Impacts Of Forest Disturbance On Small Mammal Distribution

Allyson Lenora Degrassi

University of Vermont

Follow this and additional works at: https://scholarworks.uvm.edu/graddis

Part of the Biology Commons, Ecology and Evolutionary Biology Commons, and the Natural Resources and Conservation Commons

Recommended Citation

Degrassi, Allyson Lenora, "Impacts Of Forest Disturbance On Small Mammal Distribution" (2016). Graduate College Dissertations and Theses. 640.
https://scholarworks.uvm.edu/graddis/640

This Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks @ UVM. It has been accepted for inclusion in Graduate College Dissertations and Theses by an authorized administrator of ScholarWorks @ UVM. For more information, please contact donna.omalley@uvm.edu.
IMPACTS OF FOREST DISTURBANCE ON SMALL MAMMAL DISTRIBUTION

A Dissertation Presented

by

Allyson Lenora Degrassi

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy Specializing in Department of Biology

October, 2016

Defense Date: August 3rd, 2016

Dissertation Examination Committee:

Nicholas J. Gotelli, Ph.D., Advisor
Therese M. Donovan, Ph.D., Chairperson
Alison K. Brody, Ph.D., Committee Member
C. William Kilpartick, Ph.D., Committee Member
Cynthia J. Forehand, Ph.D., Dean of the Graduate College
ABSTRACT

Natural habitat in the eastern United States has diminished over the past century because of the effects of invasive species. Both plant and animal invaders can alter habitat structure and may decrease survival of native species. The degree to which an invasive species alters ecosystem function depends on the functional characteristics of affected species and the resulting cascading effects. The loss of important native species, such as foundation species, can potentially influence the structure and distribution of animal communities because of the foundation species’ unique ecosystem roles. The foundation species concept is relatively new to the terrestrial ecology and the impact on animal communities resulting in the loss of terrestrial foundation species is generally unknown.

Eastern hemlock (*Tsuga Canadensis*), a foundation species in the eastern United States, is declining in abundance due to the invasive sap-sucking insect, hemlock woolly adelgid (*Adelges tsugae*, Annand). The loss of hemlock may impact the distribution and microhabitat associations of dependent species such as small mammals. I hypothesized that the distribution, population size, community composition, and microhabitat associations of small mammal species differ in response to canopy disturbance from the effects of logging and invasive species.

In this dissertation, Chapter One provides an exploration of the past research conducted on 1) invasive species and how they affect habitat structure, 2) foundation species and how they affect ecosystem function, 3) small mammal habitat associations and population cycling, 4) occupancy modeling and its usefulness and limitations in the analysis of local occupancy, colonization rates, and extinction rates. Chapter Two presents a large-scale experiment on how the hemlock woolly adelgid impacts distribution and community assembly of small mammals. Chapter Three presents how forest disturbance, food resources, and habitat structure effects local colonization and extinction patterns of southern red-backed voles. Chapter Four presents how a paper published in 2005 brought the foundation species concept to terrestrial research and how the foundation species concept can be misleading in research.
ACKNOWLEDGEMENTS

I would like to acknowledge and extend my sincere heartfelt gratitude to the following persons who have made the completion of this dissertation possible: My adviser Dr. Nick Gotelli, for all the work he put into helping me be a better researcher and for supporting my research ideas, even when they were “too ambitious”. My committee members Dr. Brody, Dr. Kilpatrick, and Dr. Donovan for their vital understanding and assistance on topics in which I was a novice. My unofficial mentor Dr. Aaron Ellison, for all of his advice on navigating the stormy waters of research. My REU students Elizabeth Kennett, Amy Balint, James Leitner, Joel van de Sande, and Ariel da Cruz Reis, for being brave enough to venture into the woods to study animals. My undergraduate research assistant Emma Cronin for sticking with me while we analyzed over 1,000 photos. The LTER Foundation Species crew, for assisting in The Loss of Foundation Species chapter. Manisha Patel M.S., Audrey Barker-Plotkin M.S., and Laurie Chiasson at Harvard Forest, for making my life in the “Sham” a great one. Biology Department faculty and staff, especially Dr. Joe Schall, Jennifer Jany, and Wendy Chagnon for always helping me.

Most especially, I would like to acknowledge my husband, soulmate, and best friend Chris Degrassi, for his encouragement and never wavering confidence in me. Without his support and love, I would have never believed that I could achieve such a feat.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................ iii

LIST OF TABLES ...................................................................................................................... vii

LIST OF FIGURES .................................................................................................................... viii

CHAPTER 1: REVIEW OF THE LITERATURE ........................................................................... 1

1.0 Introduction ...................................................................................................................... 1

1.1 Invasive Species .............................................................................................................. 1

1.2 Foundation Species ....................................................................................................... 3

1.3 Habitat Structure and Metapopulations ....................................................................... 5

1.4 Small Mammals in North America .............................................................................. 7

1.5 Small Mammal Colonization and Extinction in North America .................................. 9

1.7 Occupancy Modeling ................................................................................................... 14

1.8 Synthesis ....................................................................................................................... 15

1.9 Figures ........................................................................................................................... 17

CHAPTER 2: IMPACTS OF HEMLOCK WOOLLY ADELGID ON SMALL MAMMAL ............ 18

2.1 Abstract ......................................................................................................................... 18

2.2 Introduction .................................................................................................................. 20

2.3 Methods ....................................................................................................................... 26

2.4 Results ........................................................................................................................ 35

2.5 Conclusion ................................................................................................................... 40

2.6 Tables .......................................................................................................................... 44

2.7 Figures ........................................................................................................................ 55
6.I Chapter 3: Multiseason Occupancy Prep R Script ...........................................156
6.K Chapter 3: Multiseason Occupancy Models R Script...........................................167
6.L Chapter 4: Citation Analysis R Script....................................................................177
LIST OF TABLES

Table 2.1: Hypothesized influence on relative abundance of individuals within small mammal communities in Hemlock, Girdled, Logged, and Hardwood canopy treatments at HF-HeRE in Petersham, MA. 

Table 3.1: Description of covariates considered for all models, a brief description of covariate used in each parameter, and formulas used in model sets (n=38).

Table 3.2: Exploratory model set for detection parameter. The $\psi$ (Elevation + Habitat) $\gamma$ (Landscape + Food) $\varepsilon$ (Winter) $p(.)$ varies with model, number of parameters, AIC, $\Delta$AIC, AIC weight, and negative log-likelihood.

Table 3.3: Model ranking based. The $\psi$ (Elevation + Habitat) $p$(night soil temperature) are held constant, but $\gamma$ and $\varepsilon$ formula structure vary. The number of parameters, AIC, $\Delta$AIC, AIC weight, and negative log-likelihood from explored covariates (Table 1). *Models could not converge on a solution and were not included in the model set.
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>17</td>
</tr>
<tr>
<td>Conceptual model of shifts in micoclimate, understory growth, canopy coverage, and small mammal assemblage (descending order from most abundant to least abundant) from eastern hemlock forests (top) to hemlock woolly adelgid forest (bottom-left), or to preemptive logging managed forest (bottom-right). Model has been modified from Ellison et al. 2005.</td>
<td></td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>45</td>
</tr>
<tr>
<td>Mean (± SE) percent cover of microhabitat ground cover characteristics (left to right, top to bottom; rock, leaf litter, soil, vegetation, woody debris, and fungi) among hemlock (green), girdled (orange), logged (yellow), and hardwood (blue) treatments. Results of randomized block analysis of variance (treatment and block) for each characteristic indicated top-center of each graph. Letters above treatment are result of Tukey’s HSD grouping.</td>
<td></td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>46</td>
</tr>
<tr>
<td>Mean (± SE) percent cover of microhabitat canopy cover characteristics (top to bottom; open canopy, high canopy, and low canopy) among hemlock (green), girdled (orange), logged (yellow), and hardwood (blue) treatments. Results of randomized block analysis of variance (treatment and block) for each characteristic indicated top-center of each graph. Letters above treatment are result of Tukey’s HSD grouping.</td>
<td></td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>47</td>
</tr>
<tr>
<td>Rank relative abundance graph of small mammals among canopy treatments (right to left: Hemlock, Girdled, Logged, and Hardwood). Each bar and color represents a different species. The height of the bar is the relative abundance of the specie in each treatment.</td>
<td></td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>48</td>
</tr>
<tr>
<td>Average PIE and 95% confidence intervals among canopy treatments (Hemlock, Girdled, Logged, and Hardwood). There is no significant difference in average evenness among treatments (P=0.79, df=3).</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.5 .................................................................49
Chao1 estimated species richness (dots) with lower and upper 95% confidence intervals (error bars) for Hemlock, Girdled, Logged, and Hardwood canopy treatments for years 2012. Letters indicate groupings based on CI overlap where same letters different letters indicate significantly different groups. The estimated species richness in Hemlock (a) control differs significantly from Girdled (c), but not from Logged(b) and Hardwood (a). Girdled treatment (c) differs significantly from Hardwood (a), but not Logged (c).

Figure 2.6 .........................................................................50
Chao1 Shared with Hemlock control estimated species richness (dots) with lower and upper 95% confidence intervals (error bars) for Girdled, Logged, and Hardwood canopy treatments.

Figure 2.7 .........................................................................51
Schnebel estimated population (N-hat) with lower and upper 95% confidence intervals (error bars) for Hemlock, Girdled, Logged, and Hardwood canopy treatments for deer mice, southern red-backed voles, and white-footed mice (top to bottom).

Figure 3.1 .........................................................................77
Simes Tract containing the Long-Term Ecological Research (LTER) Hemlock Removal Experiment (HF-HeRE) located in north-central Petersham, Massachusetts, USA (panel A). The replicated blocks, Ridge (yellow) and Valley (green), contain four ~90 x 90-m (0.81-ha) canopy treatments plots, Hemlock (He), Girdled (G), Logged (L), and Hardwood (Hw) (panel B). Sampling grids (panel C) spanned 70m². Trap (circles) were placed 10-m apart by pacing in a 7 × 7 array within each of the two Hemlock, Hardwood, Girdled, and Logged plots (n=392). Figure modified from Ellison et al. 2005.

Figure 3.2 .........................................................................78
Probability of detection (95% CI) of southern red-backed voles as a function of the nightly average soil temperature (°C).

Figure 3.3 .........................................................................79
Probability of site occupancy (95% CI) as a function of block (Ridge and Valley) and habitat treatment (Hemlock, Girdled, Logged, Hardwood) of southern red-backed voles.

Figure 3.4 .........................................................................80
Probability of site colonization (95% CI) of southern red-backed voles as a function of food (the average seed fall mass (g/m²)).
Figure 3.5 .................................................................81
Probability of site colonization (95% CI) of southern red-backed voles as a function of food (vegetation cover %).

Figure 3.6 .................................................................82
Probability of site extinction (95% CI) of southern red-backed voles as a function of food (the average seed fall mass (g/m²)).

Figure 3.7 .................................................................83
Probability of site extinction (95% CI) of southern red-backed voles as a function of food (percent vegetation ground cover).

Figure 4.1 .................................................................101
Comparison of characteristics that differentiate commonly used terms that describe common and/or abundant species

Figure 4.2 .................................................................102
Summary of general trends in the study of foundation species based on the results of the Ellison et al. 2005 study (solid lines represent the pool of studies from one analyzed question to the next; dotted lines indicate side information on how results of questions were broken down of each result; filled circles show direction of significance based on Chi² results; p-values from Chi²).

Figure 4.2 .................................................................103
Geographic map of the number of studies (circle size) that identified research organism as foundation species (FS) (green) and the studies that did not identify foundation species (blue).

Figure 4.3 .................................................................104
Suggested approach to foundation species research and how topics are connected in the scope of this paper (arrows indicate the direction and relationship to topics).
CHAPTER 1:
REVIEW OF LITERATURE

Allyson L. Degrassi

1.1 Introduction

1.1 Invasive Species

Natural habitat in the eastern United States has diminished over the past century because of intensive human activities and invasive species. Habitat loss and invasive species are the two most common threats to biodiversity in the United States, and invasive species are predicted to overcome habitat loss as the single largest threat to biodiversity (Wilcove et al. 2010). The number of invasive species is continuously growing, and consequently the damage to native ecosystems also is increasing. Invading plants and animals impact the environment, landscape, and native species differently. Plant invaders typically alter abiotic processes of native ecosystem function (e.g. Vitousek et al. 1986, Ehrenfeld 2003) while animal invaders can compete with native species for resources (e.g. Gurnell et al. 2004). Animal invaders can cause extinctions of native species through species direct interactions (e.g. predation and competition) whereas, plant invaders often cause shifts in biogeochemical processes that alter nutrient cycling and local hydrology (review by Mack et al. 2000 and papers within). Even though the routes of disturbance caused by invasive plants and animals differ, both plant and animal invaders can alter habitat structure (Lizarralde et al. 2004, Simberloff 2009) and decrease survival of native species (review by Mack et al. 2000). Regardless of whether invasive species are the “driver” or “passenger” of ecological change and
extinctions (e.g. Gurevitch and Padilla 2004, MacDougall and Turkington 2005, Didham et al. 2005), disturbance occurs, ecosystem function is disrupted, and ecosystem state changes follows.

Ecosystems response to species losses depends on both the number of species lost and their identities. The rivet hypothesis (Ehrlich and Ehrlich 1981, chapter 5) states that an ecosystem may withstand the loss of one or a few species without any discernable effect on ecosystem function, just as a plane would still function if a few rivets were removed. However, if too many species in an ecosystem are, then ecosystem functions may become unstable, just as a plane will fall apart if too many rivets were removed. This hypothesis would likely be supported if there were no difference in species function: all species had the same function in the ecosystem similar to all rivets have the same function in holding the plane together. However, not all species within an ecosystem preform the same function. Within a given ecosystem, there may be multiple species that carry out similar functional roles while few species may contribute relatively little to the overall ecosystem function (review by Hopper et al. 2005). Ecosystem stability is strongly influenced by particular species’ function roles properties. The loss of important native species, such as foundation species, due to habitat loss and invasive species, can potentially influence the structure of plant and animal communities because of the unique roles foundation species play in ecosystem function (Dayton 1972, Ellison et al. 2005a).
Foundation Species

Foundation species (sensu Dayton 1972, see Ellison et al. 2005a). may function as structural species (Huston 1994), dominant species (Grime 1984), core species (Hanski 1982), keystone species (Paine 1966), and as ecosystem engineers’ (Jones et al. 2010) (sensu Dayton 1972, see Ellison et al. 2005a). Foundation species are often abundant primary producers that provide habitat for other species, control population dynamics, and create locally stable conditions for other species. They support communities by modulating and stabilizing fundamental ecosystem process (Dayton 1972; Ellison et al. 2005a). As abundant primary producers that provide habitat and stabilize biogeochemical processes, foundation species often elicit a community wide bottom-up trophic cascade when they are removed (e.g. Carpenter et al. 1985, Polis 1999, Persson 1999, Scherber et al. 2010, Baiser et al. 2013). Foundation species loss can impact nutrient fluxes (Jenkins et al. 1999), microclimate conditions (Snyder et al. 2002), food web structure (Baiser et al. 2013), and biodiversity (e.g. Tingley et al. 2002, Snyder et al. 2002, Ellison et al. 2005b).

Eastern hemlock (Tsuga canadensis (L.) Carrière) functions as a foundation species throughout the northeastern United States. However, their populations have declined dramatically in the eastern United States because of damage inflicted on them by invasive insect pests (McClure 1991, Orwig and Foster 1998, Orwig et al. 2002, see Kizlinski et al. 2002; Ellison et al. 2005a). The hemlock woolly adelgid (Adelges tsugae, Annand 1928) is native to Japan and was introduced to the United States in the 1950’s.
The hemlock woolly adelgid is a sap-sucking insect that defoliates trees (Orwig et al. 2008) and causes rapid hemlock death (McCulre 1991). These aphid-like insects are effective dispersers and are introduced to new ranges by wind, birds, deer, and by humans through logging (McCulre 1990). Because the hemlock woolly adelgid threatens much of the old growth forests in the eastern United States, forest management regimes such as preemptive logging are being considered to decrease the adelgid threat and to conserve late successional forests (Foster and Orwig 2006).

The damage caused by hemlock woolly adelgid creates a unique mosaic of a graveyard-like forest that is characterized by having a reduced canopy and standing dead trees. The decrease in canopy covers increases the amount of light that reaches the forest floor, which allows for early successional hardwood species to flourish in the understory (Farnsworth et al. 2014). A woolly-adelgid infested forest has different characteristics than a logged forest, which also has an open canopy, but woolly-adelgid infested forest usually has coarse woody debris littering the forest floor and a slower onset of vegetative understory growth compared to logged forests (Farnsworth et al. 2014). Forest management practices, such as preemptive logging, involve the removal of trees that are newly infested, vulnerable to infestation, or are economically valuable. Trees are removed before an infestation occurs to decrease the spread of infestation and to extract economic value (Foster and Orwig 2006). These forest management regimes also affect forest seed-banks, tree regeneration, and forest dynamics (e.g. Graae and Sunde 2000, Decocq et al. 2004, Farnsworth et al. 2014).
The dramatic changes in forest structure and forest stability caused by logging or invasion can shift ecosystem processes followed by shifts in biodiversity. These shifts within the ecosystem affect taxonomic groups differently. For example, loss of eastern hemlocks results in an initial increase in local ant species diversity (Ellison et al. 2005b), but a decrease in regional bird population composition and distribution (Tingley et al. 2002). In contrast, in long-term studies, loss of hemlock does not appear to impact ants, beetles, or spiders (Sackett et al. 2011), at least at a local scale. These inconsistent responses to hemlock loss over varying temporal and spatial scales make it difficult to predict how species across multiple taxa will prevail after the loss of hemlocks, which are not expected to recover from hemlock woolly adelgid invasion (Foster 2014). This imminent loss of these widespread foundation species will cause state changes within the forest, but how this change in habitat structure will impact particular species is unknown.

1.3 Habitat Structure and Metapopulations

Heterogeneous changes in habitat structure (e.g. changes caused by invasive species) can create patches of suitable and unsuitable habitat. These variations in habitat or patch quality may influence the distribution of organisms (i.e. the number of sites occupied in a particular area by a species; Hanski 1982), which affects estimation of site occupancy (the probability that a particular species is present at a site; MacKenzie et al. 2002).
The long-term survival of species depends on patchy habitat (Hanski 1998; Wilcove et al. 1986). Although habitat patchiness can result from natural processes (e.g. disturbances) and human activities (e.g. land use, agricultural activity, perturbation), it happens in an increasing extent due to more extensive landscape fragmentation which is particularly important in the studies of population dynamics, community ecology, wildlife conservation and management (Hanski 1998a, 1999, Opdam et al. 2003, Wiens et al. 1993).

Several studies examined the effects of fragmentation and habitat degradation on microtine rodent populations in North America (e.g. Collins & Barrett 1997; Harper et al. 1993; Wolff et al. 1997) and in Europe (e.g. Bjørnstad et al. 1998; Paillat and Butet 1996; Paradis and Croset 1995; Paradis 1995). A number of studies using experimental model systems to investigate the effects of both fragmentation (Andreassen et al. 1998; Bjørnstad et al. 1998) and habitat destruction (Andreassen and Ims 1998; Johannesen et al. 2003; Johannesen & Ims 1996) on space use and demographic parameters have been applied. The effects of habitat fragmentation also affect litter sex ratio (Aars et al. 1995), movement patterns (Andreassen et al. 1996a, 1996b), natal dispersal (Gundersen & Andreassen 1998), foraging behavior (Hovlandet al. 1999), and reproductive synchrony (Johannesen et al. 2000) were investigated in many mammal species. The analysis of short-term and long-term data are important in studying the vulnerability of small mammals that are threatened by the effects of spatial heterogeneity, as well as in exploring how extinction and colonization dynamics affect community structure (de
1.4 Small Mammals in North America

In the family Cricetidae, deer mice (*Peromyscus maniculatus*, Wagner 1845) and white-footed mice (*P. leucopus*, Rafinewque 1818) are congeneric partly arboreal species that occur sympatrically in areas of the eastern United States, but deer mice are more active in large tree stands than white-footed mice (Graves et al. 1987). Deciduous woodland forests are the optimal habitat for white-footed mice (Krohne 1989, Wolf and Batzli 2002). White-footed mice occur in a variety of habitats; they are opportunisitcs and their popualtions are no impacted by habitat disturbance (Henein et al. 1998, Linzey et al. 2012).

Voles are mostly herbivorous (Lobo and Millar 2014) and have been known to reduce vegetative growth as severely as large ungulates (Howe et al. 2006). The vole’s vegetative diet would suggest they require or prefer habitats with near-ground vegetation. Voles generally are associated with a range of soil moisture and leaf litter coverage. For example, southern red-backed voles (*Myodes gapperi*, Vigors 1830 [formerly *Clethrionomys gapperi*]) are often associated with boreal mixed-forest stands in Canada characterized by having downed woody debris, a dense shrub layer, a coniferous understory and coniferous litter, and moist conditions (Merritt 1981, Vanderwel et al. 2009). In contrast, the woodland vole (*Microtus pinetorum*, LeConte 1830) requires well-drained soil, but also prefers dense vegetation (Smolen 1981).
In the family Sciridae, southern flying squirrels (*Glycomaus Volans*, Linnaeus 1758) and chipmunks (*Tamias striatus*, Linnaeus 1758) are more commonly associated with hardwood forests in contrast to northern flying squirrels (*Glaucomys sabrinus*, Shaw 1801) and American red squirrels (*Tamiasciurus hudsonicus*, Erxleben 1777) which are limited to old growth conifer stands (Ransome and Sullivan 1997). Southern flying squirrels are dynamic foragers that exploit hardwood nuts (Thomas and Weigl 1998) throughout the year. Similar to southern flying squirrels, eastern chipmunks are one of the major granivores in eastern deciduous forests, and they prefer a diet with white-acorn nuts (Pyare et al. 1993).

In the order Soricomorpha and family Soricidae, shrews are predators that feed on earthworms (Hamilton 1941), arthropods (Hamilton 1941), and other small mammals (Eadie 1948). Although shrew species are not as diverse in their feeding guilds and habitat preferences as rodents, shews do tend to differ in their sub-fossorial habitat associations. For example, the Short-tailed shrew (*Blarina brevicauda*, Gray 1838) is associated with moist, deep leaf litter (George et al. 1986) and the Masked Shew (*Sorex cinereus*, Kerr 1792) is associated with moist soil and abundant vegetation (Getz 1961, Brown 1967). Forest structure, from leaf litter to canopy cover, determines which areas rodents and shrews will occupy, but their site occupancy is influenced by several other environmental and biological factors.
1.5 Small Mammal Colonization and Extinction in North America

Rodents demonstrate diverse patterns of colonization and extinction that result in population fluctuations that vary from non-cyclic (e.g. Predavec et al. 2001) to highly cyclic (e.g. Jett and Nichols 1987, Krebs 1996, Stenseth et al. 1996, Krebs et al. 2002, Boonstra et al. 2012, Krebs 2013 and research therein). The numerous and diverse underlying mechanisms that drive these colonization and extinction patterns are continuously up for debate (Norrdahl 1995). Although several geographical factors (e.g., latitude and elevation) and biological factors (e.g., density dependence, food abundance, life history traits) can influence cycling patterns, the addition of changing habitats further complicated colonization and extinction trends. Currently, forests in the eastern United States are radically changing due to the loss of a forest foundation species (e.g. Dayton 1972, Ellison et al. 2005), the eastern hemlock (*Tsuga canadensis* (L.) Carrière) and it is unclear how forest disturbance plays on the already complex patterns of colonization and extinction of small mammals.

Mechanisms that change occupancy yield cyclic patterns of occupancy through time are of special interest to small mammal ecologists because of confusing results from mass literature (e.g. Krebs 2013). Local colonization rates (i.e., the probability that an unoccupied site in year $t$ will become occupied in year $t + 1$) and local extinction rates (i.e., the probability that an occupied site in year $t$ will be unoccupied in year $t + 1$) may vary depending on landscape characteristics (e.g. elevation, habitat type and surrounding habitat type), food resources (e.g. seed and vegetation), site structure (e.g. woody debris
and leaf litter) and climate (e.g. overwinter conditions).

Observational and experimental data support the vital roles of landscape (e.g. Barrett and Peles 1999 and papers within), habitat structure (e.g. Abramsky 1988, Stapp 1997, Brown 1988, Drickamer 1990, Fauteux et al. 2012), and habitat selection (e.g. Abramsky et al. 1990, Vickery and Rivest 1992, Morris 1996) in determining species composition for several mammal species. Landscape characteristics such as elevation, habitat type, and surrounding habitat quality are known to influence colonization and extinction patterns. The heterogeneous habitats within the landscape result in migration. Small mammal populations that occupy areas adjacent to high-quality habitats may show different local cycling patterns than populations that are adjacent to low-quality habitats due to the influx of colonists from the high-quality habitat (Hestbeck 1982). Local community structure of species can be a product of the type of adjacent habitat, which suggests that, regardless of the local habitat’s resources (internal within site), the local population would persist as a result of migration from the habitat rich area (Pulliam 1988).

Alternatively, local site colonization rates may vary as a function of food resources, which are believed to contribute to population growth and cycling. Multiple studies suggest that populations sizes of seed predators (e.g. birds, mice, and voles) correlate with the synchrony of seed production (seed masting) of many tree species (e.g. Wolff 1996, Hanski and Henttonen 1996, Selas 2000, Selas et al. 2002, Schmidt
Winter conditions, such as snow cover and onset of snow melt are often correlated with population sizes of small mammals (e.g. Hansson 1984, Hansen et al. 1999, Duchesne et al. 2011, Bierman et al. 2006, Hoset et al. 2009). Dispersal patterns can change seasonally when young individuals migrate from old growth forests where winter survival is highest to young growth forests where breeding is high, but winter survival is low (Ecke et al. 2002). This dispersal patterns suggests that small mammals do have the ability to assess habitat quality and that their assessment influences their dispersal patterns. Dispersal strengthens the role of habitat variation and its effect on small mammal demographics (e.g. reproduction and migration), behavior (e.g. competition and habitat selection), and ecological factors (e.g. food and habitat resources). Immigrants move more frequently among patches than residents (e.g. Pusenius et al. 2000). Therefore, variation in the probability that an occupied site will become unoccupied in year $t+1$ is higher in poor quality habitat than in high quality habitats.

Within-site factors (i.e. habitat structure) also affect small mammal habitat associations that influence local site colonization and extinction patterns. Within-site habitat structure, such as woody debris and leaf litter, have contrasting associations with small mammals. There are several biological hypotheses that describe the rate of population growth. The rate of population increase is 1) determined only by reproductive
output, 2) inversely related to the age at sexual maturity (see Krebs 2013 and papers within Kalela 1957, Koshkina 1965, Keller and Krebs 1970), and 3) positively related to the length of the breeding season (see Krebs 2013 and papers within Hamilton 1937, Krebs 1964, Hansson 1984). Many population studies assume closed populations (i.e. no migration, only births and deaths), but migration also influences population density and population cycling in addition to biogeographical and life history traits of many small mammals.

There are several hypotheses of density dependence that describe rodent dispersal. Density can be positively or negatively correlated dispersal. A positive correlation would predict that leaving high density areas or poor condition habitats will increase individual fitness (e.g. Waser 1985, Porter and Dooley 1993). A negative correlation would predict that leaving high density areas will increase competition and aggression in the newly migrated areas so dispersal is effectively reduced (e.g. Hestbeck 1982). Therefore, dispersal is based on social interactions (e.g. Boyce and Boyce 1988), population density (Lidicker 1975), and relatedness among individuals (Charnov and Finerty 1980, Lambin and Krebs 1991).

The “social fence” hypotheses also makes certain predictions regarding density dependence, dispersal, and habitat quality (Hestbeck 1982). The “social fence” hypotheses predicts that populations that occupy areas adjacent to high-quality or low-quality habitats will show different local cycling patterns because migration will vary
depending on the adjacent habitat quality (Hestbeck 1982). High emigration occurring from the high-quality and low emigration occurring from the low-quality habitat is predicted (Hestbeck 1982). Low-quality habitat will not reach a high enough population density to block immigration from occurring from high-quality habitats which have reached high population densities (Hestbeck 1982).

These hypotheses predict that local community structure can be a product of the type of adjacent habitat. Regardless of the local habitat’s resources (poor quality habitat), the local population would persist as a result of migration a higher quality habitat (Pulliam 1988). Species whose dispersal depends on climatic conditions (abiotic factors) would demonstrate source-sink dynamics whereas species with the ability to assess habitat quality would demonstrate balanced dispersal dynamics (reviews by McPeek and Holt 1992, Diffendorfer 1998). Occurrence in some studies support the balanced dispersal model (Diffenforfer 1998), which implies small mammals do have the ability to assess habitat quality and that their assessment influences their dispersal patterns (see Fretwell and Lucas 1970). This “free will” dispersal pattern strengthens the role of habitat variation and its effect on small mammal demographics (e.g. reproduction and migration), behavior (e.g. competition and habitat selection), and ecological factors (e.g. food and habitat resources).

In contrast, some small mammal species’ dispersal patterns fit a source-sink model in which source areas are occupied by residents and sinks are occupied by
immigrants and immigrants move more frequently among patches than residents (e.g. Pusenius et al. 2000). Dispersal in source-sink pattern can change seasonally where young individuals migrate from old growth forests, where winter survival is highest, to young growth forests, where breeding is high, but winter survival is low (Ecke et al. 2002).

1.7 Occupancy Modeling

Occupancy is a state variable that refers to the proportion of area (or a fraction of landscape) occupied by a species where the species is present (MacKenzie et al. 2006). Occupancy modeling is most commonly used in wildlife studies and is often used to infer species-habitat relationships (MacKenzie et al. 2006). Occupancy methods are used to generate models of species probability of occupancy ($\psi$) across multiple geographical locations while taking into account the probability of detecting an individual at any given site ($p$). In general, the sampling protocol commonly used for occupancy modeling involves repeated site visitation, recordings of species detection and non-detection, and recording habitat characteristics (variables that are used site and observation covariates). Occupancy modeling approaches can improve models of species distributions. While estimating occupancy is a useful tool, it is not often used in experimental manipulation of habitats as in this study.

A large variety of hierarchal occupancy models (MacKenzie et al. 2002 and 2003) are used to estimate site occupancy of a single species (e.g. Sadoti et al. 2013) and
multiple species (e.g. Fauteux et al. 2012; Kalies et al. 2012; Carrillo-Rubio et al. 2014) over a single season (e.g. Rinehart et al. 2009; Long et al. 2011, Murdoch et al. 2013) or multiple seasons (Hines et al. 2014). Site occupancy models provide useful information and help explain factors (represented as covariates) that influence animal distribution and abundance (Stanley and Royle 2005), especially when there is heterogeneity in the probability of species detection (MacKenzie et al. 2006: pp161-165, Gibson 2011).

In classic metapopulation and island-biogeography models, local colonization and extinction probability rates were estimated directly from site occupancy or the presence-absence of a species in a particular area (Levins 1970, Hanski 1982, Gotelli 1991), but the probability of actually detecting a species was not accounted for until recently (MacKenzie et al. 2002). However, occupancy models are variable, difficult to fit, and difficult to interpret when abundance varies over sites (Welsh et al. 2013), as they are hypothesized to do when logging and invasive species disrupt particular areas of the landscape and not others. Because small mammal populations can be very cyclic and occupancy estimates are based on population sizes, it can be difficult to capture the environmental factors that contribute to local colonization and extinction patterns.

1.8 Synthesis

Will small mammal’s communities be impacted if hemlock forest structure is altered by hemlock woolly adelgid invasion or preemptive logging? Will changes in the hemlock forest landscape, food resources, and winter conditions affect the colonization
and extinction of particular species? A disturbed hemlock forests may alter habitat structure and possibly food resource availability for some small mammals while others may not be affected. The disturbance in forest structure driven by hemlock loss can be enough to cause a shift in small mammal species distribution, because many small mammals have different habitat requirements and microhabitat associations. Changes in forest structure triggered by hemlock die-offs caused by the hemlock woolly adelgid and by preemptive logging management can affect the dispersal of small mammals in eastern forests differently because of the eastern hemlock’s unique abilities as a foundation species to stabilize biogeochemical processes (e.g. soil moisture and local climate) and community structure, which both contribute to small mammal dispersal (Figure 1).
Figure 1. Conceptual model of shifts in micoclimate, understory growth, canopy coverage, and small mammal assemblage (descending order from most abundant to least abundant) from eastern hemlock forests (top) to hemlock woolly adelgid forest (bottom-left), or to preemptive logging managed forest (bottom-right). Model has been modified from Ellison et al. 2005.
CHAPTER 2:

IMPACTS OF HEMLOCK WOOLLY ADELGID
AND PREEMTIVE LOGGING ON
SMALL MAMMAL COMMUNITIES AND POPULTIONS

Allyson L. Degrassi

2.1 Abstract

Eastern hemlock forests are declining in abundance from the effects of the invasive sap-sucking insect hemlock woolly adelgid (HWA). Preemptive logging management practices have been considered to stop the spread of the adelgid, but both disturbances caused by logging and HWA affect hemlock forest and associated animal communities. I censused small mammal communities in experimental plots at Harvard Forest to quantify and predict the effects of forest disturbance caused by HWA and preemptive logging. The Harvard Forest Long-Term Ecological Research experiment (LTER) is a replicated two-block design that includes four 0.81ha canopy treatments in each block: Hemlock Control, in which hemlocks are dominant, Hardwood Control, in which young hemlocks are present, but mid-successional hardwoods are dominant, Girdled Treatment, in which hemlock trees were girdled and killed to simulate a woolly adelgid invasion, and Logged Treatment, in which hemlocks and commercial hardwood species were removed to simulate the effect of preemptive forest management. Small mammal trapping grids spanning 0.49ha and consisting of 49 Sherman live-traps were placed within both blocks of each treatment and were set from June to July 2012. Small scale habitat characteristics (e.g. woody debris, leaf litter, canopy cover, etc) was estimated per treatment. Species richness, the probability of interspecific encounter (PIE), and mark-recapture of population size were estimated for each plot. There was a significant difference in percent ground cover of leaf litter, vegetation, and woody debris, and percent canopy cover among treatments. Estimated species richness was higher in the girdled treatment (9), but not significantly different among all other treatments (6 to 7 species). In hemlock and hardwood controls, deer mice, and southern-flying squirrels were captured more frequently than southern red-backed voles. However, in logged and adelgid, southern red-backed voles and eastern chipmunks were captured more frequently than mice and southern-flying squirrels, but PIE did not differ significant among treatments Disturbance caused by girdling and preemptive logging for forest management affected site occupancy, estimated abundance, and composition of small mammal communities, but did not affect species richness. The was no effect of treatment on deer mice and white-footed mice populations among treatments, but there was a positive affect of logging and girdling on southern red-backed vole populations. The effects on small mammal distribution did not differ significantly between the girdled and
logged treatments, which suggest that preemptive logging is as detrimental to small mammal distribution as the woolly adelgid invasion. Because eastern hemlocks are not expected to recover from the adelgid invasion, there may be widespread changes in the abundance and composition of small mammal assemblages.
2.2 Introduction

Natural habitat in the eastern United States has diminished over the past century because of damage caused by invasive species. Habitat loss and invasive species are the two most common threats to biodiversity in the United States and invasive species were predicted to overtake habitat loss as the single largest threat to biodiversity (Wilcove et al. 2010). Invading plants and animals impact the environment, landscape, and native species differently. Plant invaders usually alter abiotic processes of native ecosystem function (e.g. Ehrenfeld 2003) where animal invaders may compete with native species for resources. Animal invaders cause extinctions of native species through species direct interactions (e.g. predation and competition) whereas, plant invaders often cause shifts in biogeochemical processes that alter nutrient cycling, and local hydrology (review by Mack et al. 2000). Although the routes of disturbance caused by invasive plants and animals differ and the combination of both plant and animal invaders can alter habitat structure (Long 2003, Lizzarralde et al. 2004, Simberloff 2009), invasions often decrease survival of native species (review by Mack et al. 2000). Regardless of whether invasive species are the “driver” or “passenger” of ecological change and extinctions (e.g. Gurevitch and Padilla 2004, MacDougall and Turkington 2005, Didham et al. 2005), disturbance occurs, ecosystem function is disrupted, and ecosystems change.

Ecosystems respond differently to disturbance depending on what organism is or suite of organisms are being affected. The rivet hypothesis (Ehrlich and Ehrlich 1981, chapter 5) states that an ecosystem may withstand the loss of one or a few species without
any discernable effect on ecosystem function, just as a plane (i.e. ecosystem) would still 
function if a few rivets (i.e. species) were lost. However, if too many species in an 
ecosystem were lost, then ecosystem functions would become unstable, just as a plane 
would fall apart if too many rivets were lost. This hypothesize would likely be supported 
if there was no difference in species function (i.e. all species had the same function in the 
ecosystem similar to all rivets have the same function of holding the plane together). 
However, not all species within an ecosystem preform the same function. Within a given 
ecosystem, there may be multiple species that carry out similar functional roles while few 
species may contribute relatively little to the overall ecosystem function (review by 
Hopper et al. 2005). Ecosystem stability is strongly influenced by particular species’ 
function roles properties. The loss of important native species, such as foundation species 
due to habitat loss and invasive species, can potentially influence the structure of plant 
and animal communities because of the unique roles foundation species play in 
ecosystem function (Dayton 1972, Ellison et al. 2005a).

Foundation species are often abundant primary producers that provide habitat 
for other species, control population dynamics, and create locally stable conditions for 
other species. They support communities by modulating and stabilizing fundamental 
ecosystem process (Dayton 1972; Ellison et al. 2005a). Because foundation species are 
often primary producers that provide habitat and stabilize biogeochemical processes, the 
removal of foundation species often elicits a community wide bottom-up trophic cascade 
(e.g. Carpenter et al. 1985, Polis 1999, Persson 1999, Scherber et al. 2010, Baiser et al.
Foundation species loss can impact nutrient fluxes (Jenkins et al. 1999), microclimate conditions (Snyder et al. 2002), food web system (Baiser et al. 2013), and biodiversity (e.g. Tingley et al. 2002, Snyder et al. 2002, Ellison et al. 2005b).

Eastern hemlock (*Tsuga canadensis* (L.) Carrière) is such a foundation species because they are abundant, are primary producers, and connect with many other species in the food web (Ellison et al. 2005, Foster 2014). Their populations have declined dramatically in the eastern United States because of the damage inflicted on them by invasive insect pests (McClure 1991, Orwig and Foster 1998, Orwig et al. 2002, see Kizlinski et al. 2002; Ellison et al. 2005a). The hemlock woolly adelgid (*Adelges tsugae*, Annand 1928) is native to Japan and was introduced to the United States in the 1950’s (McClure 1989). The hemlock woolly adelgid is a sap sucking insect that defoliates trees (Orwig et al. 2008) and causes rapid hemlock mortality (McCulre 1991). These aphid-like insects are effective dispersers and are introduced to new ranges by wind, birds, deer, and humans through logging practices (McCulre 1990). Because the hemlock woolly adelgid threatens much of the old growth forests in the eastern United States, forest management regimes such as preemptive logging were being examined as a course of action to decrease the adelgid threat and to conserve late successional forests (Foster and Orwig 2006).

The damage caused by hemlock woolly adelgid creates a unique mosaic of a graveyard-like forest that is characterized by having a reduced canopy and standing dead
trees. The decrease in canopy causes an increase in the amount of light that reaches the forest floor, which allows for early successional hardwood species to flourish in the understory (Farnsworth et al. 2012). A woolly adelgid infested forest has different characteristics than a logged forest, which also has a relative open canopy, but woolly adelgid infested forest generally has coarse woody debris littering the forest floor and a slower onset of vegetative understory growth compared to logged forests (Farnsworth et al. 2012). Forest management practices, such as preemptive logging, involve the removal of trees that are newly infested, vulnerable to infestation, or are economically valuable before an infestation occurs with the aim to decrease the spread of infestation and to extract economic value (Foster and Orwig 2006). These forest management regimes also affect forest seed-banks, tree regeneration, and forest dynamics (e.g. Graae and Sunde 2000, Decocq et al. 2004, Farnsworth et al. 2014).

The dramatic changes in forest structure and forest stability caused by logging or invasion can shift ecosystem processes followed by shifts in biodiversity. These shifts within the ecosystem affect taxonomic groups differently. For example, loss of eastern hemlocks results in an increase in local ant species diversity (Ellison et al. 2005b) and a decrease in regional bird population composition and distribution (Tingley et al. 2002). In contrast, in later studies, loss of hemlock does not appear to impact ants, beetles, or spiders (Sackett et al. 2011), at least at a local scale. These inconsistent responses to hemlock loss over varying temporal and spatial scales make it difficult to predict how species across multiple taxa will prevail after the loss of hemlocks, which are not
expected to recover from hemlock woolly adelgid invasion (Foster 2014 hemlock book). This imminent loss of eastern hemlock foundation species will cause ecosystem changes within the forest, but how this change in habitat structure impact supported species distribution is generally unknown and may be different depending on the species’ habitat associations.

The order Rodentia is the most species rich and most diverse order of mammals. They provide many ecosystem functions across a large environmental gradient. Observational and experimental data support the vital roles of habitat structure (e.g. Abramsky 1988, Stapp 1997), habitat selection (e.g. Abramsky et al. 1990, Vickery and Rivest 1992, Morris 1996), and microhabitat characteristics (e.g. Brown 1988, Drickamer 1990, Fauteux et al. 2012) in determining species richness and composition for several mammal species. Many small mammals are omnivorous or generalists, but they do tend to prefer particular diets and are associated particular habitats.

Rodents are often used as model organisms (Barrett and Peles 1999) for ecological studies to investigate ecological studies, because they are abundant, diverse, can be used as bio-indicators for forest health (Haim and Izhaki 1994; Pearce and Veiner 2005), and because there is detailed information about their biology and natural history. This diverse order comprises of species that serve numerous important ecological functions including forest regeneration, forest range extension, vegetation facilitation, disease transmission, and they are a resource to other animals.
Questions and Hypotheses

How will small mammal communities be affected if hemlock forest structure is altered by hemlock woolly adelgid invasion or preemptive logging? A disturbed hemlock forests may diminish habitat quality for some species while others may thrive in the disturbed habitats. The disturbance in forest structure driven by hemlock loss may shift small mammal species distribution, because many small mammals have different habitat requirements and microhabitat associations. Changes in forest structure triggered by hemlock die-offs can affect the dispersal of small mammals in eastern forests differently because of the eastern hemlock’s unique abilities as a foundation species to stabilize biogeochemical processes (e.g. soil moisture and local climate) and community structure, which both contribute to small mammal dispersal. I used several standard community estimates (species richness and probability of interspecific encounter) and population estimates (mark-recapture, Schnabel) to quantify the effects of hemlock die-off caused by forest disturbance on small mammals.

Because damage to foundation species may greatly alter habitat structure and available resources, I hypothesized that widespread mortality of hemlock trees would affect small mammal distribution within these disturbed forests. However, because animals have diverse feeding guilds and have diverse habitat requirements, it is not clear how particular species will respond to forest disturbance. My objectives were to 1) briefly describe the Hemlock-Removal Experiment at Harvard Forest’s (HF-HeRE) Long Term Ecological Research LTER) and how the HF experiment impacts forest habitat structure
at a fine scale that may be associated with small mammal occurrence and 2) use community and population assessment methods to describe changes in small mammal assemblages within the experiment. The purpose of this work is to quantify the effects of eastern hemlock loss due to preemptive logging and simulated hemlock woolly adelgid invasion on 1) species richness, 2) community assemblage, and 3) population density. I hypothesized the 1) species richness, communities, and population estimates differ in canopy treatments depending on the habitat requirements of the small mammal relative to hemlock stands. I predicted that deer mice and white-footed mice would not be affected by habitat treatment, but all other species identified would (Table 1).

2.3 Methods

Site Description

My work was conducted in north-central Petersham, Massachusetts, USA (42.47–42.48°N, 72.22–72.21° W; elevation 215–300-m above sea level) within the Hemlock Removal Experiment (HF-HeRE). In 2003, HF-HeRE plots were chosen in hemlock dominated forests. The chosen plots had similar topography and similar aspect. The HF-HeRE is a replicated two-block design with four ~90 x 90-m (0.81-ha) canopy treatments plots. Two of the plots received canopy manipulations and the two plots that did not receive canopy manipulation act as canopy controls. The canopy manipulations were applied in 2005 after baseline vegetation measurements were taken.
The Girdled canopy manipulation was designed to simulate hemlock woolly adelgid (HWA) infestation. Physical damage to trees was applied by girdling all hemlock trees of a particular size (seedlings and saplings) with knives or chainsaws. The girdled trees eventually died in a similar manner to the death caused by HWA damage. Girdled hemlocks die within approximately 2 years after the treatment, but dead trees are left standing for several years’ post-mortem until they fall. Since the Girdled treatment was applied, the canopy density was reduced which resulted in a gradual increase of light availability to the understory over time (Farnsworth et al. 2014).

The Logged canopy manipulation was designed to mimic the effects of preemptive logging (as in many forest management plans) or commercial hemlock-salvage. All merchantable timber (hemlock, white pine, maple, birch, and oak) was harvested and removed. In contrast to the Girdled treatment, there was immediate light availability to the understory in the Logged plot (Farnsworth et al. 2014, Lustenhouwe et al. 2014) which allowed for new vegetative growth of early sessional plants.

The Hemlock control plot was not manipulated and trees were intact to act as a control to Girdled and Logged treatments. The Hardwood plot was not manipulated and represented the future of a Hemlock stand approximately 50 years after HWA invasion and preemptive logging. On average, the daily air and soil temperatures are 2–4 °C warmer and have greater temperature variances in the canopy manipulate plots than in the hemlock control plots (Lustenhouwer et al. 2012). At the time the experiment developed,
the hemlock woolly adelgid was not present in Massachusetts.

**Sample Grid Layout**

To examine how the reduction of the foundation species, Eastern Hemlock, affects small mammal habitat and small mammal species richness, community evenness, and populations, I utilized a grid layout. Sampling grids spanned 0.49-ha and sampling locations were placed 10-m apart by pacing in a $7 \times 7$ array within each of the two Hemlock, Hardwood, Girdled, and Logged plots (n=392). The boarders of the sampling grids were placed at least 5-m from the edge of the plot to minimize the likelihood of catching animals from outside the grid and to a count for a border strip. Grids were paced in a way to cover the most homogenous topography with the least amount of slope relief as possible.

A grid-based trapping scheme was used instead of web-based (Parmenter et al. 2003) or transect-based (Pearson and Ruggiero 2003) trapping schemes, because I wanted to maximize the effective trapping area with the minimum number of traps for the restricted area. While a web-based trapping scheme is more accurate at density estimation than a grid-based trapping scheme, a larger area is needed for a web-based trapping scheme with the relatively same number of traps (Parmenter et al. 2003). A transect-based trapping scheme results in more total captures and greater species richness than a grid-based trapping scheme, but only when there are few sampling returns (Pearson and Ruggiero 2003). There is also a larger effective trapping area needed for transect-based
scheme (Pearson and Ruggiero 2003). The goals were to have traps available within 10-m radius to 1) reduce trap competition, which increases trap, or site availability for detection and 2) catch a high fraction of present animals (e.g. Krebs 1966, Krebs et al. 2011).

**Microhabitat Characteristics**

To quantify small scale habitat characteristics that may affect small mammals, I photographed each sample location on the sample grid. From August 9th to 31st 2013, digital photographs of the ground and canopy of each trapping location were taken. Approximately 1-m² quadrates were placed over the trap location and then photographed. The camera was placed at the same location as the trap ground photographs were taken approximately 1m from the ground to capture the entire 1m² quadrat. Canopy photos were taken approximately 1m from the ground with the lens pointing to the canopy. Each ground photo (n=392) and each canopy photo (n=392) was labeled and scored using ImageJ (1.42q Java 1.6.0_version 10). Fifty points were randomly generated (Appendix A) and overlaid on each digital photograph. The points (n=50) determined which characteristics that might be important to describe small mammal distribution were recorded. Ground characteristics included 1) rock, 2) soil, 3) woody debris, 4) leaf litter, 5) fungi, and 6) vegetation. Canopy characteristics included 1) open, which was open sky or no canopy cover, 2) high, which was characterized by canopy that was relatively far from the ground and considered old growth, and 3) low, which was characterized by the canopy that was near the ground and considered new growth (Table 1.1). Tree canopy
was scored as “low” if the vegetation reached the photographers lens (approximately 1-m from the ground) and was scored “high” if the vegetation was greater than the photographers lens height (approximately 2-m and above). Each characteristic (i.e. rock, soil, vegetation, high canopy, etc…) for each sampling location (n=392) was calculated as a percent.

Randomized block analysis of variance (Randomized Block ANOVA) was used to determine significant difference of habitat characteristics among treatments. Tukey’s Honest Significant Difference (HSD) post hoc pair-wise comparison test was used to identify differences between means that were greater than the expected standard error for the particular treatment (Hemlock, Girdled, Logged, and Hardwood) (Tukey 1949, e.g. Gotelli and Ellison 2013).

Small Mammal Live Trapping

Sherman traps (H. B. Sherman, Tallahassee, FL USA) (9 x 9 x 3 inches) were then placed within approximately 0.25-m of the actual grid and trap openings were haphazardly arranged. The goal was to promote captures, but not at the cost of assuming non-random captures (see Hulbert 1985, Bowman et al. 2001). Sherman traps were used because they are more effective at capturing relatively larger small mammals than pitfall traps; however, there is a bias towards capture of Peromyscus spp. when using Sherman traps (Dizney et al. 2008). Traps were baited originally with peanut butter, oats, and sunflower seeds; however, there was a high frequency of trap disturbance from trap
predators (black bears, raccoons, and grey squirrels) within the first two trapping nights in 2012. After the first two trapping nights, I used only sunflower seeds as bait to decrease trap disturbance. Clean raw cotton was used for insulation. Traps were set late afternoon to dusk hours and traps were checked from pre-dawn to dawn hours to 1) limit sampling to nocturnal small mammals and 2) to decrease stress caused by long term captivity.

Captured animals were identified to species based on morphological characteristics. Individual rodents were marked with colored non-toxic permanent ink. The color used was chosen based on the treatment the individual was captured and individual were not uniquely marked. Individuals were released at the same trapping location in which they were captured. All traps were closed or folded down during the day and non-tapping nights to decrease the risk of accidental deaths. All handling complied with rules and regulations set forth by the Animal Welfare Act and Institutional Animal Care and Use Committee from University of Vermont (12-019) and Harvard University (12-04). Scientific collecting permits were obtained from the Massachusetts Department of Fish and Game (075.15SCM).

Animals were captured in 2012 during summer months of June and July. Trapping was conducted during full, new, and half-moon conditions. NASA’s Moonphase 3.3 (Tingstrom 2009) was used to determine the percentage of moon phase (illumination) for each trapping night. Traps were set for two consecutive nights in each
block during similar peak moon phases. For example, the Valley block was set during the moon phases reaching to full moon, which had a moon phase 99% and 100% for two nights. Then, the Ridge block was set after full moon, when the moon was falling, which had a moon phase of 99% and 98%. For the following month, the block order was reversed so that Ridge block traps were set with 99% and 100% moon phase and Valley was trapped with 99% and 98% moon phase.

There were 4,131 trapping nights and 18.7% capture success among all treatments. I trapped in the Ridge block for 12 nights and in the Valley for 10 nights. There were 2,183 traps set in the Ridge block and trapping success varied among Hemlock (17%), Girdled (22%), Logged (16%), and Hardwood (22%) treatments. There were fewer traps set in the Ridge Hardwood treatment (n=420) than other treatments in the Ridge block (n=588) due to a change in property management. In the Valley block, there were 1,948 traps set and trapping success varied among Hemlock (14%), Girdled (20%), Logged (15%), and Hardwood (24%) treatments. Although there was a slight different in the number of traps used in the Ridge Hardwood plot than in the Valley Hardwood plot, the percent trapping success was comparable.

Species Richness and Evenness

I used Chao1 (Chao 1984) abundance methods to estimate species richness among the treatments with 95% upper and lower confidence intervals. I used shared abundance methods to estimate the number of shared species between the Hemlock
control and the other treatments (Chao et al. 2000). The relative abundance or the proportion of the total assemblage that is represented by each species was calculated for each treatment. The average probability of interspecific encounter (PIE) was used to estimate species evenness for each treatment (Hurlbet 1971, Gotelli 2008). Confidence intervals for PIE were calculated from the standard deviations from the replicated treatments. My null hypotheses, which stated that there is no significant difference in species richness, shared species richness and PIE among treatments, were determined by comparing the 95% confidence interval for the difference between the treatment means. If the confidence interval between treatments did not contain zero, the null hypothesis was rejected (Knezevic 2008).

Population Estimates with Mark-Recapture

The ratio of marked individuals to not marked individuals was used to estimate the population in each treatment using Schnabel cumulative marking estimates. Two major assumptions of the Schnabel mark-recapture method are that the population remain closed throughout the study and marks are not lost. To insure that this closed population assumption was not violated, 10 out of the 12 nights were used for population estimates. Another assumption is that species were correctly identified. Correct identification in the field of *Peromyscus maniculatus* and *Peromyscus leucopus* based on external morphology alone is very difficult and can lead to misidentification (Rich et al. 1996). These species are usually distinguished by their behavior, pelage, and tail bi-coloration. Because of this identification difficulty, these species are often grouped into “*Peromyscus*
spp.” category. Other characteristics, such as using genetic markers and salivary amylase (Kilpatrick et al. 1994), are more accurate for distinguishing between the two species (Rich et al. 1996). For identification in the field, I distinguished between *P. maniculatus* and *P. leucopus* using behavior, pelage, and tail bi-coloration characteristics. In addition, I attempted to use the non-lethal, non-invasive method of saliva collection as outlined by Rich et al. (1996) to correctly distinguish between *P. maniculatus* and *P. leucopus*. Saliva was collected by rinsing the mouth of live captured *Peromyscus spp.* with 2-mL of distilled water with sterile 3-mL pipettes. The saliva sample was stored in ice cooler during field collection hours and immediately transported to -80°C freezer in Torrey Lab at Harvard Forest and back to the University of Vermont. Unfortunately, these samples were not viable for amylase analysis so deermice and white-footed mice were identified from morphological and behavioral differences.

**Software and R Packages**

All data were analyzed using R version 3.2.3 (2015). The package “reshape” version 0.8.5 (Wickham 2014) was used to restructure and aggregate data. The package “plyr” version 1.8.3 (Wickham 2015) was used for data frame manipulation. Packages “lattice” version 0.20-33 (Sarkar 2015), “ggplot2” version 2.0.0 (Wickham and Chang 2015), and “grid” version 3.2.3 (Murrell 2005) were used for graphics. The package “agricolae” version 1.2.3 (de Meniburu 2015) was used for Tukey’s HSD grouping statistical procedures for microhabitat characteristics. The package “SpadeR” version 0.1.0 (Chao et al. 2015) was used to estimate species richness (‘ChaoSpecies’) and
estimated shared species richness (‘ChaoShared’).

2.4 Results

*Microhabitat Characteristics*

There was no significant difference in the percent of rock cover among Hemlock (1.94%, SE= 0.39), Girdled (0.63%, SE= 0.33), Logged (1.04%, SE= 0.38), and Hardwood (1.29%, SE= 0.32) (P = 0.066, df = 3, residuals= 387, sum²= 0.008, mean²= 0.003, F-value= 2.412), but there was a significant difference between Ridge and Valley blocks (P< 0.00001, df= 1, F-value= 22.34; Figure 2.1, panel A). There was a significant difference in leaf litter ground cover among treatments (P< 0.0001, df= 3, residuals= 387, sum²= 5.36, mean²= 1.79, F-value= 53.62) and among blocks (P < 0.0001, df= 1, F-Value= 20.23, Figure 2.1 panel B). There was a difference in means among Hemlock (51.94%, SE= 1.90), Girdled (24.84%, SE= 1.78), and Logged (33.10%, SE= 1.81), but there was not a difference in means of percent leaf litter between Hemlock and Hardwood (51.29%, SE=2.06) (Figure 2.1, panel B).

There was a significant difference in percent soil cover among treatments (P< 0.0001, df= 3, residuals= 387, sum²= 0.92, mean²= 0.31, F-value= 22.587), but not between blocks (P= 0.42, df= 1, F-value= 0.651, Figure 2.1, panel C). There was a higher percent of soil cover on average in the Hemlock (15.69%, SE=1.63) treatments and the lowest percent of soil cover on average in the Hardwood (2.92%, SE= 0.66) treatment. There was no difference in soil means between the Girdled (5.22%, SE= 0.79) and
Logged (9.12%, SE= 1.34) treatments (Figure 2.1; panel C).

There was a significant difference in the percent of vegetation ground cover among treatments ($P< 0.0001$, df= 3, residuals= 387, $\text{sum}^2 = 8.08$, $\text{mean}^2 = 2.69$, $F$-value= 76.06) and between blocks ($P< 0.0001$, df= 1, $F$-value= 22.93). There was a higher percent of vegetation ground cover in the Girdled (45.60%, SE= 2.39) treatment, but no difference between Logged (32.92%, SE= 2.18) and Hardwood (29.06%, SE= 1.85) treatments. Hemlock (5.86%, SE=1.15) treatments had the lowest percent vegetation cover among the treatments (Figure 2.1 panel D). There was a significant difference in percent woody debris cover among treatments ($P< 0.0001$, df= 3, residuals= 387, $\text{sum}^2 = 0.72$, $\text{mean}^2 = 0.24$, $F$-value= 15.68) and between blocks ($P< 0.0001$, df= 1, $F$-value= 29.42). There was a higher percent cover in Hemlock and Logged treatments, but there was no difference in the mean of woody debris cover between Hemlock (18.67%, SE= 1.26) and Logged (20.10%, SE=1.66). There was no difference between Girdled (12.92%, SE= 1.29) and Hardwood (9.51%, SE=0.83) means (Figure 2.1; panel E). There was a significant difference in percent fungi cover among treatments ($P< 0.0001$, df= 3, residuals= 387, $\text{sum}^2 = 0.006$, $\text{mean}^2 = 0.002$, $F$-value= 7.06) and between blocks ($P< 0.0001$, df= 1, $F$-value= 8.036). There was no difference in the mean of fungi cover between Logged (0.88%, SE= 0.29) and Hardwood (0.67%, SE= 0.17) and no difference between Hemlock (0.02%, SE= 0.02) and Girdled (0.0%, SE= 0.0) (Figure 2.1; panel F). However, values were extremely low in all treatments.
There was a significant difference in percent of open canopy among treatments (P< 0.0001, df= 3, residuals= 387, sum²= 1.28, mean²= 0.42, F-value= 28.332), but not between blocks (P= 0.069, df= 1, F-value= 3.33). There was a difference in mean among Hemlock (8.22%, SE= 0.49), Girdled (23.14%, SE= 1.52), Logged (19.65%, SE= 1.82), and Hardwood (13.51%, SE= 0.62), but not between Girdled and Logged treatments (Figure 2.2, panel A). There was a significant difference in percent of high canopy cover among treatments (P< 0.0001, df= 3, residuals= 387, sum²= 27.56, mean²= 9.187, F-value= 169.02), but not between blocks (P= 0.21, df= 1, F-value= 1.56). There was not a significantly higher percent in high canopy between Hemlock (86.41%, SE= 1.08) and Hardwood (79.55%, SE= 1.73) treatments, but there was a difference lower percent of high canopy in Girdled (37.49%, SE= 2.97) and Logged (24.47%, SE=3.04) treatments with Logged having the least amount of high canopy cover (Figure 2.2, panel B). There was a significant difference in low canopy cover among treatments (P< 0.0001, df= 1, residuals= 387, sum²= 18.18, mean²= 6.06, F-value= 85.12), but not between block (P=0.05, df= 1, F-Value= 3.77). The Logged (55.84%, SE= 3.59) treatment had a higher percent canopy cover than Girdled (39.31%, SE= 3.57), Hemlock (5.37%, SE= 0.92), and Hardwood (6.90%, SE= 1.66). However, there was not a difference in low canopy percent cover between Hemlock and Hardwood (Figure 2.2, panel C).

Species Richness and Evenness

The observed small mammal species (i.e. rodents and shrews) varied slightly among treatments (Figure 2.3). There were more observed species found in the Girdled
than in the Logged (7), Hemlock (6) and Hardwood (6) treatments. Deermice, white-footed mice, southern red-backed voles, and short-tailed shrews were found among all treatments (Figure 2.3). Southern flying squirrels were most abundantly found in the control plots, but one was captured in the Girdled plots (Figure 2.3). Eastern chipmunks and masked shrews were more abundant in disturbed treatments than in controls (Figure 2.3). Chipmunks were not seen in the Hemlock control and masked shrews were not captured in the Hardwood controls (Figure 2.3). Woodland jumping mice and woodland voles were only captured in the disturbed treatments and with very low captures (Figure 2.3). Deermice realitve capture abundance ranked highest in Hemlock, Logged, and Hardwood and southern red-backed vole abundance ranked highest in the Girdled treatment (Figure 2.3). There was no significant difference in the average PIE (P=0.79, df= 3, mean²= 0.003, F-value= 0.34) among Hemlock (PIE= 0.59, lower 95%CI= 0.14, upper 95%CI= 1.00), Girdled (PIE= 0.63, lower 95%CI= 0.63, upper 95%CI= 0.74), Logged (PIE= 0.68, lower 95%CI=0.63, upper 95%CI= 0.73), and Hardwood (PIE= 0.63, lower 95%CI= 0.58, upper 95%CI= 0.68, Figure 2.4) treatments or among blocks (P= 0.22, df= 1, mean²= 0.01, F-value= 2.29).

The estimated species richness was highest in the Girdled treatment (n= 8, lower 95%CI = 8.07, upper 95%CI = 9.59, Figure 2.5), followed by the Logged treatment (n= 7, lower 95%CI = 7.0, upper 95%CI = 8.45 Figure 2.5). The estimated species richness was the same (n=6) in the Hemlock (lower 95%CI = 6.0, upper 95%CI = 7.40) and in Hardwood controls (lower 95%CI = 6.0, upper 95%CI = 6.49, (Figure 2.5). There was a
significant difference in estimate species richness between the Hemlock and Hardwood controls and Girdled treatment, but not between controls and Logged treatment (Figure 2.5). There was no significant difference between Hemlock and Hardwood controls and between Girdled and Logged treatments (Figure 2.5). There were six shared species between Hemlock and Girdled (SE= 0.57, lower 95%CI = 5.35, upper 95%CI = 7.85), five shared between Hemlock and Logged (SE= 0.46, lower 95%CI = 4.43, upper 95%CI = 6.29) and Hemlock and Hardwood (SE= 0.0, lower 95%CI = 5.0, upper 95%CI = 5.0, Figure 2.6).

Population Estimates

The population was estimated with 10 nights of trapping so there is an equal number of nights among blocks. There was a denser population of deermice in the Logged treatment (N-hat = 40.7 per 0.64ha, lower 95%CI = 27.17, upper 95%CI = 64.33) than in the Hemlock control (N-hat = 17.14 per 0.64ha, lower 95%CI = 13.19, upper 95%CI = 24.47, Figure 2.7), but all other treatments do not have overlapping error bars (Knezevic 2008). There was a denser population of southern-red backed voles in the Girdled (N-hat = 84.4 per 0.64ha, lower 95%CI = 59.80, upper 95%CI = 136.41) and Logged treatments (N-hat = 47.11 per 0.64ha, lower 95%CI = 31.15, upper 95%CI = 85.62) than the Hemlock (N-hat = 8.14 per 0.64ha, lower 95%CI = 4.85, upper 95%CI = 14.89) and Hardwood controls (N-hat= 17.2 per 0.64ha, lower 95%CI = 12.73, upper 95%CI = 25.43, Figure 2.7). There was no difference in population density of white-footed mice among Hemlock (N-hat= 9.0 per 0.64ha, lower 95%CI = 4.74, upper 95%CI
2.5 Conclusion

The purpose of this study was to 1) briefly describe microhabitat characteristics in disturbed forests that are known to influence small mammal distribution and 2) determine if simulated damage caused by hemlock woolly adelgid and preemptive logging impact small mammal communities. Even though there is a species capture bias with Sherman traps, some inferences on species richness, abundance, and evenness can be made. I found small scale microhabitat characteristics (Figures 2.1 and 2.2), small mammal species richness (Figure 2.5), and vole populations were generally affected by girdled and logged disturbance, but relative abundance (Figure 2.3), community evenness (Figure 2.4), and mice populations (Figure 2.6) did not seem affected by the disturbance.

Overall, estimated species richness did increase in the Girdled treatment relative to the Hemlock control (Figure 2.5). There were more species represented in the Girdled treatment than in the Hemlock (Figures 2.3 & 2.5). Given the number of observed species and the average estimated species richness from Chao1 for asymptotic species richness were similar, I conclude my sampling was thorough. However, not all species that were sampled were found in the Girdled treatment and several were rarely captured (Figure 2.3). Community evenness did not differ among treatments (Figure 2.4), but this could be
due to the large variation of PIE estimates in the Hemlock controls (Figure 2.4). Even though hemlock woolly adelgid simulation did not affect shared species richness or evenness, there was a difference in the species richness estimates between Girdled and Hemlock and Hardwood controls. It seems that habitat generalists (e.g. deermice, white-footed mice) may not be as impacted by hemlock woolly adelgid as habitat specialists (e.g southern flying squirrels).

Southern flying squirrels were not captured in the logged treatment. Only one southern flying squirrel was captured in the Girdled treatment and this individual was originally captured in the Hemlock control. This suggests that the presence of southern flying squirrels or site occupancy may decrease as hemlock woolly adelgid continues to spread and destroy hemlock forests in New England. Given that no southern flying squirrels were found in the logged treatments it seems safe to assume that preemptive logging management would be equally devastating to these arboreal rodents as girdling form hemlock woolly adelgid. Although northern flying squirrels were not captured in this study, I hypothesize that their populations would also decrease dramatically as adelgid spreads. Unlike southern flying squirrels that utilize both hemlock and hardwood stands (primarily hardwood), northern flying squirrels depend on old growth forests (Ransome and Sullivan 1997). If the spread of the adelgid continues to increase, the northern flying squirrels may not have time to adapt to the changing forests.

Although the deermouse and white-footed mouse populations were not affected
by girdled and logged disturbances, the southern red-backed vole populations were affected (Figure 2.7).

My results indicate that the distribution of the short-tailed shrew and southern red-backed vole differ from that of previous work by DeGraff et al. (1991). Where this study shows that short-tailed shrews are captured more frequently in hemlock (softwood) stands (Figure 2.3), DeGraff et al. (1991) found that short-tailed shrews were more abundant in hardwood stands than softwood. However, these differences may be artifacts of my sampling methodology. For example, I used Sherman live-traps only, whereas DeGraff et al. (1991) used multiple sized snap traps. However, the interpretation of southern red-backed voles were similar were abundance did not differ significantly between softwood stands and hardwood stands (Figure 2.7).

These data suggest that there are varying degrees in which small mammal communities will be impacted with continued spread of hemlock woolly adelgid and that destructive management practices (preemptive logging) would impact rodent communities analogous to hemlock woolly adelgid. These changes in forest structure and small changes in small mammal species richness may have a negative feedback loop on forest dynamics. Small mammals provide many ecosystem functions across a large environmental gradient. Rodents are often used as model organisms (Barrett and Peles 1999) to investigate ecological studies, because they are abundant, diverse, can be used as bio-indicators for forest health (Haim and Izhaki 1994, Leis et al. 2008), and there is
detailed information about their biology and natural history (Barrett and Peles 1999). This diverse order comprises of species that serve numerous important ecological functions including forest regeneration, forest range extension, vegetation facilitation, disease transmission, and they are a resource to other animals. Granivorous small mammals are often seed dispersers that can increase forest range (e.g. Steele et al. 2006, Beck and Vander Wall 2010, Yu et al. 2013). North American mice, squirrels, and chipmunks, are generally referred to as granivorous and have been known to influence seed fate through seed foraging, dispersal, caching, and hoarding (e.g. Steele et al. 2006, Beck and Vander Wall 2010, Steele et al. 2011). Herbivorous small mammals can facilitate vegetation growth in forests and fields (Ostfeld and Canham 1993, Howe et al. 2006). They are also important food resources for many vertebrates such as birds, snakes, and mammals (e.g. Sullivan et al. 2004, Sundell et al. 2013). Small mammals are hosts to diverse groups of parasites (Kuhnen et al. 2012) and they contribute to the biodiversity by being hosts to a variety of endoparasites such as nematodes, protozoans, and cestodes, (e.g. Pedersen 2005, Vandergrift et al. 2009, Pedersen and Antonovics 2013) as well as ectoparasites (Rand et al. 1993) such as botflies (e.g. Burns et al. 2005, Cramer and Cameron 2006 & 2007) and ticks (e.g. Van Buskirk and Ostfeld 1995, Awerbuch and Sandberg 1995).
Table 2.1. Hypothesized influence on relative abundance of individuals within small mammal communities in Hemlock, Girdled, Logged, and Hardwood canopy treatments at HF-HeRE in Petersham, MA.

<table>
<thead>
<tr>
<th>Habitat Treatment</th>
<th>Predicted influence of HWA on species relative to Hemlock Treatment</th>
<th>Supporting Literature</th>
</tr>
</thead>
</table>
| Deer mouse  
*Peromyscus maniculatus* | 0 | Wolff 1985 |
| White-footed mouse  
*Peromyscus leucopus* | ↓ | Henein et al. 1998 |
| Woodland jumping mouse  
*Napaeozapus insignis* | ↑ | Vickery et al. 1992 |
| Southern red-backed vole  
*Myodes gapperi* | ↓ | Merritt 1981 |
| Woodland vole  
*Microtus pinetorum* | ↓ | Smolen 1981 |
| Southern flying squirrel  
*Glaucomys volans* | ↑ | Taulman 2000 |
| American red squirrel  
*Tamiasciurus hudsonicus* | ↑ | Ransome et al. 1997 |
| Eastern chipmunk  
*Tamias striatus* | ↓ | Pyare et al. 1993 |
| Short-tailed shrew  
*Blarina brevicauda* | ↑ | George et al. 1986; Ford and Rodrigue 2001 |
| Smokey shrew  
*Sorex fumeus* | 0 ↓ | Ford and Rodrigue 2001 |
| Masked shrew (Common shrew)  
*Sorex cinereus* | 0↓ | Getz 1961, Brown 1967; Ford and Rodrigue 2001 |
Figure 2.1. Mean (± SE) percent cover of microhabitat ground cover characteristics of rock (A), leaf litter (B), soil (C), vegetation (D), woody debris (E), and fungi (F) among Hemlock (green), Girdled (orange), Logged (yellow), and Hardwood (blue) treatments. Results of randomized block ANOVA for each characteristic indicated top-center of each graph. Lower case letters result of Tukey’s HSD grouping.
Figure 2.2. Mean (± SE) percent cover of microhabitat canopy cover characteristics open canopy (A), high canopy (B), and low canopy (C) among Hemlock (green), Girdled (orange), Logged (yellow), and Hardwood (blue) treatments. Results of randomized block ANOVA for each characteristic indicated top-center of each graph. Lower case letters above treatment are result of Tukey’s HSD grouping.
Figure 2.3. Rank relative abundance graph of small mammals in 2012 among canopy treatments (right to left: Hemlock, Girdled, Logged, and Hardwood). Each bar and each color represents a different species. The height of the bar is the relative abundance of the species in each treatment.
Figure 2.4. Average PIE and 95% confidence intervals among canopy treatments (P= 0.79, df=3).
Figure 2.5. Chao1 estimated species richness (dots) with lower and upper 95% confidence intervals (error bars) for Hemlock, Girdled, Logged, and Hardwood canopy treatments for years 2012. Letters indicate groupings based on CI overlap where same letters different letters indicate significantly different groups. The estimated species richness in Hemlock (a) control differs significantly from Girdled (c), but not from Logged (b) and Hardwood (a). Girdled treatment (c) differs significantly from Hardwood (a), but not Logged (c).
Figure 2.6. Chao1 Shared with Hemlock control estimated species richness (dots) with lower and upper 95% confidence intervals (error bars) for Girdled, Logged, and Hardwood canopy treatments.
Figure 2.7. Schnebel estimated population (N-hat) with lower and upper 95% confidence intervals (error bars) for Hemlock, Girdled, Logged, and Hardwood canopy treatments for deermice, southern red-backed voles, and white-footed mice (top to bottom).
Appendix. The first two principle component scores from the PCA on microhabitat characteristics.
LOCAL OCCUPANCY, COLONIZATION, AND EXTINCTION OF SOUTHERN RED-BACKED VOLEs IN DISTURBED EASTERN HEMLOCK FORESTS

Allyson L. Degrassi

3.1 Abstract

Currently, forests in the eastern United States are radically changing due to the loss of a forest foundation species, the eastern hemlock. It is unclear how forest disturbance influences the already complex patterns of colonization and extinction of small mammals. We developed 38 different multi-season occupancy models which reflected our hypotheses that patterns of local colonization and extinction of southern red-backed voles would be a function of environmental characteristics (landscape, habitat type, neighboring habitat, site structure, food resources, and overwintering temperatures) and would be influenced by ramifying effects of disturbance in eastern hemlock forests. Detection/non-detection data for four southern red-backed voles were collected for 10 nights from 392 trapping sites and four habitats from June and July 2012-2014 in the Harvard Forest’s Long-Term Ecological Research experiment (LTER). The HF-LTER experiment is a replicated two-block design with slightly different elevations (ridge and valley) which includes four 0.81ha canopy treatments: 1) hemlock control, in which hemlocks trees are dominant, 2) hardwood control, in which mid-successional hardwoods are dominant, 3) girdled treatment, in which hemlock trees have been girdled and killed to simulate the impact of hemlock woolly adelgid (HWA), and 4) logged treatment, in which hemlocks and commercial hardwood species have been removed. Small mammal trapping grids spanning 0.49ha and consisting of 49 Sherman live-traps were set in June and July from 2012-2014 within each canopy treatment. We found that 1) sites lower in elevation had consistently higher site occupancy probability than sites at higher elevations, 2) gridled and logged disturbed sites has a significantly higher probability of site occupancy than site within hemlock controls, 3) food resources (seed and vegetation) was a function of the probability of site colonization, 4) no covariate we analyzed influenced the probability of extinction and extinction was a constant among all sites, and 5) the average soil temperature during trapping events increased the detection probability. Our data suggest that as forest disturbance caused by HWA continues to spread through New England, southern red-backed voles will colonize and occupy these disturbed forests. We think this model approach can be used for estimating patterns in occupancy and colonization over time for southern red-backed voles, but patterns of extinction may be better explained by biological factors (e.g. social behavior) that are not associated with state occupancy models.
3.2 Introduction

It is well accepted that rodents demonstrate varying patterns of colonization and extinction that result in population fluctuations that vary from non-cyclic (e.g. Predavec et al. 2001) to highly cyclic (e.g. Jett and Nichols 1987, Krebs 1996, Stenseth et al. 1996, Krebs et al. 2002, Boonstra et al. 2012, Krebs 2013 and research therein). The numerous and diverse underlying mechanisms that drive these colonization and extinction patterns are continuously up for debate (Norrdahl 1995). Although several geographical factors (e.g., latitude and elevation) and biological factors (e.g., density dependence, food abundance, life history traits) can influence cycling patterns, changes in habitats further masks these colonization and extinction trends. Currently, forests in the eastern United States are radically changing due to the loss of the eastern hemlock (*Tsuga canadensis* (L.) Carrière), a forest foundation species (e.g. Dayton 1972, Ellison et al. 2005), and it is unclear how loss if these trees and associated habitat will affect the already complex patterns of colonization and extinction of small mammals.

Mechanisms that determine a change in occupancy pattern in a site, and whether these mechanisms yield cyclic patterns of occupancy through time are of special interest to population ecologists (e.g. Krebs 2013). Local colonization rates (i.e., the probability that an unoccupied site in year $t$ will become occupied in year $t + 1$) and local extinction rates (i.e., the probability that an unoccupied site in year $t$ will stay unoccupied in year $t + 1$) may vary depending on landscape characteristics (e.g. elevation, habitat type and surrounding habitat type), food resources (e.g. seed and vegetation), site structure (e.g.
woody debris and leaf litter) and climate (e.g. overwinter conditions). Although all characteristics are related at some level, the diverse mixture of site characteristics shape and impact the probability of local site occupancy, colonization, and extinction of animals, in particular small mammals.

At any given time, observational and experimental data support the vital roles of landscape (e.g. Barrett and Peles 1999 and papers within), habitat structure (e.g. Abramsky 1988, Stapp 1997, Brown 1988, Drickamer 1990, Fauteux et al. 2012), and habitat selection (e.g. Abramsky et al. 1990, Vickery and Rivest 1992, Morris 1996) in determining species composition for several mammal species. Landscape characteristics such as elevation, habitat type, and surrounding habitat quality can influence colonization and extinction patterns. The heterogeneous habitats within the landscape result in migration. Small mammal populations that occupy areas adjacent to high-quality habitats may show different local cycling patterns than populations that are adjacent to low-quality habitats due to the influx of colonists from the high-quality habitat (Hestbeck 1982). Local community structure of species can be a product of the type of adjacent habitat, which suggests that regardless of the local habitat’s resources (internal within site), the local population would persist as a result of migration from the habitat rich area (Pulliam 1988).

Alternatively, local site colonization rates may vary as a function of food resources, which are believed to contribute to population growth and cycling. Multiple
studies suggest that seed predators (e.g. birds, mice, and voles) population’s growth and cycling correlates with the synchrony of seed production (seed masting) of various tree species within the landscape (e.g. Wolff 1996, Hanski and Henttonen 1996, Selas 2000, Selas et al. 2002, Schmidt 2003, Elias et al. 2006, Zwolak et al. 2016).

Overwinter conditions, such as snow cover and onset of snow melt, correlate with small mammal population numbers (e.g. Hansson 1984, Hansen et al. 1999, Duchesne et al. 2011, Bierman et al. 2006, Hoset et al. 2009). Dispersal patterns can change seasonally where young individuals migrate from old growth forests where winter survival is highest to young growth forests where breeding is high, but winter survival is low (Ecke et al. 2002). This seasonal dispersal patterns suggests that small mammals have the ability to assess habitat quality and their assessment influences their dispersal (see Fretwell and Lucas 1970) and strengthens the role of habitat variation and its effect on small mammal demographics (e.g. reproduction and migration), behavior (e.g. competition and habitat selection), and ecological factors (e.g. food and habitat resources). Immigrants move more frequently among patches than residents (e.g. Pusenius et al. 2000), therefore the variation in the probability that a site will become locally extinct in year $t+1$ is higher in poor quality habitat than in high quality habitats.

Within-site factors (i.e. habitat structure) also affect small mammal habitat associations that influence local site colonization and extinction patterns. Within-site habitat structure, such as woody debris and leaf litter, have contrasting associations with
small mammals. In the family Cricetidae, voles are generally herbivorous and have been known to reduce vegetative growth as severely as large ungulates (Howe et al. 2006). Voles generally are associated with a range of soil moisture and leaf litter coverage. In Canada, southern red-backed voles (Myodes gapperi, Vigors 1830 [formerly Clethrionomys gapperi]) are often associated with boreal mixed forest stands characterized by having downed woody debris, dense shrub layer, coniferous understory and coniferous litter, and moist conditions (Merritt 1981, Vanderwel et al. 2009, Vanderwel et al. 2010). They are positively associated with areas that contain high volumes of coarse woody debris at a fine scale (e.g. Fauteux et al. 2012, Sullivan and Sullivan 2012, Fauteux et al. 2013).

Eastern hemlock populations have declined dramatically in the eastern United States because of the damage inflicted on them by the invasive hemlock woolly adelgid (McClure 1991, Orwig and Foster 1998, Orwig et al. 2002, see Kizlinski et al. 2002; Ellison et al. 2005a). The effect of disturbance is particularly great because eastern hemlock is a foundation species that supports communities by modulating and stabilizing fundamental ecosystem process (sensu Dayton 1972; Ellison et al. 2005a). The damage creates a unique mosaic of a graveyard-like forest that is characterized by having a reduced canopy and standing dead trees. Proposed preemptive logging forest management practices used to stop the spread of non-native species result in disturbed forests that differ in characteristics from an adelgid infested forest. Both disturbed forests have open canopies, but woolly adelgid infested forest generally has coarse woody debris
littering the forest floor and a slower onset of vegetative understory growth compared to logged forests (Farnsworth et al. 2014). These disturbances (invasion and management regimes) can also affect forest seed-banks, tree regeneration, and forest dynamics (e.g. Graae and Sunde 2000, Decocq et al. 2004, Farnsworth et al. 2014), that determine where small mammals live within these forests and may influence population trajectories.

Ultimately, disturbance caused by hemlock woolly adelgid and preemptive logging management alter 1) landscape characteristics such as habitat type and surrounding habitat types, 2) food resources such as seed fall and vegetation, 3) within-habitat structures such as woody debris and leaf litter, and 4) local climate temperatures such as overwinter temperatures. All of these may impact the probability of colonization and extinction of southern red-backed voles in New England. The purpose of this study was to model and predict landscape, habitat structures, food resources, and winter conditions that drive colonization and extinction probabilities for southern red-backed voles in disturbed areas with varying habitat types, habitat structures, food resources, and overwintering temperatures. We used multi-season occupancy modeling (MacKenzie et al. 2003) to test multiple hypotheses (represented by models) that colonization and extinction patterns of southern red-backed voles are functions of habitat characteristics in healthy hemlock forests, disturbed hemlock forest, and hardwood forests. Our objectives were to 1) describe the study system and experimental design, 2) distinguish characteristics among habitat treatments and their link to habitat variables being modeled, 3) assess and relate the variables to colonization and extinction of southern red-backed
voles, and 4) define and test each model within our model set that represent our hypotheses of local habitat characteristics which best explain the probability of small mammal yearly presence in varying habitats.

3.3 Methods

Site Description

This study was conducted in north-central Petersham, Massachusetts, USA (42.47°–42.48°N, 72.22°–72.21° W; elevation 215–300-m above sea level; Figure 1). Animals were captured in the Hemlock Removal Experiment (HF-HeRE), part of Harvard Forest’s Long-Term Ecological Research (LTER) experiment. Study plots are considered to be within the “hemlock-hardwood” transition region of the forest (Keman 1980). The HF-HeRE is a replicated two-block design with slightly varying elevations (Ridge and Valley). Each block had four ~90 x 90-m (0.81-ha) canopy treatments plots (Figure 1). The canopy manipulations were applied in 2005 within hemlock forests after baseline vegetation measurements were taken. For full detailed methodology on the HF-HeRE treatments, refer to Ellison et al. (2010).

Experimental Design: Habitat Treatments

The first treatment, the Girdled canopy manipulation, was designed to simulate hemlock woolly adelgid (HWA) infestation. Physical damage to trees was applied by girdling all hemlock trees of a particular size (seedlings and saplings) with knives or chainsaws. The girdled trees eventually died in a similar manner to the death caused by
HWA damage. Since the Girdled treatment was applied, the canopy density was reduced which resulted in a gradual increase of light availability to the understory over time (Farnsworth et al. 2014). Second, the Logged canopy manipulation was designed to mimic the effects of preemptive logging (as in many forest management plans) or commercial hemlock-salvage. All merchantable timber (hemlock, white pine, maple, birch, and oak) was harvested and removed. In contrast to the Girdled treatment, there was immediate light availability to the understory in the Logged plot (Farnsworth et al. 2014, Lustenhouwe et al. 2014) which allowed for new vegetative growth of early sessional plants. Third, the Hemlock control plot was not manipulated and trees were left intact to act as a control to Girdled and Logged treatments. Fourth, the Hardwood plot was also not manipulated. The Hardwood plot was intended to represent the future of a Hemlock stand approximately 50 years after HWA invasion and preemptive logging. On average, the daily air and soil temperatures are 2–4 °C warmer and have greater temperature variances in the canopy manipulate plots than in the hemlock control plots (Lustenhouwer et al. 2012). Intact hemlock forest surrounds the Hemlock control, Girdled treatment, and the Logged treatments.

*Experimental Design: Small Mammal Captures*

Sampling grids spanned 0.49-ha and sampling locations were placed 10-m apart by pacing in a 7 × 7 array within each of the two Hemlock, Hardwood, Girdled, and Logged plots (n=392) (Figure 1). The borders of the sampling grids were placed at least 5-m from the edge of the plot. Grids were paced to cover the most homogenous
topography with the least amount of slope relief as possible. Animals were captured and released from 2012-2014 during summer months of June and July, as described below. All handling complied with rules and regulations set forth by the Animal Welfare Act and Institutional Animal Care and Use Committee from University of Vermont (12-019) and Harvard University (12-04). Scientific collecting permits were obtained from the Massachusetts Department of Fish and Game.

Sherman traps (H. B. Sherman, Tallahassee, FL USA) (9 x 9 x 3 inches) were placed within approximately 0.25-m of the paced location and trap openings were haphazardly arranged. The goal was to promote captures, but not at the cost of assuming non-random captures (see Hulbert 1985, Bowman et al. 2001). Sherman traps were used because they are more effective at capturing relatively larger small mammals than pitfall traps; however, there is a bias towards capturing *Peromyscus spp.* when using Sherman traps (Dizney et al. 2008).

Traps were baited originally with peanut butter, oats, and sunflower seeds; however, there was a high frequency of trap disturbance from trap predators (black bears, raccoons, and grey squirrels) within the first two trapping nights in 2012. After the first two trapping nights, only sunflower seeds to decrease trap disturbance. Clean raw cotton was used for within-trap insulation. Traps were set late afternoon to dusk hours, and traps were checked from pre-dawn to dawn hours to 1) limit sampling to nocturnal small mammals and 2) decrease stress caused by long term captivity. All traps were closed or
folded down during the day and non-trapping nights to decrease the risk of accidental deaths. Used traps were removed from the field site, washed with water, dipped in 10% bleach solution, rinsed with tap water, and allowed to dry before returning them to the field. Traps were set for two consecutive nights in each treatment, but traps were not set in each block on the same trapping night. This was the desired trapping scheme, but it did not always occur because weather conditions such as wind >15mph and heavy rain prevented us from setting traps as the risk of falling hemlocks in Girdled treatment was high and trap mortality increases for some species with rainy nights and colder nights (Shonefield et al. 2013).

**Statistical Analysis Framework**

We used a multi-season occupancy framework (MacKenzie et al. 2003) to test hypotheses regarding site occupancy ($\psi$), local extinction ($\varepsilon$), and local colonization rates ($\gamma$) represented by 39 models (Table 1). We used the capture data to create detection and non-detection data for each of the 392 trap locations (sites) for each of three summers in 2012-2014. This framework accounts for detection probability ($p$), which in this study is the probability of capturing an animal given it is present at a site. However, the probability of detection is considered a “nuisance” parameter and thus, an additional model set was to select the covariate most likely responsible for the detection of southern red-backed voles (Table 1). A key assumption of this framework is that, within a summer, site occupancy patterns remains fixed while occupancy patterns can change between summers. Although sites were separated by 10-m and individuals could travel
among traps within a treatment, we believe that we have sufficiently represented the occupancy state of these species as we are estimating occupancy within a treatment and these sites are subsamples of the treatment. Furthermore, a large percentage of sites remained unoccupied throughout the duration of this work with 51% of sites (traps) remaining unoccupied by southern red-backed voles.

**Detection, Occupancy, Colonization, and Extinction Characteristics**

We considered six covariates, in addition to the null model, that may affect the probability of detection of a target species. Daily variation in the local weather patterns has been known to influence small mammal activity which consequently affects detection. Small mammals are generally less active on 1) bright nights (Lockard and Owings 1974, Bengsen et al. 2010, Brown et al. 1988, Orrock et al. 2004, Fanson 2010, see Barnet and Dutton 1995), 2) cold nights (Getz 1968, Vickery and Bider 1978), 3) heavy rain nights (Mystkowska and Sidorowicz 1961) and 4) nights with increased risk of predation (Orrock et al. 2003). Therefore, we recorded the percent illumination, cloud cover (clear sky to rain), air and soil temperature, and percent trap disturbance during trapping nights. The percent of illumination during trapping nights was recorded with the use of NASA’s Moonphase 3.3 (Tingstrom 2009). Cloudiness was scored as clear (0% cloudiness) to rainy (100% cloudiness) nightly by physical observation. The soil temperatures were provided by Harvard Forest Archive (hf108-04). Please refer to the “Air and Soil Temperate in Hemlock Removal Experiment at Harvard Forest since 2004” metafile for detailed descriptions on collection process. The presence of trap predators
(mostly black bears, raccoons, and gray squirrels) was calculated as a fraction of the number of disturbed traps in each plot treatment per trapping night and the number of traps set (Table 1).

The landscape effects of elevation (ridge and valley blocks) and habitat treatment (Hemlock, Girdled, Logged, and Hardwood) were included in all models as occupancy covariates (Table 1). “Surrounding habitat” was identified as the type of forest that neighbored the habitat treatments; i.e. Hemlock, Girdled, and Logged treatments were surrounded by hemlock forests and Hardwood treatment was surrounded by hardwood forests. Because landscape characteristics play an important role in colonization and extinction, we assumed that habitat treatment (type) and surrounding habitat type represented a type of landscape habitat quality or suitability for the southern red-backed vole (Table 1).

Site structure is represented by the percent of wood and leaf litter cover at each site. From August 9th to 31st 2013, photographs of the ground surrounding each trap location were taken to quantify site structure variables. Approximately 1-m² quadrates were placed over each trap location and then photographed. The camera was placed at the same location as the trap ground photographs were taken approximately 1-m from the ground to capture the entire 1-m² quadrat. Each photograph (n=392) was analyzed using ImageJ (1.42q Java 1.6.0_version 10) by randomly generating fifty overlaid points (Appendix A) on each digital photograph and scoring the dots that landed on wood or
leaf litter. Each site has a percent woody debris cover and a percent leaf litter cover to represent site structure at each trap location.

Food availability contributes to colonization and extinction patterns of southern red-backed voles. Seed and vegetation are food resources utilized by these voles (Table 1). Yearly seed fall (g/m²) from 2011-2013 was calculated using the Harvard Forest Archive (hf105-05). Please refer to the “Seed Bank in Hemlock Removal Experiment at Harvard Forest 2001-2010” (Ellison et al. 2005b) metafile for detailed descriptions on collection process. Average seed mass from one year previous to the trapping year was calculated from hf105-05. Vegetation was calculated in the same manner as woody debris and leaf litter.

Winter conditions, such as snow cover and onset of snow melt, correlate with population cycling of small mammals where decrease in snow cover decreases over winter survival (e.g. Hansson 1984, Bierman et al. 2006, Hoset et al. 2009). Unfortunately, we were unable to record snow depth and snow melt during the course of the study. Instead, we averaged and calculated the variation of overwinter soil temperatures from the first day of winter to the last day of winter each year (Table 1) using the Harvard Forest Archive (hf108-04). Therefore, the winter covariate is both the average and the range of overwinter soil temperature.
Models

We used a multi-season occupancy modeling framework (MacKenzie et al. 2013) from the package “unmarked” in R (Fiske and Chandler 2011) to examine the influence of landscape characteristics (habitat type and surrounding habitat), site structure (woody debris and leaf litter), food resources (seed and vegetation), and winter conditions (average and range of soil temperatures) on the probability of colonization and extinction on southern red-backed voles while taking into account factors that influence the probability of detection (illumination, air and soil temperature, cloud cover, and trap predators). The NULL model assumes that the effects of all covariates on occupancy, colonization, extinction, and detection do not differ. The NULL model was used in both model sets (Table 1).

Models were ranked by their Akaike Information Criterion (AIC) scores and weighted AIC as the probability of being the best model in the entire model set (Burnham and Anderson 2010). We used a goodness-of-fit test on the most parameterized model in each set to examine how well the model fit the observed data (Hilborn and Mangel 1997, MacKenzie and Bailey 2004), or the model that explained species detection histories given site and covariates, using a parametric bootstrap procedure with 1000 simulations. The Pearson’s (1900) Chi² (\(\chi^2\)) of the observed data and \(\chi^2\) bootstrap were recorded (MacKenzie and Bailey 2004). The null hypothesis for the goodness-of-fit test stated that the observed results from the multi-season occupancy model were not random. If the null hypothesis cannot be rejected, then the estimated parameters from the multi-season model
were not random and therefore a good model fit. If the results from the observed multi-
season model was greater than 5% and less than 95% of the \( \chi^2 \) results from the 
bootstrapped distribution, the null hypothesis could not be rejected and a good model fit
was assumed.

\textit{R Packages}

All data were analyzed using RStudio version 3.2.3 (R Core Team 2013). The 
package “unmarked” version 0.11-0 (Fiske and Chandler 2011 maintained by Royle 
USGS) and packages within were used to fit hierarchical model of animal occurrence. 
The function “coext” was used for multi-season occupancy models. The package 
“reshape” version 0.8.5 (Wickham 2015) was used to restructure and aggregate data. The 
package “plyr” version 1.8.3 (Wickham 2016) was used for data frame manipulation. 
Packages “lattice” version 0.20-33 (Sarkar 2015) and “ggplot2” version 2.0.0 (Wickham 
and Chang 2015) were used for graphics.

3.4 Results

\textit{Goodness of Fit: Southern Red-backed Voles}

The model that was used to test goodness of fit was FOOD model \( \psi(\text{elevation +} 
\text{habitat}) \gamma(\text{food}) \varepsilon(.) p(\text{nightly soil temperature}) \). The \( \chi^2 \) of the observed data was 9175 and 
probability of this value occurring in the bootstrapped models was 75%, which was in the 
25% quantiles of the bootstrap models. There is no evidence to suggest that the observed
data do not fit the multi-season model and we conclude that the observed data fit the model.

**Model Selection**

There was a significant difference in the probability of site occupancy among Hemlock, Girdled, and Logged (upper and lower 95% CI did not overlap and the difference did not contain zero, Knezevic 2008) within blocks (ridge and valley). There was no significant difference in site occupancy difference between Logged and Hardwood treatments (difference in upper and lower 95% CI did include 0 value). FDetection as a function of nightly soil temperature was the top ranking AIC model in the DETECTION model set (Table 2). The nightly soil temperature covariate was used in all following models in the detection parameter. Colonization and extinction probabilities as a function of food and colonization as a function of habitat type was the top ranking AIC model (Table 3). The expected probability of detection of southern red-backed voles increased as the soil temperature increased (Figure 2). The predicted probability of occupancy for southern red-backed voles varied among elevation (ridge and valley) and habitat treatments (Figure 3). There was a higher probability of occupancy at lower elevation in the valley than higher elevation on the ridge among all treatments (Figure 2). The highest probability of occupancy was in the Girdled treatment ($\psi = 0.89$, SE = 0.06, lower 95% CI =0.69, upper 95% CI =0.96) followed by the Logged treatment ($\psi = 0.63$, SE = 0.10, lower 95% CI =0.10, upper 95% CI =0.41) and the Hardwood control ($\psi = 0.59$, SE = 0.09, lower 95% CI =0.41, upper 95% CI =0.76). The lowest probability of
site occupancy for southern red-backed voles was in the Hemlock control ($\psi = 0.24$, SE = 0.07, lower 95% CI =0.12, upper 95% CI =0.41) (Figure 3). Food for voles is a function of seed mass and percent vegetation ground cover, but probability of colonization decreased as seed rain mass increased (Figure 4) and the probability of colonization increased as vegetation cover increased (Figure 5). Although the food covariate (seed + vegetation) was the top ranking model while in the colonization and extinction parameters (Table 3), food resources (seed mass and vegetation) had no noticeable influence on the probability of extinction (Figures 6 & 7 respectively), therefore, the probability of extinction was equal among all sites. Probability of extinction was 97.8% among all sites (SE = 0.01, lower 95% CI =0.75, upper 95% CI =1.22, Figures 6 & 7).

### 3.5 Conclusion

In this study, we hypothesized that patterns of local colonization and extinction of southern red-backed voles would be a function of environmental characteristics and would be influenced by ramifying effects of disturbance in eastern hemlock forests. We developed and tested 38 multi-season models that reflected our hypotheses to identify characteristics that explain account history patterns from 2012-2014 (Table 1). We found that 1) lower elevation sites had a higher probability of site occupancy by southern red-backed voles than higher elevation sites (Figure 1), 2) there was an effect of treatment on site occupancy where there was a higher probability of site occupancy in girdled and logged treatments and in non-disturbed hemlock and hardwood controls, 3) food resources (seed and vegetation) were a function of the probability of site colonization.
(Figure 3), 4) no covariate we analyzed influenced the probability of extinction and extinction was a constant for all sites, and 5) the average increase of soil temperature during trapping events increased the detection probability (Figure 1) most likely explained presence/not detection patterns observed by southern red-backed voles.

Girdling and logging of trees resulted in dramatic changes in eastern hemlock forest (Ellison et al. 2010). The shift in habitat characteristics (habitat structure, food, microclimate, etc…) likely increased the probability that southern red-backed voles would occupy these distributed forests than non-disturbed hemlock forests (Figure 2). The disturbed treatments have characteristics that voles are generally more associated with such as low growth shrub vegetation (Vanderwel et al. 2010), decaying woody debris (Fauteux et al. 2013) and leaf litter (Vanderwel et al. 2010) (Figure 1), but these site structures did not influence the colonization of southern red-backed voles. Interestingly, Vanderwel et al. (2010) also found a strong association between southern red-backed voles and coniferous understory. Although understory vegetation was not identified to species in this study, Farnsworth et al. (2012) found the understory growth in the disturbed treatments had very little coniferous growth.

Food (seed mass and vegetation) did influence colonization (Figures 3 & 4). Mast seeding events act as a resource pulse for granivores (and food web linked organisms) and this pulse affect community dynamics (Ostfeld and Kessing 2002). Our results for southern red-backed voles suggest that the probability of colonization
decreases with an increase in seed rain (g) (Figure 3). We think that the factors which contribute to colonization in this system are complex because we are experimenting in coniferous, deciduous, and disturbed forests which are partly coniferous and deciduous. Rodent responses to pulses in food resources is more complex in coniferous forests than in deciduous forests (Lobo and Millar 2013). In addition, conifer seeds are not a major component of the southern red-backed vole’s diet (Martell 1981) and are insufficient food resources for these voles (Lobo and Millar 2011). New vegetation and berries make up the southern red-backed vole’s seasonal diet, but they primarily consume fungi and lichen (Martell 1981). We attempted to quantify the above ground fungi percent cover, but there was none to be found during the time we collected these data.

We were surprised that none of the covariates we examined could explain the probability of site extinction (Figures 6 & 7). Neither landscape, habitat treatment, site structure, food, nor overwinter temperatures seemed to influence the local extinction for southern red-backed voles (Table 3). While our $\gamma$(food) $\epsilon$(food) model was ranked the highest in our model set, the influence of food (seed + vegetation) on extinction was insignificant as there was a 100% of the probability of extinction regardless of food resources, habitat structure, etc…. As mentioned, many factors, both biological and environmental, contribute to patterns in distribution, colonization, and extinction and we evaluated combinations of environmental factors only. It is very possible that extinction in this system is influenced by other biological factors, such as density dependence. Density dependent dispersal can be positively or negatively correlated. Positively
correlated density dispersal predicts that an individual that leaves a high density area or poor condition habitat will have an increased fitness over an individual that does not disperse (Waser 1985, Porter and Dooley 1993). Negatively correlated density dispersal typically predicts that individuals leaving high density areas will be faced with increased aggression and competition from locals in the newly migrated areas which suggests that dispersal is effectiveness reduced because migrators will have a decreased fitness (e.g. Hestbeck 1982). Therefore, dispersal is based on social interactions (e.g. Boyce and Boyce 1988), population density (Lidicker 1975), and relatedness among individuals (Charnov and Finerty 1980, Lambin and Krebs 1991) and it seems likely that the covariates we chose to study did not reflect the behaviors in population because we focused on the landscape and site, not population dynamics. It is interesting that even though there was 100% extinction from 2012-2013, varying food resources influenced the recolonization of these sites from 2013-2014.

Occupancy models are variable, can be difficult to fit, and difficult to interpret when abundance varies over sites (Welsh et al. 2013). However, they provide useful insight in identifying factors that influence animal distribution (Stanley and Royle 2005), especially when there is heterogeneity in the probability of species detection (MacKenzie et al. 2006, Gibson 2011). We investigated the link between detection/non-detection patterns and environmental characteristics to simplify the diverse underlying mechanisms that drive these patterns of colonization and extinction for southern red-backed voles in disturbed New England forests. Elevation, disturbance (girdled and logged treatments),
and food resources played major roles in site occupancy and colonization. Our results suggest that southern red-backed voles will continue to occupy and colonize areas in New England as hemlock forests decline from hemlock woolly adelgid infestation. We think this model approach can be used for estimating patterns in occupancy and colonization over time for southern red-backed voles, but patterns of extinction may be better explained by biological factors (e.g. social behavior) that are not associated with state occupancy models.

3.6 Acknowledgements

We would like to thank the hardworking NSF-REU undergraduate researchers that contributed to the data collection of this work Elizabeth Kennett, James Leitner, Amy Balint, Joel van de Sande, and Ariel de Cruz Reis, in addition to Emma Cornin (UVM) and Jefferson Franca de Jesus (Saint Michael’s College). We would like to acknowledge our funding support for this project: NSF-GRFP 2013-2016, American Society of Mammologists Grant-In-Aid 2013, and the Northeastern States Research Cooperative (NSRC) Graduate Researcher Grant, a partnership of Northern Forest states (New Hampshire, Vermont, Maine, and New York), in coordination with the USDA Forest Service.
Table 3.1. Description of covariates considered for all models, a brief description of covariate used in each parameter, and formulas used in model sets (n=38).

<table>
<thead>
<tr>
<th>Covariate Considered</th>
<th>Parameter Description</th>
<th>Formulas Included in Model Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>NULL</td>
<td>Occupancy, colonization, extinction, and detection probabilities are not influenced by any site or observation covariates and are equal.</td>
<td>$\psi$ (.). $\gamma$ (.). $\epsilon$ (.). $p$(.) $\psi$ (parametrized). $\gamma$ (.). $\epsilon$ (.). $p$(parametrized)</td>
</tr>
<tr>
<td>DETECTION</td>
<td>Detection is a function of the percent moon illumination, percent of cloud cover, average air temperature, average soil temperature, and percent trap predation during trapping nights (n=6). Detection is a standalone exploratory model set. The covariate from the top ranking AIC model was used in all other models.</td>
<td>$\psi$ (parameterized). $\gamma$ (parameterized). $\epsilon$ (parameterized)</td>
</tr>
<tr>
<td>LANDSCAPE</td>
<td>Occupancy, colonization, and extinction are a function of the replicated blocks with two elevation differences (Ridge and Valley) the habitat treatment (Hemlock, Girdled, Logged, and Hardwood) and the surrounding habitat type (Hemlock or Hardwood) (n=7). “Landscape” comprises of “habitat + surrounding.”</td>
<td>$\psi$ (~elevation + habitat). $\gamma$ (~landscape), (~habitat), (~surrounding), (.). $\epsilon$ (~landscape), (~habitat), (~surrounding), (.). $p$ (~best from DETECTION)</td>
</tr>
<tr>
<td>FOOD RESOURCES</td>
<td>Colonization and extinction are a function of food seed mass (g) and percent ground cover of vegetation (n=3).</td>
<td>$\psi$ (~elevation + habitat). $\gamma$ (~food), (.). $\epsilon$ (~food), (.). $p$ (~best from DETECTION)</td>
</tr>
<tr>
<td>SITE STRUCTURE</td>
<td>Colonization and extinction are a function of the percent woody debris and leaf litter ground cover (n=3).</td>
<td>$\psi$ (~elevation + habitat). $\gamma$ (~structure), (.). $\epsilon$ (~structure), (.). $p$ (~best from DETECTION)</td>
</tr>
<tr>
<td>WINTER TEMPERATURE</td>
<td>Colonization and extinction are a function of the overwinter average soil temperature and the range of soil temperature (n=3).</td>
<td>$\psi$ (~elevation + habitat). $\gamma$ (~winter), (.). $\epsilon$ (~winter), (.). $p$ (~best from DETECTION)</td>
</tr>
<tr>
<td>COMBINATIONS</td>
<td>Colonization and extinction are a function of multiple combinations of landscape, food, site structure, and winter temperature (n=23). Habitat was not used in the same parameter as structure or food because the relationship between habitat treatment.</td>
<td>Example $\psi$ (~elevation + habitat). $\gamma$ (~structure + food). $\epsilon$ (~winter). $p$ (~best from DETECTION)</td>
</tr>
</tbody>
</table>
Table 3.2. Exploratory model set for detection parameter. The $\psi$ (Elevation + Habitat) $\gamma$ (Landscape + Food) $\varepsilon$ (Winter) $p(.)$ varies with model, number of parameters, AIC, $\Delta$AIC, AIC weight, and negative log-likelihood.

<table>
<thead>
<tr>
<th>Formula Structure</th>
<th>No. Parameters</th>
<th>AIC</th>
<th>$\Delta$AIC</th>
<th>AIC weight</th>
<th>-LogLike</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p (\sim \text{night soil temperature})$</td>
<td>14</td>
<td>2764.46</td>
<td>0.00</td>
<td>8.7e-01</td>
<td>1368.23</td>
</tr>
<tr>
<td>$p (\sim \text{night air temperature})$</td>
<td>14</td>
<td>2768.44</td>
<td>3.97</td>
<td>1.2e-01</td>
<td>1370.22</td>
</tr>
<tr>
<td>$p (\sim \text{trap predation})$</td>
<td>14</td>
<td>2797.28</td>
<td>32.82</td>
<td>6.5e-08</td>
<td>1384.64</td>
</tr>
<tr>
<td>$p (\sim \text{illumination})$</td>
<td>14</td>
<td>2804.33</td>
<td>39.87</td>
<td>1.9e-09</td>
<td>1388.16</td>
</tr>
<tr>
<td>$p (\sim)$</td>
<td>13</td>
<td>2804.92</td>
<td>40.46</td>
<td>1.4e-09</td>
<td>1389.46</td>
</tr>
<tr>
<td>$p (\sim \text{cloud cover})$</td>
<td>14</td>
<td>2806.46</td>
<td>42.00</td>
<td>6.6e-10</td>
<td>1389.23</td>
</tr>
<tr>
<td>$\psi (\sim) \gamma (\sim) \varepsilon (\sim) p(.)$</td>
<td>4</td>
<td>2878.26</td>
<td>113.80</td>
<td>1.7e-25</td>
<td>1435.13</td>
</tr>
</tbody>
</table>
Table 3.3. Model ranking based. The $\psi$ (Elevation + Habitat) $p$ (night soil temperature) are held constant, but $\gamma$ and $\varepsilon$ formula structure vary. The number of parameters, AIC, $\Delta$AIC, AIC weight, and negative log-likelihood from explored covariates (Table 1). *Models could not converge on a solution and were not included in the model set.

<table>
<thead>
<tr>
<th>Formula Structure</th>
<th>No. Parameters</th>
<th>AIC</th>
<th>$\Delta$AIC</th>
<th>AIC weight</th>
<th>-LogLike</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$ (Food) $\varepsilon$ (Food)</td>
<td>13</td>
<td>2767.89</td>
<td>0.00</td>
<td>2.94E-01</td>
<td>1370.94</td>
</tr>
<tr>
<td>$\gamma$ (Food) $\varepsilon$ (Surrounding)</td>
<td>12</td>
<td>2767.99</td>
<td>0.09</td>
<td>2.86E-01</td>
<td>1371.99</td>
</tr>
<tr>
<td>$\gamma$ (Surrounding + Food) $\varepsilon$ (Habitat)</td>
<td>15</td>
<td>2768.05</td>
<td>0.15</td>
<td>2.73E-01</td>
<td>1369.02</td>
</tr>
<tr>
<td>$\gamma$ (Surrounding + Food) $\varepsilon$ (.)</td>
<td>12</td>
<td>2770.77</td>
<td>2.87</td>
<td>6.99E-02</td>
<td>1373.38</td>
</tr>
<tr>
<td>$\gamma$ (Food) $\varepsilon$ (Habitat)</td>
<td>14</td>
<td>2772.00</td>
<td>4.10</td>
<td>3.78E-02</td>
<td>1372.00</td>
</tr>
<tr>
<td>$\gamma$ (Food) $\varepsilon$ (Structure)</td>
<td>13</td>
<td>2773.20</td>
<td>5.30</td>
<td>2.08E-02</td>
<td>1373.60</td>
</tr>
<tr>
<td>$\gamma$ (Food) $\varepsilon$ (Landscape)</td>
<td>15</td>
<td>2773.99</td>
<td>6.09</td>
<td>1.40E-02</td>
<td>1371.99</td>
</tr>
<tr>
<td>$\gamma$ (Food) $\varepsilon$ (.)</td>
<td>11</td>
<td>2774.73</td>
<td>6.83</td>
<td>9.64E-03</td>
<td>1376.36</td>
</tr>
<tr>
<td>$\gamma$ (Surrounding + Food) $\varepsilon$ (Winter)</td>
<td>14</td>
<td>2793.86</td>
<td>25.97</td>
<td>6.75E-07</td>
<td>1382.93</td>
</tr>
<tr>
<td>$\gamma$ (Food) $\varepsilon$ (Winter)</td>
<td>13</td>
<td>2793.93</td>
<td>26.03</td>
<td>6.52E-07</td>
<td>1383.96</td>
</tr>
<tr>
<td>$\gamma$ (Winter) $\varepsilon$ (.)</td>
<td>11</td>
<td>2820.44</td>
<td>52.55</td>
<td>1.14E-12</td>
<td>1399.22</td>
</tr>
<tr>
<td>$\gamma$ (Habitat) $\varepsilon$ (Surrounding)</td>
<td>13</td>
<td>2823.50</td>
<td>55.60</td>
<td>2.48E-13</td>
<td>1398.75</td>
</tr>
<tr>
<td>$\gamma$ (Landscape) $\varepsilon$ (.)</td>
<td>13</td>
<td>2823.50</td>
<td>55.60</td>
<td>2.47E-13</td>
<td>1398.75</td>
</tr>
<tr>
<td>$\gamma$ (Habitat) $\varepsilon$ (Winter)</td>
<td>14</td>
<td>2824.00</td>
<td>56.10</td>
<td>1.93E-13</td>
<td>1398.00</td>
</tr>
<tr>
<td>$\gamma$ (Winter) $\varepsilon$ (Winter)</td>
<td>13</td>
<td>2824.43</td>
<td>56.53</td>
<td>1.55E-13</td>
<td>1399.21</td>
</tr>
<tr>
<td>$\gamma$ (Winter) $\varepsilon$ (Food)</td>
<td>13</td>
<td>2824.46</td>
<td>56.565</td>
<td>1.53E-13</td>
<td>1399.23</td>
</tr>
<tr>
<td>$\gamma$ (Landscape) $\varepsilon$ (Structure)</td>
<td>15</td>
<td>2826.07</td>
<td>58.17</td>
<td>6.86E-14</td>
<td>1398.03</td>
</tr>
<tr>
<td>$\gamma$ (Habitat) $\varepsilon$ (Habitat)</td>
<td>15</td>
<td>2827.50</td>
<td>59.6</td>
<td>3.34E-14</td>
<td>1398.75</td>
</tr>
<tr>
<td>$\gamma$ (Habitat) $\varepsilon$ (Surrounding + Food)</td>
<td>15</td>
<td>2827.51</td>
<td>59.61</td>
<td>3.34E-14</td>
<td>1398.75</td>
</tr>
<tr>
<td>$\gamma$ (Landscape) $\varepsilon$ (Food)</td>
<td>15</td>
<td>2827.51</td>
<td>59.62</td>
<td>3.33E-14</td>
<td>1398.75</td>
</tr>
<tr>
<td>$\gamma$ (Winter) $\varepsilon$ (Landscape)</td>
<td>15</td>
<td>2828.44</td>
<td>60.54</td>
<td>2.09E-14</td>
<td>1399.22</td>
</tr>
<tr>
<td>$\gamma$ (Landscape) $\varepsilon$ (Landscape)</td>
<td>17</td>
<td>2830.07</td>
<td>62.17</td>
<td>9.27E-15</td>
<td>1398.03</td>
</tr>
<tr>
<td>$\gamma$ (Structure) $\varepsilon$ (.)</td>
<td>11</td>
<td>2836.41</td>
<td>68.51</td>
<td>3.90E-16</td>
<td>1407.20</td>
</tr>
<tr>
<td>$\gamma$ (.) $\varepsilon$ (Winter)</td>
<td>11</td>
<td>2839.99</td>
<td>72.10</td>
<td>6.49E-17</td>
<td>1408.99</td>
</tr>
<tr>
<td>$\gamma$ (Structure) $\varepsilon$ (Structure)</td>
<td>13</td>
<td>2840.41</td>
<td>72.52</td>
<td>5.26E-17</td>
<td>1407.20</td>
</tr>
<tr>
<td>$\gamma$ (Structure) $\varepsilon$ (Landscape)</td>
<td>15</td>
<td>2844.41</td>
<td>76.51</td>
<td>7.14E-18</td>
<td>1407.20</td>
</tr>
<tr>
<td>$\gamma$ (Structure) $\varepsilon$ (Food)</td>
<td>13</td>
<td>2847.17</td>
<td>79.27</td>
<td>1.79E-18</td>
<td>1410.58</td>
</tr>
<tr>
<td>$\gamma$ (.) $\varepsilon$ (Food)</td>
<td>11</td>
<td>2849.28</td>
<td>81.38</td>
<td>6.25E-19</td>
<td>1413.64</td>
</tr>
<tr>
<td>$\gamma$ (Surrounding) $\varepsilon$ (Habitat)</td>
<td>13</td>
<td>2850.46</td>
<td>82.56</td>
<td>3.46E-19</td>
<td>1412.23</td>
</tr>
<tr>
<td>$\gamma$ (.) $\varepsilon$ (Surrounding + Food)</td>
<td>12</td>
<td>2851.26</td>
<td>83.36</td>
<td>2.32E-19</td>
<td>1413.63</td>
</tr>
<tr>
<td>$\gamma$ (.) $\varepsilon$ (Surrounding + Habitat)</td>
<td>13</td>
<td>2853.29</td>
<td>85.39</td>
<td>8.43E-20</td>
<td>1413.64</td>
</tr>
</tbody>
</table>

* $\gamma$ (.) $\varepsilon$ (.) $p$ (.)

4 | 2878.26 | 110.37 | 3.18E-25 | 1435.13 |
Figure 3.1. Simes Tract containing the Long-Term Ecological Research (LTER) Hemlock Removal Experiment (HF-HeRE) located in north-central Petersham, Massachusetts, USA (panel A). The replicated blocks, Ridge (yellow) and Valley (green), contain four ~90 x 90-m (0.81-ha) canopy treatments plots, Hemlock (He), Girdled (G), Logged (L), and Hardwood (Hw) (panel B). Sampling grids (panel C) spanned 70m². Trap (circles) were placed 10-m apart by pacing in a 7 × 7 array within each of the two Hemlock, Hardwood, Girdled, andLogged plots (n=392). Figure modified from Ellison et al. 2005.
Figure 3.2. Probability of detection (95% CI) of southern red-backed voles as a function of the nightly average soil temperature (°C).
Figure 3.3. Probability of site occupancy (95% CI) as a function of block (Ridge and Valley) and habitat treatment (Hemlock, Girdled, Logged, Hardwood) of southern red-backed voles.
Figure 3.4. Probability of site colonization (95% CI) of southern red-backed voles as a function of food (the average seed fall mass (g/m²)).
Figure 3.5. Probability of site colonization (95% CI) of southern red-backed voles as a function of food (vegetation cover %).
Figure 3.6. Probability of site extinction (95% CI) of southern red-backed voles as a function of food (the average seed fall mass (g/m²)).
Figure 3.7. Probability of site extinction (95% CI) of southern red-backed voles as a function of food (percent vegetation ground cover).
CHAPTER 4:
THE LOSS OF FOUNDATION SPECIES REVISITED

Allyson L. Degrassi¹*, Steven Brantley², Carrie R. Levine², Robert J. Miller⁴,
Jacqueline Mohan⁵, Sydne Record⁶, and Aaron M. Ellison⁷

Author’s affiliation addresses:
¹ Department of Biology, University of Vermont, 120 Marsh Life Science, Burlington VT 05468 USA
² Joseph W. Jones Ecological Research Center, 3988 Jones Center Drive, Newton, GA 39870 USA
³ Department of Environmental Science, Policy, and Management, University of California, Berkeley, 130 Mulford Hall, Berkeley, CA 94720 USA
⁴ Marine Science Institute, University of California Santa Barbara, Santa Barbara CA 93109-6150 USA
⁵ Odum School of Ecology, University of Georgia Athens, GA 30601 USA
⁶ Department of Biology, Bryn Mawr College, Park 209 Bryn Mawr, PA 19010 USA
⁷ Harvard University, Harvard Forest, 324 North Main Street, Petersham, MA 01366 USA

Corresponding author’s e-mail address:
* E-mail: adegrass@uvm.edu

84
4.1 Abstract

Ecologists and environmental scientists often prioritize research efforts with conservation importance. Dominant, widespread, or locally abundant species at low risk of extinction receive relatively little attention unless they are invasive. Native foundation species create habitats and environmental conditions that support many associated species and modulate local-scale ecosystem processes, but the generally high local or regional abundance of foundation species may lead to less research about them. We used citation analysis (2005-2014) to examine research following from a suggestion to identify and study foundation species while they were still common and not threatened. We explored the use and expanding definition of the foundation species concept, as well as the trajectory and ecological focus of research on foundation species throughout the world in 378 papers published in this nine-year span. Contemporary authors who cite key papers defining a foundation species pay little attention to its actual definition and species studied in this context rarely were identified as foundation species. Although functions and roles of foundation species, such as creating unique microclimates or supporting dependent species, are being studied, less research is focused on identifying them before they are threatened or lost from the ecosystem that they otherwise define. Invasive species were identified as the most common threat to foundation species. Our citation analysis and synthesis provides a new conceptual framework linking identification of and research about foundation species with their functional roles and our ability to manage emerging threats to them.
4.2 Introduction

Ecologists and environmental scientist often rank species in order of conservation importance (Mace et al. 2007) and target for research or monitoring those species that are rare (Courchamp et al. 2006, Angulo and Courchamp 2009), endangered (Caro and Sherman 2010), or occupy habitat biodiversity hotspots (Myers et al. 2000). Not surprisingly, species that are dominant, abundant, or are not in immediate danger of population loss are studied less frequently by conservation biologists (Gaston and Fuller 2007) than endangered or threatened species.

The assumption that abundant species are not a priority for conservation is unwarranted: common species often are ecologically important as structural, dominant, or foundation species, and commonness itself is rare (Gaston and Fuller 2007). As a consequence, abundant species – and especially foundation species – may receive little attention from conservation biologists until their populations are threatened and a compelling need arises to understand their life history, their roles and functions in their ecosystem, and the factors that threaten these roles. Yet, understanding how foundation species interact with their environment and other species could allow for much better forecasts of the cascading consequences of population declines and enable early adoption of strategies to ameliorate those consequences.

Dayton (1972), working in a benthic marine system, described a foundation
species as, “a single species that defines much of the structure of a community by creating locally stable conditions for other species and by modulating and stabilizing fundamental ecosystem processes.” In applying the foundation species concept to terrestrial ecosystems, Ellison et al. (2005) identified foundation species as (usually) primary producers that occupy low trophic levels, are locally abundant, regionally common, and create stable habitat conditions that are necessary for survival of dependent species (see Baiser et al. 2013). The loss of foundation species can impact energy and nutrient fluxes (Jenkins et al. 1999), microclimate conditions (Snyder et al. 2002), food webs (Baiser et al. 2013), and biodiversity (Tingley et al. 2002; Ellison et al. 2005; Sackett et al. 2011). Foundation species thus fundamentally shape both community structure and ecosystem function.

Other categories, such as ecosystem engineer (Jones et al. 2010), core species (Hanski 1982), dominant species (Grime 1984), and structural species (Huston 1994) describe particular aspects of foundation species (Ellison et al. 2005). However, foundation species are distinct from these other types of species because they also have unique sets of traits that are functionally irreplaceable in a given ecosystem and that, coupled with a foundation species’ system-wide dominance and high abundance, define that ecosystem (Figure. 1). However, foundation species appear to be studied less than either rare species or other types of “important” species: a title-only search in Web of Science (run on 1 July 2015 for papers published between 1972 and 2014) recovered “foundation species” in only 54 papers, compared to 473 for “rare species”, 202 for
Foundation species are not often monitored and, as with other common species, any population changes likely go unnoticed until there is a sudden or dramatic decline in their abundance or range (Gaston and Fuller 2007). For example, whitebark pine (*Pinus albicaulis* Engelm.), a foundation species in many western North American high-elevation forests, is threatened by the introduced fungal pathogen *Cronartium ribicola* (J. C. Fisch.). If this foundation tree species had been better understood when it was abundant, preventing its loss or mitigating the negative effects of reductions in its population may have been possible.

In 2005, Ellison et al. published an article in Frontiers of Ecology and the Environment emphasizing the importance of identifying foundation species before their populations become threatened. Ellison et al. (2005) argued that as of the early 21st century, understanding the consequences of foundation species loss was based on only a small number of case studies and these case studies were conducted after the species had declined. The lack of data on how foundation species, while still abundant and widespread, structured and supported ecological systems led to an incomplete understanding of their overall role in these systems. Thus, Ellison et al. (2005) also called on scientists to fill knowledge gaps on how foundation species respond to environmental changes and biotic threats. Since its publication, Ellison et al. (2005) has been cited nearly 500 times in primary articles, review articles, and book chapters; here, we ask whether these citations actually reflect increasing identification or study of
foundation species.

We reviewed papers published through the end of 2014 that cited Ellison et al. (2005) and assessed whether these studies 1) adequately or accurately defined foundation species; 2) identified a particular foundation species; 3) identified an ecological role associated with foundation species; and 4) identified a threat to foundation species populations. We synthesized our results to develop a framework for studying foundation species that emphasizes how identifying and studying them can improve both our understanding of the roles of these species and our ability to manage effectively emerging threats to them.

4.3 Methods

Data collection was restricted to a citation analysis of Ellison et al. (2005) because that review not only introduced the concept to terrestrial ecology, but also specifically addressed the importance of studying foundation species and encouraged research on them. We recognize that many other studies of foundation species have highlighted their importance, but because Ellison et al. (2005) is the most highly cited paper about foundation species and emphasized an agenda for future research, we were interested in whether it has acted as a catalyst for increasing research on foundation species.

Using several research platforms and article databases (Web of Science, JSTOR, Google Scholar, Pub Med), we found that Ellison et al. (2005) was cited in at least 446
papers through December 2014 (number of citations varied among the databases). We reviewed 378 of these papers to determine the main focus of the original research described and its relationship to the key questions proposed by Ellison et al. (2005). Review papers, book chapters, commentaries, and all other non-primary literature were excluded from the present study (n=47).

Questions for Data Collection

We developed a set of six of questions to assess research on foundation species published since 2005 and used that information to compare cohesiveness between individual studies and the goals of Ellison et al.’s. The raw data are available from the Harvard Forest Data Archive, file HF-259 (http://harvardforest.fas.harvard.edu/data-archive).

Question 1: Was a foundation species precisely or accurately defined and what definition was used?

It is important to know whether other studies recognized or differentiated foundation species from other similar, but distinct, species roles. The definition of a foundation species definition found in each paper was placed into one of five categories: 1) Ellison et al.’s (2005) definition; 2) Dayton’s (1972) definition; 3) Dayton and Ellison’s definitions combined; 4) neither Dayton or Ellison’s definition (i.e., “Other”); or 5) not defined. If categorized as “Other”, then the alternative definition was recorded. If other definitions included multiple terms, each term was counted, so that a definition
could be classified with multiple terms.

*Question 2: Was a foundation species explicitly studied?*

We recorded as a single binary variable (yes/no) indicating whether or not any single focal study species in the study was explicitly considered a “foundation species.”

*Question 3: What was the main role of the foundation species that were studied?*

Two broad roles of foundation species were distinguished: *direct support of other species* (*e.g.*, effects on associated species or assemblages); and *modulation and stabilization of fundamental ecosystem processes* (*e.g.*, effects on abiotic or biogeochemical processes). We classified each paper as focusing on support for associated species (“Community”), modulation/stabilization (“Ecosystem”), both, or neither.

*Question 4: Were threats to foundation species identified?*

We identified six broad classes of threats to foundation species: “climate change” (*e.g.*, changes in atmospheric composition, temperature, or hydrological flow); “invasive species” (*i.e.*, nonnative or invasive species); “habitat degradation” (*e.g.*, pollution, habitat loss, human disturbance,); “exploitation” (*i.e.*, over-use by humans or increased herbivory or predation by non-human species); “disease or pathogen” (*e.g.*, fungal, bacterial, and viral causes) ; or “no threat”. Note that studies could be classified into more than one of the threat categories.
**Question 5: Where were experiments on foundation species done?**

We counted the number of studies on foundation species done in each country. We recognize that these data were biased toward journals printed in English and that national or regional resources will influence where foundation species are studied. However, as a first pass of the citation record, identifying geographic location of the studies allowed us to identify regions where the study of foundation species is focused.

**Question 6: To what extent did Ellison et al. (2005) influence research on foundation species?**

We inferred strength of influence from the results of three of the previous questions. Influence was based on 1) whether the definition of foundation species followed Ellison et al. (2005) (question 1); 2) if the foundation species was identified as the main study organism (question 2); and 3) identification of possible threats to foundation species loss (question 4). Studies that contained all three qualities were categorized as “Strongly Influenced.” Studies that contain two qualities in any combination were categorized as “Moderately Influenced.” Studies that contain one quality were categorized as “Marginally Influenced.”

### 3.3. Data Analysis

All data were analyzed using RStudio version 3.0.2 (R Core Team 2013; Appendix D). The packages “maps” (Brownrigg and Minka 2014), “plotrix” (Lemon et al. 2015), and “rworldmap” (South 2013) were used to display geographic locations...
of surveyed studies. The package “plyr” (Wickham 2014) was used for data frame manipulation. Because our sample size was large and no experiment was conducted (Gotelli and Ellison 2013), we coded the answers to our Questions as categorical data and analyzed them using Pearson’s chi-square statistic (Pearson 1900) in the R package “MASS” (Ripley et al. 2013).

### 4.4 Results and Conclusion

Papers citing Ellison et al. (2005) came from 15 countries on 6 continents. Most of the studies were conducted in the United States, while papers on foundation species from mainland Asia were notably absent (Fig. 3). These data suggest that the reach of, interest in, or concern for foundation species applies mainly to the Americas, and that loss of foundation species is not yet a global concern.

Foundation species was not mentioned in every paper and 43% (143) of the studies reviewed did not define the concept (Fig. 2). When it was defined, Ellison et al.’s definition was cited 42% of the time and more frequently than Dayton’s (2%), the combination of Dayton’s and Ellison’s (3%), or other definitions (10%) (Fig. 2). These last 33 papers defined foundation species as something other than the original concept or used multiple defining terms, including: ecosystem engineer (7), keystone (7), a definer, driver, or supporter of forest structure (9), dominant species (8), trees (2), framework species (2), long-lived and widespread (2), or foundation genus (1). Another 16 authors cited for definitions of foundation species include Whitham et al. 2006 (2), Grime 1998
Study organisms were identified as a foundation species in 50% of the reviewed papers that cited Ellison et al. (2005). There was no significant difference in the number of studies that did or did not identify the study organism as a foundation species (Fig. 2). The remaining papers did not specifically identify a study organism as a foundation species (Fig. 2) or only mentioned the concept in passing. These data suggest either that foundation species were not being researched, or that species being studied were not identified as such.

Among studies that did identify foundation species, 34% studied their role in community interactions, 32% studied both community interactions and ecosystem processes, and 22% studied ecosystem processes alone (Fig. 2). The remaining 12% did not identify any specific role of foundation species in the study system (Fig. 2). These data suggest that community ecologists either may be more familiar with or show greater...
Eighty-four percent of the studies identified a threat or potential threat to a foundation species (Fig. 2). The most frequently reported threat to foundation species was invasive species (24%), followed by climate change (18%), disease or pathogens (16%), habitat loss or degradation (16%), and exploitation (10%) (Fig. 2). These data suggest that foundation species are being studied during or after population loss has already begun. We note that the emphasis on threats to foundation species by nonnative species contrasts with threats identified for rare species. In the latter, the vast majority (81%) were reported to be threatened by habitat loss, whereas only 57% were reported to be threatened by invasive species (Wilcove et al. 1998). These data suggest that research on foundation species has not followed the recommendations to study them before they were threatened. We conclude that Ellison's suggestions to increase study of foundation species and leverage the opportunity to study foundation species before decline have been largely ignored for many (though not all) species, and that research on foundation species is still lagging except in cases where species are threatened (e.g., Prevèy et al. 2010, Garneau et al. 2012, Vose et al. 2013).

Finally, there were nearly 1.5 times more papers in the present dataset that were “Marginally Influenced” by Ellison et al. (2005) than were “Strongly Influenced” (Fig. 2), suggesting that Ellison et al. (2005) was being cited for reasons other than supporting research on foundation species. This may not be unexpected as Ellison et al. (2005) used...
several case studies to illustrate the importance of foundation species loss including eastern hemlock (*Tsuga canadensis* (L.) Carr.), whitebark pine and America chestnut (*Castanea dentate* (Marsh.) Borkh.). Focus on content related to these species may have been higher than for the overarching message of the paper. Indeed, many of the citations to Ellison et al. (2005) in the first few years after it was published were focused primarily on the case studies contained in the paper and not specifically on the concept of foundation species. In many such cases, the term "foundation species" was never mentioned in the citing paper and the general concept was not discussed (Fig. 2.1). Alternatively, the citations could have been 1) “ambiguous”, “empty”, or “not supported” (Todd et al. 2007), 2) that Ellison et al. (2005) was mis-cited or misprinted (Simkin and Roychowdhury 2003), or 3) that it was not completely read (Ball 2002, Simkin and Roychowdhury 2003).

It also is possible that, despite a high citation rate, there is little interest in the foundation species concept itself, the concept is not considered useful, or there has been a failure to distinguish “foundation species” from other common species classifications (Fig. 1). The likelihood that the foundation species concept is underrepresented is supported by examination of species excluded from the citation analysis through personal experience of the authors. For example, longleaf pine (*Pinus palustris* Mill.) displays all the critical characteristics of a foundation species and has received considerable attention in the literature (e.g., Van Lear et al. 2005, Kirkman et al. 2013).
Like American chestnut, eastern hemlock, and a number of other species in papers covered by our citation analysis, longleaf pine was a dominant species in its original range, was abundant throughout a wide geographic area, and possessed specific characteristics that supported unique communities and controlled ecosystem processes (Van Lear et al. 2005, Butler et al. 2014). Also like American chestnut, the species has mostly disappeared from its historic range (Van Lear et al. 2005, Butler et al. 2014). Yet, this species is completely absent from the citation analysis because researchers who study it have not classified it as a foundation species.

To further advance foundation species research, we suggest an integrated framework that tracks the research cycle from definition and scoping through conservation and management (Fig. 2.4). We intend this framework to both improve the recognition of foundation species and provide a general workflow for prioritizing research and/or conservation conditional on threats to a particular foundation species. The definition of the foundation species concept is directly related to the correct identification of a foundation species. The correct identification allows researchers to identify the foundation species’ role in the ecosystem, which allows for quantification of the foundation species’ ecosystem services. The interaction between ecosystem services and specific ecosystem roles provides information on how foundation species’ roles in supporting species and stabilizing microclimate may influence ecosystem services at different levels. The ability to identify vulnerabilities to foundation species will allow researchers to identify ecosystem change in response to loss. Conservation management
strategies could be studied before a threat to a particular foundation species becomes a problem. The increase of foundation species research will help to define and continue to stress the importance of foundations species in ecosystem function (Fig 2.4). Because one of the more interesting take home messages from this analysis is that foundation species were not identified as such, we encourage researchers to distinguish “foundation species” from other categories of important species so that their research can find a place in this framework and contribute additional and cumulative knowledge of foundation species research (Fig. 2.1).

We also think this conceptual diagram will be particularly useful for ecosystem and community ecologists studying species for which threats have yet to be identified (Fig. 2.4). Ecosystem science tends to focus on total system fluxes and, by necessity, simplify ecosystems using stand-wide parameters (e.g. leaf area index) regardless of individual species characteristics. In such cases, the system is treated as the subject rather than the species, even when system processes may be highly species-dependent. Examples of the unique role that foundation species can have in undisturbed conditions may identify characteristics that make ecosystems either vulnerable or resilient to change.

Lastly, we believe this framework will help land managers discover commonalities between their species of interest and other foundations species. These commonalities might include threats to ecosystems and/or lessons learned about the effectiveness of specific management techniques applied to a given situation. These could
be particularly useful for conservationists who are looking for case-studies of restoration to use as examples for species that are becoming more vulnerable as disturbances increase. For the land manager interested in restoration, these studies can also provide insight into the possible desired future conditions of other ecosystems being considered for restoration. Thus, to fully account for the influence of foundation species, there is a need to communicate the importance of foundation species to the broader scientific community so that important studies on stable systems or systems that have been successfully restored can be included (e.g. through keywords, etc.) and further our understanding of the role of foundation species in ecosystem structure, function and resilience.

We do not suggest that we have identified all potential foundation species through our citation analysis and call on other scientists, especially ecosystem scientists, to consider whether they are studying a foundation species and identify those species as such. Nor do we mean to suggest that scientists are unaware that they are studying important species; on the contrary, having studied foundation species it seems likely that their importance is valued. We hope that in the future foundation species will be universally recognized as such and identified in the literature whenever appropriate so that we can coordinate efforts to understand and conserve them. Such species, and the systems that depend on them, may serve as valuable models of resistant and resilient ecosystems and the lessons learned can be applied to areas experiencing similar change.
4.5 Acknowledgements

This work is the result of a Foundation Species Working Group (“Foundation Species in North America”) organized at the 2012 Long-Term Ecological Research-All Science Meeting (LTER-ASM) in Estes Park, CO.
**Figure 4.1.** Comparison of characteristics that differentiate commonly used terms that describe common and/or abundant species.
Figure 4.2. Summary of general trends in the study of foundation species based on the results of the Ellison et al. 2005 study (solid lines represent the pool of studies from one analyzed question to the next; dotted lines indicate side information on how results of questions were broken down of each result; filled circles show direction of significance based on Chi$^2$ results; p-values from Chi$^2$).
Figure 4.3. Geographic map of the number of studies (circle size) that identified research organism as foundation species (FS) (green) and the studies that did not identify foundation species (blue).
Figure 2.5. Suggested approach to foundation species research and how topics are connected in the scope of this paper (arrows indicate the direction and relationship to topics).
COMPREHENSIVE BIBLIOGRAPHY


Angulo, E., and F. Courchamp. 2009. Rare species are valued big time. PLOS. 4: (4) e5125 1-6.


Pearson, K. 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variable is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science, Fifth Series 50:157-172.


117


Wickham, H. 2015. Package ‘reshape’. Pages 1-20 CRAN.


APPENDICES

APPENDIX A

Chapter two: Source script Chao1 Species Estimates
# Chao1 Estimates based on:
# http://www.mothur.org/wiki/Chao

getChao1Estimate <- function(sObs, f1, f2){
  if(f2 > 0){
    return(sObs + f1^2/(2*f2))
  } else {
    return(sObs + f1*(f1-1)/2)
  }
}

getChao1Variance <- function(sObs, f1, f2, n){
  if(f1 > 0 & f2 > 0){
    fRatio <- f1/f2
    return(f2*(0.5*fRatio^2+fRatio^3+0.25*fRatio^4))
  } else if(f1 == 0 & f2 >= 0){
    return(sObs * exp(-n/sObs)*(1-exp(-n/sObs)))
  } else if(f1 > 0 & f2 == 0){
    chao1Estimate <- getChao1Estimate(sObs, f1, f2)
    return(f1/2*(f1-1)+f1/4*(2*f1-1)^2-(f1^4)/(4*chao1Estimate))
  }
}

cGet <- function(sObs, chao1Estimate, chao1Variance){
  return(exp(1.96*sqrt(log(1+chao1Variance/(chao1Estimate - sObs)^2))))
}

getP <- function(sObs, n){
  return(exp(-n/sObs))
}

getChao1Lower95 <- function(sObs, chao1Estimate, chao1Variance, f1, n){
  if(f1 > 0){
    c <- getC(sObs, chao1Estimate, chao1Variance)
    return(sObs + (chao1Estimate - sObs)/c)
  } else {
    p <- getP(sObs, n)
    return(max(sObs, (sObs/(1-p) - 1.96* sqrt((sObs*p)/(1-p))))))
  }
}  

121
getChao1Upper95 <- function(sObs, chao1Estimate, chao1Variance, f1, n){
  if(f1 > 0){
    c <- getC(sObs, chao1Estimate, chao1Variance)
    return(sObs + c*(chao1Estimate - sObs))
  } else {
    p <- getP(sObs, n)
    return(sObs/(1-p) - 1.96 * sqrt((sObs*p)/(1-p)))
  }
}
APPENDIX B

Chapter two: source script for multiplot
library(gcookbook)
library(ggplot2)
library(grid)

# for plotting multiple graphs in ggplot.

multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  library(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }

  if (numPlots==1) {
    print(plots[[1]])
  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                       layout.pos.col = matchidx$col))
    }
  }
}


APPENDIX C

Chapter two: Random point placement for microhabitat photos for chapter one: Impacts of disturbance caused by hemlock woolly adelgid and logging on small mammals distribution and microhabitat associations.

ImageJ 1.42q Java 1.6.0_version 10

//******* Configuration *******
samples = 50;
dotSize = 50;
fontSize = 50;
lineWidth = 8;
// for the color, leave the 0x and replace the remaining part from a color table
lineColor = 0xffff00;
//****************************

width = getWidth();
height = getHeight();
currentFont = getInfo("font.name");
setFont(currentFont, fontSize);
setLineWidth(lineWidth);
setColor(lineColor);
random('seed', getTime());

for (i=1; i<=samples; i++) {
    w = dotSize;
    h = dotSize;
    x = random()*width-w/2;
    y = random()*height-h/2;

    drawOval(x, y, w, h);
    drawString(i, x,y);
}

for (i=1; i<=samples; i++) {
    w = dotSize;
    h = dotSize;
    x = random()*width-w/2;
    y = random()*height-h/2;

    drawOval(x, y, w, h);
    drawString(i, x,y);
}
APPENDIX D

Chapter two: Script for microhabitat analysis
library(plyr)
library(ggplot2)
library(agricolae)
library(grid)

# Checking for differences in treatment and blocks site-covs only
setwd("C:/Users/Ally/Documents/UVM/Projects/Rodent_Harvard/HF_Project/")
source("20-Analysis/R/functions/multiplotFunction.R")

# load microhabitat data
Microhabitat <- read.csv("10-PreProcessing/Dataset/Microhabitat/Microhabitat_Master_v7.csv")

std.err <- function(input){
  return(sd(input)/sqrt(length(input)))
}

Tx.ColorPalette <- c("#009E73", "#E69F00", "#F0E442", "#56B4E9")

#---- ROCK
# Stats, Randomized block ANOVA
Rock.ANOVA <- aov(Rock ~ PlotID + BlockID, data=Microhabitat)
summary(Rock.ANOVA)
Rock.HSD <- TukeyHSD(Rock.ANOVA)
Rock.HSD

# Tukey HSD grouping
Rock.HSD.gp <- HSD.test(Rock.ANOVA, "PlotID", group=TRUE)
Rock.HSD.gp

# calcuate SE
Rock.se <- aggregate(Rock*100 ~ PlotID, data=Microhabitat, FUN = std.err)
Rock.mean <- aggregate(Rock*100 ~ PlotID, data=Microhabitat, FUN = mean)

# Create Average Data Frame
Rock.df <- data.frame(Habitat = Rock.se$PlotID, SE=Rock.se$Rock, Rock=Rock.mean$Rock)
#Treatments in order to plot
treatments <- factor(Rock.df$R habitat, levels= c("Hemlock", "Girdled", "Logged", "Hardwood"))

# Groups of Tukey's HSD group results to plot
Rock.Hemlock.gp <- "a"
Rock.Girdled.gp <- "b"
Rock.Logged.gp <- "ab"
Rock.Hardwood.gp <- "ab"

Rock.Tx.pvalue <- "Treatments (P= 0.066)"
Rock.Bk.pvalue <- "Block (P< 0.0001)"

# Plot
Rock.plot<-ggplot(Rock.df, aes(x=Treatments, y=Rock)) +
  geom_bar(stat = "identity", aes(color=Treatments, fill=Treatments)) +
  geom_errorbar(aes(ymin=Rock-SE, ymax=Rock+SE), width=.2) +
  geom_text(x=1, y=10, aes(label=Rock.Hemlock.gp), size=5) +
  geom_text(x=2, y=10, aes(label=Rock.Girdled.gp), size=5) +
  geom_text(x=3, y=10, aes(label=Rock.Logged.gp), size=5) +
  geom_text(x=4, y=10, aes(label=Rock.Hardwood.gp), size=5) +
  geom_text(x=2.5, y=100, aes(label=Rock.Tx.pvalue), size = 5) +
  geom_text(x=2.5, y=95, aes(label=Rock.Bk.pvalue), size = 5) +
  scale_colour_manual(values= Tx.ColorPalette) +
  scale_fill_manual(values= Tx.ColorPalette) +
  expand_limits(y=c(0,100)) +
  xlab ("") +
  ylab ("Rock Cover (\%)") +
  theme(axis.title.x = element_blank(),
         axis.text.x = element_blank(),
         axis.title.y = element_text(size=20),
         axis.text.y= element_text(size=18)) +
  theme(legend.title=element_text(size=10),
         legend.text=element_text(lineheight = .9, size= 8),
         legend.key.height=unit(1,"cm"),
         legend.key.width= unit(1, "cm")) +
  theme(legend.position = c(.9, .5))
  multiplot(Open.plot, High.plot, Low.plot, cols= 1)

dev.off()
APPENDIX E

Chapter two: Script or Chao1
# Chao1 Estimates based on:
# http://www.mothur.org/wiki/Chao
gETCHAO1ESTIMATE <- function(sObs, f1, f2){
  if(f2 > 0){
    return(sObs + f1^2/(2*f2))
  } else {
    return(sObs + f1*(f1-1)/2)
  }
}
gETCHAO1VARIANCE <- function(sObs, f1, f2, n){
  if(f1 > 0 & f2 > 0){
    fRatio <- f1/f2
    return(f2*(0.5*fRatio^2+fRatio^3+0.25*fRatio^4))
  } else if(f1 == 0 & f2 >= 0){
    return(sObs * exp(-n/sObs)*(1-exp(-n/sObs)))
  } else if(f1 > 0 & f2 == 0){
    chao1Estimate <- getCHAO1ESTIMATE(sObs, f1, f2)
    return(f1/2*(f1-1)+f1/4*(2*f1-1)^2-(f1^4)/(4*chao1Estimate))
  }
}

getC <- function(sObs, chao1Estimate, chao1Variance){
  return(exp(1.96*sqrt(log(1+chao1Variance/(chao1Estimate - sObs)^2))))
}

getP <- function(sObs, n){
  return(exp(-n/sObs))
}

gETCHAO1LOWER95 <- function(sObs, chao1Estimate, chao1Variance, f1, n){
  if(f1 > 0){
    c <- getC(sObs, chao1Estimate, chao1Variance)
    return(sObs + (chao1Estimate - sObs)/c)
  } else {
    p <- getP(sObs, n)
    return(max(sObs, (sObs/(1-p) - 1.96* sqrt((sObs*p)/(1-p)) )))
  }
}
getChao1Upper95 <- function(sObs, chao1Estimate, chao1Variance, f1, n){
  if(f1 > 0){
    c <- getC(sObs, chao1Estimate, chao1Variance)
    return(sObs + c*(chao1Estimate - sObs))
  } else {
    p <- getP(sObs, n)
    return(sObs/(1-p) - 1.96 * sqrt((sObs*p)/(1-p)))
  }
}

# AD & CD
# Species Capture, then species rank, by plot
# community analysis for 2014
# 2014-05-17

# goal is to produce a graph with plots on x-axis, captured individuals on y-axis, then
# rank by abundance, try to make transparent colors for less dominant species

#### Prep work: libraries, assign variables, set up speciesID colors and common names

#libraries
library(ggplot2)
library(plyr)
library(gcookbook)
library(reshape)
library(agricolae)
library(grid)
library(devtools)
library(SpadeR)

# Set WD and Get Data
#setwd("~/ally/HF_Project/")
setwd("C:/Users/Ally/Documents/UVM/Projects/Rodent_Harvard/HF_Project/")

source("20-Analysis/R/functions/Functions_Species_Richness_Estimate.R")
source("20-Analysis/R/functions/multiplotFunction.R")

masterData <- read.csv("10-PreProcessing/Dataset/Degrassi_HF_Master_v14.csv", stringsAsFactors = FALSE)

# Remove unwanted species
masterDataPreProcess <- subset(masterData, LegalExcluded != "Y")
masterDataPreProcess <- subset(masterData, SpeciesID != "PE")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "SHREW")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "Mouse")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "SOFU")

# Subset by Year
Year <- "2012"

SubsetDataByYear <- subset(masterDataPreProcess, YearID == Year)

# subset by capture
SubsetDataByCapture <- subset(SubsetDataByYear, Capture == "CAPTURE")
SubsetDataByCapture$CaptureCt <- 1

# S.obs = Number of species observed
getSpeciesCount <- function(x) {
  speciesCount <- length(unique(x))
  return (speciesCount)
}

S.obs <- aggregate(SpeciesID ~ YearID + PlotID, data = SubsetDataByCapture, FUN = getSpeciesCount)
S.obs$Obs <- S.obs$SpeciesID
S.obs$SpeciesID <- NULL

# Aggregate data to obtain Capture of species by plot and total captures by plot
CaptureBySpecies <- aggregate(CaptureCt ~ PlotID + SpeciesID, data = SubsetDataByCapture, FUN = sum)

PivotCaptureBySpecies <- cast(CaptureBySpecies, SpeciesID ~ PlotID)
PivotCaptureBySpecies[is.na(PivotCaptureBySpecies)] <- 0

##################################################
Chao.He <- data.frame(Hemlock = PivotCaptureBySpecies$Hemlock)
ChaoSpecies(Chao.He, datatype="abundance", k=10, conf=0.95)
Chao.Gi <- data.frame(Girdled = PivotCaptureBySpecies$Girdled)
ChaoSpecies(Chao.Gi, datatype="abundance", k=10, conf=0.95)

Chao.Lo <- data.frame(Logged = PivotCaptureBySpecies$Logged)
ChaoSpecies(Chao.Lo, datatype="abundance", k=10, conf=0.95)

Chao.Ha <- data.frame(Hardwood = PivotCaptureBySpecies$Hardwood)
ChaoSpecies(Chao.Ha, datatype="abundance", k=10, conf=0.95)

Chao.df <- data.frame(PlotID = c("Hemlock", "Girdled", "Logged", "Hardwood"),
                       Estimate = c(6,8,7, 6),
                       SE = c(.481,.544,.50,.223),
                       Lower95 = c(6,8,7,6),
                       Upper95 = c(7.40,9.59,8.45,6.49))

###########################
###CI overlap test
#Ridge hemlock vs logged
RHeAvg <- Chao.df[1,2]
RHeSE <- Chao.df[1,3]
RGiAvg <- Chao.df[2,2]
RGiSE <- Chao.df[2,3]
RLoAvg <- Chao.df[3,2]
RLoSE <- Chao.df[3,3]
RHaAvg <- Chao.df[4,2]
RHaSE <- Chao.df[4,3]

#---------------------
# hemlock vs girdled
#determine if error bars cross zero. They don't so reject the Ho.
(RHeAvg - RGiAvg) + 1.96 * sqrt(RHeSE^2 + RGiSE^2) #(-0.57)
(RHeAvg - RGiAvg) - 1.96 * sqrt(RHeSE^2 + RGiSE^2) #(-3.42)

#---------------------
# hemlock vs logged
#determine if error bars cross zero. They DO so CAN'T reject the Ho.
(RHeAvg - RLoAvg) + 1.96 * sqrt(RHeSE^2 + RLoSE^2) #(0.35)
(RHeAvg - RLoAvg) - 1.96 * sqrt(RHeSE^2 + RLoSE^2) #(-2.36)

#---------------------
# hemlock vs hardwood
#determine if error bars cross zero. They DO so CAN'T reject the Ho.
(RHeAvg - RHaAvg) + 1.96 * sqrt(RHeSE^2 + RHaSE^2) #(1.03)
(RHeAvg - RHaAvg) - 1.96 * sqrt(RHeSE^2 + RHaSE^2) #(-1.03)
# girdled vs logged
# determine if error bars cross zero. They DO so CAN’T reject the Ho.
(RGiAvg - RLoAvg) + 1.96 * sqrt(RGiSE^2 + RLoSE^2) #(2.44)
(RGiAvg - RLoAvg) - 1.96 * sqrt(RGiSE^2 + RLoSE^2) #(-0.44)

# girdled vs hardwood
# determine if error bars cross zero. They don’t so can reject the Ho.
(RGiAvg - RHaAvg) + 1.96 * sqrt(RGiSE^2 + RHaSE^2) #(3.15)
(RGiAvg - RHaAvg) - 1.96 * sqrt(RGiSE^2 + RHaSE^2) #(0.847)

# logged vs hardwood
# determine if error bars cross zero. They DO so CAN’T reject the Ho.
(RLoAvg - RHaAvg) + 1.96 * sqrt(RLoSE^2 + RHaSE^2) #(2.07)
(RLoAvg - RHaAvg) - 1.96 * sqrt(RLoSE^2 + RHaSE^2) #(-0.17)

# so, there is a significant difference between, He and all others, Gi and Lo.

Hemlock.gp <- "ab"
Girdled.gp <- "c"
Logged.gp <- "bc"
Hardwood.gp <- "a"

Treatments <- factor(Chao.df$PlotID, levels=c("Hemlock", "Girdled", "Logged", "Hardwood"))

ggplot(Chao.df, aes(x=Treatments, y=Estimate, fill= Treatments)) +
  #geom_bar(stat="identity", position = "dodge") +
  geom_errorbar(aes(ymin=Lower95, ymax=Upper95), width = 0.25) +

  #geom_point(aes(x=Treatments, y= Obs), size = 10) +
  geom_point(size =7) +

  #geom_text for the sig groupings
  geom_text(x=1, y=7.5, aes(label=Hemlock.gp), size=5) +
  geom_text(x=2, y=9.74, aes(label=Girdled.gp), size=5) +
  geom_text(x=3, y=8.6, aes(label=Logged.gp), size=5) +
  geom_text(x=4, y=6.64, aes(label=Hardwood.gp), size=5) +

  expand_limits(y=c(0,10)) +
  xlab (""") +
  ylab("Estimated Species Richness") +
theme_bw() +
theme(axis.line = element_line(colour = "black"),
      panel.grid.major = element_blank(),
      panel.grid.minor = element_blank(),
      panel.background = element_blank()) +

theme(axis.title.x = element_text(size=20),
      axis.text.x = element_text(size= 20),
      axis.title.y = element_text(size=20),
      axis.text.y= element_text(size=18)) +

theme(legend.position = "none")

#2012 Shared

#He and Gi
HeGiShared <- data.frame(Hemlock = PivotCaptureBySpecies$Hemlock,
                          Girdled = PivotCaptureBySpecies$Girdled)
ChaoShared(HeGiShared, datatype="abundance", se=TRUE, nboot=200, conf=0.95)

#He and Lo
HeLoShared <- data.frame(Hemlock = PivotCaptureBySpecies$Hemlock,
                          Logged = PivotCaptureBySpecies$Logged)
ChaoShared(HeLoShared, datatype="abundance", se=TRUE, nboot=200, conf=0.95)

#He and Gi
HeHaShared <- data.frame(Hemlock = PivotCaptureBySpecies$Hemlock,
                          Hardwood = PivotCaptureBySpecies$Hardwood)
ChaoShared(HeHaShared, datatype="abundance", se=TRUE, nboot=200, conf=0.95)

Shared12 <- data.frame(PlotID = c("Girdled", "Logged", "Hardwood"),
                        Estimate = c(6,5,5),
                        SE= c(.57,.46,0.0),
                        Lower95 = c(5.35,4.43,5.0),
                        Upper95 = c(7.84,6.29,5.0))
SharedTreatments <- factor(Shared12$PlotID, levels=c("Girdled", "Logged", "Hardwood"))

ggplot(Shared12, aes(x=SharedTreatments, y=Estimate, fill = SharedTreatments)) +
  #geom_bar(stat="identity", position = "dodge") +
  geom_errorbar(aes(ymin=Lower95, ymax=Upper95), width = 0.25) +
  #geom_point(aes(x=Treatments, y= Obs), size = 10) +
  geom_point(size =10) +

  expand_limits(y=c(0,10)) +
  xlab("") +
  ylab("Estimated Shared Species Richness with Hemlock") +

  theme_bw() +
  theme(axis.line = element_line(colour = "black"),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank()) +

  theme(axis.title.x = element_text(size=20),
        axis.text.x = element_text(size= 20),
        axis.title.y = element_text(size=20),
        axis.text.y = element_text(size=18)) +

  theme(legend.position = "none")

#########################################################################
png(filename="SharedRichness.png",
     type="cairo",
     units="in",
     width=10,
     height=15,
     pointsize=12,
     res=300)
multiplot(S.est12Plot, S.est13Plot, S.est14Plot, cols= 1)

dev.off()
#########################################################################
data(DiversityDataAbu)
HeDiversity <- Diversity(PivotCaptureBySpecies$Hemlock, datatype = "abundance", q=NULL)
GiDiversity <- Diversity(PivotCaptureBySpecies$Girdled, datatype = "abundance", q= NULL)
LoDiversity <- Diversity(PivotCaptureBySpecies$Logged, datatype = "abundance", q= NULL)
HaDiversity <- Diversity(PivotCaptureBySpecies$Hardwood, datatype = "abundance", q= NULL)

SimilarityMult(PivotCaptureBySpecies, q = 0, nboot = 200)

Similarity.df <- PivotCaptureBySpecies[,2:5]
data(SimilarityMultDataAbu)
SimilarityMult(Similarity.df, q=2, nboot=500)
APPENDIX F

Chapter two: Script for relative abundance
# AD & CD
# Species Capture, then species rank, by plot
# community analysis for 2014
# 2014-05-17

# goal is to produce a graph with plots on x-axis, captured individuals on y-axis, then
rank by abundance, try to make transparent colors for less dominate specie

#### Prep work: libraries, assign variables, set up speciesID colors and common names

#libraries
library(ggplot2)
library(plyr)
library(reshape)
library(grid)
library(gridExtra)

# Set WD and Get Data
setwd("~/ally/HF_Project/")
setwd("C:/Users/Ally/Documents/UVM/Projects/Rodent_Harvard/HF_Project/")

source("20-Analysis/R/functions/Functions_Species_Richness_Estimate.R")
source("20-Analysis/R/functions/multiplotFunction.R")

masterData <- read.csv("10-PreProcessing/Dataset/Degrassi_HF_Master_v14.csv", stringsAsFactors = FALSE)

# Remove unwanted species
masterDataPreProcess <- subset(masterData, LegalExcluded != "Yes")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "PE")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "SHREW")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "Mouse")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "SOFU")

# Subset by Year
Year <- "2012"

SubsetDataByYear <- subset(masterDataPreProcess, YearID == Year)

#subset by capture
SubsetDataByCapture <- subset(SubsetDataByYear, Capture == "CAPTURE")
SubsetDataByCapture$CaptureCt <- 1

# aggregate to calculate the sum of captures per species per plot
CaptureBySpecies <- aggregate(CaptureCt ~ PlotID + SpeciesID, data =
SubsetDataByCapture, FUN = sum)

# aggregate to calculate the total number of captures per plot
CaptureByPlot <- aggregate(CaptureCt ~ PlotID, data = SubsetDataByCapture, FUN =
sum)

# subset by species
He.df <- subset(CaptureBySpecies, PlotID == "Hemlock")
# subset by plot
HePlot.df <- subset(CaptureByPlot, PlotID == "Hemlock")
# put the plot capture count into df
He.df$PlotCt <- HePlot.df$CaptureCt
# calculate the relative abundance
He.df$Abundance <- He.df$CaptureCt/He.df$PlotCt
# reorder the abundance so it is descending for the species graph
He.df <- He.df[order(-He.df$Abundance),]

# do for Gi, Lo, and Ha
Gi.df <- subset(CaptureBySpecies, PlotID == "Girdled")
GiPlot.df <- subset(CaptureByPlot, PlotID == "Girdled")
Gi.df$PlotCt <- GiPlot.df$CaptureCt
Gi.df$Abundance <- Gi.df$CaptureCt/Gi.df$PlotCt
Gi.df <- Gi.df[order(-Gi.df$Abundance),]

Lo.df <- subset(CaptureBySpecies, PlotID == "Logged")
LoPlot.df <- subset(CaptureByPlot, PlotID == "Logged")
Lo.df$PlotCt <- LoPlot.df$CaptureCt
Lo.df$Abundance <- Lo.df$CaptureCt/Lo.df$PlotCt
Lo.df <- Lo.df[order(-Lo.df$Abundance),]

Ha.df <- subset(CaptureBySpecies, PlotID == "Hardwood")
HaPlot.df <- subset(CaptureByPlot, PlotID == "Hardwood")
Ha.df$PlotCt <- HaPlot.df$CaptureCt
Ha.df$Abundance <- Ha.df$CaptureCt/Ha.df$PlotCt
Ha.df <- Ha.df[order(-Ha.df$Abundance),]

Capture.df <- rbind(He.df, Gi.df, Lo.df, Ha.df)

# for graphing
SpeciesIDs <- c("PEMA", "PELE", "MYGA", "MIPE", "NAIN", "GLVO", "TAST", "BLBR", "SOCI")
CommonSpeciesNames <- c("deermouse", "white-footed mouse", "southern red-backed vole", "woodland vole", "woodland jumping mouse", "southern flying squirrel", "eastern chipmunk", "short-tailed shrew", "masked shrew")
#colors I like
#SpeciesColors <- c("steelblue4", "steelblue2", "navajowhite1", "burlywood2", "rosybrow1", "darkolivegreen4", "darkolivegreen3", "red4", "red3")
#colorchallenged friendly
#SpeciesColors<- c("#000000", "#999999","#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")
SpeciesColors<- c("#0072B2", "#56B4E9", "#D55E00", "#E69F00", "#009E73", "#F0E442", "#CC79A7", "#999999", "#000000")

# Create Empty data frame for species names; it is going to be used as a refernce to rename speciesID from PEMA to Deer mouse and assoicate a consitant color with it blue
SpeciesAppearance <- data.frame(t(rep(NA,3)))
names(SpeciesAppearance) <- c("SpeciesID","CommonSpeciesName", "Color")

# Adds the id and the common names to the table
SpeciesAppearance[1:length(SpeciesIDs),1] <- SpeciesIDs
SpeciesAppearance[1:length(SpeciesIDs),2] <- CommonSpeciesNames
SpeciesAppearance[1:length(SpeciesIDs),3] <- SpeciesColors

# Add common names and colors to the abundance table
AbundanceData <- join(Capture.df, SpeciesAppearance, by = "SpeciesID")

# We do this before factor because ggplot does not take "factor" vectors for colors
graphSpeciesColors <- unique(AbundanceData$Color)
graphSpeciesColors <- graphSpeciesColors[match(SpeciesColors, graphSpeciesColors)]
graphSpeciesColors <- graphSpeciesColors[!is.na(graphSpeciesColors)]

#subset out again cause making one plot
He <- subset(AbundanceData, PlotID == "Hemlock")
Gi <- subset(AbundanceData, PlotID == "Girdled")
Lo <- subset(AbundanceData, PlotID == "Logged")
Ha <- subset(AbundanceData, PlotID == "Hardwood")

#factor to get the order I want

#Graph with ggplot
yMax = 1
titleFace = "bold"
titleSize = 20

xaxisSize = 15

yaxisFace = "bold"
yaxisSize = 15

barwidth = 0.90

He$CommonSpeciesName <- factor(He$CommonSpeciesName, levels=He$CommonSpeciesName)
He.plot <- ggplot(He, aes(x= CommonSpeciesName, y= Abundance)) +
  geom_bar(aes(fill=CommonSpeciesName, width= barwidth), position = "dodge", stat="identity", color = "black") +
  xlab("") +
  ylab("Relative Capture Abundance") +
  labs (title = "Hemlock") +
  scale_fill_manual(values= He$Color) +
  expand_limits(ymin =0, ymax=yMax) +
  theme(legend.position="none") +
  theme(  
    axis.text.x = element_text(colour="black",size= xaxisSize, hjust=1, vjust=.5, angle = 90),
    axis.text.y = element_text(colour="black",size= yaxisSize),
    axis.title.y = element_text(color = "black", size = yaxisSize),
    axis.title.x = element_blank())
########################################################################
Gi$CommonSpeciesName <- factor(Gi$CommonSpeciesName, levels=Gi$CommonSpeciesName)

Gi.plot <- ggplot(Gi, aes(x= CommonSpeciesName, y= Abundance)) +
  geom_bar(aes(fill=CommonSpeciesName, width= barwidth), position = "dodge", stat="identity", color = "black") +
  xlab("") +
  ylab("") +
  labs (title = "Girdled") +
  scale_fill_manual(values= Gi$Color) +
  expand_limits(ymin =0, ymax=yMax) +
  theme(legend.position="none") +
  theme(  
    axis.text.x = element_text(colour="black",size= xaxisSize, hjust=1, vjust=.5, angle = 90),
    axis.text.y = element_text(colour="black",size= yaxisSize),
    axis.title.y = element_text(color = "black", size = yaxisSize),
    axis.title.x = element_blank())
########################################################################
axis.text.x = element_text(colour="black", size=xaxisSize, hjust=1, vjust=.5, angle = 90),
axis.text.y = element_text(colour="black", size=yaxisSize),
axis.title.y = element_text(color = "black", size=yaxisSize),
axis.title.x = element_blank())
###############################
Lo$CommonSpeciesName <- factor(Lo$CommonSpeciesName, levels=Lo$CommonSpeciesName)
Lo.plot <- ggplot(Lo, aes(x=CommonSpeciesName, y=Abundance)) +
  geom_bar(aes(fill=CommonSpeciesName, width=barwidth), position = "dodge", stat="identity", color = "black") +
xlab("") +
ylab("") +
labs(title = "Logged") +
scale_fill_manual(values= Lo$Color) +
expand_limits(ymin =0, ymax=yMax) +
theme(legend.position="none") +
  theme(
    axis.text.x = element_text(colour="black", size=xaxisSize, hjust=1, vjust=.5, angle = 90),
    axis.text.y = element_text(colour="black", size=yaxisSize),
    axis.title.y = element_text(color = "black", size=yaxisSize),
    axis.title.x = element_blank())
###############################
Ha$CommonSpeciesName <- factor(Ha$CommonSpeciesName, levels=Ha$CommonSpeciesName)
Ha.plot <- ggplot(Ha, aes(x=CommonSpeciesName, y=Abundance)) +
  geom_bar(aes(fill=CommonSpeciesName, width=barwidth), position = "dodge", stat="identity", color = "black") +
xlab("") +
ylab("") +
labs(title = "Hardwood") +
scale_fill_manual(values= Ha$Color) +
expand_limits(ymin =0, ymax=yMax) +
theme(legend.position="none") +
  theme(
    axis.text.x = element_text(colour="black", size=xaxisSize, hjust=1, vjust=.5, angle = 90),
    axis.text.y = element_text(colour="black", size=yaxisSize, hjust=1, vjust=.5, angle = 90),
    axis.title.y = element_text(color = "black", size=yaxisSize, hjust=1, vjust=.5, angle = 90),
    axis.title.x = element_text(color = "black", size=xaxisSize, hjust=1, vjust=.5, angle = 90),
    axis.title.x = element_blank())

139
axis.text.y = element_text(colour="black",size= yaxisSize),
axis.title.y = element_text(color = "black", size = yaxisSize),
axis.title.x = element_blank())

################
grid.arrange(He.plot, Gi.plot, Lo.plot, Ha.plot, nrow = 1)
##
# for the legend
ggplot(Gi, aes(x= CommonSpeciesName, y= Abundance)) +
  geom_bar(aes(fill=CommonSpeciesName, width= barwidth), position = "dodge", stat=
  "identity", color = "black") +
  xlab("") +
  ylab("") +
  labs (title = "Girdled") +
  scale_fill_manual(values= Gi$Color) +
  expand_limits(ymin =0, ymax=yMax) +
  theme(
    axis.text.x = element_text(colour="black",size= xaxisSize, hjust=1, vjust=.5, angle =
    90),
    axis.text.y = element_text(colour="black",size= yaxisSize),
    axis.title.y = element_text(color = "black", size = yaxisSize),
    axis.title.x = element_blank())

  
ggplot(Lo, aes(x= CommonSpeciesName, y= Abundance)) +
  geom_bar(aes(fill=CommonSpeciesName, width= barwidth), position = "dodge", stat=
  "identity", color = "black") +
  xlab("") +
  ylab("") +
  labs (title = "Logged") +
  scale_fill_manual(values= Lo$Color) +
  expand_limits(ymin =0, ymax=yMax) +
  theme(legend.title=element_text("Species"))+
  theme(
    axis.text.x = element_text(colour="black",size= xaxisSize, hjust=1, vjust=.5, angle =
    90),
    axis.text.y = element_text(colour="black",size= yaxisSize),
    axis.title.y = element_text(color = "black", size = yaxisSize),
    axis.title.x = element_blank())
APPENDIX G

Chapter two: Script for PIE analysis
# AD & CD
# Species Capture, then species rank, by plot
# community analysis for 2014
# 2014-05-17
#

#### Prep work: libraries, assign variables, set up speciesID colors and common names

#libraries
library(ggplot2)
library(plyr)
library(gcookbook)

# Set WD and Get Data
setwd("C:/Users/Ally/Documents/UVM/Projects/Rodent_Harvard/HF_Project/")
masterData <- read.csv("10-PreProcessing/Dataset/Degrassi_HF_Master_v14.csv", stringsAsFactors = FALSE)

# Remove unwanted species
masterDataPreProcess <- subset(masterData, LegalExcluded != "Y")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "PE")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "SHREW")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "Mouse")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "SOFU")

# Subset by Year
Year = "2012"
SubsetDataByYear <- subset(masterDataPreProcess, YearID == Year)

#subset by capture
SubsetDataByCapture <- subset(SubsetDataByYear, Capture == "CAPTURE" )
SubsetDataByCapture$CaptureCt <- 1

#Aggregate data to obtain Capture of species by plot
CaptureBySpecies <- aggregate(CaptureCt ~ BlockID + PlotID + SpeciesID, data = SubsetDataByCapture, FUN = sum)
CaptureTotalCt <- aggregate(CaptureCt ~ BlockID + PlotID, data = SubsetDataByCapture, FUN = sum)
CaptureTotalCt$N <- CaptureTotalCt$CaptureCt
CaptureTotalCt$CaptureCt <- NULL

CaptureData <- join(CaptureBySpecies, CaptureTotalCt, by=c("PlotID", "BlockID"), type="left", match="all")
CaptureData$pi <- CaptureData$CaptureCt / CaptureData$N

PIE <- function(N.species){
  N <- sum(N.species)
  p_i <- N.species/N
  pie <- (N/(N-1))*(1-sum(p_i^2))
  pie
}

PIE.Est <- aggregate(CaptureCt ~ BlockID + PlotID, data=CaptureData, FUN=PIE)
PIE.Est$pie <- PIE.Est$CaptureCt
PIE.Est$CaptureCt <- NULL

PIE.SD <- aggregate(pie ~ PlotID, data=PIE.Est, FUN=sd)
PIE.SD$pieSD <- PIE.SD$pie
PIE.SD$pie <- NULL

PIE.Avg <- aggregate(pie ~ PlotID, data=PIE.Est, FUN=mean)
PIE.Avg$pieAvg <- PIE.Avg$pie
PIE.Avg$pie <- NULL

PIE.Avg$Lower95 <- PIE.Avg$pieAvg - (1.96)*PIE.SD$pieSD
PIE.Avg$Upper95 <- PIE.Avg$pieAvg + (1.96)*PIE.SD$pieSD
PIE.Avg[3,4] <- 1

########################################
#Graph with ggplot
Treatments <- factor(PIE.Avg$PlotID, levels=c("Hemlock", "Girdled", "Logged", "Hardwood"))

ggplot(PIE.Avg, aes(x= Treatments, y=pieAvg)) +
geom_errorbar(aes(ymin=Lower95, ymax=Upper95), width = 0.25) +

  #geom_point(aes(x=Treatments, y= Obs), size = 10) +
geom_point(size =8) +
xlab("") +
ylab("Average PIE") +
expand_limits(ymin =0, ymax=1) +
theme_bw() +
theme(axis.line = element_line(colour = "black"),
       panel.grid.major = element_blank(),
       panel.grid.minor = element_blank(),
       panel.background = element_blank()) +

theme(axis.title.x = element_text(size=20),
       axis.text.x = element_text(size = 20),
       axis.title.y = element_text(size=20),
       axis.text.y= element_text(size=18)) +

theme(legend.position = "none")

fit <- aov(pie ~ PlotID + BlockID, data=PIE.Est)
summary(fit)
APPENDIX H

Chapter two: Script for Schnabel Population Estimates
library(reshape)
library(ggplot2)
library(gplots)

#----------------------------- FUNCTION DEFINITIONS -----------------------------

Schnabel = function(masterData, year, numberOfNights){

# Subset by Year
YearRodentData <- subset(masterData, YearID == year & !is.na(SpeciesID))

# Get the species and plots for this subset
SpeciesNames <- unique(YearRodentData$SpeciesID)
PlotNames <- unique(YearRodentData$PlotID)

# Subset the Columns from Rodent data: BlockID, PlotID, Night, SpeciesID,
# FirstMarked, Recapture
MarkRecapFieldData <- YearRodentData[, c("BlockID", "PlotID", "Night",
"SpeciesID", "FirstMarked", "Recapture")]
MarkRecapFieldData$RecapBinary <- NA
MarkRecapFieldData$FirstMarkedBinary <- NA

#View(MarkRecapFieldData)

#Converts Y -> 1 and N -> 0; Multiply by 1 to convert TRUE to 1 and FALSE to 0
MarkRecapFieldData$RecapBinary <- (MarkRecapFieldData$Recapture == "Y") * 1
MarkRecapFieldData$FirstMarkedBinary <- (MarkRecapFieldData$FirstMarked ==
"Y") * 1

#Sort By Night
MarkRecapFieldData <- MarkRecapFieldData[order(MarkRecapFieldData$Night),]

View(MarkRecapFieldData)

# Create Empty Results Table
SchnabelResults <- data.frame(t(rep(NA,15)))
names(SchnabelResults) <- c("PlotID","SpeciesID","N", "Upper95N", "Lower95N")
# Removes all the empty rows and only keeps field names
SchnabelResults <- SchnabelResults[-1,]

for (SelectedSpecies in SpeciesNames){

#print("*****************************************
**")
#print(SelectedSpecies)

for (SelectedPlotID in PlotNames){

#print(SelectedPlotID)

# Selects by PlotID, Species and excludes NA from "Recapture" field (keeps Y and N)
SpeciesPlotSubset <- subset(MarkRecapFieldData, PlotID == SelectedPlotID & SpeciesID == SelectedSpecies & Recapture != "NA")
SpeciesPlotSubset <- subset(MarkRecapFieldData, PlotID == SelectedPlotID & SpeciesID == SelectedSpecies & !is.na(Recapture) & !is.na(FirstMarkedBinary))

# Calculate number of unique nights since we have replicates with same nights
NumberOfUniqueNights = length(unique(SpeciesPlotSubset$Night))

#Must have at least 2 nights
if(NumberOfUniqueNights > 1) {
    #print(SpeciesPlotSubset)

    # Calculate Rt
    Rt <- aggregate(RecapBinary ~ Night, data = SpeciesPlotSubset, FUN = sum)
    Rt <- Rt[order(Rt$Night),]

    # Calculate Ut
    Ut = aggregate(FirstMarkedBinary ~ Night, data = SpeciesPlotSubset, FUN = sum)
    Ut <- Ut[order(Ut$Night),]

    # Create summary table
    SchnabelSummary <- data.frame(
        Night = Rt$Night,
        PlotID = SelectedPlotID,
        SpeciesID = SelectedSpecies,
        Rt = Rt$RecapBinary,
        Ut = Ut$FirstMarkedBinary)

    # Only keeps "numberOfNights" events or n rows if there are already less nights
    numberOfEvents <- min(nrow(SchnabelSummary), numberOfNights)
    SchnabelSummary <- SchnabelSummary[1:numberOfEvents, ]
SchnabelSummary$Ct <- SchnabelSummary$Rt + SchnabelSummary$Ut
SchnabelSummary$Mt <- 0

# Calculate Mt
for(index in 2:length(SchnabelSummary$Night))
{
    SchnabelSummary[index, "Mt"] = SchnabelSummary[index - 1, "Ut"] +
    SchnabelSummary[index - 1, "Mt"]
}

SchnabelSummary$MCt <- SchnabelSummary$Ct * SchnabelSummary$Mt

SchnabelSummary$Nights <- numberOfEvents

SumOfRt = sum(SchnabelSummary$Rt)
SumOfMCt = sum(SchnabelSummary$MCt)

print(""
print("------------------------------------------------------------------")
print(SchnabelSummary)

# IF SumOfRt is 0 then all the calculations would fail to NaN (division by 0)
if(SumOfRt == 0){
    N = 0
    Lower95N = 0
    Upper95N = 0
} else {
    # Estimate population (Krebs, p. 36 - Eq. 2.9 & 2.10)
    N = SumOfMCt/(SumOfRt)
    fractionOfPopulation = SchnabelSummary$Ct/N
    fractionOfTotalPopulation = SchnabelSummary$Mt/N

    if(fractionOfPopulation < 0.1 && fractionOfTotalPopulation < 0.1){
        N = SumOfMCt/(SumOfRt + 1)
    }

    # Calculate Upper and Lower CI
    if(SumOfRt <= 50){
        # Use Poisson limits for Rt < 50 (Krebs, p. 37, example on p. 38)
        poissonLimits = CalculatePoissonLimitsFunction(SumOfRt, 0.95)
        Lower95N = SumOfMCt/poissonLimits[2]
        Upper95N = SumOfMCt/poissonLimits[1]
    } else {

147

# Begin of Schnabel Method for Variance (Krebs, p.37)
Variance1OverN = SumOfRt/(SumOfMCt^2)
StandardErrorOfVariance1OverN = sqrt(Variance1OverN)
# End of Schnabel Method for Variance

# Begin Confidence Intervals for Rt > 50 (Krebs, p.37 - Eq 2.16)
degreesOfFreedom = length(SchnabelSummary$MCt) -1
$ t = qt(0.975, df=degreesOfFreedom)
Upper951OverN = (1/N) - t * StandardErrorOfVariance1OverN
Lower951OverN = (1/N) + t * StandardErrorOfVariance1OverN
Upper95N = 1/Upper951OverN
Lower95N = 1/Lower951OverN

row <- cbind(SchnabelSummary[length(SchnabelSummary$PlotID), c("PlotID", "SpeciesID", "Nights")], N, Upper95N, Lower95N)

#SchnabelResults <- rbind(SchnabelResults, SchnabelSummary[length(SchnabelSummary$PlotID), c(2,3)])
SchnabelResults <- rbind(SchnabelResults, row)

# Show Plot x Species Summary
#if(SelectedSpecies == "CLGA")
#
#SchnabelSummary
#View(SchnabelSummary)
#
#
#SchnabelResults = SchnabelResults[order(SchnabelResults$PlotID),]
row.names(SchnabelResults) <- seq(nrow(SchnabelResults))

return(SchnabelResults)

}

CalculatePoissonLimitsFunction = function(estimatedPopulationSize, confidence){
  alpha = 1-confidence
  chiAlpha = alpha/2
  return (c( qchisq(chiAlpha, 2*estimatedPopulationSize)/2, qchisq(1-chiAlpha, 2*(estimatedPopulationSize+1))/2 ));
}

#Pivot the Schnabel results for plotting
PivotPlotVsSpeciesFunction = function(SchnabelResults, PlotNames, AllSpeciesNames){
  PivotResults <- cast(SchnabelResults, PlotID ~ SpeciesID, value="N")
  PivotResults <- PivotResults[c(3, 2, 1, 4),]

  # Gets species from Schnebel and organizes them based on the "AllSpeciesName" order
  PresentSpecies <- unique(SchnabelResults$SpeciesID)
  OrganizedPresentSpecies <- PresentSpecies[match(AllSpeciesNames, PresentSpecies)]
  CleanedPresentSpecies <- as.character(OrganizedPresentSpecies[!is.na(OrganizedPresentSpecies)])

  # Organizes rows by given PlotNames and Columns by given SpeciesNames
  orderedRowsPivot <- PivotResults[match(PlotNames, PivotResults$PlotID),]
  orderedPivot <- orderedRowsPivot[c("PlotID", CleanedPresentSpecies)]

  # Removes row.names column
  row.names(orderedPivot) <- seq(nrow(orderedPivot))

  return(orderedPivot)
}

GeneralPivotPlotVsSpeciesFunction = function(SchnabelResults, PlotNames, AllSpeciesNames, FieldName){
  PivotResults <- cast(SchnabelResults, PlotID ~ SpeciesID, value=FieldName)

  # Specifically orders the rows
  #PivotResults <- PivotResults[c(3, 2, 1, 4),]
  #print(PivotResults)

  # Gets species from Schnebel and organizes them based on the "AllSpeciesName" order
  PresentSpecies <- as.character(unique(SchnabelResults$SpeciesID))
  OrganizedPresentSpecies <- PresentSpecies[match(AllSpeciesNames, PresentSpecies)]
  CleanedPresentSpecies <- as.character(OrganizedPresentSpecies[!is.na(OrganizedPresentSpecies)])

  # Organizes rows by given PlotNames and Columns by given SpeciesNames
  orderedRowsPivot <- PivotResults[match(PlotNames, PivotResults$PlotID),]
  orderedPivot <- orderedRowsPivot[c("PlotID", CleanedPresentSpecies)]

  # Removes row.names column
  row.names(orderedPivot) <- seq(nrow(orderedPivot))

  return(orderedPivot)
}

PlotEstimatedPopulationFunction = function(PivotedDataSet, Colors){


plotData <- data.matrix(PivotedDataSet[, 2:ncol(PivotedDataSet)])
rownames(plotData) <- PivotedDataSet$PlotID

barplot(t(plotData),
        legend.text = colnames(plotData),
        xlab = "Treatments",
        ylab = "Estimated Population",
        ylim = c(0, 160),
        axis.lty = 1,
        axes = TRUE,
        col = Colors,
        beside = TRUE
    )

PlotEstimatedPopulationWithCIFunction = function(PivotedPopulation, PivotedLower95, PivotedUpper95, Colors) {

    plotData <- data.matrix(PivotedPopulation[, 2:ncol(PivotedPopulation)])
    rownames(plotData) <- PivotedPopulation$PlotID

    plotDataUpper <- data.matrix(PivotedUpper95[, 2:ncol(PivotedPopulation)])
    rownames(plotDataUpper) <- PivotedUpper95$PlotID

    plotDataLower <- data.matrix(PivotedLower95[, 2:ncol(PivotedPopulation)])
    rownames(plotDataLower) <- PivotedLower95$PlotID

    barplot2(t(plotData),
             legend.text = colnames(plotData),
             xlab = "Treatments",
             ylab = "Estimated Population",
             ylim = c(0, 160),
             axis.lty = 1,
             axes = TRUE,
             col = Colors,
             beside = TRUE,
             plot.ci = TRUE,
             ci.u = t(plotDataUpper),
             ci.l = t(plotDataLower)
        )

    }

PracticePopulationFunction = function(PivotedDataSet, Colors) {

}
practiceData <- data.matrix(PivotedDataSet[,2:ncol(PivotedDataSet)])
rownames(practiceData) <- PivotedDataSet$SpeciesID

barplot(practiceData,
    legend.text = PivotedDataSet$PlotID,
    xlab = "Species",
    ylab = "Estimated Population",
    ylim = c(0, 160),
    axis.lty = 1,
    col = Colors,
    beside = TRUE)
}
}

PlotSpeciesPopulationWithCIFunction = function(PivotedPopulation, PivotedLower95, PivotedUpper95, Colors, SpeciesGraphNames) {

    plotData <- data.matrix(PivotedPopulation[, 2:ncol(PivotedPopulation)])
    rownames(plotData) <- PivotedPopulation$PlotID

    plotDataUpper <- data.matrix(PivotedUpper95[, 2:ncol(PivotedPopulation)])
    rownames(plotDataUpper) <- PivotedUpper95$PlotID

    plotDataLower <- data.matrix(PivotedLower95[, 2:ncol(PivotedPopulation)])
    rownames(plotDataLower) <- PivotedLower95$PlotID

    par(cex = 2)
    barplot2(plotData,
        #legend.text = PivotedPopulation$PlotID,
        #legend = TRUE,
        #arg.legend = list(x = "topleft"),
        #names.arg = SpeciesGraphNames,
        #cex.names = 0.7,
        #cex.axis = .75,
        #cex.lab = .75,
        #cex.main = 2,
        #font = 3,
        #main = "Schnabel Population Estimates",
        xlab = "Species",
        ylab = "Estimated Population",
        ylim = c(0, 160),
        axis.lty = 1,
        col = Colors,
        beside = TRUE,
        plot.ci = TRUE,
        ci.u = plotDataUpper,
ci.1 = plotDataLower

legend("topleft",
    PivotedPopulation$PlotID,
    fill = Colors,
    bty = "n",
    cex = 0.7)

masterDataPreProcess <- subset(masterData, LegalExcluded != "Yes")
masterDataPreProcess <- subset(masterData, SpeciesID != "PE")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "SHREW")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "Mouse")
masterDataPreProcess <- subset(masterDataPreProcess, SpeciesID != "SOFU")

# calvuate Schnabel by year and night (1-12)
MarkRecaptureResults = Schnabel(masterData, 2012, 10)
MarkRecaptureResults[is.na(MarkRecaptureResults)] <- 0
#View(MarkRecaptureResults)

SpeciesIDs <- c( "PEMA", "PELE", "MYGA")
CommonSpeciesNames <- c("deermouse", "white-footed mouse", "southern red-backed vole")
#SpeciesIDs <- c( "PEMA", "PELE", "MYGA", "MIPE", "NAIN", "GLVO", "TAST", "BLBR", "SOCI")
#CommonSpeciesNames <- c("deermouse", "white-footed mouse", "southern red-backed vole", "woodland vole", "woodland jumping mouse", "southern flying squirrel", "eastern chipmunk", "short-tailed shrew", "masked shrew")

#colors I like
SpeciesColors <- c("steelblue4", "steelblue2", "navajowhite1", "burlywood2", "rosybrown1", "darkolivegreen4", "darkolivegreen3", "red4", "red3")
#colorchallenged friendly
SpeciesColors <- c("#000000", "#999999", "#E6F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")
SpeciesColors <- c("#0072B2", "#56B4E9", "#D55E00", "#E6F00", "#009E73", "#F0E442", "#CC79A7", "#999999", "#000000")
SpeciesColors <- c("#0072B2", "#56B4E9", "#D55E00")

151
# Create Empty data frame for species names; it is going to be used as a reference to rename speciesID from PEMA to Deer mouse and associate a consistent color with it blue
SpeciesAppearance <- data.frame(t(rep(NA, 3)))
names(SpeciesAppearance) <- c("SpeciesID", "CommonSpeciesName", "Color")

# Adds the id and the common names to the table
SpeciesAppearance[1:length(SpeciesIDs), 1] <- SpeciesIDs
SpeciesAppearance[1:length(SpeciesIDs), 2] <- CommonSpeciesNames
SpeciesAppearance[1:length(SpeciesIDs), 3] <- SpeciesColors

# Plot colors
PlotID <- c("Hemlock", "Girdled", "Logged", "Hardwood")
# cbbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")
PlotColors <- c("#E69F00", "#009E73", "#D55E00", "#0072B2")

PlotAppearance <- data.frame(t(rep(NA, 2)))
names(PlotAppearance) <- c("PlotID", "PlotColors")
PlotAppearance[1:length(PlotID), 1] <- PlotID
PlotAppearance[1:length(PlotID), 2] <- PlotColors

# Add common names and colors to the Capture table
CaptureData <- join(MarkRecaptureResults, SpeciesAppearance, by = "SpeciesID")
CaptureData <- join(CaptureData, PlotAppearance, by = "PlotID")

# We do this before factor because ggplot does not take "factor" vectors for colors
graphSpeciesColors <- unique(CaptureData$Color)
graphSpeciesColors <- graphSpeciesColors[match(SpeciesColors, graphSpeciesColors)]
graphSpeciesColors <- graphSpeciesColors[!is.na(graphSpeciesColors)]

PEMA <- subset(CaptureData, SpeciesID == "PEMA")
MYGA <- subset(CaptureData, SpeciesID == "MYGA")
PELE <- subset(CaptureData, SpeciesID == "PELE")

# Graph with ggplot
yMax = 150
titleFace = "bold"
titleSize = 20
xaxisSize = 15
yaxisFace = "bold"
yaxisSize = 15
barwidth = 0.90
pd <- position_dodge(0.5)
pointSize<-5
textSize <- 5

Treatments <- factor(PEMA$PlotID, levels=c("Hemlock", "Girdled", "Logged","Hardwood"))
PEMA.Text <- "deermice"

PEMA.plot <- ggplot(PEMA, aes(x= Treatments, y= N, fill=CommonSpeciesName)) +
   geom_errorbar(aes(ymin=Lower95N, ymax=Upper95N, fill=CommonSpeciesName),
   width=.1) +
   geom_point(aes(color=CommonSpeciesName, fill=CommonSpeciesName),
   size=pointSize) +

   # Affects points with shaped symbols
   scale_fill_manual(values= PEMA$Color) +
   scale_colour_manual(values= PEMA$Color) +

   # add text
   geom_text(x= 1, y=yMax, label = PEMA.Text, size = textSize) +

   xlab("") +
   ylab("Estimated (N)") +

   expand_limits(ymin =0, ymax=yMax) +

   theme_bw() +
   theme(axis.line = element_line(colour = "black"),
       panel.grid.major = element_blank(),
       panel.grid.minor = element_blank(),
       panel.background = element_blank()) +

   theme(axis.title.x = element_blank(),
       axis.text.x = element_blank(),
       axis.title.y = element_text(size=20),
       axis.text.y= element_text(size=15)) +

   theme(legend.position = "none")

#############################
Treatments <- factor(MYGA$PlotID, levels=c("Hemlock", "Girdled", "Logged","Hardwood"))
MYGA.Text <- "southern red-backed voles"
MYGA.plot <- ggplot(MYGA, aes(x = Treatments, y = N, fill=CommonSpeciesName)) +
  geom_errorbar(aes(ymin=Lower95N, ymax=Upper95N, fill=CommonSpeciesName),
                width=.1) +
  geom_point(aes(color=CommonSpeciesName, fill=CommonSpeciesName),
             size=pointSize) +

  # Affects points with shaped symbols
  scale_fill_manual(values= MYGA$Color) +
  scale_colour_manual(values= MYGA$Color) +

  # add text
  geom_text(x = 1, y = 150, label = MYGA.Text, size = textSize) +

  xlab("") +
  ylab("Estimated (N)") +
  expand_limits(ymin =0, ymax=yMax) +

  theme_bw() +
  theme(axis.line = element_line(colour = "black"),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank()) +

  theme(axis.title.x = element_blank(),
        axis.text.x = element_blank(),
        axis.title.y = element_text(size=20),
        axis.text.y = element_text(size=15)) +

  theme(legend.position = "none")

#################################################################
Treatments <- factor(PELE$PlotID, levels=c("Hemlock", "Girdled",
                                       "Logged","Hardwood"))
PELE.Text <- "white-footed mice"

PELE.plot <- ggplot(PELE, aes(x = Treatments, y = N, fill=CommonSpeciesName)) +
  geom_errorbar(aes(ymin=Lower95N, ymax=Upper95N, fill=CommonSpeciesName),
                width=.1) +
  geom_point(aes(color=CommonSpeciesName, fill=CommonSpeciesName),
             size=pointSize) +

  # Affects points with shaped symbols
  scale_fill_manual(values= PELE$Color) +
  scale_colour_manual(values= PELE$Color) +
# add text
geom_text(x= 1, y= 150, label = PELE.Text, size = textSize) +
xlab(""") +
ylab("Estimated (N)") +
expand_limits(ymin =0, ymax=yMax) +
theme_bw() +
theme(axis.line = element_line(colour = "black"),
panel.grid.major = element_blank(),
panel.grid.minor = element_blank(),
panel.background = element_blank()) +
theme(axis.title.x = element_blank(),
axis.text.x = element_text(size = 20),
axis.title.y = element_text(size= 20),
axis.text.y= element_text(size=15)) +
theme(legend.position = "none")

grid.arrange(PEMA.plot, MYGA.plot, PELE.plot, ncol = 1)

PivotedPopulation = GeneralPivotPlotVsSpeciesFunction(MarkRecaptureResults, PlotNames, AllSpeciesNames, "N")
PivotedUpper95 = GeneralPivotPlotVsSpeciesFunction(MarkRecaptureResults, PlotNames, AllSpeciesNames, "Upper95N")
PivotedLower95 = GeneralPivotPlotVsSpeciesFunction(MarkRecaptureResults, PlotNames, AllSpeciesNames, "Lower95N")
View(PivotedPopulation)
APPENDIX I

Chapter three: Script for Multiseason Occupancy prep

library(reshape)
library(plyr)

getMultiSeasonVariableHistory <- function(surveyData, variableName, timeVariableName) {
  covariateHistoryColumns <- c("TrapUID", timeVariableName, variableName)
  covariateSubset <- subset(surveyData, select=covariateHistoryColumns)
  covariateHistory <- reshape(covariateSubset, direction = "wide", idvar = c("TrapUID"),
                             timevar = timeVariableName)

  return(covariateHistory)
}

getPreProcessedMasterData <- function(masterData, byColumns) {
  # adds a column called CaptureBin where 1 = capture and 0 = no capture
  masterDataEnhanced <- masterData
  masterDataEnhanced$Hx <- (masterDataEnhanced$Capture == "CAPTURE") * 1

  # adds a column for percDisturbed traps
  masterDataEnhanced$Disturbed <- (masterDataEnhanced$TrapStatus ==
                                  "DISTURBED") * 1

  numberOfDisturbedTrapsPerNight <- aggregate(Disturbed ~ Night + PlotID + BlockID + YearID,
                                              masterDataEnhanced, FUN=sum)
  numberOfDisturbedTrapsPerNight$SumOfDisturbed <- numberOfDisturbedTrapsPerNight$Disturbed
  numberOfDisturbedTrapsPerNight$Disturbed <- NULL

  numberOfTrapsPerNight <- aggregate(TrapUID ~ Night + PlotID + BlockID + YearID,
                                      masterDataEnhanced, FUN=length)
  numberOfTrapsPerNight$SumOfTraps <- numberOfTrapsPerNight$TrapUID
  numberOfTrapsPerNight$TrapUID <- NULL

  percentDisturbedDataFrame <- join(numberOfDisturbedTrapsPerNight,
                                     numberOfTrapsPerNight, by=c("Night", "PlotID", "BlockID", "YearID"), type="left",
                                     match="all")
  percentDisturbedDataFrame$PercDisturbed <- percentDisturbedDataFrame$SumOfDisturbed /
                                          percentDisturbedDataFrame$SumOfTraps
  percentDisturbedDataFrame$SumOfDisturbed <- NULL
  percentDisturbedDataFrame$SumOfTraps <- NULL
masterDataEnhanced <- join(masterDataEnhanced, percentDisturbedDataFrame, by=c("Night", "PlotID", "BlockID", "YearID"), type="left", match="all")

# subset by year and block filter; if the provided blockID is "" the all the blocks are returned
preProcessedMasterData <- subset(masterDataEnhanced, select=byColumns)

# add Night_Year identifier
preProcessedMasterData$NightYear <- paste(preProcessedMasterData$Year, preProcessedMasterData$Night, sep="_")

countNightsByYear <- aggregate(Night ~ YearID, preProcessedMasterData, FUN=max)
lastNight <- min(countNightsByYear$Night)

preProcessedMasterData <- subset(preProcessedMasterData, subset=(Night <= lastNight))

return(preProcessedMasterData)
}

definitionBaseDataFrame <- function(preProcessedMasterData) {
  # setup base merge dataframe with traps UIDs only
  sortedUniqueTrapUIDs <- sort(unique(preProcessedMasterData$TrapUID))
  baseDataFrame <- data.frame(TrapUID = as.character(sortedUniqueTrapUIDs), stringsAsFactors=FALSE)
  return(baseDataFrame)

}

definitionSpeciesHistoryData <- function(speciesSubset, nightsData){

  speciesHistoryData <- getMultiSeasonVariableHistory(speciesSubset, "Hx", "NightYear")

  ### Get the missing nights and adds the columns
  # get the nights in the species history just created
  existingFieldNames <- colnames(speciesHistoryData)
  # creates a filter by finding which columns do not exist by comparing with the
  # HistoryFieldNames in the nightsData
  columnFilter <- (match(nightsData$HistoryFieldName, existingFieldNames, nomatch = 0) == 0)
  # applies the filter to get the column names of the missing nights
  missingNights <- nightsData$HistoryFieldName[columnFilter]
  # adds the missing nights columns and sets them empty
  speciesHistoryData[missingNights] <- 0

# recreates the dataset with ordered columns
speciesHistoryData <- data.frame(TrapUID = 
as.character(speciesHistoryData$TrapUID),
speciesHistoryData[ nightsData$HistoryFieldName], stringsAsFactors=FALSE)

return(speciesHistoryData)
}

getYearlySiteCovariates <- function(preProcessedMasterData) {
  years <- sort(unique(preProcessedMasterData$YearID))
  T <- length(years)

  sitesSubset <- subset(preProcessedMasterData,
    select=c("TrapUID","NightYear","BlockID"))
  sitesAggregate <- aggregate(NightYear ~ TrapUID + BlockID, data=sitesSubset, length)

  n <- length(sitesAggregate$TrapUID)

  yearlySiteCovariates <- data.frame(matrix(rep(years, each=n), n, T))
  yearlySiteCovariates <- data.frame(lapply(yearlySiteCovariates, as.factor))

  names(yearlySiteCovariates) <- sub("X", "Year", names(yearlySiteCovariates))

  return(yearlySiteCovariates)
}

createMultiSeasonalOccupancyDataFrameBySpecies <- function(masterData, speciesID){
  byColumns <- c("YearID", "SpeciesID", "BlockID", "PlotID", "TrapUID","Night",
               "Sex", "Age", "RS", "PercDisturbed",
               "site.mast.sqm.oneYearLag", "site.mast.sqm.twoYearLag",
               "winter.air.avg", "winter.air.range", "winter.soil.avg", "winter.soil.range")

  preProcessedMasterData <- getPreProcessedMasterData(masterData, byColumns)

  # get yearly site covariates
  yearlySiteCovs <- getYearlySiteCovariates(preProcessedMasterData)

  oneYearLagSeedRainHistory <-
  getMultiSeasonVariableHistory(preProcessedMasterData, "site.mast.sqm.oneYearLag",
                                 "YearID")
  twoYearLagSeedRainHistory <-
getMultiSeasonVariableHistory(preProcessedMasterData, "site.mast.sqm.twoYearLag", "YearID")

winterAirAvgHistory <- getMultiSeasonVariableHistory(preProcessedMasterData, "winter.air.avg", "YearID")
winterAirRangeHistory <- getMultiSeasonVariableHistory(preProcessedMasterData, "winter.air.range", "YearID")
winterSoilAvgHistory <- getMultiSeasonVariableHistory(preProcessedMasterData, "winter.soil.avg", "YearID")
winterSoilRangeHistory <- getMultiSeasonVariableHistory(preProcessedMasterData, "winter.soil.range", "YearID")

# gets array of unique nights
sortedUniqueNights = sort(unique(preProcessedMasterData$NightYear))
nightsData <- data.frame(Night = as.character(sortedUniqueNights), stringsAsFactors = FALSE)
nightsData$HistoryFieldNames <- paste("Hx.", nightsData$Night, sep="")

baseDataFrame <- getBaseDataFrame(preProcessedMasterData)

# subsets by species ID
speciesSubset <- subset(preProcessedMasterData, SpeciesID %in% speciesID)
# create species history
speciesHistoryData <- getSpeciesHistoryData(speciesSubset, nightsData)

##### survey covariants subsets and histories
illuminationHistoryData <- getMultiSeasonVariableHistory(preProcessedMasterData, "Illumination", "NightYear")
skyHistoryData <- getMultiSeasonVariableHistory(preProcessedMasterData, "Sky", "NightYear")
airTemperatureHistoryData <- getMultiSeasonVariableHistory(preProcessedMasterData, "AvgAirTemp", "NightYear")
soilTemperatureHistoryData <- getMultiSeasonVariableHistory(preProcessedMasterData, "AvgSoilTemp", "NightYear")
percDisturbedHistoryData <- getMultiSeasonVariableHistory(preProcessedMasterData, "PercDisturbed", "NightYear")

# This is only to get unique TrapUID and BlockID columns together by just counting the nights
blockIDSubset <- subset(preProcessedMasterData, select=c("TrapUID", "NightYear", "BlockID"))
blockIDNamesSubset <- aggregate(NightYear ~ TrapUID + BlockID, data=blockIDSubset, length)
blockIDNamesSubset$NightYear <- NULL
# This is only to get unique TrapUID and PlotID columns together by just counting the nights
plotIDSubset <- subset(preProcessedMasterData, select=c("TrapUID", "NightYear", "PlotID"))
plotIDNamesSubset <- aggregate(NightYear ~ TrapUID + PlotID, data=plotIDSubset, length)
plotIDNamesSubset$NightYear <- NULL
plotIDNamesSubset$Surrounding <- ifelse(plotIDNamesSubset$PlotID == "Hardwood", "Hardwood", "Hemlock")

# merging base trap UIDs with survey and site covariants
# the all.x=TRUE parameter means that the ALL trap UIDs are preserved, even those with no captures on all nights
occupancyDataFrame <- join(baseDataFrame, speciesHistoryData, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, blockIDNamesSubset, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, plotIDNamesSubset, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, illuminationHistoryData, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, skyHistoryData, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, airTemperatureHistoryData, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, soilTemperatureHistoryData, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, percDisturbedHistoryData, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, oneYearLagSeedRainHistory, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, twoYearLagSeedRainHistory, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, winterAirAvgHistory, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, winterAirRangeHistory, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, winterSoilAvgHistory, by="TrapUID", type="left", match="all")
occupancyDataFrame <- join(occupancyDataFrame, winterSoilRangeHistory, by="TrapUID", type="left", match="all")
winterSoilRangeHistory, by="TrapUID", type="left", match="all")

occupancyDataFrame <- occupancyDataFrame[order(occupancyDataFrame$TrapUID),]
occupancyDataFrame[is.na(occupancyDataFrame)] <- 0

occupancyDataFrame <- cbind(occupancyDataFrame, yearlySiteCovs)

return(occupancyDataFrame)

getMultiSeasonalOccupancyDataFrameBySpecies <- function(masterData, microHabitat, seedRain, winterTemperatures, nightTemperatures, speciesID) {

  seedDataLagOneSubset <- subset(seedRain, select = c(BlockID, PlotID, EffectOneYearLag, site.mast.sqm))

  names(seedDataLagOneSubset)[names(seedDataLagOneSubset)=="EffectOneYearLag"] <- "YearID"
  names(seedDataLagOneSubset)[names(seedDataLagOneSubset)=="site.mast.sqm"] <- "site.mast.sqm.oneYearLag"

  seedDataLagTwoSubset <- subset(seedRain, select = c(BlockID, PlotID, EffectTwoYearLag, site.mast.sqm))

  names(seedDataLagTwoSubset)[names(seedDataLagTwoSubset)=="EffectTwoYearLag"] <- "YearID"
  names(seedDataLagTwoSubset)[names(seedDataLagTwoSubset)=="site.mast.sqm"] <- "site.mast.sqm.twoYearLag"

  winterTempSubset <- subset(winterTemperatures, select = c(YearID, BlockID, PlotID, winter.air.avg, winter.air.range, winter.soil.avg, winter.soil.range))

  nightTempSubset <- subset(nightTemperatures, select = c(YearID, BlockID, PlotID, Night, AvgAirTemp, AvgSoilTemp))

  mergedMasterData <- join(masterData, seedDataLagOneSubset, by=c("YearID", "BlockID", "PlotID"), type = "left", match="all")
  mergedMasterData <- join(mergedMasterData, seedDataLagTwoSubset, by=c("YearID", "BlockID", "PlotID"), type = "left", match="all")
  mergedMasterData <- join(mergedMasterData, winterTempSubset, by=c("YearID", "BlockID", "PlotID"), type = "left", match="all")
  mergedMasterData <- join(mergedMasterData, nightTempSubset, by=c("YearID", "BlockID", "PlotID", "Night"), type = "left", match="all")

  return(mergedMasterData)
}
occupancyBase <-
createMultiSeasonalOccupancyDataFrameBySpecies(mergedMasterData, speciesID)
occupancyDataFrame <- join(occupancyBase, microHabitat[,3:ncol(microHabitat)], by= "TrapUID", type = "left", match="all")
return(occupancyDataFrame)
}
APPENDIX J

Chapter three: Script for UNMARKED dataframe
# this script is used to generate the unmarked data frame used in the multiseason occupancy model
# file found in OUTPUT
#AD/CD

setwd("C:/Users/Ally/Documents/UVM/Projects/Rodent_Harvard/HF_Project/")
#setwd("~/ally/HF_Project/")

source("20-Analysis/R/functions/Functions_Occupancy_Multi_Season_v5.R") # rearragne data
source("20-Analysis/R/functions/multiplotFunction.R") #multiplot

library(unmarked)
library(reshape)
library(lattice)
library(Rcpp)
library(plyr)
library(ggplot2)

# load master data files (captures and microhabitat
masterData <- read.csv("10-PreProcessing/Dataset/Degrassi_HF_Master_v14.csv", stringsAsFactors = FALSE)
microHabitat <-read.csv("10-PreProcessing/Dataset/Microhabitat/Microhabitat_Master_v9.csv", stringsAsFactors = FALSE)
nightTemp <- read.csv("10-PreProcessing/Dataset/NightTemp/NightTemp_master.csv", stringsAsFactors = FALSE)
seedRain <- read.csv("10-PreProcessing/Dataset/SeedRain/Mast_Master_2010_2014.csv", stringsAsFactors = FALSE)
winterTemp <- read.csv("10-PreProcessing/Dataset/WinterTemp/Winter_Temp_master_v1.csv", stringsAsFactors = FALSE)

# Current masters have NO duplicates row for the double trapping. The animal caputers (n= 3, Rhe-B7a PEMA, VHe-A7a MYGA, VLo-A5a PEMA)
Species <- "MYGA"

MasterOccupancyDataFrame <- getMultiSeasonalOccupancyDataFrameBySpecies(masterData, microHabitat, seedRain, winterTemp, nightTemp, Species)

#Site history
HxColumns <- grep("Hx", colnames(MasterOccupancyDataFrame))
SurveyHx <- MasterOccupancyDataFrame[,HxColumns]

str(SurveyHx)
save(SurveyHx, file = "SurveyHx_MYGA.csv")

# yearly covariates
YearlyCovsColumns <- grep("Year\d", colnames(MasterOccupancyDataFrame))
YearlyCovs <- MasterOccupancyDataFrame[,YearlyCovsColumns]

WinAirAvgCovsColumns <- grep("winter.air.avg.",
colnames(MasterOccupancyDataFrame))

WinAirRangeCovsColumns <- grep("winter.air.range.",
colnames(MasterOccupancyDataFrame))

WinSoilAvgCovsColumns <- grep("winter.soil.avg.",
colnames(MasterOccupancyDataFrame))

WinSoilRangeCovsColumns <- grep("winter.soil.range.",
colnames(MasterOccupancyDataFrame))

MastOneYearLagCovsColumns <- grep("site.mast.sqm.oneYearLag.",
colnames(MasterOccupancyDataFrame))
Mast.OneYear.Lag.Covs <-
MasterOccupancyDataFrame[,MastOneYearLagCovsColumns]

MastTwoYearLagCovsColumns <- grep("site.mast.sqm.twoYearLag.",
colnames(MasterOccupancyDataFrame))
Mast.TwoYear.Lag.Covs <-
MasterOccupancyDataFrame[,MastTwoYearLagCovsColumns]

YearlySiteCovs <- list(Year = YearlyCovs, # detection
  Mast.OneYear.Lag = Mast.OneYear.Lag.Covs, # col
  Mast.TwoYear.Lag = Mast.TwoYear.Lag.Covs) # col

View(YearlySiteCovs)
# Now without standardized covariates
SiteCovs <- data.frame(Block= MasterOccupancyDataFrame$BlockID, #occ
                        Habitat = MasterOccupancyDataFrame$PlotID, #occ
                        Rock= MasterOccupancyDataFrame$Rock,
                        Soil= MasterOccupancyDataFrame$Soil,
                        Wood= MasterOccupancyDataFrame$Wood,
                        Litter= MasterOccupancyDataFrame$LeafLitter,
                        Fungi= MasterOccupancyDataFrame$Fungi,
                        Veg= MasterOccupancyDataFrame$Veg,
                        Open= MasterOccupancyDataFrame$Open,
                        High= MasterOccupancyDataFrame$High,
                        Low= MasterOccupancyDataFrame$Low,
                        Tree= MasterOccupancyDataFrame$TreeDistance_m,
                        Surounding = MasterOccupancyDataFrame$Surounding) #col

# Now observation covariates
illuminationColumns <- grep("Illumination.", colnames(MasterOccupancyDataFrame))
illuminationCovs <- MasterOccupancyDataFrame[,illuminationColumns]

skyColumns <- grep("Sky.", colnames(MasterOccupancyDataFrame))
skyCovs <- MasterOccupancyDataFrame[,skyColumns]

nightAirTempColumns <- grep("AvgAirTemp.",
colnames(MasterOccupancyDataFrame))
nightAirTempCovs <- MasterOccupancyDataFrame[,nightAirTempColumns]

nightSoilTempColumns <- grep("AvgSoilTemp.",
colnames(MasterOccupancyDataFrame))
nightSoilTempCovs <- MasterOccupancyDataFrame[,nightSoilTempColumns]

disturbedColumns <- grep("PercDisturbed.", colnames(MasterOccupancyDataFrame))
disturbedCovs <- MasterOccupancyDataFrame[,disturbedColumns]

ObsCovs <- list(illumination = illuminationCovs,
                sky = skyCovs,
                nightAirTemp = nightAirTempCovs,
                nightSoilTemp = nightSoilTempCovs,
                disturbed = disturbedCovs)

################
# unmarkedMultFrame ( y= Observed data matrix,
#     siteCovs = site covaraiates,
#     obsCovs= obseration covariates that vary within the site,
#     numPrimary = primary time periods or season in the multisean model,
#     yearlySiteCovs = convaraietate at the site year level)

Occ.umf <- unmarkedMultFrame(y = SurveyHx,
siteCovs = SiteCovs,
obsCovs = ObsCovs,
numPrimary = 3,
yearlySiteCovs = YearlySiteCovs)

save(Occ.umf, file = "30-
Output/R/UNMARKED/MultiSeason/Data_Files/MultiSeason_MYGA_v1.RData")
Chapter three: Script for Multiseason Occupancy Models

```r
setwd("C:/Users/Ally/Documents/UVM/Projects/Rodent_Harvard/HF_Project/"
load("30-
Output/R/UNMARKED/MultiSeason/Data_Files/MultiSeason_MYGA_v1.RData")

library(ggplot2)
library(unmarked)
library(reshape)
library(lattice)
library(Rcpp)
library(plyr)

#################################
# testing the detection first to include on all the models

(fm.test <- colext(psiformula = ~Block + Habitat,
    gammaformula = ~Mast.OneYear.Lag + Veg,
    epsilonformula = ~1,
    pformula = ~nightSoilTemp,
    data = Occ.umf))

# Model fitting
(fm1.Null <- colext(~1, ~1, ~1, ~1, Occ.umf))
backTransform(fm1.Null, type="psi")

#####
# model set to determine detection (p)

(fm.Illum <- colext(psiformula = ~Block + Habitat,
    gammaformula = ~Mast.OneYear.Lag + Veg + Surrounding,
    epsilonformula = ~Winter.Air.Avg + Winter.Air.Range,
    pformula = ~illumination,
    data = Occ.umf))

(fm.Cloud <- colext(psiformula = ~Block + Habitat,
    gammaformula = ~Mast.OneYear.Lag + Veg + Surrounding,
    epsilonformula = ~Winter.Air.Avg + Winter.Air.Range,
    pformula = ~sky,
    data = Occ.umf))
```
(fm.nightAir <- coext(pformula = ~Block + Habitat,
gammaformula = ~Mast.OneYear.Lag + Veg + Surrounding,
epsilonformula = ~Winter.Air.Avg + Winter.Air.Range,
pformula = ~nightAirTemp,
data = Occ.umf))

(fm.nightSoil <- coext(pformula = ~Block + Habitat,
gammaformula = ~Mast.OneYear.Lag + Veg + Surrounding,
epsilonformula = ~Winter.Air.Avg + Winter.Air.Range,
pformula = ~nightSoilTemp,
data = Occ.umf))

(fm.disturbed <- coext(pformula = ~Block + Habitat,
gammaformula = ~Mast.OneYear.Lag + Veg + Surrounding,
epsilonformula = ~Winter.Air.Avg + Winter.Air.Range,
pformula = ~disturbed,
data = Occ.umf))

(fm.pNull <- coext(pformula = ~Block + Habitat,
gammaformula = ~Mast.OneYear.Lag + Veg + Surrounding,
epsilonformula = ~Winter.Air.Avg + Winter.Air.Range,
pformula = ~1,
data = Occ.umf))

Fitted.p.MS <- fitList(
  "psi(.)c(.)e(.)p(.)" = fm1.Null,
  "psi(occ)c(col)e(ext)p(illumination)" = fm.Illum,
  "psi(occ)c(col)e(ext)p(sky)" = fm.Cloud,
  "psi(occ)c(col)e(ext)p(airTemp)" = fm.nightAir,
  "psi(occ)c(col)e(ext)p(soilTemp)" = fm.nightSoil,
  "psi(occ)c(col)e(ext)p(disturbed)" = fm.disturbed,
  "psi(occ)c(col)e(ext)p(.)" = fm.pNull)

###
# Rank them by AIC
(p.MS <- modSel(Fitted.p.MS))

# Do stuff
coef(p.MS)
# co-efficient of the Occ Ridge Model Selection
p.MS.ToExport <- as(p.MS, "data.frame")
View(p.MS.ToExport)
# Null model

(fm1.Null <- coext(~1, ~1, ~1, ~1, Occ.umf))
backTransform(fm1.Null, type="psi")

(fm2.Null2 <- coext(psiformula = ~Block + Habitat,
                    gammaformula = ~1,
                    epsilonformula = ~1,
                    pformula = ~nightSoilTemp,
                    data = Occ.umf)) #NAs produced

# model
# landscape  = block, habitat, surrounding habitat, winter, seed
# structure= within site = veg, wood, leaf litter

(fm3.Landscape1 <- coext(psiformula = ~Block + Habitat,
                          gammaformula = ~Habitat + Surrounding,
                          epsilonformula = ~Habitat + Surrounding,
                          pformula = ~nightSoilTemp,
                          data = Occ.umf))

(fm4.Landscape2 <- coext(psiformula = ~Block + Habitat,
                          gammaformula = ~Habitat + Surrounding,
                          epsilonformula = ~1,
                          pformula = ~nightSoilTemp,
                          data = Occ.umf))

(fm5.Landscape3 <- coext(psiformula = ~Block + Habitat,
                          gammaformula = ~1,
                          epsilonformula = ~Habitat + Surrounding,
                          pformula = ~nightSoilTemp,
                          data = Occ.umf))

(fm6.Landscape4 <- coext(psiformula = ~Block + Habitat,
                          gammaformula = ~Habitat,
                          epsilonformula = ~Habitat,
                          pformula = ~nightSoilTemp,
                          data = Occ.umf))

(fm7.Landscape5 <- coext(psiformula = ~Block + Habitat,
                          gammaformula = ~Surrounding,
epsilonformula = ~Surrounding,
pformula = ~nightSoilTemp,
data = Occ.umf)) #NA produced

(fm8.Landscape6 <- coext(psiformula = ~Block + Habitat,
gammaformula = ~Habitat,
epsilonformula = ~Surrounding,
pformula = ~nightSoilTemp,
data = Occ.umf))

(fm9.Landscape7 <- coext(psiformula = ~Block + Habitat,
gammaformula = ~Surrounding,
epsilonformula = ~Habitat,
pformula = ~nightSoilTemp,
data = Occ.umf))

#########
# Food models
(fm10.Food1 <- coext(psiformula = ~Block + Habitat,
gammaformula = ~Mast.OneYear.Lag + Veg,
epsilonformula = ~Mast.OneYear.Lag + Veg,
pformula = ~nightSoilTemp,
data = Occ.umf))

(fm11.Food2 <- coext(psiformula = ~Block + Habitat,
gammaformula = ~Mast.OneYear.Lag + Veg,
epsilonformula = ~1,
pformula = ~nightSoilTemp,
data = Occ.umf))

(fm12.Food3 <- coext(psiformula = ~Block + Habitat,
gammaformula = ~1,
epsilonformula = ~Mast.OneYear.Lag + Veg,
pformula = ~nightSoilTemp,
data = Occ.umf))

#########
# Winter

(fm13.Winter1 <- coext(psiformula = ~Block + Habitat,
gammaformula = ~Winter.Soil.Avg + Winter.Soil.Range,
epsilonformula = ~Winter.Soil.Avg + Winter.Soil.Range,
pformula = ~nightSoilTemp,
data = Occ.umf))

170
(fm14.Winter2 <- coext(psiformula = ~Block + Habitat,  
gammaformula = ~Winter.Soil.Avg + Winter.Soil.Range,  
epsilonformula = ~1,  
pformula = ~nightSoilTemp,  
data = Occ.umf))

(fm15.Winter3 <- coext(psiformula = ~Block + Habitat,  
gammaformula = ~Winter.Soil.Avg + Winter.Soil.Range,  
epsilonformula = ~1,  
pformula = ~nightSoilTemp,  
data = Occ.umf))

############################
#Structure

(fm16.SiteStructure1 <- coext(psiformula = ~Block + Habitat,  
gammaformula = ~Wood + Litter,  
epsilonformula = ~Wood + Litter,  
pformula = ~nightSoilTemp,  
data = Occ.umf))

(fm17.SiteStructure2 <- coext(psiformula = ~Block + Habitat,  
gammaformula = ~Wood + Litter,  
epsilonformula = ~1,  
pformula = ~nightSoilTemp,  
data = Occ.umf))

(fm18.SiteStructure3 <- coext(psiformula = ~Block + Habitat,  
gammaformula = ~1,  
epsilonformula = ~Wood + Litter,  
pformula = ~nightSoilTemp,  
data = Occ.umf)) #NA produced

############################
#Combinations

(fm19.LandscapeFood1 <- coext(psiformula = ~Block + Habitat,  
gammaformula = ~Habitat + Surrounding,  
epsilonformula = ~Mast.OneYear.Lag + Veg,  
pformula = ~nightSoilTemp,  
data = Occ.umf))

(fm20.LandscapeFood2 <- coext(psiformula = ~Block + Habitat,
gammaformula = ~Mast.OneYear.Lag + Veg,
epsilonformula = ~Habitat + Surrounding,
pformula = ~nightSoilTemp,
data = Occ.umf))

(fm21.LandscapeFood3 <- colext(psiformula = ~Block + Habitat,
gammaformula = ~Mast.OneYear.Lag + Veg,
epsilonformula = ~Habitat,
pformula = ~nightSoilTemp,
data = Occ.umf))

(fm22.LandscapeFood4 <- colext(psiformula = ~Block + Habitat,
gammaformula = ~Mast.OneYear.Lag + Veg,
epsilonformula = ~Surrounding,
pformula = ~nightSoilTemp,
data = Occ.umf))

####

(fm23.LandscapeWinter1 <- colext(psiformula = ~Block + Habitat,
gammaformula = ~Habitat + Surrounding,
epsilonformula = ~Winter.Soil.Avg + Winter.Soil.Range,
pformula = ~nightSoilTemp,
data = Occ.umf)) #NA produced

(fm24.LandscapeWinter2 <- colext(psiformula = ~Block + Habitat,
gammaformula = ~Winter.Soil.Avg + Winter.Soil.Range,
epsilonformula = ~Habitat + Surrounding,
pformula = ~nightSoilTemp,
data = Occ.umf))

(fm25.LandscapeWinter3 <- colext(psiformula = ~Block + Habitat,
gammaformula = ~Habitat,
epsilonformula = ~Winter.Soil.Avg + Winter.Soil.Range,
pformula = ~nightSoilTemp,
data = Occ.umf))

##

(fm26.LandscapeSite1 <- colext(psiformula = ~Block + Habitat,
gammaformula = ~Wood + Litter,
epsilonformula = ~Habitat + Surrounding,
pformula = ~nightSoilTemp,
data = Occ.umf))

(fm27.LandscapeSite2 <- colext(psiformula = ~Block + Habitat,
gammaformula = ~Habitat + Surrounding,
epsilonformula = ~Wood + Litter,
pformula = ~nightSoilTemp,
data = Occ.umf))

####

(fm28.FoodSite1 <- coext(piformula = ~Block + Habitat,
gammaformula = ~Mast.OneYear.Lag + Veg,
epsilonformula = ~Wood + Litter,
pformula = ~nightSoilTemp,
data = Occ.umf))

(fm29.FoodSite2 <- coext(piformula = ~Block + Habitat,
gammaformula = ~Wood + Litter,
epsilonformula = ~Mast.OneYear.Lag + Veg,
pformula = ~nightSoilTemp,
data = Occ.umf))

(fm30.FoodWinter1 <- coext(piformula = ~Block + Habitat,
gammaformula = ~Mast.OneYear.Lag + Veg,
epsilonformula = ~Winter.Soil.Avg + Winter.Soil.Range,
pformula = ~nightSoilTemp,
data = Occ.umf))

(fm31.FoodWinter2 <- coext(piformula = ~Block + Habitat,
 gammaformula = ~Winter.Soil.Avg + Winter.Soil.Range,
 epsilonformula = ~Mast.OneYear.Lag + Veg,
 pformula = ~nightSoilTemp,
 data = Occ.umf))

(fm32.FoodLandscape1 <- coext(piformula = ~Block + Habitat,
 gammaformula = ~Surrounding + Mast.OneYear.Lag + Veg,
 epsilonformula = ~1,
 pformula = ~nightSoilTemp,
 data = Occ.umf))

(fm33.FoodLandscape1 <- coext(piformula = ~Block + Habitat,
 gammaformula = ~1,
 epsilonformula = ~Surrounding + Mast.OneYear.Lag + Veg,
 pformula = ~nightSoilTemp,
 data = Occ.umf))

(fm34.FoodLandscape2 <- coext(piformula = ~Block + Habitat,
 gammaformula = ~Surrounding + Mast.OneYear.Lag + Veg,
 epsilonformula = ~Habitat,
pformula = ~nightSoilTemp,
  data = Occ.umf))

(fm35.FoodLandscape2 <- coext(psiformula = ~Block + Habitat,
  gammaformula = ~Habitat,
  epsilonformula = ~Surrounding + Mast.OneYear.Lag + Veg,
  pformula = ~nightSoilTemp,
  data = Occ.umf))

(fm36.LandFoodWinter1 <- coext(psiformula = ~Block + Habitat,
  gammaformula = ~Surrounding + Mast.OneYear.Lag + Veg,
  epsilonformula = ~Winter.Soil.Avg + Winter.Soil.Range,
  pformula = ~nightSoilTemp,
  data = Occ.umf))

(fm37.LandFoodWinter2 <- coext(psiformula = ~Block + Habitat,
  gammaformula = ~Surrounding + Mast.OneYear.Lag + Veg +
  Winter.Soil.Avg + Winter.Soil.Range,
  epsilonformula = ~Surrounding + Mast.OneYear.Lag + Veg +
  Winter.Soil.Avg + Winter.Soil.Range,
  pformula = ~nightSoilTemp,
  data = Occ.umf)) # did not converge high SE

(fm38.LandFoodWinterStructure <- coext(psiformula = ~Block + Habitat,
  gammaformula = ~Surrounding + Mast.OneYear.Lag + Veg +
  epsilonformula = ~Surrounding + Mast.OneYear.Lag + Veg +
  pformula = ~nightSoilTemp,
  data = Occ.umf)) #did not converge

Fitted.MS <- fitList(
  "psi(.)c(.)e(.)p(.)" = fm1.Null,
  "psi(B + H)c(Surrounding + Habitat)e(Surrounding + Habitat)p(soilTemp)" = fm3.Landscape1,
  "psi(B + H)c(Surrounding + Habitat)e(.).p(soilTemp)" = fm4.Landscape2,
  "psi(B + H)c(.).e(Surrounding + Habitat)p(soilTemp)" = fm5.Landscape3,
  "psi(B + H)c(Habitat)e(Habitat)p(soilTemp)" = fm6.Landscape4,
  "psi(B + H)c(Habitat)e(Surrounding)p(soilTemp)" = fm8.Landscape6,
  "psi(B + H)c(Surrounding)e(Habitat)p(soilTemp)" = fm9.Landscape7,
  "psi(B + H)c(Food)e(Food)p(soilTemp)" = fm10.Food1,
  "psi(B + H)c(Food)e(.)p(soilTemp)" = fm11.Food2,
  "psi(B + H)c(.).e(Food)p(soilTemp)" = fm12.Food3,
"psi(B + H)c(Winter)e(Winter)p(soilTemp)" = fm13.Winter1,
"psi(B + H)c(Winter)e(.)p(soilTemp)" = fm14.Winter2,
"psi(B + H)c(.)e(Winter)p(soilTemp)" = fm15.Winter3,
"psi(B + H)c(Structure)e(Structure)p(soilTemp)" = fm16.SiteStructure1,
"psi(B + H)c(Structure)e(.)p(soilTemp)" = fm17.SiteStructure2,
"psi(B + H)c(Landscape)e(Food)p(soilTemp)" = fm19.LandscapeFood1,
"psi(B + H)c(Food)e(Landscape)p(soilTemp)" = fm20.LandscapeFood2,
"psi(B + H)c(Food)e(Habitat)p(soilTemp)" = fm21.LandscapeFood3,
"psi(B + H)c(Food)e(Surrounding)p(soilTemp)" = fm22.LandscapeFood4,
"psi(B + H)c(Winter)e(Landscape)p(soilTemp)" = fm24.LandscapeWinter2,
"psi(B + H)c(Habitat)e(Winter)p(soilTemp)" = fm25.LandscapeWinter3,
"psi(B + H)c(Structure)e(Landscape)p(soilTemp)" = fm26.LandscapeSite1,
"psi(B + H)c(Landscape)e(Structure)p(soilTemp)" = fm27.LandscapeSite2,
"psi(B + H)c(Food)e(Structure)p(soilTemp)" = fm28.FoodSite1,
"psi(B + H)c(Structure)e(Food)p(soilTemp)" = fm29.FoodSite2,
"psi(B + H)c(Food)e(Winter)p(soilTemp)" = fm30.FoodWinter1,
"psi(B + H)c(Winter)e(Food)p(soilTemp)" = fm31.FoodWinter2,
"psi(B + H)c(Surrouinding + Food)e(.)p(soilTemp)" = fm32.FoodLandscape1,
"psi(B + H)c(.)e(Surrounding + Food)p(soilTemp)" = fm33.FoodLandscape1,
"psi(B + H)c(Surrouinding + Food)e(Habitat)p(soilTemp)" = fm34.FoodLandscape2,
"psi(B + H)c(Habitat)e(Surrounding + Food)p(soilTemp)" = fm35.FoodLandscape2,
"psi(B + H)c(Sourrounding + Food)e(Winter)p(soilTemp)" = fm36.LandFoodWinter1

###

# Rank them by AIC
(MS <- modSel(Fitted.MS))

# Do stuff
coef(MS)
# co-efficient of the Occ Ridge Model Selection
MS.ToExport <- as(MS, "data.frame")
View(MS.ToExport)

# for copy and paste into table in paper
View(MS.ToExport$model)
View(MS.ToExport$formula)
View(MS.ToExport$nPars)
View(MS.ToExport$AIC)
View(MS.ToExport$delta)
View(MS.ToExport$AICwt)
View(MS.ToExport.negLogLike)

# chi^2 Test

###

# chi^2 Test
chisq <- function(fm) {
  umf <- getData(fm)
  y <- getY(umf)
  sr <- fm@sitesRemoved
  if(length(sr)>0)
    y <- y[-sr,,drop=FALSE]
  fv <- fitted(fm, na.rm=TRUE)
  y[is.na(fv)] <- NA
  sum((y-fv)^2/(fv*(1-fv)))
}

#pb.gof <- parboot(fm.1, statistic=chisq, nsim=1000)
#plot(pb.gof)

#########################################################################
APPENDIX L

Chapter four: Script for citation analysis

# Foundation Species citation Meta-Analysis
# Product of LTER Working Group 2012
# Collaborative project with:
# Allyson Degrassi <ademgrass@uvm.edu>,
# Steven Brantley <sbrantle@umn.edu>,
# Robert Miller <miller@msi.ucsb.edu>,
# Carrie R Levine <crlevine@berkeley.edu>,
# Sydne Record <sydne.record@gmail.com>,
# Jacqueline Mohan <jmohan@uga.edu>,
# Aaron Ellison <aellison@fas.harvard.edu>

# Date: 30 October 2014 - 30 June 2015
# Primary: A. Degrassi

# MASTER SCRIPT
# no graphs
# no annotation (for annotations see specific script
# just quick runs

library(plyr) # to merge data frame

# Attach meta analysis pre-processing csv file
# file.choose()

FSMeta <- read.csv("C:\Users\Ally\Documents\UVM\Projects\FS Meta Analysis\10-PreProcessing\FSMeta_HF Archive_v2.csv")

# Find out literature types: Reviews, Commentary, Letters, ETC
FSNonPrimary <- aggregate(FSMeta$LiteratureType, by=list(FSMeta$LiteratureType), FUN=length)
names(FSNonPrimary)[names(FSNonPrimary)="Group.1"] <- "Literature Type"
names(FSNonPrimary)[names(FSNonPrimary)="x"] <- "Studies"
FSNonPrimary

# Select primary articles only
FSMetaPrimary <- subset(FSMeta, LiteratureType == "Primary")
nrow(FSMetaPrimary) #should be 331

######################################################################## Find out definitions from FSDefined
FSDefine <- aggregate(FSMetaPrimary$FSDefined, by=list(FSMetaPrimary$FSDefined), FUN=length)
names(FSDefine)[names(FSDefine)=="Group.1"] <- "Definition"
names(FSDefine)[names(FSDefine)=="x"] <- "Studies"

chi.test(FSDefine$Studies)

#percents
names(FSDefine)[names(FSDefine)=="V3"] <- "Percent"
FSDefine$Percent <- FSDefine$Studies/sum(FSDefine$Studies)
FSDefine

######################################################################## Find out FS Claim ########################################################################
FSClaim <- aggregate(FSMetaPrimary$FSClaim, by=list(FSMetaPrimary$FSClaim), FUN=length)
names(FSClaim)[names(FSClaim)=="Group.1"] <- "FSClaim"
names(FSClaim)[names(FSClaim)=="x"] <- "Studies"
chisq.test(FSClaim$Studies)

#percents
names(FSClaim)[names(FSClaim)=="V3"] <- "Percent"
FSClaim$Percent <- FSClaim$Studies/sum(FSDefine$Studies)
FSClaim

######################################################################## Find out strength of Influence ##############################
Strong <- subset(FSMetaPrimary, Strong == 1)
Moderate <- subset(FSMetaPrimary, Moderate == 1)
Marginal <- subset(FSMetaPrimary, Marginal == 1)
Strong <- aggregate(Strong$Strong, by=list(Strong$Strong), FUN=length)
Moderate <- aggregate(Moderate$Moderate, by=list(Moderate$Moderate), FUN=length)
Marginal <- aggregate(Marginal$Marginal, by=list(Marginal$Marginal), FUN=length)
Strong[,1] = "Strong"
Moderate[,1] = "Moderate"
Marginal[,1] = "Marginal"
Influence <- join_all(list(Strong, Moderate, Marginal), by = 'Group.1', type = 'full')
names(Influence)[names(Influence)=="Group.1"] <- "Strength"
names(Influence)[names(Influence)=="x"] <- "Studies"
Influence
chisq.test(Influence$Studies)

#percents
names(Influence)[names(Influence)=="V3"] <- "Percent"
Influence$Percent <- Influence$Studies/sum(Influence$Studies)
Influence

#########################################################################
# here filter for primary papers and papers that claimed to study foundation species
# so FS=YES
FSMetaPrimaryFoundation <- subset(FSMetaPrimary, FSClaim == "Foundation Species")
#View(FSMetaPrimaryFoundation)
nrow(FSMetaPrimaryFoundation)

#########################################################################
# Find out FS Role
FSRole <- aggregate(FSMetaPrimaryFoundation$FSRole,
by=list(FSMetaPrimaryFoundation$FSRole), FUN=length)
names(FSRole)[names(FSRole)=="Group.1"] <- "Role"
names(FSRole)[names(FSRole)=="x"] <- "Studies"

chisq.test(FSRole$Studies)

#percents
names(FSRole)[names(FSRole)=="V3"] <- "Percent"
FSRole$Percent <- FSRole$Studies/sum(FSRole$Studies)
FSRole

#########################################################################
# Find out Threat to FS

179
Climate <- subset(FSMetaPrimaryFoundation, ClimateChange == 1)
InvasiveSpecies <- subset(FSMetaPrimaryFoundation, InvasiveSpp == 1)
HabitatDeg <- subset(FSMetaPrimaryFoundation, HabitatDegradation == 1)
Exploitation <- subset(FSMetaPrimaryFoundation, Exploitation == 1)
DiseasePathogen <- subset(FSMetaPrimaryFoundation, DiseasePathogen == 1)
NoThreat <- subset(FSMetaPrimaryFoundation, NoThreat == 1)
Climate <- aggregate(Climate$ClimateChange, by=list(Climate$ClimateChange), FUN=length)
InvasiveSpecies <- aggregate(InvasiveSpecies$InvasiveSpp, by=list(InvasiveSpecies$InvasiveSpp), FUN=length)
HabitatDeg <- aggregate(HabitatDeg$HabitatDegradation, by=list(HabitatDeg$HabitatDegradation), FUN=length)
Exploitation <- aggregate(Exploitation$Exploitation, by=list(Exploitation$Exploitation), FUN=length)
DiseasePathogen <- aggregate(DiseasePathogen$DiseasePathogen, by=list(DiseasePathogen$DiseasePathogen), FUN=length)
NoThreat <- aggregate(NoThreat$NoThreat, by=list(NoThreat$NoThreat), FUN=length)
Climate[,1] = "Climate Change"
InvasiveSpecies[,1] = "Invasive Species"
HabitatDeg [,1] = "Habitat Degradation"
Exploitation [,1] = "Exploitation"
DiseasePathogen [,1] = "Disease or Pathogen"
NoThreat [,1] = "No Threat"
Threat <- join_all(list(Climate, InvasiveSpecies, HabitatDeg, Exploitation, DiseasePathogen, NoThreat), by = 'Group.1', type = 'full')
names(Threat)[names(Threat)=="Group.1"] <- "Threat"
names(Threat)[names(Threat)=="x"] <- "Studies"

chisq.test(Threat$Studies)

# percents
names(Threat)[names(Threat)=="V3"] <- "Percent"
Threat$Percent <- Threat$Studies/sum(Threat$Studies)
Threat

# analysis not used in the paper

FSJournalID <- aggregate(FSMetaPrimaryFoundation$JournalID, by=list(FSMetaPrimaryFoundation$JournalID), FUN=length)
names(FSJournalID)[names(FSJournalID)=="Group.1"] <- "Journal"
names(FSJournalID)[names(FSJournalID)=="x"] <- "Studies"
FSJournalID <- FSJournalID[order(-FSJournalID$Studies),]
FSJournalID

180
# MAP

```r
source("C:\Users\Ally\Documents\UVM\Projects\FS Meta Analysis\40-Software\R Source Scripts\projectFSfunctions_FSMap.R")
```

```r
library(maps)
library(plyr)
#library(mapproj)
library(rworldmap)
library(plotrix)
```

# make sure this is primary literature only still

```r
# select studies that did claim the organism was a FS
FSClaimYes <- subset(FSMeta, FSClaim == "Foundation Species")
ClaimYesCount <- nrow(FSClaimYes) # is the number of studies that did claim FS
ClaimYesCount
```

```r
ClaimYesMap <- subset(FSClaimYes, CountryID1 != "North and South America, Europe, Asia and New Zealand") # n = 1
ClaimYesMap <- subset(ClaimYesMap, CountryID1 != "global") # n = 1
ClaimYesMap <- subset(ClaimYesMap, CountryID1 != "Most of Europe") # n = 1

YesClaimMapAttributes1 <- aggregate(ClaimYesMap$CountryID1, by=list(ClaimYesMap$CountryID1), FUN=length)
YesClaimMapAttributes2 <- aggregate(ClaimYesMap$CountryID2, by=list(ClaimYesMap$CountryID2), FUN=length)
```

```r
# join the two data frames together
YesClaimMapAttributes <- join_all(list(YesClaimMapAttributes1, YesClaimMapAttributes2), by = 'Group.1', type = 'full')
```

# Second I wanted to change the names of the headings of the new table created
```r
names(YesClaimMapAttributes)[names(YesClaimMapAttributes)=="Group.1"] <- "Country"
names(YesClaimMapAttributes)[names(YesClaimMapAttributes)=="x"] <- "FSYesStudies"
```

```r
# select studies that did NOT claim the organism was a FS
FSClaimNo <- subset(FSMeta, FSClaim == "Not Foundation Species")
ClaimNoCount <- nrow(FSClaimNo) # number of studies that did not claim FS
```

```r
# out of 154 that defined FS only 115 claimed there species as a FS
```
ClaimNoMap <- subset(FSClaimNo, CountryID1 != "global" & CountryID1 != "Most of Europe" & CountryID1 != "NA")

NoClaimMapAttributes1 <- aggregate(ClaimNoMap$CountryID1, by=list(ClaimNoMap$CountryID1), FUN=length)
NoClaimMapAttributes2 <- aggregate(ClaimNoMap$CountryID2, by=list(ClaimNoMap$CountryID2), FUN=length)

# join the two data frames together
NoClaimMapAttributes <- join_all(list(NoClaimMapAttributes1, NoClaimMapAttributes2), by = 'Group.1', type = 'full')

# Second I wanted to change the names of the headings of the new table created
names(NoClaimMapAttributes)[names(NoClaimMapAttributes)=="Group.1"] <- "Country"
names(NoClaimMapAttributes)[names(NoClaimMapAttributes)=="x"] <- "FSNoStudies"

#View(NoClaimMapAttributes)

# join Yes map and No map together
ClaimMapAttributes <- join_all(list(YesClaimMapAttributes, NoClaimMapAttributes), by = 'Country', type = 'full')

# replace NA's with 0's
ClaimMapAttributes[is.na(ClaimMapAttributes)] <- 0
View(ClaimMapAttributes)

#################
#blank map dataset
blankmap <- getMap(resolution = "low")

mapAttributes <- data.frame(blankmap$NAME, blankmap$LON, blankmap$LAT)
names(mapAttributes) <- c("Country", "Lon", "Lat")
ClaimMapAttributes <- join_all(list(ClaimMapAttributes, mapAttributes), by = 'Country', type = 'left')

FSYesColor <- "limegreen"
FSNoColor <- "blue"
TextSize <- 1.45
plot(blankmap)
plotPieCharts(ClaimMapAttributes, 2, 10, FSYesColor, FSNoColor)

legend(x = -30, y = -30,
      legend= c("FS Studied", "FS Not Studied"),
      cex = TextSize,
      col = c("white", "white"),
      pt.cex = cex,
      box.lty = 0)

floating.pie(-25, -38, legendSlice, radius = 3, col = FSYesColor)
floating.pie(-25, -46, legendSlice, radius = 3, col = FSNoColor)

legend(x = -180, y = 35,
       legend= c("# Studies"),
       cex = TextSize,
       pt.cex = cex,
       box.lty = 0)

legendSlice <- c(1)
legendColor <- "grey51"

floating.pie(legendLat, -65, legendSlice, radius = 10, col = legendColor)
legend(legendLat, -57, legend = "223", cex = TextSize, pt.cex = cex, box.lty = 0, bg = "transparent")

floating.pie(legendLat, -43, legendSlice, radius = 8, col = legendColor)
legend(legendLat, -35, legend = "13", cex = TextSize, pt.cex = cex, box.lty = 0, bg = "transparent")
floating.pie(legendLat, -25, legendSlice, radius = 7, col = legendColor)
legend(legendLat+2, -17, legend = "8", cex = TextSize, pt.cex = cex, box.lty = 0, bg = "transparent")

floating.pie(legendLat, -8, legendSlice, radius = 6, col = legendColor)
legend(legendLat+2, 1.5, legend = "3", cex = TextSize, pt.cex = cex, box.lty = 0, bg = "transparent")

floating.pie(legendLat, 5, legendSlice, radius = 4, col = legendColor)
legend(legendLat+1.5, 14, legend = "2", cex = TextSize, pt.cex = cex, box.lty = 0, bg = "transparent")

floating.pie(legendLat, 15, legendSlice, radius = 2, col = legendColor)
legend(legendLat, 25, legend = "1", cex = TextSize, pt.cex = cex, box.lty = 0, bg = "transparent")

######
dev.off()