, 7,

| Correlation coefficients | Percent Field | Percent Forest | Average Distance |
|---|---|---|---|
| Cover | -0.9420 | 0.5372 | -0.6329 |

Table 1: *Correlation coefficients for least squares lines from scatter plots.*

and 8, along with a table of linear correlation coefficients from the individual linear regressions.

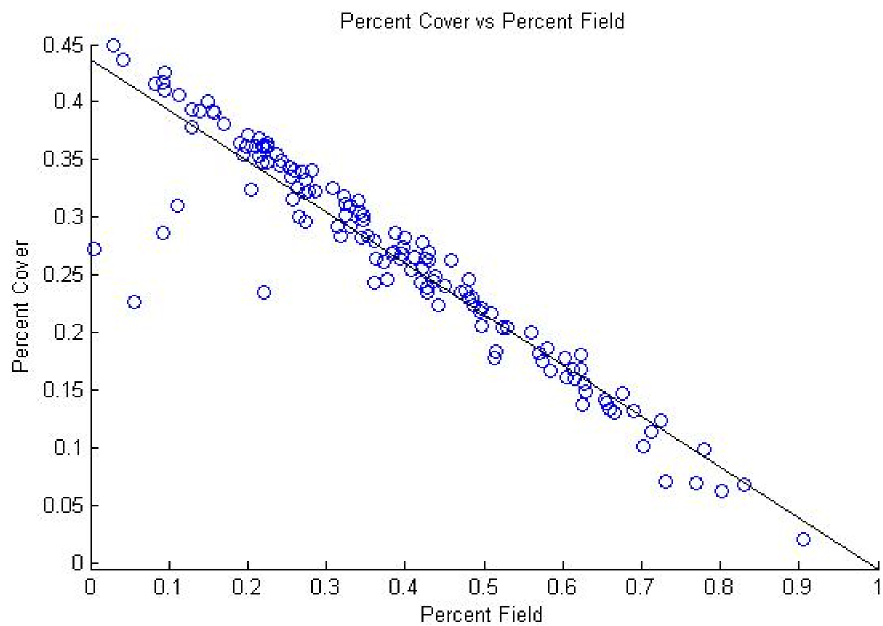



Figure 6: *Percent Open Field vs. Cover*

Percent Cover is strongly correlated with percent field, and significantly correlated to percent forest and average distance to nearest landscape edge.

The negative correlation coefficient associated with percent field implies that as the percentage of the landscape composed of open field increases, the percent cover decreases. This makes intuitive sense as garlic mustard cannot penetrate very far into open fields. The positive correlation coefficient associated with percent forest implies

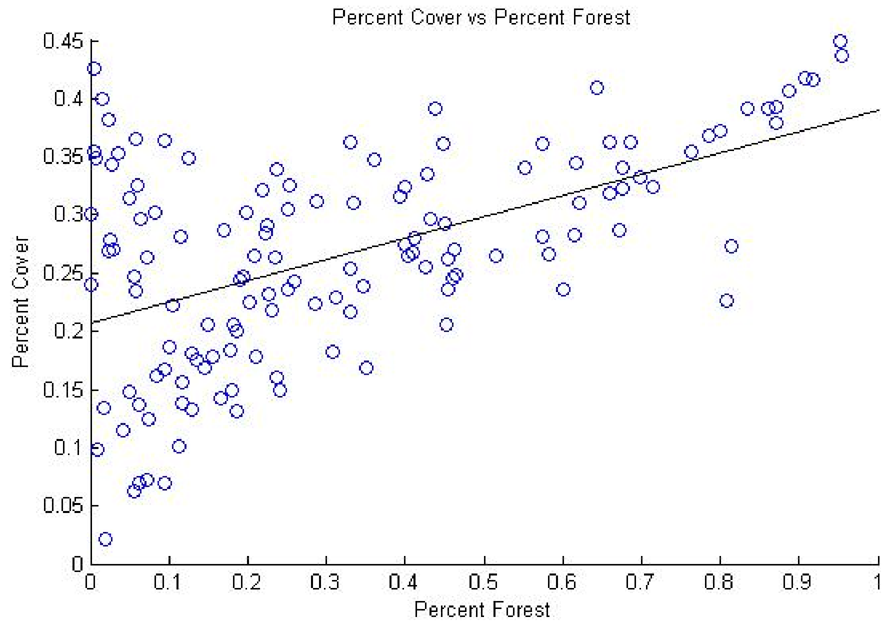



Figure 7: *Percent Forest vs. Cover*

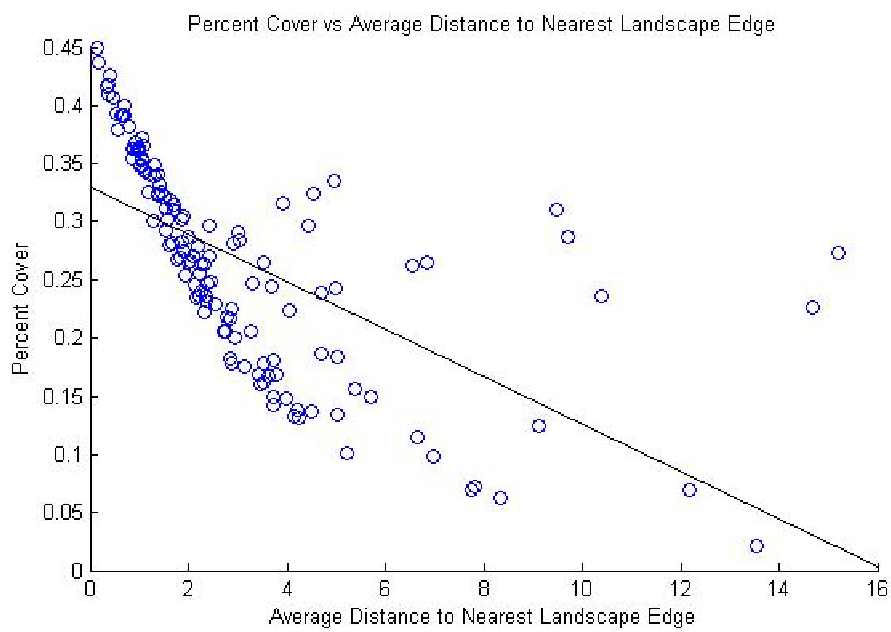



Figure 8: *Average Distance vs. Cover*

that as the percentage of landscape composed of forest increases, the cover increases. Again, this makes intuitive sense because garlic mustard can penetrate deeply into forests. Forest is the most susceptible landscape, and hence as the percentage of the landscape that is forest increases, infection will also increase. The negative correlation coefficient associated with average distance to nearest landscape edge implies that as the average distance to nearest landscape edge increases, the cover increases. This also makes intuitive sense because the probability of spread decreases with increased distance.

When considering the percent cover versus percent forest scatter plot in Figure 7, it should be noted that the plot has a funnel shape. This funnel shape implies that the variance in landscape covers changes as percent forest increases. The variance in percent cover starts off very large for small values of percent forest and decreases as percent forest increases. This change in variance can be explained by considering the way that forests can be distributed across landscapes. If a small percentage of the landscape is composed of forest then there are many ways that these forest cells can be distributed across the landscape. For example, the forest cells can be connected, disconnected, concentrated in one location, spread out uniformly across the landscape, connected in a straight line, connected in the form of a circle, or various other configurations. Recall that garlic mustard has the greatest probability of spreading inside the boundaries of forests. Therefore, when the configurations of forest cells are highly variable, the existence of susceptible locations that have high probabilities of spread is also highly variable. This explains the high variance in percent covers for small values of percent forest.

As the value of percent forest increases, the number of possible forest configurations gets smaller. There is a higher probability that forest cells will begin to clump into groups. Since forest clumps are highly susceptible locations for garlic mustard invasion, the probabilities of spread are very high. This causes a decrease in the variability of percent cover since there are few forest configurations that yield low percent covers.

Similarly, there is a slight funnel shape in the percent cover versus average distance to nearest landscape edge (henceforth known as ADNE) scatter plot in Figure 8. The funnel implies that for small values of ADNE there is very little variance in percent cover. However, as ADNE increases, the variance in percent cover greatly increases. The explanation for this is similar to the percent cover versus percent forest case. The highest probability for the spread of garlic mustard occurs at locations that are close to landscape edges. Hence there should be high values of percent cover when the ADNE is small. When considering the distribution of landscape edges, there are very few ways to distribute landscape edges such that the ADNE will be high and the percent cover will be low. But when the ADNE is large there are a lot of possible edge distributions that will yield high percent covers and a lot of possible edge distributions that will yield small percent covers. Hence the variance in percent cover will be much higher for large ADNE values.

## 3.2 Multiple Regression Models and Generalized Linear Models

Generalized linear models, known as GLMs, are very similar to multiple regression models. Multiple regression models are used to model a normally distributed response variable as a linear combination of predictor variables. For example, the goal of this study is to model the response variable percent cover as a function of predictor variables percent field, percent forest, and ADNE. In general, if $y$ is a normally distributed response variable that depends on predictor variables $x_1, x_2$, and $x_3$ then the multiple regression model is described by,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

where $\beta_0$ can be interpreted as the intercept along the response variable axis. Although multiple regression models are very useful, it is not always the case that the response variable $y$ is normally distributed. In cases where the response isnt normally distributed it is sometimes possible to find a function, called a linking function, that will normalize the response variable. For example, if a response variable $y$ comes from a Poisson distribution, then $y$ is log-normal. This means that although $y$ isnt normally distributed, $\log(y)$ is normally distributed. Hence log is a linking function that normalizes the response variable $y$. Since $\log(y)$ is normally distributed, a multiple regression model can be applied to it yielding,

$$\log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

However, this implies that

$$y = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3}.$$

The incorporation of a linking function to make the response variable normally distributed so that a multiple regression model can be used is called a generalized linear model. An added benefit of using GMLs is that cleverly chosen linking functions can restrict the response variable space. For example, if a response variable $y$ represents a quantity that cannot be negative, it may be possible to choose a linking function that will normalize $y$ and also force $y$ to take on positive values. This is a very helpful property in many cases.

### 3.2.1 Normality of Cover

A generalized linear model, GLM, was fit to the landscape data and simulated garlic mustard spread. This model has the form of $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ where $y$ represents the percent cover response, $x_1$ represents the percent field factor, $x_2$ represents the percent forest factor, and $x_3$ represents the average distance factor. The $\beta_i$'s represent the corresponding coefficients. The response variable percent cover was checked for normality in Figure 9. Figure 10 is a histogram and norm plot for cover.

The Poisson linking function, $\log(x)$, was considered. However, this linking function introduced two problems. The first issue is that it transformed the data to be less normal, Figure 11. The second issue is that the response variable data, the percent covers, all take on values between zero and one. However, $\log(x)$ is negative on this domain. This means that the log linking function will yield a very inaccurate model

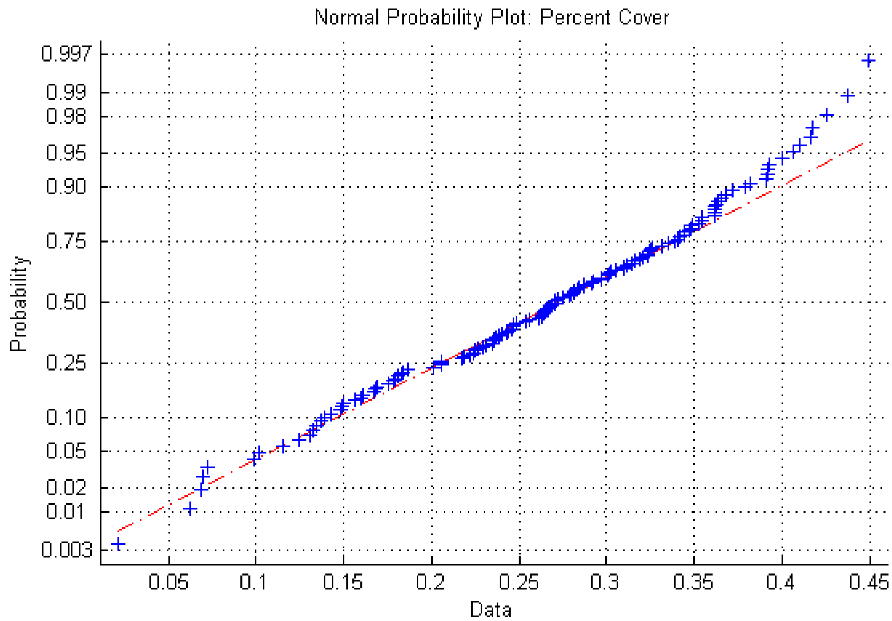



Figure 9: *Norm Plot of Cover*

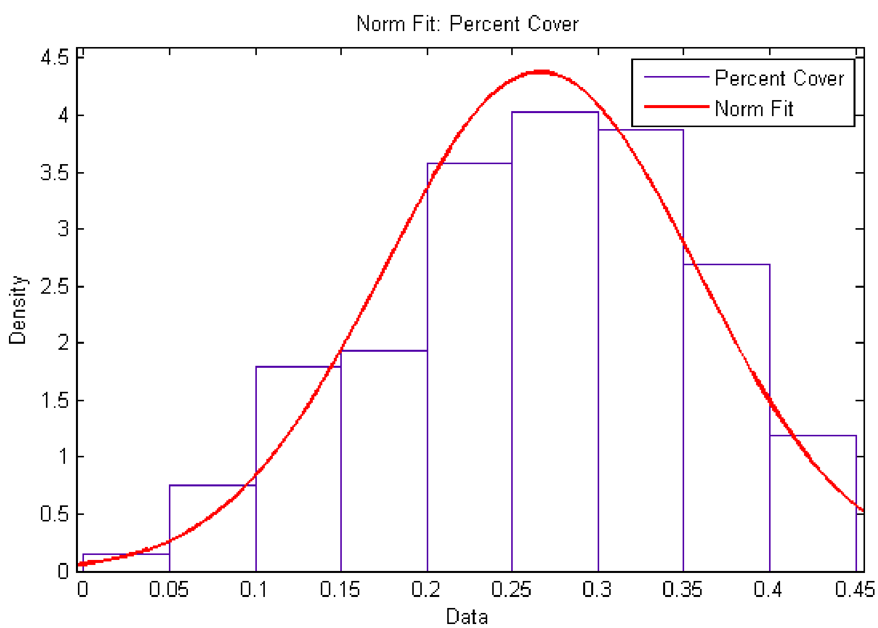



Figure 10: *Histogram of Cover with normal fit.*

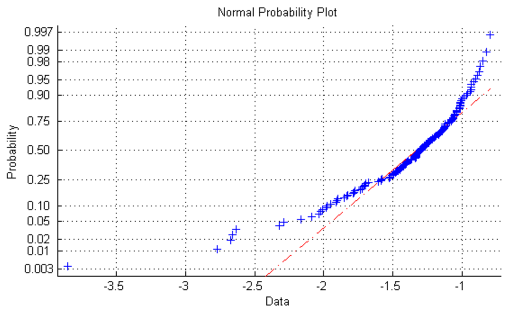
with the response space restricted to negative values, Figure 11.

In order to resolve this negativity issue all of the percent cover data was multiplied by 100 to yield true percentages, no longer in decimal form. This forced the domain of $\log(x)$ to be positive. However, when a GLM with a Poisson linking function was fit to the data, the model was less accurate then the model with no linking function, Figure 11. In order to be thorough the percent forest and percent field predictors were also multiplied by 100 to yield true percentages. A GLM was again fit to the new data set, where percent cover, percent forest, and percent field were all true percentages. Again, the model was less accurate than the model with no linking function, Figure 11.

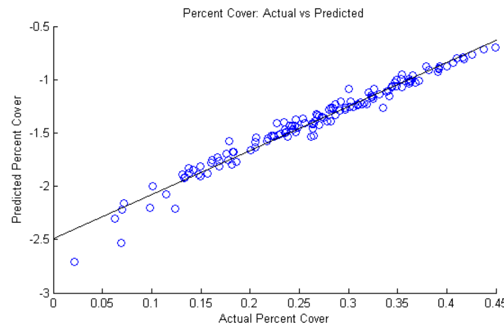
Although future research will focus on finding an appropriate linking function, a GLM with a normal linking function was chosen for this study. It should be noted that a GLM with a normal linking function is equivalent to a multiple regression model, meaning that the normal linking function is the "do nothing" linking function. From Figure 10 it is clear that the simulated cover data is not perfectly normal. The data is skewed to the left. However, for the purposes of fitting a GLM with a normal linking function, this data is sufficiently normal.

### 3.2.2 Model Fit

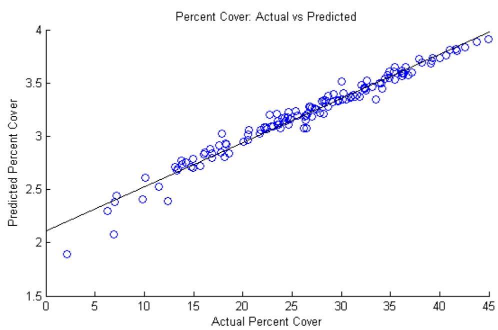
The student version of MATLAB, along with the MATLAB Statistical Toolbox, was used to fit a GLM to the simulated data. Below are the model and parameters that MATLAB returned,



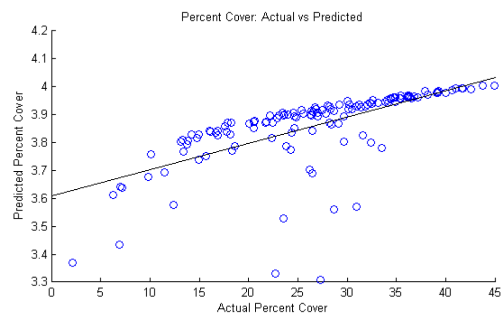

(a) *Normality plot of* log*(percent cover).*




(b) *A scatter plot of the predicted versus actual cover for the GLM run with the Poisson linking function.*




(c) *A scatter plot of the predicted versus actual cover for the GLM run with the Poisson linking function where percent cover was given as a true percent.*




(d) *A scatter plot of the predicted versus actual cover for the GLM run with the Poisson linking function where percent cover, percent forest, and percent field were given as a true percents.*

Figure 11: *Results of Poisson linking function analysis.*

$$y = 0.4457 - 0.3837x_2 + 0.0009x_3 - 0.01x_4.$$

Using a value of $\alpha = 0.05$, all of the parameters are significant except for percent forest.

Figure 12 is a plot actual percent cover versus the percent cover predicted by

| | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ |
|---|---|---|---|---|
| $\beta$ value | 0.4457 | -0.3837 | 0.0009 | -0.01 |
| p-value | $\approx$ 0.0000 | $\approx$ 0.0000 | 0.8951 | $\approx$ 0.000 |

Table 2: *GLM results*

the GLM above. This plot helps evaluate the accuracy of the GLM and how well it predicts the response variable percent cover.

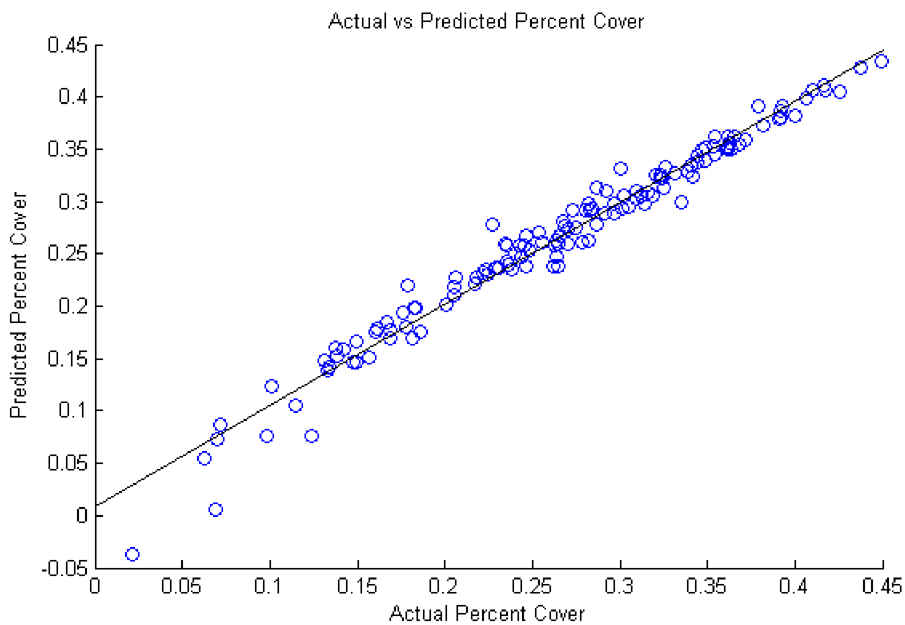



Figure 12: *Predicted Cover vs. the Actual Cover*

The correlation coefficient of linear regression line fit in Figure 12 is $r = 0.9843$. This is strong evidence supporting the accuracy of the GLM and its ability to predict the percent cover response.

# 4 Conclusions

From analyzing the above graphs and the results of the GLM it is clear that landscape distribution has a direct effect on the ability of garlic mustard to spread across a given landscape. Specifically, the percentage of the landscape that is composed of open field and the average distance to nearest edge of the landscape dramatically impact garlic mustards ability to spread across a landscape.

It has been shown that there is a positive linear relationship between the amount of landscape infected with garlic mustard and the amount of landscape that is made up of forest. It has also been shown that there is a negative linear relationship between the amount of landscape infected with garlic mustard and the amount of landscape that is made up of open field. Caution must be given when interpreting these results. It would be easy to fall into the trap of interpreting these results to simply imply that landscapes completely composed of forest are at the greatest risk for garlic mustard invasion. However, recall that garlic mustard only grows on or near edges (composed of transition zones between forest and open field). Without some open field, there would be no edges, and hence the most important resource for garlic mustard would be absent. Therefore, it must be the case that some balance between forest and open field will optimize the spread of garlic mustard. This balance depends as much on percent field and forest as it does the physical distribution of the fields and forests. The identification of this optimal balancing point will be the subject of future study. Identifying this balancing point is an essential aspect of being able to determine the susceptibility of landscape distributions in general. Once these balancing characteristics are determined, a formal method of classifying particular landscapes as being at risk for garlic mustard invasion can be developed.

The work carried out in this study in an essential first step to determining the optimal balancing point for garlic mustard invasion. It will be the goal of future study to also consider additional landscape characteristics and their effect on garlic mustard spread. Percent field, percent forest, and ADNE are clearly important characteristics to consider, but there are still many potentially important characteristics that need consideration.

# References

[1] Kevin Burls and Charles McClaugherty. Landscape position influences the distribution of garlic mustard, an invasive species. *Northeastern Naturalist*, 15(4):541–556, 2008.

[2] Thomas Nagylaki. The diffusion model for migration and selection in a plant population. *Journal of Mathematical Biology*, 35:409–431, 1997.

[3] Eleanor A. Pardini, John M. Drake, Jonathan M. Chase, and Tiffany M. Knight. Complex population dynamics and control of the invasive biennial *Alliaria petiolata* (garlic mustard). *Ecological Applications*, 19(2):387–397, 2009.

[4] Kimberly A. Rebek and Robert J. O'Neil. Impact of simulated herbivory on *Alliaria petiolata* survival, growth, and reproduction. *Biological Control*, 34:283–289, 2005.

[5] Vikki L. Rodgers, Kristina A. Stinson, and Adrien C. Finzi. Ready or not, garlic mustard is moving in: *Alliaria petiolata* as a member of eastern North American forests. *BioScience*, 58(5):426–436, May 2008.

[6] Geoffrey R. Smith, Haley A. Dingfelder, and David A. Vaala. Effect of plant size and density on garlic mustard reproduction. *Northeastern Naturalist*, 10(3):269–276, 2003.

[7] David J. Susko and Lesley Lovett-Doust. Patterns of seed mass variation and their effects on seedling traits in *Alliaria petiolata* (brassicaceae). *American Journal of Botany*, 87(1):56–66, 2000.

[8] Erik Welk, Konstanze Schubert, and Matthias H. Hoffmann. Present and potential distribution of invasive garlic mustard (*Alliaria petiolata*) in north america. *Diversity and Distributions*, 8:219–233, 2002.

[9] Juliette Winterer, Megan C. Walsh, Minakshi Poddar, Jeffrey W. Brennan, and Sarah M. Primak. Spatial and temporal segregation of juvenile and mature garlic mustard plants (alliaria petiolata) in a central Pennsylvania woodland. *The American Midland Naturalist*, 153(2):209–216, 2005.