ABSTRACT

BIGSBY, KEVIN M. Anthropogenic Drivers of Gypsy Moth Dispersal. (Under the direction of Dr. Erin O. Sills.)

Gypsy moth (*Lymantria dispar:* Linnaeus) is a polyphagous non-native forest pest first introduced to Medford, Massachusetts in 1869. It has since spread as far south as North Carolina and as far west as Wisconsin. Gypsy moth is responsible for defoliating approximately 100,000 hectares of forest annually, resulting in mortality in a small percentage of trees, averting behavior by recreators, and creating a nuisance to the general public. Limiting the spread of gypsy moth is beneficial because it delays the onset of costs and losses associated with quarantine, tree defoliation and mortality, and nuisance. Gypsy moth is believed to disperse naturally up to 2.5 km/yr (e.g. early instar ballooning) but has been observed to disperse much greater distances. The scientific consensus is that this longer distance dispersal occurs through anthropogenic vectors (e.g. egg masses being transported on firewood). Despite the resources that United States Department of Agriculture and state agencies dedicate to eliminating and managing new infestations resulting from long distance dispersal, there has been limited empirical research on the relationship between the dispersal of gypsy moth and the movement of people and their goods. This thesis develops a conceptual framework of the anthropogenic factors that could affect dispersal, measures these factors with secondary data at the county level from a variety of sources, and estimates the presence or absence of gypsy moth using logistic regression models. The dependent variable is drawn from trap catch records archived by the United States Department of Agriculture Forest Service program, Slow-the-Spread, in areas distal to the 1 moth/trap line between 1999 and 2007. Through step-wise logistic regression estimating sub-

models that include variables representing each broad anthropogenic factor, a final empirical model is specified. The variables of the model are estimated independently for each year from 1999 to 2007, resulting in a mean Pseudo R square of 0.568. Consistently significant ($\alpha < 0.05$) anthropogenic variables are the number of households using wood for heating fuel and mean household income. These findings are discussed with regard to invasion theory and quarantine policy. One key implication is the continual importance of regulating and raising awareness about the risk of moving firewood from infested to uninfested zones.

Anthropogenic Drivers of Gypsy Moth Dispersal

by
Kevin Michael Bigsby

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Forestry

Raleigh, North Carolina

2009

APPROVED BY:

Dr. H.M. Cheshire

Dr. F.P. Hain

Dr. E.O. Sills
Chair of Advisory Committee

Dr. P.C. Tobin

BIOGRAPHY

B.A. Anthropology and Geography, University of Colorado, 2004

B.S. Forestry, North Carolina State University, 2009

ACKNOWLEDGMENTS

Thanks to Dr. Heather Cheshire, Dr. Fred Hain, Dr. Patrick Tobin, and Dr. Erin Sills for their help teaching me the art of science.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 - INTRODUCTION

1.1 Introduction to gypsy moth

Gypsy moth (*Lymantria dispar:* Linnaeus) was first introduced to Medford, MA in 1869 (Liebhold et al. 1989) and is now established in 19 states (Code of Federal Regulations 2008). The gypsy moth caterpillar is capable of feeding on more than 300 deciduous and coniferous tree species in North America (Elkinton and Liebhold 1990). Significant resources are invested in management of this species because first, it imposes very significant out-of-pocket costs as well as losses in quality of life (Tuthill et al. 1984, Jakus et al. 1994, Leuschner et al. 1996, Pimentel et al. 2005), and second, it only occupies one-third of its potential range in North America (Morin et al. 2005). Public management efforts can be divided into three categories: 1) eradication in the uninfested region (led by the United States Department of Agriculture Animal and Plant Health Inspection Service), 2) outbreak suppression in the infested region (led by states in cooperation with United States Department of Agriculture Forest Service); and, 3) slowing the spread in the transition region (led by United States Department of Agriculture Forest Service) (Gypsy Moth Program Manual 2004, Tobin et al. 2004). This analysis focuses on the transition region where management limits the successful establishment of newly arrived gypsy moth colonies, delaying the onset of costs and losses in the uninfested region.

Mack et al. (2000) noted that ongoing invasions must be assessed to gain a better understanding of policy alternatives and the science of invasions. The sections that follow provide background on the ongoing gypsy moth invasion, providing context for my research objective and approach, which are outlined in the final section of this chapter.

1.2 Slow-the-Spread

The Appalachian Integrated Pest Management Program began in 1987 (Reardon et al. 1991). The goal of this project was to study the feasibility of barrier zone management for gypsy moth (Sharov and Liebhold 1998). This approach had been suggested much earlier at a 1922 conference in Albany, NY (Gerardi and Grimm 1979). However, the initial program was unsuccessful, as it did not utilize a floating or migrating barrier zone that is sensitive to fluctuations in local populations. The Appalachian Integrated Pest Management Program, however, did utilize a moving barrier zone and was able to answer a number of important questions about project feasibility nationwide. For example, the Appalachian Integrated Pest Management Program established an optimal treatment regime to reduce spread to the target rate of 10 km/yr (Sharov and Liebhold 1998). The target rate was established based on a study by Leuschner et al. (1996) documenting the net losses that would occur at varying spread rates due to discounting of the impacts on the residential, recreation, and timber sectors in each county expected to eventually become infested with the gypsy moth. Leuschner et al. (1996) concluded that Slow-the-Spread would save between \$3.8, \$2.3, or \$0.8 billion in costs and losses over a 25 year period under greatest, medium, and least benefit scenarios. In all scenarios the residential impacts are substantially greater than the impacts on the timber or recreation sector, comprising 83% of all losses.

In 1999, Slow-the-Spread was initiated with the goal of reducing gypsy moths spread rate to 10 km/yr, half the mean spread rate between 1966 and 1990 (Liebhold et al. 1992). Slow-the-Spread operations take place in the transition zone where Slow-the-Spread works with states to implement a systematic trapping program. Based on the trap catch data, the boundary of the generally infested zone and transition zone is defined by traps that generally catch 10 moths, or the 10 moth/trap line (For a detailed explanation of the Slow-the-Spread program see Tobin et al. 2004). The generally infested zone is defined as the region behind the 10 moth/trap line in which counties are generally in quarantine. The transition area, for operational purposes, generally extends up to 170 km ahead of the 10 moth/trap line since most new colonies arise within this area (Sharov and Liebhold 1998).

In the transition area approximately 100,000 pheromone-baited traps are placed at varying grid intervals to establish gypsy moth population densities within the transition area. Slow-the-Spread traps are placed on grid intervals between 500 m and 8 km. Wider trapping intervals in the evaluation area of the transition zone are used to derive population boundary estimates. Tighter trapping intervals in the distal portion of the transition zone, known as the action area, are used to delineate isolated populations. The primary management objective in the action area is to detect and prevent isolated populations of gypsy moth from coalescing. Evaluations of the Slow-the-Spread program suggest that this has reduced spread rates from 20.8

km/yr between 1966 and 1990 (Liebhold et al. 1992) to 5.52 km/yr between 1999 and 2007 (Tobin et al. 2007a).

In addition to records from Slow-the-Spread, trap records from nearby states (e.g. Tennessee, Iowa) are included in the Slow-the-Spread database. These states may use some combination of a gridded trapping protocol and "priority traps." Priority traps are placed in locations believed to have a higher probability of gypsy moth, for example rest areas or campgrounds.

1.4 Quarantine

Counties have been quarantined for gypsy moth dating back to 1900 (Figure 1.1, Appendix 1). The states and the Animal and Plant Health Inspection Service generally declare quarantine in counties that the 10 moth/ trap line intersects, and thus the quarantined counties are often used as a proxy for the generally infested area. Quarantine requires the inspections of, and permits for the movement of objects known to harbor gypsy moth egg masses.

The Animal and Plant Health Inspection Service specifies that the following articles are regulated (Code of Federal Regulations 2008). This already suggests some key anthropogenic factors in dispersion.

> *(1) Associated equipment - Articles associated and moved with mobile homes and recreational vehicles, such as, but not limited to, awnings, tents, outdoor furniture, trailer blocks, and trailer skirts.*

(2) Bark - The tough outer covering of the woody stems of trees, shrubs, and other woody plants as distinguished from the cambium and inner wood.

(3) Bark product - Products containing pieces of bark including bark chips, bark nuggets, bark mulch, and bark compost.

(4) Mobile home - Any vehicle, other than a recreational vehicle, designed to serve, when parked, as a dwelling or place of business.

(5) Move (movement, moved)- Shipped, offered for shipment to a common carrier, received for transportation or transported by a common carrier, or carried, transported, moved, or allowed to be moved by any means. ''Movement'' and ''moved'' shall be construed in accordance with this definition.

(6) Outdoor household articles - Articles associated with a household that have been kept outside the home such as awnings, barbecue grills, bicycles, boats, dog houses, firewood, garden tools, hauling trailers, outdoor furniture and toys, recreational vehicles and associated equipment, and tents.

(7) Recreational vehicles - Highway vehicles, including pickup truck campers, one-piece motor homes, and travel trailers, designed to serve as temporary places of dwelling.

(8) Regulated articles - (1) Trees without roots (e.g., Christmas trees), trees with roots, and shrubs with roots and persistent woody stems, unless they are greenhouse grown throughout the year. (2) Logs, pulpwood, and bark and bark products. (3) Mobile homes and associated equipment. (4) Any other products, articles, or means of conveyance, of any character whatsoever, when it is determined by an inspector that any life stage of gypsy moth is in proximity to such articles and the articles present a high risk of artificial spread of gypsy moth infestation and the person in possession thereof has been so notified (Plant Protection Act 7 U.S.C 7711, 7712, 7714, 7751, and 7754).

1.5 Gypsy moth spread

Gypsy moth disperses in a stratified manner in which both short (i.e. neighborhood or local) and long distance (often anthropogenically influenced) dispersal occur simultaneously (Hengeveld 1989, Shigesada et al. 1995). Stratified dispersal has been observed and quantified in many invasive species (e.g. Mack 1985, Hengeveld 1989, Okubo 1989, Andow et al. 1990). In the case of gypsy moth, founder colonies can be the result of unintentional introduction by people. Liebhold et al. (1992) estimated the natural diffusion of gypsy moth to be 2.5 km/yr based on larval dispersal research by Mason and McManus (1981). They estimated much faster historical spread rates, ranging from 2.82 km/yr to 20.78 km/yr in different time period. This suggested stratified dispersal involving accidental long range introductions. Tobin and Blackburn (2008) also discuss such long distance dispersal in the case of the initial gypsy moth invasion of Wisconsin.

Sharov et al. (1995, 1996, 1997, 1998, 1999), Tobin et al. (2007a, 2007b), Tobin (2007c) and Liebhold et al. (1992, 2006) demonstrate that gypsy moth spread is dynamic and depends on numerous parameters. For example, stochasticity, the Allee effect, forest susceptibility and, winter temperature have all been suggested as determinants of spread rates. While many researchers have noted that anthropogenic factors probably have a significant effect on the dispersal of gypsy moth (McFadden and McManus 1991, Liebhold et al. 1992, Sharov 2004), Lippitt et al. (2008) is the only previous study that empirically models anthropogenic factors in the spread of gypsy moth. Lippitt et al. (2008) model the probability of gypsy moth being introduced and established in counties in the uninfested area using biotic, abiotic, and anthropogenic variables. To the best of my knowledge, no previous research has included both biophysical and anthropogenic factors in models of gypsy moth dispersal in the transition area.

1.6 Objective and approach

The objective of this research is to assess the role of anthropogenic factors in the dispersal of gypsy moth in a sample of counties up to700 km from the generally infested area. This “expanded transition area” (or equivalently, the expanded transition *zone* or *region*) is defined as any county distal to the 1 moth/trap line that has at least one gypsy moth trap in a given year. I first develop a conceptual framework identifying anthropogenic factors (e.g. sectors regulated by quarantine) and biophysical factors (e.g. susceptible basal area and distance to infestation) expected to affect the spread of gypsy moth. Second, I assess the spatial patterns in mean trap counts and presence/absence of gypsy moth in counties in the expanded transition area, focusing on the degree of spatial autocorrelation (which would violate the assumption that data are independently distributed). Third, I conduct exploratory data analysis using step-wise logistic regression to identify the best variables to represent each anthopogenic factor. Finally, based on both the conceptual framework and the exploratory data analysis, I specify and estimate logistic models of the presence/absence of gypsy moth in each county in the expanded transition area in each year from 1999 and 2007 logistic regression.

1.7 References

Andow, D., Kareiva, P., Levin, S.A., and Okubo, A. (1990). Spread of invading organisms. *Landscape Ecology.* 4 p177-188

Code of Federal Regulations. Title 7, Agriculture, Part 301- Domestic quarantine notices. 73 FR 32432, June 9, 2008

Elkinton and Liebhold (1990). Population dynamics of gypsy moth in North America. *Annual review of entomology*, 35, p571-596.

Gerardi, M.H. and Grimm, J.K. (1979). The history, biology, damage, and control of the gypsy moth (Porthetria Dispar (L.). Associated University Press. p230.

Gypsy Moth Program Manual. 2004. United States Department of Agriculture, Marketing and Regulatory Programs, Animal and Plant Health Inspection Service, Plant Protection and Quarantine.

Hengeveld, B. (1989). *Dynamics of biological invasions*. Chapman and Hall, London. p160.

Jackus, P.M. (1994). Averting Behavior in the Presence of Public Spillovers: Household Control of Nuisance Pests. *Land Economics*, 70;3 p273-285.

Leuschner, W.A., Young, J.A., and Ravlin, W.F. (1996). Potential Benefits of Slowing the Gypsy Moth's Spread. *Society of American Foresters*, 20;2 p65-73.

Liebhold, A. M., Mastro, V., and Schaefer, P. W. (1989). Learning from the legacy of Leopold Trouvelot. *Bull. Entomol. Soc. Am.* 35, p20–21.

Liebhold, A.M., Halverson, J.A., and Elmes, G.A. (1992). Gypsy moth invasion in North America: a quantitative analysis. *Journal of Biogeography*(1992). 19;5, p513-520.

Liebhold, A.M. and Tobin, P.C. (2006). Growth of newly established alien populations: Comparison of North American gypsy moth colonies with invasion theory. *Population Ecology.* 48;4, p253-262.

Lippitt C.D., Rogan, J., Toledano, J., Sangermano, F., Eastman, R.J., Mastro, V., Sawyer, A. (2008). Incorporating anthropogenic variables into species distribution model to map gypsy moth risk. *Ecological Modelling*. 210;3, p339-350.

Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, H., Clout, M., Bazzaz, F.A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*. 10;3, p689-710.

Mason, C.J., and McManus, M.L. (1981). Larval dispersal of the gypsy moth. *The gypsy moth: research toward an integrated pest management*. Editors: Doane, C.C. and McManus, M.L. *United States Department of Agriculture Forest Service Technical Bull*. p161-202.

McFadden, M.W. and McManus, M.E. (1991). An insect out of control? The potential for spread and establishment of the gypsy moth in new forest areas in the United States, pp. 172-186. In: Y.N. Baranchikov, W.J. Mattson, F.P. Hain, and T.L. Payne (eds.), Forest insect guilds: patterns of interaction with host trees. U.S. For. Serv. Gen. Tech. Rep. NE-153. p172-186

Morin, R.S., Liebhold, A.M., Luzader, E.R., Lister, A.J., Gottschalk, K.W., Twardus, D.B. (2005). Mapping Host-Species Abundance of Three Major Exotic Forest Pests United States Department of Agriculture Research Paper NE-726, United States Department of Agriculture Forest Service, Newtown Square, Pennsylvania.

Okubo, A., Maini, P.K., Williamson, M.H., and Murray, J.D. (1989). On the spatial spread of the gray squirrel in Britain. *Proceedings of the Royal Society of London B, Biological Sciences*. 238;1291 p113-125

Pimentel, D., Zuniga, R., and Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, 52;3, p273-288.

Reardon, R.C. (1991). Appalachian gypsy-moth integrated pest-management project. *Forest Ecology and Management*. 39;1-4, p107-112.

Sharov, A.A. Roberts, E.A., and Liebhold, A.M. (1995). Gypsy moth (Lepidoptera: Lymantriidae) spread in the central Appalachians: Three methods for species boundary estimation. *Environmental Entomology*. 24;6, p1529-1538.

Sharov, A.A. Liebhold, A.M., and Roberts, E.A. (1996). Spread of gypsy moth (Lepidoptera: Lymantriidae) in the central Appalachians: Comparison of population boundaries obtained from male moth capture, egg mass counts, and defoliation records. *Environmental Entomology*. 25;4, p783-792.

Sharov, A.A., Liebhold, A.M., and Roberts, E.A. (1997). Methods for monitoring the spread of gypsy moth (Lepidoptera: Lymantriidae) populations in the Appalachian mountains. *Journal of Economic Entomology* . 90;5, p1259-1266.

Sharov, A.A. and Liebhold, A.M. (1998). Model of slowing the spread of gypsy moth (Lepidoptera: lymantridae) with barrier zones. *Ecological Applications,* 8;4, p1170-1179.

Sharov, A.A., Pjanowski, B.C., Liebhold, A.M., and Gage, S.H. (1999). What affects the rate of gypsy moth (Lepodoptera: Lymantriidae) spread: winter temperature or forest susceptibility? *Agricultural and Forest Entomology*, 1, p37-45.

Sharov, A.A. (2004). Bioeconomics of Managing the Spread of Exotic Pest Species with Barrier Zones. *Risk Analysis*, 24;4, p879-892.

Tobin, P.C., Sharov, A.A., Liebhold, A.A., Leonard, D.S., Roberts, E.A., and Learn, M.R. (2004). Management of the Gypsy Moth through a Decision Algorithm under the STS Project. *American Entomologist*, 50:4, p 200-209.

Tobin, P.C., Liebhold, A.M., and Roberts, E.A. (2007a). Comparison of methods for estimating the spread of a non-indigenous species. *Journal of Biogeography* 34, p305-312.

Tobin, P.C., Whitmire, S.L., Johnson., D.M. Bjornstad, O.N., and Liebhold, A.M. (2007b). Invasion speed is affected by geographical variation in the strength of Allee effects. *Ecology Letters.* 10, p36-43.

Tobin, P.C. and Blackburn, L.M. (2008). Long-Distance Dispersal of the Gypsy Moth (Lepidoptera: Lymantriidae) Facilitated its Initial Invasion of Wisconsin. *Environmental Entomology.* 37;1, 87-93

Tobin, P.C. (2007c). Space-time patterns during the during the establishment of non-indigenous species. *Population Ecology*. 49, p257-263.

Tuthill, R.W., Canada, A.T., Wilcock, K., Etkind, P.H., O'Dell, T.M., & Shama, S.K. (1984). An Epidemiologic Study of Gypsy Moth Rash, *American Journal of Public Health*, 74;8, p799- 803.

Appendix 1

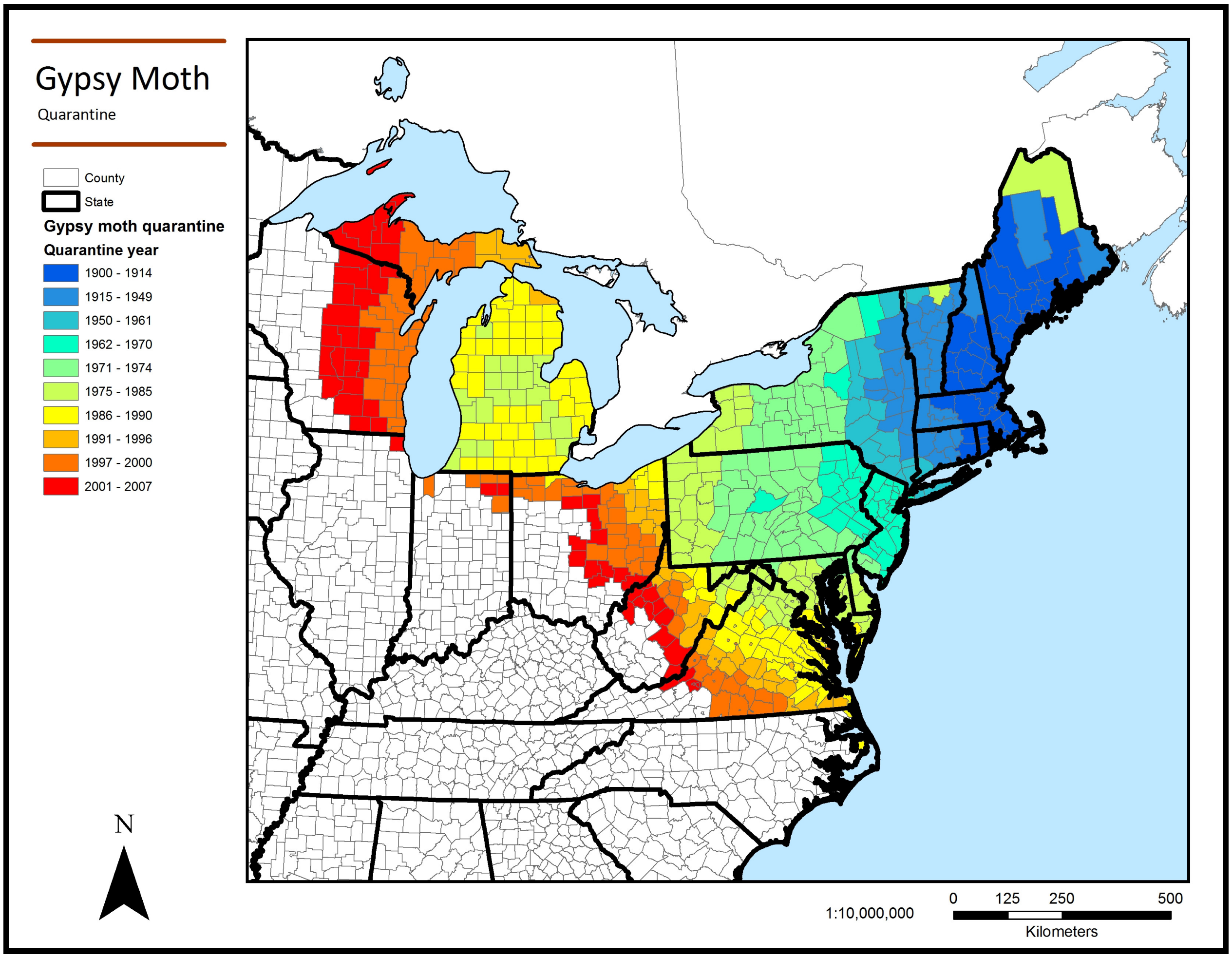


Figure 1.1 – Gypsy moth quarantine from 1900 - 2007.

CHAPTER 2 – LITERATURE REVIEW

2.1 Introduction to the literature review

The spread of invasive species has been studied and modeled dating back to Fisher's (1937) and Skellam's (1951) research on the dispersal of living organisms. The first theoretical models considered the dispersal of organisms stochastically as a function of their population density and fecundity (Skellam 1951). Over the years other research built upon this basic idea to include additional variables and apply new statistical methods (Hastings et al. 2005). More recent models includes measures of evolution (Ellstrand and Schierbeck 2000), mutualism (Clergeau and Mandon-Dalger 2001), spatial heterogeneity (Jules et al. 2002), and temporal and spatial variability of the invaded environment (Davis et al. 2000). However, most attempts to model the spread of gypsy moth have not explicitly included one of the most significant contributors to the dispersal of invasive species: people. The literature that has addressed human mediated dispersal of non-native forest pests is reviewed in section 2.2. Then I review the literature on spread of gypsy moth in section 2.3 first presenting the concept of spread via stratified dispersal (section 2.4). Sections 2.5 and 2.6 present how gypsy moth trap catch data is derived and applied. Finally, section 2.7 introduces the gravity model, including the origin of the concept and previous applications to the spread of invasive species in aquatic and forest ecosystems. Section 2.8 concludes this chapter by presenting the summary chapter summary.

2.2 Anthropogenic introduction and spread of non-native species

The invasion process is defined by at least three distinct processes: 1) introduction, 2) establishment, and 3) spread. Introduction occurs when a non-native species first arrives in a novel environment. Spread occurs after the species has successfully established and begins to disperse into new regions. The introduction rate of non-native species into novel environments has increased over the past 200 years (di Castri 1989). This increase is empirically correlated with the increased movement of people and goods (di Castri 1989, Mack et al. 2000). Introductions also tend to follow the flow of transportation and commodity routes. A positive correlation has been established between the direction of transport routes and the arrival of invasive species (Lockwood et al. 2006, Meyerson and Mooney 2007, Westphal et al. 2008). For example, islands tend to receive more invasive species than mainlands because, in general, there are more imports to islands than exports from islands (Lockwood et al. 2006). Invasive species use ship ballast water, crop seeds, domestic animals, and nursery stock as vectors for transport into novel environments (Sakai et al. 2001). Dunnage and other solid wood packaging materials are also known to transport wood boring insects such as the emerald ash-borer (*Agrilus planipennis: Fairmaire*) and bark beetles (Family: *Curculionidae*) (e.g., Brockerhoff et al. 2006). The Asian gypsy moth (*Lymantria dispar asiatica:* Vnukovskij) has been found multiple times in North American, often detected in close proximity to ports and successfully eradicated (Animal and Plant Health Inspection Service 2006). Other non-native species have been intentionally imported and then accidentally (e.g., the gypsy moth) or intentionally (e.g., the Asian lady beetle[1] - *Harmonia axyridis: Pallas*) released.

Once established in a novel environment the spread of a non-native species is hypothesized to correlate with local transportation patterns, hitch-hiking on vectors that are often species dependent. Invasive species interceptions usually occur on high traffic roads on country borders and around shipping ports (Work et al. 2004, Brockerhoff et al, 2006). Though there is little quantitative evidence linking the spread of non-native forest pests to anthropogenic pathways there is plenty of anecdotal evidence and examples of non-native forest pests being transported on cars or firewood (Code of Federal Regulations 2008, McFadden and McManus 1991). In order to examine this process I turn to the literature on the spread of invasive species that examines anthropogenic influences or suggests anthropogenic influences impact species dispersal.

The zebra mussel (*Dreissena polymorpha*: Pallas) is a well studied aquatic species that was introduced to the Great Lakes in the mid 1980s. The mussel's range expansion has been directly linked to anthropogenic factors. Previous research established the probability of zebra mussel becoming introduced into an uninfested lake as a function of boater traffic, distance traveled by boaters, and the attractiveness of a lake as a boating destination (Leung 2004 and 2006, Bossenbroeck et al. 2001, Schneider et al. 1998).

---

[1] First introduced as early as 1916 to control pest insects, the Asian lady beetle was reintroduced in 1980s to control aphid populations on pecan plantations in the Southeast. It was not until 1988 that the Asian lady beetle was documented as being established near New Orleans, Louisana. It has since spread to other regions of the United States where it continues to be beneficial to farmers but is viewed as pest to many because of its uncanny ability to overwinter inside buildings (United States Department of Agriculture 2009).

Sudden oak death (*Phytophthora ramorum)* is a fungus that causes tree dieback in infected trees and has a current range of California and southern Oregon. Meentmeyer et al. (2008) modeled the susceptible habitat for sudden oak death based on a suite of biotic and abiotic variables. Though his models performed quite well, he suggested that human movement of infected nursery stock and other material could increase the likelihood of the long-distance dispersal of sudden oak death, hindering the model's ability to account for stratified dispersal.

In the case of the emerald ash borer (*Agrilus planipennis:* Fairmaire), a non-native forest pest originally reported in the Detroit area in the summer of 2002 (Poland and McCullough 2006), dispersal is often linked to the movement of firewood, nursery stock, or other wood products (BenDor et al. 2006, Muirhead et al. 2006). The results of BenDor's (2006) model simulations indicate that a quarantine on firewood would effectively slow the spread of emerald ash borer. Muirhead et al. (2006) also found that human population was positively correlated with the probability of emerald ash borer being detected at the sub-county level. Including population density in Muirhead et al. (2006) model also significantly improved the model fit.

The horse chestnut leaf miner (*Cameraria ohridella:* Deshka & Dimic) is a non-native forest pest that was introduced to Macedonia in 1985 and has since invaded Central and Western Europe (Gilbert et al. 2004). This rapid range expansion suggests that anthropogenic movement greatly influenced its dispersal. Gilbert et al. (2004) found that human population density is positively correlated with a higher probability of detecting the horse chestnut leaf miner.

In the case of gypsy moth, it is widely believed that egg masses are often unintentionally transported by people (McFadded and McMannus 1991, Liebhold et al. 1992, Sharov 2004).

Egg masses are normally found in dark areas inside crevices (e.g. under loose bark) on the trunks of highly susceptible host trees (Leonard 1981). However, during outbreaks, egg masses may be laid in many unconventional sites. Egg masses have been spotted on unsusceptible tree species, cars, rocks, outdoor playground equipment, grills, and nursery stock (Animal and Plant Health Inspection Service 2006). Gypsy moth can spread by the movement of any infested item from the generally infested area to the uninfested area. The egg mass is the diapause stage (i.e. dormant) of the gypsy moth life cycle. This stage occurs for approximately eight months (Leonard 1968), increasing the temporal likelihood that this life stage is transported and gypsy moth's ability to survive transport, and increased fecundity after introduction[2].

2.3 Gypsy moth spread: an overview of empirical analysis

Empirical analyses of the spread of gypsy moth can be categorized into two general methodologies using two different types of data (Tobin et al. 2007a). The first method uses county level presence/absence data; the second method makes use of more detailed data on trap catch density to estimate the displacement in yearly population density boundaries. Other researchers (Liebhold et al. 1992; 2006, Johnson et al. 2006, Sharov et al. 1995; 1996; 1997) have also used these methodologies to empirically measure the spread of gypsy moth.

[2] Egg masses can survive winter temperatures as low as -28° C and each egg mass may contain between 75-450 larvae, thus increasing reproduction success after introduction when compared to moving a single larvae or moth that would be without a reproductive partner (McManus and Zerillo 1978).

2.4 Spread via stratified dispersal

Range expansion of invasive species is dominated by stratified dispersal in which both short (i.e. neighborhood or local) and long distance dispersal occur congruently (Shigesada et al. 1995). Stratified dispersal deviates from Skellam's (1951) and Fisher's (1937) proposal of linear dispersal, which is dependent on the population of the invading species and an assumed Fickian diffusion (see Holmes et al. 1994). Instead stratified dispersal suggests invasive species disperse non-linearly, or in a stratified manner, dominated by local and long distance dispersal (Hengeveld 1989). Stratified dispersal has been observed in many invasive species. Shigesada et al. (1995) reviews the application of stratified dispersal to invasive species. Shigesada et al. (1995) first document that Skellam's model was successful when applied to the North American muskrat (*Ondatra zibethica: Linnaeus*) (Skellam 1951) and the sea otter (*Enhydra Lutris: Linnaeus*) (Lubina and Levin 1988). However, other studies (e.g. Hengeveld 1989, Okubo 1989, Andow et al. 1990, Mack 1985) found that this model is not applicable because other species disperse by establishing founder colonies ahead (e.g. long distance dispersal) of their established range. In the case of gypsy moth and other invasive insects founder colonies can be the result of unintentional introduction by people.

Stratified dispersal in gypsy moth occurs through short distance dispersal (e.g. natural spread 2.5km/yr) and long distance dispersal (e.g. anthropogenically influenced dispersal). Liebhold et al. (1992) estimated the natural diffusion of gypsy moth to be 2.5 km/yr. They found this be curious as they estimated much greater spread rates for gypsy moth historically (2.82

km/yr, 7.61 km/yr, and 20.78 km/yr) and because organisms generally spread at constant radial velocities (Levin, 1989). Liebhold et al. (1992) suggest that models should account for habitat heterogeneity and more complex population growth models. Tobin and Blackburn (2008) found that long distance dispersal facilitated the initial gypsy moth invasion of Wisconsin, contributing to the literature supporting stratified dispersal.

2.5 Gypsy moth county level presence/absence data

County level presence/absence data for gypsy moth provide the longest continuous record documenting the expanding range of any non-native forest pest in the USA. Despite the desirable temporal extent of these data, applications for ecological analysis are limited by the spatial resolution of the data (i.e. gypsy moth population levels cannot be determined to make inferences about finer scale ecological processes). County level presence/absence data can be compiled from multiple sources: 1) gypsy moth quarantine records maintained by the United States Department of Agriculture (Code of Federal Regulations, Title 7, Chapter III, Section 301.45); 2) cooperative agreement between the United States Department of Agriculture Animal and Plant Health Inspection Service and Cooperative Agriculture Pest Survey; and, 3) historic Animal and Plant Health Inspection Service eradication records dating back to 1967. These data sources cover the entire United States.

Gypsy moth quarantine data from the United States Department of Agriculture are available from 1934 to present for the generally infested zone. Between 1900 and 1934 data are available from other published sources (Anon; 1907a, 1907b, Burgess; 1913,1915, 1930; Liebhold et al. 1992). Though quarantine criteria have varied over time, it is generally accepted that one or more life stages had to be sighted multiple times for a county to enter quarantine (Liebhold et al. 1992). County quarantine can be used as a variable for each county and each year, indicating whether gypsy moth was present or absent. Once a county enters quarantine (i.e. gypsy moth is present), it remains in quarantine to comply with the Domestic Plant Quarantine Act of 1912 (Weber 1930). Quarantine data could also be used to construct a variable indicating the year that gypsy moth arrived in every county in the generally infested zone by using the year quarantined.

Quarantine data have been used for ordinary least-squares regression models of the historical spread rate of gypsy moth (Liebhold et al. 1992, Tobin et al. 2007a). For example, Liebhold et al. (1992) determined three unique spread rates over three time periods (Table 2.1, Appendix 2, modified from Tobin et al. 2007a). They measured spread by defining a fixed point, such as the introduction epicenter of gypsy moth, and measuring the distance between the fixed point and the quarantined county. This distance is then regressed against the year a county was quarantined. Tobin et al. (2007a) developed a modified quarantine boundary displacement approach to measure spread. This approach also uses the introduction epicenter and the distal boundary of quarantined counties. The boundary displacement approach, however, uses transects radiating from the introduction epicenter to measure the distance between the boundary around quarantined counties in *year a,* and the boundary around quarantined counties in *year a+1.*

$$X = \frac{(\sum a) - (\sum a + 1)}{t} \qquad \text{[Equation 2.1]}$$

These differences are then averaged, based on the number of transects, to establish an annual spread distance, X. Results from Tobin et al. (2007a) parallel those of Liebhold et al. (1992) but include more recent years and a regional gradient (Table 2.1, Appendix 2, from Tobin et al. 2007a).

Animal and Plant Health Inspection Service eradication data are available dating back to 1967, while the Animal and Plant Health Inspection Service and Cooperative Agriculture Pest Survey pest survey data are available from 1986. Presence data can be derived from both sources by defining gypsy moth as present in a county in a given year if eradication was attempted or if at least one moth was caught in a trap. Gypsy moth can become established multiple times in the same county in the uninfested zone. This occurs if gypsy moth is introduced, eradicated, and then re-introduced. As a result, there is no equivalent to "year introduced" in counties in the uninfested area. The Cooperative Agriculture Pest Survey was created with the goal of protecting American agriculture from foreign pest introduction and establishment. This data has been used in various applications. For example, Lippitt et al. (2008) used the Cooperative Agriculture Pest Survey data to model the risk of gypsy moth being introduced to and establishing in counties in the uninfested zone. Douce et al. (1992) used Cooperative Agriculture Pest Survey data to generate a published report in the early years of the project, documenting the survey results.

2.6 Gypsy moth trap catch data

Gypsy moth trap catch data provide greater spatial resolution and identify finer gradations of infestation than the count level presence/absence data. This allows inferences to be made about finer scale ecological processes and development of different methods for measuring the spread of gypsy moth. These data, however, are limited temporally and by the spatial extent of the trapping area.

Trap catch data were first used to decide whether or not to quarantine counties on the edge of the generally infested area (Liebhold et al. 1992). However, the only information available from these original traps is the annual quarantine status of each county. More recently, trap catch data have been employed to study the ecology of gypsy moth (Liebhold and Bascompte 2003, Tobin et al. 2007b; 2007c, Whitmire and Tobin 2006), including their spread (Sharov et al., 1995; 1996; 1997, 2002 Tobin et al. 2007a).

Determining the spread rate of gypsy moth was necessary to evaluate the effectiveness of the Appalachian Integrated Pest Management Gypsy Moth Demonstration Project (Reardon 1991) and determine the feasibility of Slow-the-Spread (Sharov et al. 1995, 1996, 1997, 2004). The Appalachian Integrated Pest Management Program used a 2-3km gridded trap surface to monitor gypsy moth populations.

Sharov et al. (1995, 1997) introduced methodologies for defining gypsy moth population boundaries and three methods for monitoring spread rates. Sharov et al. (1995) first differentiated “irregular” and “regular” boundaries. The irregular boundary contains complexities such as islands (i.e. an isolated colony of gypsy moth) or gaps (i.e. areas where gypsy moth is absent in the generally infested area). The regular boundary is defined as one that is perpendicular to the general boundary direction and crosses the boundary only once (i.e. a continuous line without islands or gaps). Trap catch records were kriged based on an omni-directional semivariogram to create a lattice surface of gypsy moth populations. The lattice surface was used with three different methods to estimate population boundaries: 1) best cell classification, 2) first occurrence, and 3) regression. These methods yielded similar boundaries and annual spread rates (Sharov et al. 1995).

The best cell classification method was used by Sharov et al. (1997) to refine the methodology for monitoring the spread of gypsy moth. This method minimizes the number of grid cells that are misclassified and is used to generate population boundaries along which traps can be expected to catch 1, 3, 10, 30, 100, and 300 moths. The residuals from an ordinary least-squares regression of boundary points with time show that the 10 moth/trap line yields the lowest variability and the 1 moth/trap line has the highest variability (Sharov et al. 1997). Because of its low variability, the 10 moth/trap line is used to define the boundary between the uninfested and generally infested areas. It is also the primary moth boundary used to determine the boundaries of Slow-the-Spread and is one of the boundaries used to measure spread. Sharov et al. (1997) hypothesized that the 1 moth/trap and 3 moth/trap lines have the highest variability because they represent migrant gypsy moth populations (i.e. populations that are explicitly isolated from the generally infested area) from wind-borne dispersal. While it remains uncertain if wind-borne dispersal is the primary vector of migrant populations (e.g. Tobin and Blackburn

2008), it has become clear that the establishment of successful migrant populations through stratified dispersal aid in the spread of gypsy moth (Liebhold et al. 1992, Sharov and Liebhold 1998).

The population boundary displacement approach for modeling spread with trap catch data, as described by Tobin et al. (2007a), is similar to the boundary displacement approach for quarantine data as described previously. In the population boundary displacement approach, the annual spread distance is the displacement between a fixed point in space to the population boundary (as defined by Sharov et al. 1997) in *year a* and the displacement in *year a+1*. Fixed transects radiating from the fixed point are used for measuring the average of (*year a)* – (*year a+1)* (Tobin et al, 2007a). Historic spread rates based on the population boundary method are presented in Table 2.1, Appendix 2.

Trap catch data allow the detection of isolated colonies of gypsy moth ahead of the generally infested zone. Sharov and Liebhold (1998) modeled the spread of gypsy moth in the transition zone as a function of the distance of isolated colonies from the generally infested zone and the population of these colonies. The probability of isolated colonies establishing ahead of the generally infested zone decreases as distance to the generally infested zone increases. Whitmire and Tobin (2006) modeled the persistence of isolated colonies in the transition zone. They found that ecological variables, such as the density of preferred host species and land-use, were not statistically significant determinants of colony persistence. Rather, there were strong regional differences in colony persistence, which they attributed to regional variation in the strength of the Allee effect. The Allee effect is defined as a positive relationship between any component of fitness of a species and either numbers or density of conspecifics[3] (Stephens et al. 1999).

Isolated colonies of gypsy moth are periodically found throughout North America. Lippitt (2008) mapped the risk (i.e. the probability of introduction and establishment) of gypsy moth in counties in the uninfested zone as a function of biotic, abiotic, and anthropogenic variables. The dependent variable was presence/absence of gypsy moth, presence being defined as 5 or more moths captured in a county and absence as zero moths captured. The statistically significant anthropogenic variables in the preferred model specification (household movement from quarantined counties, percent of population emigrated from defoliated counties, population density, and road accessibility) provide further evidence for the hypothesis suggested by McFadden and McManus (1991), Liebhold et al. (1992), and Sharov (2004) that the movement people and our goods (e.g. cars, firewood, nursery plants, etc.) are responsible for recent increases in gypsy moth spread rates. However, determining the most likely vector (e.g. is gypsy moth more likely to be transported on nursery stock, cars, or firewood?) or sector of industry in quarantine to move gypsy moth egg masses is still unknown.

---

[3] The Allee effect is named after Warder Clyde Allee who noticed that in smaller populations the reproduction and survival of individuals decreased while in larger populations the reproduction and survival of individuals increased. One mechanism of the Allee effect is the failure of finding a mate in sparse populations (Stephenson et al 1999).

2.7 Gravity models

The spread of invasive species can be studied with models that are statistical or mathematical in nature. This section considers how statistical and mathematical gravity models are derived, including the underlying assumptions for each type of model.

The gravity model can be used to predict the movement of people or goods between two places (e.g., in international trade, Thomas and Huggett 1980). Gravity models can be used to specify regression models or serve as the basis for mathematical simulations (e.g. production constrained or attraction constrained models), as long as the distance between any two locations significantly affects behavior.

The mathematical gravity models originate from Newton's Law of Universal Gravitation:

$$F_{ij} = G\frac{M_i M_j}{D_{ij}^2}$$

[Equation 2.2]

where F_{ij} is the attractive force between object *i* and object *j*; M_i and M_j are the masses of the respective objects; D_{ij} is the distance between the two objects; and *G* is the gravitational constant.

The gravity model, when applied to the movement of people or goods, takes the same form, but some of its parameters are defined differently:

$$F_{ij} = G\frac{M_i^{\alpha}\ M_j^{\beta}}{D_{ij}^{\theta}}$$

[Equation 2.3]

where F_{ij} and D_{ij} are the same as above; and M_i^{α} and M_j^{β} are the respective economic sizes (usually measured with the gross domestic product, gross national income, or population) of locations *i* and *j*.

This conceptual gravity model has been applied broadly in the social sciences and is beginning to gain ground in other fields. Specifically, gravity models have been used to model the dispersal of invasive species by replacing gross domestic product with the attractiveness of the destination (e.g. water quality of a popular boating lake). Next, we consider applications to dispersal of invasive species, both aquatic (Schneider et al. 1998, Bossenbroeck et al. 2001, Muirhead and MacIsaac 2005, Leung et al. 2006) and non-native forest pests (Gilbert et al. 2004, Muirhead et al. 2006).

Schneider et al. (1998) used a mathematical gravity model to assess the spread of zebra mussel in Illinois. The model takes on a similar form to those presented above:

$$T_{ij} = \frac{(A_i O_i B_j D_j)}{(c_{ij})^{\alpha}} \qquad \text{[Equation 2.4]}$$

where T_{ij} is the relative flow of boats between origin and destination, A_i and B_j are scaling constants ensuring that all boats arriving at a point also leave that point, O_i is the total number of boats leaving the origin D_j, c_{ij} is the distance between the origin and destination and α is a power function related to the relative likelihood of travel over short and long distances. The model was used in a predictive manner to quantify priority areas for awareness and conservation.

Bossenbroeck et al. (2001) argues that gravity models are useful when the attractiveness between a source and a destination is important (e.g. zebra mussels are more likely to be introduced to more popular lakes). In contrast, a diffusion model is effective when the movement rate of an organism is known or it does not disperse in a stratified manned. Bossenbroeck et al. (2001) parameterized a gravity model based on the probability of a boat traveling to a zebra mussel source and then traveling to an uninfested lake and the probability of zebra mussel becoming established at an uninfested lake. The gravity model was compared to observed zebra mussel infestations and performed well, successfully predicting the invasion of a majority of lakes and forecasting future spread.

Muirhead and MacIsaac (2005) used a gravity model to predict the dispersal of the non-indigenous spiny waterflea to inland lakes of Ontario. The objective of their research was to determine the lakes that would become hubs for the spiny waterflea. The authors used survey data from interviewed boaters at various boat-launches to determine lake use preferences and quantify the number of people using specific lakes. They were able to establish an invasion network, linking well trafficked lakes (called hubs, that are likely to be the source of invasions) to smaller lakes as a result of boater traffic from the hubs.

Leung et al. (2006) hypothesized that the introduction of zebra mussel into an unifested lake is a function of the distance traveled to the lake, the proportion of boaters using the lake, the proportion of boaters using a source lake (i.e. a lake where zebra mussel is present), and the boat traffic to a specific lake. They developed a mechanistic mathematical model as their *a priori* hypothesis based on the above metrics and compared their model, using least squares regression, to a model based on the results of a boater survey that empirically documented these metrics. The mechanistic model was able to explain 80% of the variation in this process. The authors note that the importance of this model is its simplicity and limited data demands (i.e. it does not rely on surveys but rather uses easily obtained data). They suggest future work could strive to increase the model fit by introducing model parameters that better suit the ecological process.

Gilbert et al. (2004) modeled the spread of the horse chestnut leaf miner (*Cameraria ohridella:* Deschka & Dimic) using three models. Their stratified dispersal model used a type of statistical gravity model. They used logistic regression in the stratified dispersal model to predict the probability of the horse chestnut leaf miner occurring in a given grid cell in a given year. They found that including human population density in the model improved the model fit, increasing the R square value from 0.379 to 0.468. They also concluded that the fat-tailed dispersal model performed better than the Gaussian model indicating the importance of stratified dispersal in the spread of invasive species. This conclusion is also supported by many other researchers years before (e.g. Liebhold et al. 1992, Sharov and Liebhold 1998)

Muirhead et al. (2006) predicted the long distance dispersal of emerald ash borer (*Agrilus planipennis*: Fairmaire) using two types of gravity models: a logistic regression and a mathematical simulation gravity model. For the logistic regression they tested two models: a spatial “null” specification, using only the distance from sub-counties to the point of introduction and a specification that included both distance and sub-county population. The gravity model was parameterized based on firewood interceptions in Ontario’s state parks. The surveillance program gathered information on the origin and destination of the firewood and allowed

Muirhead et al. (2006) to create a model based on firewood interceptions, the number of camp sites, and the shortest road distance from the parks to the invasion epicenter. The destination of firewood from the gravity model correlated with the observed detection data, although this was only documented with maps in the publication. The logistic regression that included the human population at the sub-county level increased the model fit over the null model, correctly classifying 64.6% of sites that were invaded and 99.4% of sites that were not invaded.

2.8 Conclusion/Summary

Much of the literature reviewed found anthropogenic factors to be significant (Schneider 1998, Bossenbroeck 2001, Gilbert et al. 2004, Leung 2004;2006, Muirhead and MacIsaac 2005, BenDor 2006, Lippitt 2008), or hypothesized to be significant (McFadden and McManus 1991, Liebhold et al. 1992, Sharov 2004). With the exception of Lippitt et al. (2008), there are no published results empirically linking anthropogenic factors to the spread of gypsy moth. Lippitt's (2008) research focused on mapping the risk of introduction and establishment of gypsy moth in the uninfested zone, excluding counties in the transition area. This thesis examines the dispersal of gypsy moth in the transition zone where both natural dispersal and anthropogenic transport of gypsy moth are likely to play important roles in its spread.

2.9 References

Animal and Plant Health Inspection Service (2006). The Gypsy Moth Program Manual. United States Department of Agriculture.

Andow, D., Kareiva, P., Levin, S.A., and Okubo, A. (1990). Spread of invading organisms. *Landscape Ecology*. 4 p177-188.

Anon. (1907a) *Second annual report of the superintendent for suppressing the gypsy and brown-tail moths*. Massachusetts Public Document Number 73. Wright and Potter, Boston, MA.

Anon. (1907b) *Report of the commissioner for the suppression of gypsy and brown-tail moths*. State of Rhode Island and Providence Plantations. Freeman Co., Providence, RI.

BenDor, T.K., Metcalf, S.S., Fontenot, L.E., Sangunett, B., and Hannon, B. (2006). Modeling the spread of the Emerald Ash Borer. *Ecological Modelling* 197, p221-236.

Bossenbroek, J.M., Kraft, C.E., and Nekola, J.C. (2001). Prediction of long-distance dispersal using gravity models: Zebra mussel invasion of inland lakes. *Ecological Applications*. 11;6, p1778-1788.

Brockerhoff, E.G., Bain, J., Kimberley, M., and Knizek, M. (2006). Interception frequency of exotic bark and ambrosia beetles (Coleoptera: Scolytinae) and relationship with establishment in New Zealand and worldwide. *Canadian Journal of Forest Resources,* 36, p289-298.

Burgess A.F. (1913) *The dispersion of the gipsy [sic] moth*. Bulletin Number 119. US Department of Agriculture, Washington, DC.

Burgess, A.F. (1915) *Report on the gipsy [sic] moth work in New England*. Bulletin Number 204. US Department of Agriculture, Washington, DC.

Burgess, A.F. (1930) *The gipsy [sic] moth and the brown-tail moth*. Farmers' Bulletin 1623 US Department of Agriculture, Washington, DC.

Clergeau, P. and Mandon-Dalger, I. (2001). Fast colonization of an introduced bird: the case of *Pycnonotus jocosus* on the Mascarene Islands. *Biotropica*, 33;3, p542-546.

Code of Federal Regulations. Title 7, Agriculture, Part 301- Domestic quarantine notices. 73 FR 32432, June 9, 2008.

Davis, M.A., Grime, J.P., and Thompson, K. (2000). Fluctuating resources in plant communities: a general theory of invasibility. *J. Ecol.*, 88;3, p528-534.

di Castri (1989). *History of biological invasions with emphasis on the Old World.* Drake, F. di Castri, R. Groves, F. Kruger, H.A. Mooney, M. Rejmanek, and Williamson, M., editors. Biological invasions: a global perspective. Wiley, New York, New York, USA. p.1-30.

Douce, G.K., Hamilton, R.D., Clement, G.L. (1992). Gypsy-moth programs in the southeast. *Journal of entomological science*. 29;3, p381-397.

Ellstrand, N.C. and Schierenbeck, K.A. (2000). Hybridization as a stimulus for the evolution of invasiveness in plants? *Proc. Natl Acad. Sci.* 97;13, p7043-7050.

Fisher, R.A. (1937). The wave of advance of advantageous genes. *Ann. Eugenics*, 7, p355-369.

Gilbert, M., Gregoire, J.-C., Freise, J.F., and Heitland, W. (2004). Long-distance dispersal and human population density allow the prediction of invasive patterns in the horse chestnut leafminer Cameraria ohridella. *Journal of Animal Ecology*. 73;3, p459-468.

Hastings, A, Cuddington, K., Davies, K.F., Dugam, C.J., Elmendorf, S., Freestone, A., Harrison, S., Holland, M., Lambrinos, J., Malvadkar, U., Melbourne, B.A., Moore, K., Taylor, C., and Thompson, D. (2005). The spatial spread of invasions: new developments in theory and evidence. *Ecology letters*. 8;1, p91-101.

Hengeveld, B. (1989). *Dynamics of biological invasions*. Chapman and Hall, London. p160.

Holmes, E.E., Lewis, M.A., Banks, J.E., and Veit, R.R. (1994). Partial differential equations in ecology: spatial interactions and population dynamics. *Ecology* 75:1 p17-29.

Johnson, D.M., Liebhold, A.M., Tobin, P.C., and Bjørnstad, O.N. (2006). Allee effects and pulsed invasion by the gypsy moth. *Nature.* 444, p361-363.

Jules, E.S., Kauffman, M.J., Ritts, W.D., and Carroll, A.L. (2002). Spread of an invasive pathogen over a variable landscape: a nonnative root rot on Port Orford cedar. *Ecology*, 83;11, p 3167-3181.

Leonard, D.E. (1968). Diapause in the Gypsy Moth. *Journal of Economic Entomology*. 61;3, p596-&

Leonard, D.E. (1981). *Bioecology of the gypsy moth*. Doance, C.C. and McManus, M.L., editors. The gypsy moth: research toward integrated management. U.S. Department of Agriculture, Washington D.C. p 9-30.

Leung, B., J.M. Drake & D.M. Lodge. (2004). Predicting Invasions: Propagule Pressure and the Gravity of Allee Effects. *Ecology,* 85;6, p1651-1660.

Leung, B. Bossenbroek, J.M., and Lodge, D.M. (2006). Boats, pathways, and aquatic biological invasions: Estimating dispersal potential with gravity models. *Biological Invasions*. 8; 2, p241-254.

Lubina, J.A. and Levin, S.A. (1988). The spread of a reinvading species: range expansion in the California sea otter. *American Naturalist*. 131;4 p526-543.

Levin, S.A. (1989). Analysis of risk for invasions and control programs. *Biological invasions: a global perspective* (editors Drake, J.A., Mooney, H.A., di Castri, F., Groves, R.H., Kruger, F.J., Rejmanek, and Williamson, M.) SCOPE 37. Wiley, Chichester p425-435.

Liebhold, A.M., Halverson, J.A., and Elmes, G.A. (1992). Gypsy moth invasion in North America: a quantitative analysis. *Journal of Biogeography*. 19;5, p513-520.

Liebhold A.M. and Bascompte J. (2003). The Allee effect, stochastic dynamics and the eradication of alien species. *Ecology Letters*. 6;2, p133-140.

Lippitt C.D., Rogan, J., Toledano, J., Sangermano, F., Eastman, R.J., Mastro, V., Sawyer, A. (2008). Incorporating anthropogenic variables into species distribution model to map gypsy moth risk. *Ecological Modelling*. 210;3, p339-350.

Lockwood, J.L., Hoopes, M.F., and Marchetti, M.P. (2006) *Invasion Ecology*. Blackwell Pub. Malden, MA. 304 p.

Mack, R.N. (1985). Invading plants; their potential contribution to population biology. (editor White, J.) *Studies in plant demography*. Academic Press, London p127-142

Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, H., Clout, M., Bazzaz, F.A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*. 10;3, p689-710.

McFadden, M.W. and McManus, M.E. (1991). An insect out of control? The potential for spread and establishment of the gypsy moth in new forest areas in the United States, pp. 172-

186. In: Y.N. Baranchikov, W.J. Mattson, F.P. Hain, and T.L. Payne (eds.), Forest insect guilds: patterns of interaction with host trees. U.S. For. Serv. Gen. Tech. Rep. NE-153.

McManus, M.L. and Zerillo, R.T. (1978). The Gypsy Moth: An Illustrated Biography. *The Gypsy Moth Handbook*. United States Department of Agriculture Combined Forest Pest Research and Development Program Home and Garden Bulletin No. 225

Meentemeyer, R.K., Anacker, B.L., Mark, W., and Rizzo, D.M. (2008). Early detection of emerging forest disease using dispersal estimation and ecological niche modeling. *Ecological Applications*. 18;2, p377-390.

Meyerson, L.A. and Mooney, H.A. (2007). Invasive alien species in an era of globalization. *Frontiers in Ecology and the Environment*. 5;4, p199-208.

Muirhead, J.R. and MacIsaac, H.J. (2005). Development of inland lakes as hubs in an invasion network. *The Journal of Applied Ecology*. 42;1, p80-90.

Muirhead, J.R., Leung, B., van Overdijk, C., Kelly, D.W., Nandakumar, K., Marchant, K.R., and MacIsaac, H.J. (2006). Modelling local and long-distance dispersal of invasive emerald ash borer Agrilus planipennis (Coleoptera) in North America. *Diversity and Distributions*. 12;1, p71-79.

Okubo, A., Maini, P.K., Williamson, M.H., and Murray, J.D. (1989). On the spatial spread of the gray squirrel in Britain. *Proceedings of the Royal Society of London B, Biological Sciences*. 238;1291 p113-125.

Poland, M.T. and McCullough, D.G. (2006). Emerald Ash Borer: Invasion of the Urban Forest and the Threat to North America's Ash Resources. *Journal of Forestry*. 10;3, p118-124.

Reardon, R.C. (1991). Appalachian gypsy-moth integrated pest-management project. *Forest Ecology and Management*. 39;1-4, p107-112.

Sakai A.K., Allendorf, F.W., Holt, J.S., Lodge, D.M., Molofsky, J., With, K.A., Baughman, S., Cabin, R.J., Cohen, J.E., Ellstrand, N.C., McCauley, D.E., O'Neil, P., Parker, I.M., Thompson, J.N., and Weller, S.G. (2001). The Population Biology of Invasive Species. *Annual review of ecology and systematic*. 32, p305-332.

Schneider, D.W., Ellis, C.D., and Cummings, K.S. (1998) A Transportation Model Assessment of the Risk to Native Mussel Communities from Zebra Mussel Spread. *Conservation Biology*. 12;4, p788-800.

Sharov, A.A. Roberts, E.A., and Liebhold, A.M. (1995). Gypsy moth (Lepidoptera: Lymantriidae) spread in the central Appalachians: Three methods for species boundary estimation. *Environmental Entomology*. 24;6, p1529-1538.

Sharov, A.A. Liebhold, A.M., and Roberts, E.A. (1996). Spread of gypsy moth (Lepidoptera: Lymantriidae) in the central Appalachians: Comparison of population boundaries obtained from male moth capture, egg mass counts, and defoliation records. *Environmental Entomology*. 25;4, p783-792.

Sharov, A.A., Liebhold, A.M., and Roberts, E.A. (1997). Methods for monitoring the spread of gypsy moth (Lepidoptera: Lymantriidae) populations in the Appalachian mountains. *Journal of Economic Entomology* . 90;5, p1259-1266.

Sharov, A.A. and Liebhold, A.M. (1998). Model of slowing the spread of gypsy moth (Lepidoptera: lymantridae) with barrier zones. *Ecological Applications,* 8;4, p1170-1179.

Sharov, A.A. (2002). "Slow the Spread": A national program to contain the gypsy moth. *Journal of Forestry*. 100;5, p30-35.

Sharov, A.A. (2004). Bioeconomics of managing the spread of exotic pest species with barrier zones. *Risk Analysis*, 24;4, p879-892.

Shigesada, N. Kawasaka, K, and Takeda, Y. (1995). Modeling stratified diffusion in biological invasions. *American Naturalist*. 146;2 p229-251.

Skellam, J.G. (1951). Random Dispersal in Theoretical Populations. *Biometrika*. 38;1-2, p196-218.

Stephens, P.A., Sutherland, W.J. (1999). Consequences of the Allee effect for behavior, ecology and conservation. *Trends in ecology & evolution.* 14;10, p401-405.

Thomas, R.W. and Huggett, R.J. (1980). *Modeling in geography: a mathematical approach*. Methuen, London, England.

Tobin, P.C., Liebhold, A.M., and Roberts, E.A. (2007a). Comparison of methods for estimating the spread of a non-indigenous species. *Journal of Biogeography* 34, p305-312.

Tobin, P.C., Whitmire, S.L., Johnson., D.M. Bjornstad, O.N., and Liebhold, A.M. (2007b). Invasion speed is affected by geographical variation in the strength of Allee effects. *Ecology Letters.* 10, p36-43.

Tobin, P.C. and Blackburn, L.M. (2008). Long-Distance Dispersal of the Gypsy Moth (Lepidoptera: Lymantriidae) Facilitated its Initial Invasion of Wisconsin. *Environmental Entomology.* 37;1, 87-93.

United States Department of Agriculture – News & Events (2009). February 5th, 2009. http://www.ars.usda.gov/is/br/lbeetle/

Weber, G.A. (1930) *The plant quarantine and control administration: its history, activities, and organization.* Brookings inst (Washington, DC), Inst. For Gov. Res. Serv. Mon of the U.S. Gov. No. 59, 198 pp.

Westphal, M.I., Browne, M., MacKinnon, K., and Noble, I. (2008) . The link between international trade and the global distribution of invasive alien species. *Biological Invasions.* 10;4, p391-398.

Whitmire, S.L. and Tobin, P.C. (2006). Persistence of invading gypsy moth populations in the United States. *Oecologia* 147;2, p230-237.

Work, T.W., McCullough, D.G., Cavey, J.F., and Komsa, R. (2004). Arrival rate of nonindigenous species into the United States through foreign trade. *Biological Invasions.* 7;2, p323-332.

Appendix 2

Table 2.1 Updated from Tobin et al. 2007a, spread rates (km/yr ± se) of gypsy moth based on presence/absence data and trap catch data using the boundary displacement approach and least-squares regression. From Sharov et al. 1995, spread rates based on the regression, first occurrence, and best cell classification method.

	Presence/Absence Data		Trap Catch
Region, time span (Tobin)	Least-squares regression	Boundary displacement	Trap catch boundary displacement
ME, NH, VT, MA, CT, RI, NY 1900-1915	3.0 (0.3)	2.6 (1.0)	NA
NY, PA, OH, NJ, DE, MD, VA, WV 1966-1990	19.5 (0.9)	21.1 (6.3)	NA
MI, lower Peninsula, 1981-1995	9.5 (1.9)	9.9 (5.0)	14.6 (3.4)
Central WV and VA, 1989-2005	7.8 (1.9)	7.6 (3.7)	5.8 (2.6)
MI, upper Peninsula, and WI, 1994-2005	28.6 (3.3)	15.9 (5.9)	18.0 (5.9)
WI only, 1998-2005	16.1 (2.2)	19.6 (6.6)	16.0 (6.1)

Region, time span (Sharov)	Regression	First Occurrence	Best Cell Classification
VA, WV 1989	22.6	19.4	19.6
VA, WV 1990	8.8	7.5	7.2
VA, WV 1991	3.8	3.9	4.9
VA, WV 1992	12.3	12.0	11.9

CHAPTER 3 – CONCEPTUAL FRAMEWORK FOR GYPSY MOTH DISPERSAL

3.1 Conceptual framework introduction

This chapter provides a conceptual framework of gypsy moth spread, focusing on the role of anthropogenically facilitated dispersal. Insights are drawn from a review of literature on gypsy moth as well as previous research on the role of anthropogenic factors in the spread of other invasive species. The experiences and regulatory approaches of agencies that monitor and seek to control gypsy moth also provide insight into the factors that might influence the spread of gypsy moth.

3.2 Conceptual framework for gypsy moth dispersal

The framework describes the process of gypsy moth dispersal (Figure 3.1, Appendix 3). This process is defined as the movement of a gypsy moth life stage from the infested zone, or sending zone, to the uninfested zone, or receiving zone. In this analysis the receiving zone is limited to the expanded transition area, approximately 25 to 700 km from the infested zone. First instar gypsy moth can disperse naturally up to 2.5km/yr (Liebhold et al. 1992). Thus, short-distance spread can occur by male moths flying to a new location, or caterpillars ballooning to a new location. The spread over longer distances is most likely through human transport, typically the movement of overwintering egg-masses.

The probability of observing gypsy moth in a previously uninfested county (in the receiving zone) is the product of many intermediate probabilities. The human-mediated pathway depends on 1) the probability of an egg mass being laid on a potential transport vector, 2) the probability that the transport vector is moved to the receiving zone, 3) the probability the egg mass hatches while the transport vector is in the receiving zone, and 4) the probability gypsy moth is captured in a trap.

Likely transport vectors of egg masses include objects that are likely to be moved (e.g. recreation equipment or outdoor furniture) from the infested zone to the uninfested zone during the eight month diapause period. Eggs then hatch in the spring and can be detected by traps collected later in the summer. Vehicles may also serve as a vector (if egg masses are laid on the vehicle) or be used to transport infested material such as nursery stock. In 2006 egg masses were found in Oregon on car parts of 1967 Chevrolet that were shipped from Connecticut in the previous year. The vehicle was purchased on ebay (Oregon Department of Agriculture 2009).

Equipment that can harbor egg masses is highlighted in section 1.3. The list of equipment could include, but is not limited to, grills, doghouses, canoes, kayaks, firewood, birdhouses, vehicles, and potted plants. These objects can be moved under many different scenarios but are most likely moved by vacationers, recreators, movers, timber companies, or nurseries. For example, timber companies could move recently felled logs with egg masses from the infested to the uninfested zone. Recreational vehicles, however, could move egg masses from the infested zone to the uninfested zone or could travel from the uninfested zone to the infested zone, pick up an egg mass and then move back to the uninfested zone. This would have to occur in the right time frame. The recreation vehicle would have to be in the infested zone during oviposition, which generally occurs over a two to four week period in June and August and then travel back to the uninfested zone before first instars began to emerge in spring.

Once a vector for gypsy moth egg masses is in the uninfested zone the egg mass could remain with the vector or scraped off of the vector before it departed. Scrapping an egg mass of a vector is not an effective way to ensure that larvae will not hatch. Egg masses need to be burned or sprayed with recommended oil that suffocates the embryo. Once gypsy moth larvae hatch, they survive by feeding on the foliage of preferred tree species. The larvae are generalist that feed on over 300 different species of trees and can establish anywhere susceptible species are present (Liebhold et al. 1997) and climatic conditions are suitable (Logan et al. 2007).

Delta traps are generally used to detect gypsy moths in the transition and uninfested zones. The traps are scented with (+) disparlure, an attractant designed to mimic sex pheromones emitted by females to attract male gypsy moths. Traps in the detection area are generally placed at 2 km grid spacing, but can be as compact as 500 m for delineating isolated populations and up 8 km spacing for evaluation. The traps are effective at attracting male gypsy moths even at low densities (Plimmer et al. 1981).

Because the probabilities of any of the above processes are unknown we developed a theoretical model to capture all aspects of this process. The diffusion category captures the process of natural dispersal.[4] Source infestations represents the propagule pressure or supply from the sending zone. Visitor attraction, demand for goods, and distance represent the attractiveness of the receiving zone. The ability of gypsy moths to disperse even further depends on the transportation qualities of a county (e.g. number of airports or presence of big box retail stores). For example, one would expect a higher probability of detecting gypsy moths in counties that have a constant flow of goods (e.g. airports, a large number of retail stores, highways). Though the dispersal process is complex, we believe we have developed a theoretical model that captures the mechanisms of the process.

Significant variables determined through stepwise logistic regression ($\alpha < 0.05$) in the theoretical model were used to derive two empirical models: 1) the empirical model and 2) the empirical sector model. The two empirical models were derived by determining significant variables from each theoretical category. Significance was a product of statistical significance within a year and the frequency that a variable was significant. The empirical model differs from the empirical sector model because it only includes significant variables from the theoretical model categories whereas the sector model includes specific sectors under quarantine.

[4] Ideally the vectors for long distance dispersal would be represented with change of address data or visitor address data from state or national parks, but it is difficult to obtain those data so we represent them using proxies.

3.3 References

Liebhold, A.M., Halverson, J.A., and Elmes, G.A. (1992). Gypsy moth invasion in North America: a quantitative analysis. *Journal of Biogeography*. 19;5, p513.

Liebhold, A.M., Gottschalk, K.W., Mason, D.A., and Renate, R.B. (1997). Forest susceptibility to the gypsy moth. *Journal of Forestry*. 95;5 p20-24.

Logan, J.A., Regniere, J., Gray, D.R., and Munson, A.S. (2007). Risk assessment in the face of a changing environment: gypsy moth and climate change in Utah. *Ecological Applications*. 17;1 p101-117.

Oregon Department of Agriculture – Plant Division (2009). March 19th, 2009. http://www.oregon.gov/ODA/PLANT/plant_ann_rep06_ippm_part2.shtml#Gypsy_moth

Plimmer, J.R., Bierlleonhardt, B.A., Schwalbe, C.P., Webb, R.E. (1981). Management of gypsy moth with its sex attractant pheromone. *Abstracts of papers of the American chemical society.* 82, p43-PEST.

Appendix 3

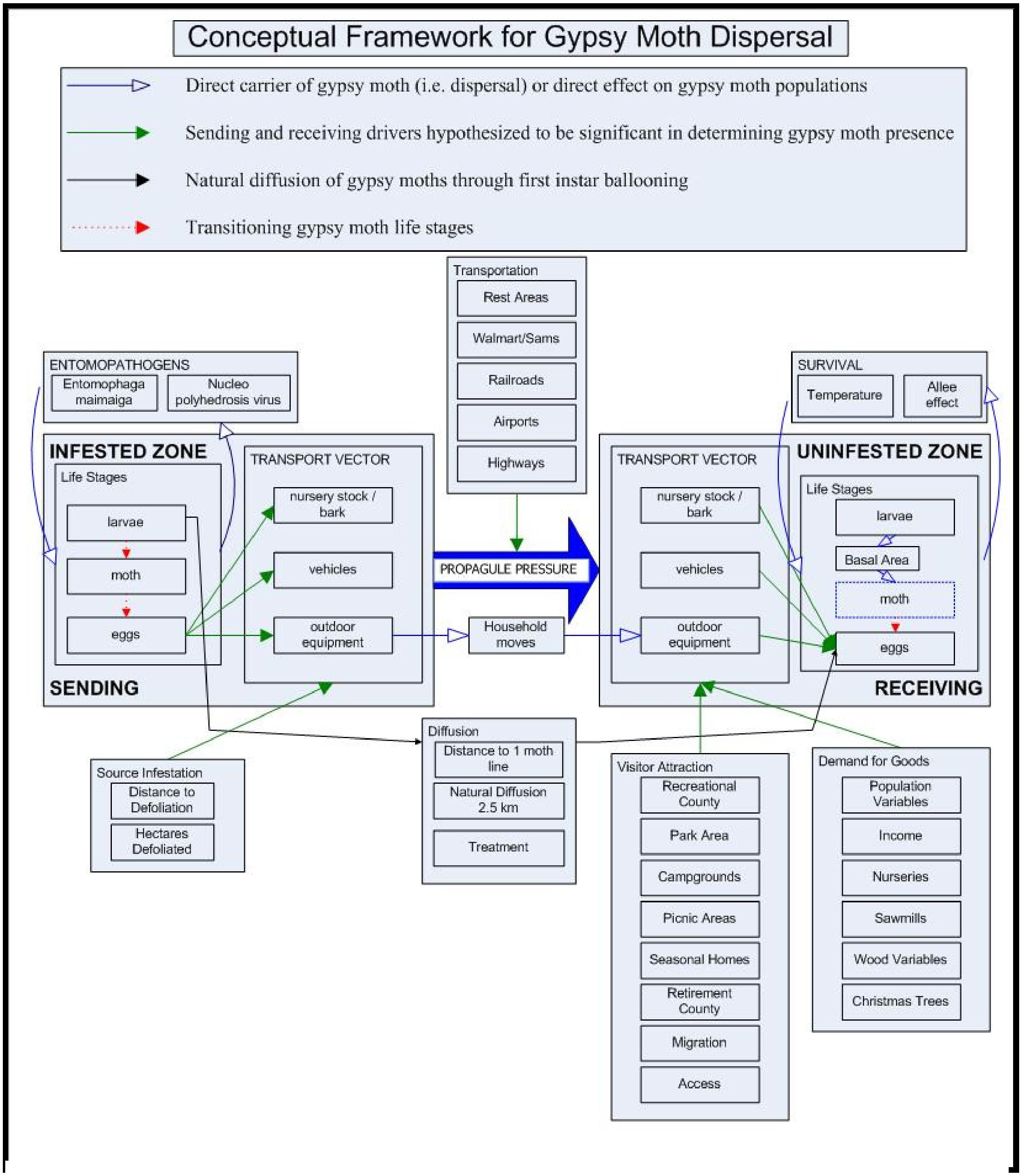


Figure 3.1 – Conceptual framework depicting the dispersal of gypsy moth from the sending to receiving zone.

CHAPTER 4 – MATERIALS AND METHODS

4.1 Introduction to materials & methods

This section provides a detailed overview of the data and methods used to analyze the hypothesized anthropogenic drivers of gypsy moth spread described in the conceptual framework. I first describe the available measures of gypsy moth dispersal and their drivers (i.e. the dependent and independent variables) in section 4.2. I then discuss the spatial and temporal extent of the analysis (section 4.3). Next, descriptive statistics for all of the variables are presented (section 4.4). Finally, I describe the analysis flow (section 4.5) and the statistical methods (section 4.5 - 4.7) used to model the anthropogenic drivers of gypsy moth spread.

4.2 Dependent and independent variables

Each year approximately 100,000 pheromone-baited traps are deployed at 2 km to 8 km grid spacing across the transition zone to estimate male moth density (Figure 4.1, Appendix 4). This data set was used to derive two dependent variables for modeling.

Only trap data distal to the 1 moth/trap line (i.e. traps ahead of the area in which at least 1 male moth is trapped annually) were selected for analysis. We aggregated the data on a county level to ensure a consistent resolution with readily available independent data. For each year in the dataset, trap records from Slow-the-Spread were used with ArcGIS tools to determine 1) the presence or absence of gypsy moth in a county (Figure 4.2, Appendix 4) and 2) the mean number of gypsy moths in a trap in a given year (Figure 4.3, Appendix 4).

The independent variables measure the biophysical and anthropogenic factors proposed in the conceptual framework: 1) source infestation, 2) biophysical processes, 3) diffusion, 4) visitor attraction, 5) transportation, and 6) demand for goods (Table 3.1, Appendix 3). A theoretical model was developed to explain the mechanistic process of dispersal and the hypothesized drivers of this system (Figure 3.1, Appendix 3). Each independent variable was categorized into one of these six categories comprising the dispersal process. Specifically, the anthropogenic variables are measures that have been used in previous research on the introduction or dispersal of invasive species (Lippitt et al. 2008, Muirhead et al. 2006, Gilbert et al. 2004) or they are the best available measures of anthropogenic factors available at the county scale for the relevant time period. Some of the variables can capture elements of multiple factors in the conceptual framework; in the following discussion, they are grouped with the most relevant factor.

Source infestation includes the size and proximity (of gypsy moth populations) to the receiving area. This is measured by: 1) size of an outbreak within 100 km of a county lagged by one year, 2) size of an outbreak within 100 km of a county lagged four years (as suggested by A. Liebhold, personal communication), 3) distance to nearest outbreak lagged 1 year, and 4) distance to nearest outbreak lagged 4 years. Outbreaks are determined by aerial sketch mapping of areas defoliated by gypsy moth. This provides a reasonable measure of the size and location of outbreaks. Outbreak size units are meters squared and distance to nearest outbreak are kilometers.

Biophysical processes are the biological and physical attributes of a county. Variables included in this category are 1) susceptible basal area of preferred host tree species (percent basal area), 2) county latitude, 3) county longitude (meters, gypsy moth projection), and 4) size of the county (acres).

Variable that influence new county infestations resulting from natural spread include 1) distance to the 1 moth/trap line (kilometers), 2) positive trap lagged by one year (binary, positive if at least one gypsy moth was found in a trap in the previous year), and 3) treatment lagged by one year (hectares, size of treatment in the previous year). Previous work (Liebhold et al. 1992) parameterized a diffusion model (i.e. Skellam's model) based upon demographic data to estimate gypsy moth natural dispersal as 2.5 km/yr. I used this rate of spread to define diffusive spread.

Visitor attraction includes services provided by a county that attract people to visit, recreate, or reside in that county. Specific variables are: 1) recreation county status as defined by the United States Department of Agriculture Economic Research Service (binary, derived based on the wage in recreation industries and number of seasonal homes) 2) total acres of state and national park in a county 3) number of campgrounds, (4) number of picnic areas, 5) number of seasonal homes , 6) retirement county statues as defined by United States Department of Agriculture Economic Research Service (binary, derived based on industry and demographics), 7) number of people migrating to and emigrating from a county; can be positive or negative, and

8) mean distance (meters) to nearest road. All of these are hypothesized to influence the likelihood of introducing gypsy moth to a county.

Transportation is defined as the attractiveness of a county for goods or people to travel a long distance to arrive at a county. Transportation variables are: 1) number of rest areas 2) number of big box retailers such as Wal-Mart and Sam's Clubs, 3) railroad distance (1000s of kilometers), 4) number of airports in a county, and 5) total distance (km) of major and minor highways and roads.

Demand for goods captures the flow of goods from sending areas into receiving areas. Variables considered include 1) population for each year in the analysis, 2) population density, 3) urban population, 4) urban population inside urban areas, 5) cluster of urban population, 6) rural population, 7) median household income, 8) number of nurseries, 9) acres of open nurseries, 10) number of sawmills, 11) number of Christmas tree farms, 12) household wood[5] (1000s of houses), 13) household wood/population, and 14) household wood/area. Household wood is defined as the number of households that use wood as a major heating fuel. These variables were hypothesized to influence the probability of new gypsy moth infestations.

A gravity variable was also created as an alternative to the population variables in the demand for goods category. The gravity variable is a function of the distance between pairs of county centroids and the population density of each county. This metric is often used to predict the movement of people or goods between multiple places in space (Thomas and Huggett 1980).

---

[5] In 1940 23% of US households relied on wood as their major heating fuel. Today, with the availability of gas and electricity, wood only comprises 1.7% of household heating source (2005 American Housing Census).

It represents the connectivity between pairs of counties, which in turn is expected to affect the probability of introducing gypsy moth (P_{ij}) from an infested county (i) to an uninfested county (j).

$$P_{ij} = \frac{\ln(pden_i)(pden_j)}{\ln(D_{ij})} \quad , \qquad \text{[Equation 4.1]}$$

Where *pden* is the population density in county i or j, and D_{ij} is the Euclidean distance between county i and j. The resulting pairwise matrix of values for P_{ij} can be multiplied by a vector of counties that were positive for gypsy moth in the previous year, P_{t-1}, to estimate the gravity variable in year t, P_t:

$$[P_t] = [P_{t-1}]*[P_{ij}] \qquad \text{[Equation 4.2]}$$

4.3 Spatial and temporal extent

The study area comprises a swath of counties along the invasion front. The invasion front for gypsy moth extends from North Carolina to Wisconsin (Figure 4.4, Appendix 4). This analysis includes any county distal to the 1 moth/trap line. The spatial extent changes from year to year, depending on the annual location of the 1 moth/trap boundary and the number of counties that have traps. As a result the size of the study area ranges between 432 and 790 counties or 650,000 and 1,100,000 km^2 annually.

The temporal extent of this analysis is 1999 to 2007. Though our analysis begins in 1999 we use independent data dating back to 1995 because of the time lag inherent in the relationship between some independent variables and the dependent variable. For example, gravid females or egg masses could be introduced to an uninfested area but the resulting male moths would not be trapped until the following year. Despite counties being added and dropped from the analysis, the sample size remains relatively consistent between years and balanced within years (Table 4.2, Appendix 4).

4.4 Descriptive statistics

Preliminary analysis includes calculating the summary statistics (Table 4.3, Appendix 4) and correlations among (proc corr in SAS, Table 4.4, Appendix 4) independent variables. The correlation matrix is useful in assessing collinearity among variables. Collinearity, when present between two variables, makes it difficult to identify both variables coefficients in the same model.

4.5 Analysis flow

Several steps are involved in specifying, estimating, and interpreting a final model of gypsy moth dispersal:

1) Both stepwise and no selection logistic regression models were estimated for each year with the variables in each category of factors defined in the conceptual framework. This supports selection of the best variables to empirically represent that category.
2) Spatial autocorrelation in the dependent variables was identified using both the semivariogram to examine anisotropy in the dependent variable and Moran's I statistic for spatial autocorrelation.
3) A semivariogram was created from the residuals of general linear and stepwise logistic regression models, with and without the latitude and longitude of the counties (and higher order polynomials of latitude and longitude), to assess whether including these coordinates controls for spatial autocorrelation.
4) Based on steps 1 and 3, two specifications were defined: (a) empirical sector model(including specific sectors under quarantine, e.g. sawmills, nurseries, Christmas trees, and X-Y coordinates), and (b) empirical model (including only statistically significant variables in each of the six model categories and X-Y coordinates).
5) Spatial autocorrelation of the final models was assessed by creating semivariograms from the residuals of the general linear and logistic regressions to examine the level of anisotropy of the model residuals, and Moran's I statistic was generated from the Pearson Chi-square residuals.

Though this process seems circular it was necessary to first determine significant variables from each model category and then examine the effects of X and Y coordinate and polynomial models independently. This allowed us to omit any unnecessary model covariates that did not improve our ability to control for spatial autocorrelation.

4.6 Spatial autocorrelation

Ecological data are often spatially correlated, violating the key assumption of independence underlying standard statistical tests and regression models (Rossi et al. 1992). Therefore, it is important to test for spatial autocorrelation. I employ a two-pronged approach to determine the level of spatial autocorrelation inherent in the dependent data and empirical models. The first approach is the analysis of the semivariograms of the dependent variable and then analysis of the semivariogram from a general linear model (SAS proc glm) and logistic regression (SAS proc logit). Second is analysis of Moran's I statistic for the dependent variable and residuals from the stepwise logistic regression model.

Moran's I statistic was calculated using an ArcGIS tool (Spatial autocorrelation – Moran's I) for the mean number of moths per trap in each county in a given year. Moran's I is a measure of spatial autocorrelation within a variable. This measure is two dimensional because it takes into account both spatial proximity and the observational value of the variable of interests. Moran's I is defined as:

$$Z = \frac{N}{\sum_i \sum_j W_{ij}} \frac{\sum_i \sum_j W_{ij}(D_i - \bar{D})(D_j - \bar{D})}{\sum_i (D_i - D)^2} \qquad \text{[Equation 4.3]}$$

Where N is the total number of observational values (counties with trap data), W_{ij} is a matrix of spatial weights (the inverse distance between observations of variable), D is mean number of moths in a county, D_i is the value of the i th observation of D and, D_j is the j th observation of D, and $\bar{D}$ is the mean of variable D. Z values greater than two indicate positive spatial autocorrelation while values less than negative two indicate negative spatial autocorrelation. Values between two and negative two indicate that no spatial autocorrelation is present.

Robust semivariograms were constructed for each year's mean trap record throughout an entire county by spatially referencing trap records to county centroids. The semivariogram is defined as:

$$2\gamma_z(h) = \frac{1}{|N(h)|} \sum_{(i,j)\varepsilon N(h)} (Z(r_i) - Z(r_j))^2 \qquad \text{[Equation 4.4]}$$

$|N(h)|$ is the total number of pairs (*i, j*) within lag distance h. r_i is the value at location i and r_j is the value at location j. (Rossie et al. 1992, Srivastava and Isaaks 1991).

The semivariogram can be used to determine spatial dependence and directional semivariograms can be used to detect anisotropy, which occurs when spatial dependence of a process is not homogenous in all directions. The semivariance can be measured by observing the value of the sill and the range. The sill is defined as the value if the semivariance tends toward spatial dependence. The range is the distance at which spatial dependence is present (see Rossie et al. 1992).

One method for controlling spatial autocorrelation is to include additional variables in the model that take into account spatial dependence. For example, measures of proximity, connectivity, area, or distance are often used as covariates (Rossi et al. 1992, Koenig et al. 1999). This approach was explored by generating semivariograms, first using only the observed mean trap counts in a county (Figures 5.1 – 5.9 A, Appendix 5), then using residual values from a general linear model and logistic regression of county trap records using the X and Y coordinate model as defined above (Figures 5.1 – 5.9 B, Appendix 5).

The general linear model and logistic regression model were then expanded to include second and third order polynomial terms of the X and Y coordinates (Figures 5.1 – 5.9 C, Appendix 5). The residuals from this model were detrended (Sharov et al. 1996). That is the residuals were used to create a semivariogram that was compared to the semivariogram of the observed data. Last, an empirical model was used to generate residuals that were analyzed in a semivariogram and compared to the above semivariograms (Figure 5.1 – 5.9 D, Appendix 5).

4.7 Statistical modeling

Both the general linear model and logistic regression model were estimated. The general linear model has greater statistical power, but it is based on an assumption of a normal error term that might be violated. The logistic regression transforms the independent data into a sigmoidal form.

The general linear model is used to predict the value of a dependent variable based on values of an independent variable(s). The general linear model assumes that 1) the sample is selected at random from a population of interest, 2) the dependent variable is continuous, and 3) the error terms follow independent and identical normal distribution (Longnecker 2000). The general linear model is defined as a linear statistic model:

$$Y = \beta_0 + \beta_i X_i + \beta_h X_h + E\,, \qquad \text{[Equation 4.5]}$$

where Y is the dependent variable, β_0 is the intercept, β_i is the coefficient for independent variable X_i where $i = 1, 2, \ldots h$ and, E is the error term

Logistic regression is used to predict the probability distribution of a dichotomous dependent variable, such as presence or absence of gypsy moth (SAS/STAT 2004). It is similar to discriminant analysis, which predicts whether or a not an event will occur (i.e. a binary outcome) but it uses the logistic function of the observed values. Logistic regression transforms the independent variables into a logistic function, taking on a sigmoidal form.

Logistic regression does not assume a normal distribution of the variables, but does assume independence of observations. As with other regression models, another assumption is that the probability distribution of the dependent variable depends on the independent variables. The logistic equation is as follows:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_i x_i + E_i \qquad \text{[Equation 4.6]}$$

where the log odds ($\log(\frac{p}{1-p})$) is the probability distribution of the dependent variable and other terms are defined as above.

To aid in model specification, six logistic regressions were estimated for each year, each with the variables representing one factor in the conceptual framework. Variables that most often had statistically significant coefficients relative to other variables in the category were selected to compose the empirical model. This selection process was also guided by the descriptive statistics and correlation matrix presented in this chapter. This generated the empirical specification which was used to estimate both a general linear and a logistic regression for each year. This model was then examined for spatial autocorrelation using Pearson Chi-square residuals in a semivariogram and Moran's I statistic. These were compared to the original statistics from the observational values to determine whether spatial autocorrelation was effectively controlled in the model.

4.8 References

Gilbert, M., Gregoire, J.-C., Freise, J.F., and Heitland, W. (2004). Long-distance dispersal and human population density allow the prediction of invasive patterns in the horse chestnut leafminer Cameraria ohridella. *Journal of Animal Ecology*. 73, 3, 459.

Liebhold, A.M., Halverson, J.A., and Elmes, G.A. (1992). Gypsy moth invasion in North America: a quantitative analysis. *Journal of Biogeography*(1992). 19;5, 513.

Lippitt C.D., Rogan, J., Toledano, J., Sangermano, F., Eastman, R.J., Mastro, V., Sawyer, A. (2008). Incorporating anthropogenic variables into species distribution model to map gypsy moth risk. *Ecological Modelling*. 210;3, p339-350.

Muirhead, J.R., Leung, B., van Overdijk, C., Kelly, D.W., Nandakumar, K., Marchant, K.R., and MacIsaac, H.J. (2006). Modelling local and long-distance dispersal of invasive emerald ash borer Agrilus planipennis (Coleoptera) in North America. *Diversity and Distributions*. 12;1, p71-79.

Koenig, W.D. (1999). Spatial autocorrelation of ecological phenomena. *TREE*. 14;1, p22-26.

Longnecker, M.T. and Ott, R. (2000). An Introduction to Statistical Methods and Data Analysis. Duxbury Pr. Massachusetts p1152.

Rossi, R.E., Mulla, D.J., Journel, A.G., and Franz, E.H. (1992). Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*. 62;2, p277-314.

SAS/STAT 9.1 User's Guide (2004). SAS Institute, Cary, NC, USA. p. 5136

Sharov, A.A. Liebhold, A.M., and Roberts, E.A. (1996). Spread of gypsy moth (Lepidoptera: Lymantriidae) in the central Appalachians: Comparison of population boundaries obtained from male moth capture, egg mass counts, and defoliation records. *Environmental Entomology*. 25;4, p783-792.

Srivastava, R.M. and Isaaks, E.H. (1991). A Reply to "A Study of 'Probabilistic' and Deterministic' Geostatistics" by Robert F. Shurtz. Mathematical Geology, 23;3 p481-495.

Thomas, R.W. and Huggett, R.J. (1980). Modeling in geography: a mathematical approach. Methuen, London, England. P338.

Appendix 4

Table 4.1 – Data definitions.

	Parameter	Definition
VISITOR ATTRACTION	Recreational Counties	A county is defined as recreation dependent if it exceeded a certain threshold that took into account: wage and salary in entertainment and recreation, personal income from the former categories, housing units for seasonal use, and hotel and motel use.
	Park Area	Total area of all state and national parks within a county
	Campgrounds	Total number of campgrounds (individual sites) for Forest Service, US Army Corp of Engineers, Bureau of Land Management, and the Tennessee Valley Authority
	Picnic Areas	Total number of campgrounds (individual sites) for Forest Service, US Army Corp of Engineers, Bureau of Land Management, and the Tennessee Valley Authority
	Seasonal Homes	housing units intended for occupancy only during certain seasons of the year and are found primarily in resort areas. This also includes housing units used for migratory labor employed during a crop season and seasonal mobile homes
	Retirement Counties	Retirement destination county
	Migration	Demographic components of change (birth, deaths, domestic and international migration) are developed as inputs to the population estimates. Data are derived from changes in group quarter population. This data source is solely for internal migration.
	Mead Distance to Nearest Road	Mean distance of any locations in a county to the nearest road (based on 30m gridded surface)
SOURCE INFESTATION	Size of outbreak within 100km (1 year lag)	Data from aerially detected surveys was used to spatial join to the centroids of each county in the U.S within 100 km on an outbreak and record the size of the outbreak. Data was lagged 1 year, e.g. defoliation in 1999 corresponds to gypsy moth trap results in 2000.
	Size of outbreak within 100km (4 year lag)	Data from aerially detected surveys was used to spatial join to the centroids of each county in the U.S within 100 km on an outbreak and record the size of the outbreak. Data was lagged 4 year, e.g. defoliation in 1999 corresponds to gypsy moth trap results in 2003.
	Lagged Distance to Outbreak (1 year)	Data from aerially detected outbreaks was used to spatial join to the centroids of each county in the U.S. and record the distance from the centroid of the county to the centroid of the outbreak. The data was lagged 1 year, e.g. defoliation in 1999 corresponds to gypsy moth trap results in 2000.
	Lagged Distance to Outbreak (4 year)	Data from aerially detected outbreaks was used to spatial join to the centroid of each county in the U.S. and record the distance from the centroid of the county to the centroid of the outbreak. The data was lagged 1 year, e.g. defoliation in 1999 corresponds to gypsy moth trap results in 2000.
BIOPHYSICAL	Susceptible Basal Area (20%)	Proportion of land area covered by susceptible species (> 20% basal area in preferred species)
	County Latitude	In decimal degrees, the latitude of a county's centroid
	County Longitude	In decimal degrees, the longitude of a county's centroid
	County Area	Total area of a county. Value derived by me using the gypsy moth projection.
DIFFUSION	Distance to 1 moth/trap line	Distance from each county's centroid to the 1 moth/trap line
	Treatment	Aerial and other treatments to eradicate gypsy moth populations each year.
	Lagged Positive Trap	Each trap that was positive for gypsy moth was buffered 2.5 km (the max natural dispersal of gm) each year. Any county that was within this 2.5 km buffer was assigned a 1.

Table 4.1 continued – Data definitions continued.

	Variable	Definition
DEMAND FOR GOODS	Gravity	A metric that relates the
	Population	Number of people living in each county in the USA
	Population Density	ratio of people per acre
	Urban Population	all territory, population, and housing units located within urbanized area or an urban cluster
	Urban Population Inside Urban Areas	densely settled territory consisting of a core census block groups or blocks that have a population density of at least 1,000 people per square mile
	Cluster of Urban Population	surrounding census blocks that have an overall density of at least 500 people per square mile
	Rural Population	all territory, population, and housing units located outside Uas and Ucs
	Income	Mean household income per county
	Nurseries	Total acres of nurseries, greenhouse, floriculture, sod, mushrooms, vegetable seeds, and propagative materials by county for areas under glass or other protection, including tunnel production
	Acres of Open Nurseries	Total acres of nurseries, greenhouse, floriculture, sod, mushrooms, vegetable seeds, and propagative materials by county
	Sawmills	Data was compiled by mill managers for the FIA. Saw mills include: sawmills, composite, plyven, postpole, pulpmill, and other (firewood, bark products, charcoal).
	Household Wood	The use of wood or wood charcoal as a fuel
	Household Wood/Population	Ratio of household wood to population in a county
	Household Wood/Area	Ratio of household wood to area of a county
	Christmas Tree Farms	Total acres of Christmas tree farms by county
TRANSPORT-ATION	Rest Areas	The total number of rest areas, visitor welcome centers, or scenic vistas in a given county
	Walmart and Sams Clubs	Count of Wal-mart and Sams clubs in a given county
	Railroad Distance	This map layer includes railroads in the conterminous United States and Alaska
	Airports	Number of airports in county.
	Highway Distance	Major and minor US highways and roads. Interstates, U.S. Highways, State Highways, major roads, and minor roads
DEPENDENT VARIABLE	Mean Trap Count	The mean number of gypsy moths in the entire county in a given year
	Presence/Absence	Presence = at least one gypsy moth in the county. Absence = no gypsy moths in the county

Table 4.1 continued – Data definitions continued, this table includes parameter units, year, and source.

	Variable	Units	Year	Source
VISITOR ATTRACTION	Recreational Counties	count	2004	USDA economic research service
	Park Area	count	1997	NORSIS
	Campgrounds	1000s of campgrounds	1997	NORSIS
	Picnic Areas	count	1997	NORSIS
	Seasonal Homes	1000s of seasonal homes	2000	Decennial Census
	Retirement Counties	Binary (yes=1, no-2)	2004	USDA economic research service
	Migration	10000s of people	1999-2007	U.S. Census Bureau, Population Division
	Mead Distance to Nearest Road	meters	2007	USGS Raymond Watts
SOURCE INFESTATION	Size of outbreak within 100km (1 year lag)	Meters squared	1999-2007	STS
	Size of outbreak within 100km (4 year lag)	Meters squared	1999-2007	STS
	Lagged Distance to Outbreak (1 year)	Kilometers	1999-2007	STS
	Lagged Distance to Outbreak (4 year)	Kilometers	1999-2007	STS
BIOPHYSICAL	Susceptible Basal Area (20%)	Percent basal area	1995	USDA FS, Sandy Liebhold
	County Latitude	Decimal Degrees	na	na
	County Longitude	Decimal Degrees	na	na
	County Area	acres	2000	STS
DIFFUSION	Distance to 1 moth/trap line	kilometers	1999-2007	STS
	Treatment	Hectares	1999-2007	STS
	Lagged Positive Trap	Binary (yes =1, no-2)	1999-2007	STS

Table 4.1 continued – Data definitions continued, this table includes parameter units, year, and source.

	Variable	Units	Year	Source
DEMAND FOR GOODS	Gravity		2000*	Decennial Census
	Population	10,000s of people	2000*	Decennial Census
	Population Density	people/acre	2000*	Decennial Census
	Urban Population	10,000s of people	2000	Decennial Census
	Urban Population Inside Urban Areas	10,000s of people	2000	Decennial Census
	Cluster of Urban Population	10,000s of people	2000	Decennial Census
	Rural Population	10,000s of people	2000	Decennial Census
	Income	1,000s of dollars	2000	Decennial Census
	Nurseries	count	2002	USDA NASS
	Acres of Open Nurseries	acres	2002	USDA NASS
	Sawmills	count	2004	USDA FS
	Household Wood	1,000s of households	2005	American Housing Survey (AHS, US census)
	Household Wood/Population	ratio	2000	AHS, US Census
	Household Wood/Area	ratio	2000	AHS, US Census
	Christmas Tree Farms	acres	2002	USDA NASS
TRANSPORT-ATION	Rest Areas	count	2007	Poi Factory
	Walmart and Sams Clubs	count	2000	Poi factory
	Railroad Distance	1000s of kilometers	2006	ESRI data and Map
	Airports	count	2004	Federal Aviation Administration
	Highway Distance	kilometers	1998	US Department of Commerce
DEPENDENT VARIABLE	Mean Trap Count	count	1999-2007	STS
	Presence/Absence	binary	1999-2007	STS

Table 4.2 – Sample size, the number of counties in the analysis.

Year	1999	2000	2001	2002	2003
N	499	534	465	472	528
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307

Year	2004	2005	2006	2007
N	559	432	769	790
N (Gypsy moth present)	290	164	441	371
N (Gypsy moth absent)	269	268	328	419

Table 4.3 – Descriptive statistics for the model parameters and dependent variables.

	Parameter	Mean	Standard Deviation	Minimum	Maximum	N
VISITOR ATTRACTION	Gravity	3.70	5.85	0.08	56.64	904.00
	Recreational Counties	NA	NA	NA	NA	NA
	Park Area	3840.61	15822.87	0.00	307770.60	904.00
	Campgrounds	0.01	0.07	0.00	0.65	904.00
	Picnic Areas	6.97	29.73	0.00	351.00	904.00
	Seasonal Homes	0.71	1.72	0.00	25.84	904.00
	Retirement Counties	NA	NA	NA	NA	NA
	Migration	0.00	0.07	-2.00	0.41	904.00
	Mead Distance to Nearest Road	298.23	188.15	0.00	3240.00	904.00
SOURCE INFESTATION	Size of outbreak within 100km (1 year lag)	4541752.21	12628112.56	62963.72	48116015.45	20.00
	Size of outbreak within 100km (4 year lag)	20293764.84	40240388.13	127363.17	135463291.00	17.00
	Lagged Distance to Outbreak (1 year)	435.82	178.97	38.23	1164.85	904.00
	Lagged Distance to Outbreak (4 year)	489.51	232.02	37.94	1077.10	904.00
BIOPHYSICAL	Susceptible Basal Area (20%)	0.19	0.18	0.00	0.71	904.00
	County Latitude	-142289.27	429123.82	-944739.00	829168.00	904.00
	County Longitude	4137164.91	386402.85	2996158.00	5224220.00	904.00
	County Area	344127.68	237852.77	1660.95	4311669.17	904.00
DIFFUSION	Distance to 1 moth/trap line	247.04	173.18	0.74	874.37	904.00
	Lagged Positive Trap	NA	NA	NA	NA	NA
	Treatment	127.80	1037.42	0.00	17661.00	904.00

Table 4.3 continued – Descriptive statistics continued for the model parameters and dependent variables.

	Parameter	Mean	Standard Deviation	Minimum	Maximum	N
DEMAND FOR GOODS	Population	6.38	20.69	0.23	537.67	904.00
	Population Density	0.23	0.57	0.00	8.73	904.00
	Urban Population	4.40	20.53	0.00	537.22	904.00
	Urban Population Inside Urban Areas	3.51	20.59	0.00	537.22	904.00
	Cluster of Urban Population	0.88	1.11	0.00	7.77	904.00
	Rural Population	1.98	1.32	0.00	10.12	904.00
	Income	35.93	7.88	15.81	71.03	904.00
	Nurseries	14.67	23.66	0.00	512.00	904.00
	Acres of Open Nurseries	216.37	893.91	0.00	22016.00	904.00
	Sawmills	2.49	3.05	0.00	24.00	904.00
	Household Wood	0.44	0.40	0.01	3.09	904.00
	Household Wood/Population	0.16	0.16	0.00	1.04	904.00
	Household Wood/Area	15.96	26.29	0.27	541.59	904.00
	Christmas Tree Farms	6.26	18.37	0.00	317.00	904.00
TRANSPORT-ATION	Rest Areas	0.46	0.90	0.00	8.00	904.00
	Walmart and Sams Clubs	1.16	2.00	0.00	33.00	904.00
	Railroad Distance	76.45	64.46	0.01	881.88	803.00
	Airports	4.86	5.06	0.00	48.00	904.00
	Highway Distance	224.44	116.00	6.78	1281.56	904.00

Table 4.3 continued – Descriptive statistics continued for the model parameters and dependent variables.

	Parameter	Mean	Standard Deviation	Minimum	Maximum	N
TRAP RECORDS	1999	0.35	2.07	0.00	32.82	499.00
	2000	0.23	1.47	0.00	28.79	534.00
	2001	0.42	3.05	0.00	49.91	465.00
	2002	0.22	1.43	0.00	28.48	472.00
	2003	0.17	0.55	0.00	10.07	528.00
	2004	0.32	3.59	0.00	84.00	559.00
	2005	0.65	6.79	0.00	135.70	432.00
	2006	0.49	7.09	0.00	186.00	769.00
	2007	0.16	1.39	0.00	38.43	790.00

Table 4.4 – Correlation matrix.

Covaraite Matrix: Entire US

The CORR Procedure

	BA_20	BA_50	pop_99	popdens99	URBAN	URBAN_ INSIDEURBANAREAS	URBAN_ URBANCLUSTER	RURAL_POP
BA_20	1.00000	0.89976	-0.07275	-0.06696	-0.08508	-0.08338	-0.03926	0.24580
		<.0001	<.0001	0.0002	<.0001	<.0001	0.0259	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
BA_50	0.89976	1.00000	-0.07111	-0.06267	-0.08047	-0.07878	-0.03890	0.19495
	<.0001		<.0001	0.0005	<.0001	<.0001	0.0273	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
pop_99	-0.07275	-0.07111	1.00000	0.35142	0.99797	0.99552	0.10856	0.27398
	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001
	3141	3141	3141	3125	3141	3141	3141	3141
popdens99	-0.06696	-0.06267	0.35142	1.00000	0.36469	0.36799	-0.04597	-0.03922
	0.0002	0.0005	<.0001		<.0001	<.0001	0.0102	0.0284
	3125	3125	3125	3125	3125	3125	3125	3125
URBAN	-0.08508	-0.08047	0.99797	0.36469	1.00000	0.99878	0.08369	0.21749
	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
URBAN_INSIDEURBANAREAS	-0.08338	-0.07878	0.99552	0.36799	0.99878	1.00000	0.03442	0.19525
	<.0001	<.0001	<.0001	<.0001	<.0001		0.0509	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
URBAN_URBANCLUSTER	-0.03926	-0.03890	0.10856	-0.04597	0.08369	0.03442	1.00000	0.46192
	0.0259	0.0273	<.0001	0.0102	<.0001	0.0509		<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
RURAL_POP	0.24580	0.19495	0.27398	-0.03922	0.21749	0.19525	0.46192	1.00000
	<.0001	<.0001	<.0001	0.0284	<.0001	<.0001	<.0001	
	3219	3219	3141	3125	3219	3219	3219	3219
MEAN_DTR	-0.06956	-0.05527	0.00740	-0.05910	0.01100	0.00662	0.08902	-0.00762
	<.0001	0.0017	0.6785	0.0009	0.5328	0.7073	<.0001	0.6655
	3219	3219	3141	3125	3219	3219	3219	3219
seas_homes	0.03661	0.04535	0.36979	0.12285	0.36241	0.35269	0.21783	0.29297
	0.0378	0.0101	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
MIG_1999_1998	0.04949	0.04358	-0.53858	-0.37720	-0.55219	-0.55786	0.08281	0.14057
	0.0055	0.0146	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3141	3141	3141	3125	3141	3141	3141	3141
INCOME	-0.08995	-0.09605	0.28125	0.13185	0.25313	0.24559	0.16719	0.34978
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3217	3217	3139	3124	3217	3217	3217	3217
hgwy_dist	-0.01526	-0.02303	0.44151	-0.00208	0.41347	0.39669	0.36633	0.58502
	0.3943	0.1988	<.0001	0.9076	<.0001	<.0001	<.0001	<.0001
	3116	3116	3116	3116	3116	3116	3116	3116
rest_areas	-0.07395	-0.06606	0.19238	-0.00482	0.17955	0.16995	0.20443	0.28344
	<.0001	0.0002	<.0001	0.7878	<.0001	<.0001	<.0001	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219

Table 4.4 – Correlation matrix continued.

Covaraite Matrix: Entire US

The CORR Procedure

	MEAN_DTR	seas_ homes	MIG_ 1999_ 1998	INCOME	hgwy_dist	rest_ areas	wal_sams	CMPG_SUM	PICNIC_ SUM
BA_20	-0.06956	0.03661	0.04949	-0.08995	-0.01526	-0.07395	-0.05288	0.09636	0.08657
	<.0001	0.0378	0.0055	<.0001	0.3943	<.0001	0.0031	<.0001	<.0001
	3219	3219	3141	3217	3116	3219	3125	3219	3219
BA_50	-0.05527	0.04535	0.04358	-0.09605	-0.02303	-0.06606	-0.04320	0.09677	0.07021
	0.0017	0.0101	0.0146	<.0001	0.1988	0.0002	0.0157	<.0001	<.0001
	3219	3219	3141	3217	3116	3219	3125	3219	3219
pop_99	0.00740	0.36979	-0.53858	0.28125	0.44151	0.19238	0.75708	0.00315	0.03336
	0.6785	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.8597	0.0616
	3141	3141	3141	3139	3116	3141	3125	3141	3141
popdens99	-0.05910	0.12285	-0.37720	0.13185	-0.00208	-0.00482	0.10801	-0.01697	-0.00845
	0.0009	<.0001	<.0001	<.0001	0.9076	0.7878	<.0001	0.3428	0.6369
	3125	3125	3125	3124	3116	3125	3125	3125	3125
URBAN	0.01100	0.36241	-0.55219	0.25313	0.41347	0.17955	0.74754	0.00225	0.02930
	0.5328	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.8986	0.0965
	3219	3219	3141	3217	3116	3219	3125	3219	3219
URBAN_INSIDEURBANAREAS	0.00662	0.35269	-0.55786	0.24559	0.39669	0.16995	0.74099	0.00038	0.02702
	0.7073	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.9829	0.1253
	3219	3219	3141	3217	3116	3219	3125	3219	3219
URBAN_URBANCLUSTER	0.08902	0.21783	0.08281	0.16719	0.36633	0.20443	0.17717	0.03787	0.04774
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0317	0.0067
	3219	3219	3141	3217	3116	3219	3125	3219	3219
RURAL_POP	-0.00762	0.29297	0.14057	0.34978	0.58502	0.28344	0.42161	0.03625	0.09940
	0.6655	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0397	<.0001
	3219	3219	3141	3217	3116	3219	3125	3219	3219
MEAN_DTR	1.00000	0.14245	0.04012	0.01591	0.12240	0.06262	0.00884	-0.00088	-0.01119
		<.0001	0.0246	0.3669	<.0001	0.0004	0.6212	0.9600	0.5258
	3219	3219	3141	3217	3116	3219	3125	3219	3219
seas_homes	0.14245	1.00000	0.10063	0.14642	0.33909	0.19877	0.39135	0.00114	0.00209
	<.0001		<.0001	<.0001	<.0001	<.0001	<.0001	0.9483	0.9055
	3219	3219	3141	3217	3116	3219	3125	3219	3219
MIG_1999_1998	0.04012	0.10063	1.00000	0.08514	-0.05496	0.04827	-0.06943	0.03218	0.04887
	0.0246	<.0001		<.0001	0.0021	0.0068	0.0001	0.0713	0.0062
	3141	3141	3141	3139	3116	3141	3125	3141	3141
INCOME	0.01591	0.14642	0.08514	1.00000	0.16766	0.19114	0.30863	-0.02775	0.02695
	0.3669	<.0001	<.0001		<.0001	<.0001	<.0001	0.1155	0.1264
	3217	3217	3139	3217	3115	3217	3124	3217	3217
hgwy_dist	0.12240	0.33909	-0.05496	0.16766	1.00000	0.44138	0.51224	0.01696	0.07255
	<.0001	<.0001	0.0021	<.0001		<.0001	<.0001	0.3438	<.0001
	3116	3116	3116	3115	3116	3116	3116	3116	3116
rest_areas	0.06262	0.19877	0.04827	0.19114	0.44138	1.00000	0.23565	-0.01151	0.02606
	0.0004	<.0001	0.0068	<.0001	<.0001		<.0001	0.5137	0.1394
	3219	3219	3141	3217	3116	3219	3125	3219	3219

Table 4.4 – Correlation matrix continued.

Covaraite Matrix: Entire US

The CORR Procedure

	WOOD	wood_pop	wood_area	DIST_1999	sawmill	ACRE_ CHRIS_ FARM	NURSERIES	nurse_ acreopen	countyarea_ acres
BA_20	0.14412	0.22652	0.18317	-0.34301	0.30129	0.10216	0.00528	-0.05112	-0.11920
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.7648	0.0037	<.0001
	3219	3219	3125	3125	3219	3219	3219	3219	3125
BA_50	0.12102	0.24636	0.13189	-0.30699	0.26375	0.07285	-0.01516	-0.04137	-0.09949
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.3899	0.0189	<.0001
	3219	3219	3125	3125	3219	3219	3219	3219	3125
pop_99	0.24623	-0.14603	0.06789	0.08092	0.01480	0.10336	0.44172	0.26856	0.02808
	<.0001	<.0001	0.0001	<.0001	0.4069	<.0001	<.0001	<.0001	0.1165
	3141	3141	3125	3125	3141	3141	3141	3141	3125
popdens99	-0.01752	-0.09029	0.15391	-0.04852	-0.03061	0.00210	0.04083	0.01832	-0.03291
	0.3275	<.0001	<.0001	0.0067	0.0871	0.9067	0.0225	0.3059	0.0658
	3125	3125	3125	3125	3125	3125	3125	3125	3125
URBAN	0.21637	-0.13994	0.05619	0.08817	-0.00804	0.08223	0.41444	0.25541	0.02775
	<.0001	<.0001	0.0017	<.0001	0.6483	<.0001	<.0001	<.0001	0.1209
	3219	3219	3125	3125	3219	3219	3219	3219	3125
URBAN_INSIDEURBANAREAS	0.20017	-0.13437	0.05334	0.08181	-0.01529	0.07448	0.40287	0.24449	0.02392
	<.0001	<.0001	0.0029	<.0001	0.3857	<.0001	<.0001	<.0001	0.1813
	3219	3219	3125	3125	3219	3219	3219	3219	3125
URBAN_URBANCLUSTER	0.34004	-0.12069	0.06138	0.13479	0.14597	0.16148	0.25810	0.23561	0.07970
	<.0001	<.0001	0.0006	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3219	3219	3125	3125	3219	3219	3219	3219	3125
RURAL_POP	0.53078	-0.08760	0.20204	-0.09528	0.40026	0.37263	0.52925	0.30162	0.01888
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.2914
	3219	3219	3125	3125	3219	3219	3219	3219	3125
MEAN_DTR	0.11313	0.18991	-0.10266	0.25469	0.02109	-0.02089	0.02749	0.02877	0.10144
	<.0001	<.0001	<.0001	<.0001	0.2315	0.2360	0.1189	0.1027	<.0001
	3219	3219	3125	3125	3219	3219	3219	3219	3125
seas_homes	0.25271	0.01941	0.06232	0.09328	0.13004	0.11201	0.34717	0.27772	0.08149
	<.0001	0.2710	0.0005	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3219	3219	3125	3125	3219	3219	3219	3219	3125
MIG_1999_1998	0.01140	0.02446	-0.00429	0.01775	0.01893	0.02600	0.04363	0.08586	0.02841
	0.5232	0.1706	0.8104	0.3212	0.2889	0.1451	0.0145	<.0001	0.1123
	3141	3141	3125	3125	3141	3141	3141	3141	3125
INCOME	0.15716	-0.14904	0.12577	-0.04219	0.01505	0.21947	0.32222	0.20195	-0.01620
	<.0001	<.0001	<.0001	0.0184	0.3935	<.0001	<.0001	<.0001	0.3653
	3217	3217	3124	3124	3217	3217	3217	3217	3124
hgwy_dist	0.43357	-0.01212	-0.02645	0.21354	0.29922	0.19827	0.40884	0.24381	0.27305
	<.0001	0.4988	0.1398	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3116	3116	3116	3116	3116	3116	3116	3116	3116
rest_areas	0.21896	-0.05684	0.00367	0.12380	0.09038	0.13714	0.25451	0.16190	0.10320
	<.0001	0.0013	0.8373	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3219	3219	3125	3125	3219	3219	3219	3219	3125

Table 4.4 – Correlation matrix continued.

Covaraite Matrix: Entire US

The CORR Procedure

	park_area	airports	TREAT_99	POINT_X	POINT_Y	TREAT_99
BA_20	-0.05873	-0.13311	0.03379	0.44793	-0.25208	0.03379
	0.0009	<.0001	0.0553	<.0001	<.0001	0.0553
	3219	3219	3219	3125	3125	3219
BA_50	-0.05047	-0.10920	0.04322	0.36708	-0.21762	0.04322
	0.0042	<.0001	0.0142	<.0001	<.0001	0.0142
	3219	3219	3219	3125	3125	3219
pop_99	0.05956	0.66152	0.00662	0.00397	-0.00165	0.00662
	0.0008	<.0001	0.7107	0.8244	0.9267	0.7107
	3141	3141	3141	3125	3125	3141
popdens99	-0.00809	0.07653	-0.00478	0.11948	0.00959	-0.00478
	0.6511	<.0001	0.7895	<.0001	0.5922	0.7895
	3125	3125	3125	3125	3125	3125
URBAN	0.05847	0.64037	0.00252	-0.01115	-0.00058	0.00252
	0.0009	<.0001	0.8863	0.5331	0.9740	0.8863
	3219	3219	3219	3125	3125	3219
URBAN_INSIDEURBANAREAS	0.05640	0.62721	0.00074	-0.00799	-0.00084	0.00074
	0.0014	<.0001	0.9663	0.6552	0.9625	0.9663
	3219	3219	3219	3125	3125	3219
URBAN_URBANCLUSTER	0.04541	0.30364	0.03600	-0.06515	0.00524	0.03600
	0.0100	<.0001	0.0411	0.0003	0.7695	0.0411
	3219	3219	3219	3125	3125	3219
RURAL_POP	0.03065	0.49046	0.06364	0.26907	-0.04828	0.06364
	0.0820	<.0001	0.0003	<.0001	0.0070	0.0003
	3219	3219	3219	3125	3125	3219
MEAN_DTR	0.09111	0.08861	-0.01600	-0.20641	0.03879	-0.01600
	<.0001	<.0001	0.3641	<.0001	0.0301	0.3641
	3219	3219	3219	3125	3125	3219
seas_homes	0.07975	0.36661	-0.00400	0.07661	-0.01101	-0.00400
	<.0001	<.0001	0.8204	<.0001	0.5383	0.8204
	3219	3219	3219	3125	3125	3219
MIG_1999_1998	0.00316	-0.10615	0.00503	-0.02032	-0.04973	0.00503
	0.8595	<.0001	0.7781	0.2561	0.0054	0.7781
	3141	3141	3141	3125	3125	3141
INCOME	0.04854	0.35689	0.03349	0.08492	0.22567	0.03349
	0.0059	<.0001	0.0576	<.0001	<.0001	0.0576
	3217	3217	3217	3124	3124	3217
hgwy_dist	0.16635	0.57671	0.02889	-0.04653	0.07249	0.02889
	<.0001	<.0001	0.1069	0.0094	<.0001	0.1069
	3116	3116	3116	3116	3116	3116
rest_areas	0.03791	0.25746	-0.00886	-0.05465	0.08400	-0.00886
	0.0315	<.0001	0.6155	0.0022	<.0001	0.6155
	3219	3219	3219	3125	3125	3219

Table 4.4 – Correlation matrix continued.

Covaraite Matrix: Entire US

The CORR Procedure

	BA_20	BA_50	pop_99	popdens99	URBAN	URBAN_ INSIDEURBANAREAS	URBAN_ URBANCLUSTER	RURAL_POP
wal_sams	-0.05288	-0.04320	0.75708	0.10801	0.74754	0.74099	0.17717	0.42161
	0.0031	0.0157	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3125	3125	3125	3125	3125	3125	3125	3125
CMPG_SUM	0.09636	0.09677	0.00315	-0.01697	0.00225	0.00038	0.03787	0.03625
	<.0001	<.0001	0.8597	0.3428	0.8986	0.9829	0.0317	0.0397
	3219	3219	3141	3125	3219	3219	3219	3219
PICNIC_SUM	0.08657	0.07021	0.03336	-0.00845	0.02930	0.02702	0.04774	0.09940
	<.0001	<.0001	0.0616	0.6369	0.0965	0.1253	0.0067	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
WOOD	0.14412	0.12102	0.24623	-0.01752	0.21637	0.20017	0.34004	0.53078
	<.0001	<.0001	<.0001	0.3275	<.0001	<.0001	<.0001	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
wood_pop	0.22652	0.24636	-0.14603	-0.09029	-0.13994	-0.13437	-0.12069	-0.08760
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
wood_area	0.18317	0.13189	0.06789	0.15391	0.05619	0.05334	0.06138	0.20204
	<.0001	<.0001	0.0001	<.0001	0.0017	0.0029	0.0006	<.0001
	3125	3125	3125	3125	3125	3125	3125	3125
DIST_1999	-0.34301	-0.30699	0.08092	-0.04852	0.08817	0.08181	0.13479	-0.09528
	<.0001	<.0001	<.0001	0.0067	<.0001	<.0001	<.0001	<.0001
	3125	3125	3125	3125	3125	3125	3125	3125
sawmill	0.30129	0.26375	0.01480	-0.03061	-0.00804	-0.01529	0.14597	0.40026
	<.0001	<.0001	0.4069	0.0871	0.6483	0.3857	<.0001	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
ACRE_CHRIS_FARM	0.10216	0.07285	0.10336	0.00210	0.08223	0.07448	0.16148	0.37263
	<.0001	<.0001	<.0001	0.9067	<.0001	<.0001	<.0001	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
NURSERIES	0.00528	-0.01516	0.44172	0.04083	0.41444	0.40287	0.25810	0.52925
	0.7648	0.3899	<.0001	0.0225	<.0001	<.0001	<.0001	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
nurse_acreopen	-0.05112	-0.04137	0.26856	0.01832	0.25541	0.24449	0.23561	0.30162
	0.0037	0.0189	<.0001	0.3059	<.0001	<.0001	<.0001	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219
countyarea_acres	-0.11920	-0.09949	0.02808	-0.03291	0.02775	0.02392	0.07970	0.01888
	<.0001	<.0001	0.1165	0.0658	0.1209	0.1813	<.0001	0.2914
	3125	3125	3125	3125	3125	3125	3125	3125
park_area	-0.05873	-0.05047	0.05956	-0.00809	0.05847	0.05640	0.04541	0.03065
	0.0009	0.0042	0.0008	0.6511	0.0009	0.0014	0.0100	0.0820
	3219	3219	3141	3125	3219	3219	3219	3219
airports	-0.13311	-0.10920	0.66152	0.07653	0.64037	0.62721	0.30364	0.49046
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3219	3219	3141	3125	3219	3219	3219	3219

Table 4.4 – Correlation matrix continued.

Covaraite Matrix: Entire US

The CORR Procedure

	MEAN_DTR	seas_ homes	MIG_ 1999_ 1998	INCOME	hgwy_dist	rest_ areas	wal_sams	CMPG_SUM	PICNIC_ SUM
wal_sams	0.00884	0.39135	-0.06943	0.30863	0.51224	0.23565	1.00000	0.03936	0.09123
	0.6212	<.0001	0.0001	<.0001	<.0001	<.0001		0.0278	<.0001
	3125	3125	3125	3124	3116	3125	3125	3125	3125
CMPG_SUM	-0.00088	0.00114	0.03218	-0.02775	0.01696	-0.01151	0.03936	1.00000	0.62352
	0.9600	0.9483	0.0713	0.1155	0.3438	0.5137	0.0278		<.0001
	3219	3219	3141	3217	3116	3219	3125	3219	3219
PICNIC_SUM	-0.01119	0.00209	0.04887	0.02695	0.07255	0.02606	0.09123	0.62352	1.00000
	0.5258	0.9055	0.0062	0.1264	<.0001	0.1394	<.0001	<.0001	
	3219	3219	3141	3217	3116	3219	3125	3219	3219
WOOD	0.11313	0.25271	0.01140	0.15716	0.43357	0.21896	0.18927	0.06124	0.07567
	<.0001	<.0001	0.5232	<.0001	<.0001	<.0001	<.0001	0.0005	<.0001
	3219	3219	3141	3217	3116	3219	3125	3219	3219
wood_pop	0.18991	0.01941	0.02446	-0.14904	-0.01212	-0.05684	-0.19999	0.05361	0.00353
	<.0001	0.2710	0.1706	<.0001	0.4988	0.0013	<.0001	0.0023	0.8413
	3219	3219	3141	3217	3116	3219	3125	3219	3219
wood_area	-0.10266	0.06232	-0.00429	0.12577	-0.02645	0.00367	0.03768	0.00191	0.01356
	<.0001	0.0005	0.8104	<.0001	0.1398	0.8373	0.0352	0.9152	0.4487
	3125	3125	3125	3124	3116	3125	3125	3125	3125
DIST_1999	0.25469	0.09328	0.01775	-0.04219	0.21354	0.12380	0.05430	0.02451	0.00883
	<.0001	<.0001	0.3212	0.0184	<.0001	<.0001	0.0024	0.1708	0.6218
	3125	3125	3125	3124	3116	3125	3125	3125	3125
sawmill	0.02109	0.13004	0.01893	0.01505	0.29922	0.09038	0.04953	0.02785	0.10592
	0.2315	<.0001	0.2889	0.3935	<.0001	<.0001	0.0056	0.1142	<.0001
	3219	3219	3141	3217	3116	3219	3125	3219	3219
ACRE_CHRIS_FARM	-0.02089	0.11201	0.02600	0.21947	0.19827	0.13714	0.08961	-0.01517	0.04016
	0.2360	<.0001	0.1451	<.0001	<.0001	<.0001	<.0001	0.3896	0.0227
	3219	3219	3141	3217	3116	3219	3125	3219	3219
NURSERIES	0.02749	0.34717	0.04363	0.32222	0.40884	0.25451	0.42741	-0.00091	0.03961
	0.1189	<.0001	0.0145	<.0001	<.0001	<.0001	<.0001	0.9589	0.0246
	3219	3219	3141	3217	3116	3219	3125	3219	3219
nurse_acreopen	0.02877	0.27772	0.08586	0.20195	0.24381	0.16190	0.31530	0.00643	0.01264
	0.1027	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.7152	0.4735
	3219	3219	3141	3217	3116	3219	3125	3219	3219
countyarea_acres	0.10144	0.08149	0.02841	-0.01620	0.27305	0.10320	0.03236	-0.00611	-0.00663
	<.0001	<.0001	0.1123	0.3653	<.0001	<.0001	0.0705	0.7328	0.7109
	3125	3125	3125	3124	3116	3125	3125	3125	3125
park_area	0.09111	0.07975	0.00316	0.04854	0.16635	0.03791	0.03809	-0.00593	-0.00760
	<.0001	<.0001	0.8595	0.0059	<.0001	0.0315	0.0333	0.7365	0.6665
	3219	3219	3141	3217	3116	3219	3125	3219	3219
airports	0.08861	0.36661	-0.10615	0.35689	0.57671	0.25746	0.69557	0.02428	0.06205
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.1685	0.0004
	3219	3219	3141	3217	3116	3219	3125	3219	3219

Table 4.4 – Correlation matrix continued.

Covaraite Matrix: Entire US

The CORR Procedure

	WOOD	wood_pop	wood_area	DIST_1999	sawmill	ACRE_ CHRIS_ FARM	NURSERIES	nurse_ acreopen	countyarea_ acres
wal_sams	0.18927	-0.19999	0.03768	0.05430	0.04953	0.08961	0.42741	0.31530	0.03236
	<.0001	<.0001	0.0352	0.0024	0.0056	<.0001	<.0001	<.0001	0.0705
	3125	3125	3125	3125	3125	3125	3125	3125	3125
CMPG_SUM	0.06124	0.05361	0.00191	0.02451	0.02785	-0.01517	-0.00091	0.00643	-0.00611
	0.0005	0.0023	0.9152	0.1708	0.1142	0.3896	0.9589	0.7152	0.7328
	3219	3219	3125	3125	3219	3219	3219	3219	3125
PICNIC_SUM	0.07567	0.00353	0.01356	0.00883	0.10592	0.04016	0.03961	0.01264	-0.00663
	<.0001	0.8413	0.4487	0.6218	<.0001	0.0227	0.0246	0.4735	0.7109
	3219	3219	3125	3125	3219	3219	3219	3219	3125
WOOD	1.00000	0.35971	0.35567	0.27204	0.35692	0.37838	0.47734	0.18309	0.13548
		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3219	3219	3125	3125	3219	3219	3219	3219	3125
wood_pop	0.35971	1.00000	0.14248	0.19620	0.19756	0.04249	-0.06942	-0.08658	0.18309
	<.0001		<.0001	<.0001	<.0001	0.0159	<.0001	<.0001	<.0001
	3219	3219	3125	3125	3219	3219	3219	3219	3125
wood_area	0.35567	0.14248	1.00000	-0.13078	0.18402	0.19896	0.18025	0.03762	-0.07508
	<.0001	<.0001		<.0001	<.0001	<.0001	<.0001	0.0355	<.0001
	3125	3125	3125	3125	3125	3125	3125	3125	3125
DIST_1999	0.27204	0.19620	-0.13078	1.00000	-0.12522	-0.02308	0.10265	0.10041	0.35080
	<.0001	<.0001	<.0001		<.0001	0.1970	<.0001	<.0001	<.0001
	3125	3125	3125	3125	3125	3125	3125	3125	3125
sawmill	0.35692	0.19756	0.18402	-0.12522	1.00000	0.30691	0.19092	0.04598	-0.00577
	<.0001	<.0001	<.0001	<.0001		<.0001	<.0001	0.0091	0.7470
	3219	3219	3125	3125	3219	3219	3219	3219	3125
ACRE_CHRIS_FARM	0.37838	0.04249	0.19896	-0.02308	0.30691	1.00000	0.58659	0.33493	-0.01308
	<.0001	0.0159	<.0001	0.1970	<.0001		<.0001	<.0001	0.4648
	3219	3219	3125	3125	3219	3219	3219	3219	3125
NURSERIES	0.47734	-0.06942	0.18025	0.10265	0.19092	0.58659	1.00000	0.70344	0.01543
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001		<.0001	0.3884
	3219	3219	3125	3125	3219	3219	3219	3219	3125
nurse_acreopen	0.18309	-0.08658	0.03762	0.10041	0.04598	0.33493	0.70344	1.00000	0.01990
	<.0001	<.0001	0.0355	<.0001	0.0091	<.0001	<.0001		0.2661
	3219	3219	3125	3125	3219	3219	3219	3219	3125
countyarea_acres	0.13548	0.18309	-0.07508	0.35080	-0.00577	-0.01308	0.01543	0.01990	1.00000
	<.0001	<.0001	<.0001	<.0001	0.7470	0.4648	0.3884	0.2661	
	3125	3125	3125	3125	3125	3125	3125	3125	3125
park_area	0.09082	0.11353	-0.02699	0.25864	-0.01180	-0.00472	0.05476	0.02725	0.78769
	<.0001	<.0001	0.1314	<.0001	0.5033	0.7889	0.0019	0.1221	<.0001
	3219	3219	3125	3125	3219	3219	3219	3219	3125
airports	0.34842	-0.07574	0.02092	0.23089	0.11709	0.23215	0.49480	0.37512	0.30090
	<.0001	<.0001	0.2424	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3219	3219	3125	3125	3219	3219	3219	3219	3125

Table 4.4 – Correlation matrix continued.

Covaraite Matrix: Entire US

The CORR Procedure

	park_area	airports	TREAT_99	POINT_X	POINT_Y	TREAT_99
wal_sams	0.03809	0.69557	0.01195	0.02935	-0.10228	0.01195
	0.0333	<.0001	0.5041	0.1010	<.0001	0.5041
	3125	3125	3125	3125	3125	3125
CMPG_SUM	-0.00593	0.02428	-0.01386	-0.04141	-0.05053	-0.01386
	0.7365	0.1685	0.4318	0.0206	0.0047	0.4318
	3219	3219	3219	3125	3125	3219
PICNIC_SUM	-0.00760	0.06205	-0.01322	0.00902	-0.03580	-0.01322
	0.6665	0.0004	0.4535	0.6143	0.0454	0.4535
	3219	3219	3219	3125	3125	3219
WOOD	0.09082	0.34842	0.03474	-0.16856	0.25479	0.03474
	<.0001	<.0001	0.0487	<.0001	<.0001	0.0487
	3219	3219	3219	3125	3125	3219
wood_pop	0.11353	-0.07574	0.01348	-0.20458	0.28366	0.01348
	<.0001	<.0001	0.4445	<.0001	<.0001	0.4445
	3219	3219	3219	3125	3125	3219
wood_area	-0.02699	0.02092	0.01734	0.21461	0.06777	0.01734
	0.1314	0.2424	0.3327	<.0001	0.0001	0.3327
	3125	3125	3125	3125	3125	3125
DIST_1999	0.25864	0.23089	-0.07555	-0.83521	0.21821	-0.07555
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	3125	3125	3125	3125	3125	3125
sawmill	-0.01180	0.11709	0.03261	0.27631	0.11160	0.03261
	0.5033	<.0001	0.0643	<.0001	<.0001	0.0643
	3219	3219	3219	3125	3125	3219
ACRE_CHRIS_FARM	-0.00472	0.23215	0.03903	0.10784	0.17191	0.03903
	0.7889	<.0001	0.0268	<.0001	<.0001	0.0268
	3219	3219	3219	3125	3125	3219
NURSERIES	0.05476	0.49480	0.02404	0.03185	0.06140	0.02404
	0.0019	<.0001	0.1726	0.0750	0.0006	0.1726
	3219	3219	3219	3125	3125	3219
nurse_acreopen	0.02725	0.37512	0.00336	-0.02829	-0.00777	0.00336
	0.1221	<.0001	0.8489	0.1138	0.6643	0.8489
	3219	3219	3219	3125	3125	3219
countyarea_acres	0.78769	0.30090	-0.00540	-0.28066	0.27240	-0.00540
	<.0001	<.0001	0.7629	<.0001	<.0001	0.7629
	3125	3125	3125	3125	3125	3125
park_area	1.00000	0.26599	-0.00516	-0.16510	0.23278	-0.00516
		<.0001	0.7696	<.0001	<.0001	0.7696
	3219	3219	3219	3125	3125	3219
airports	0.26599	1.00000	0.03920	-0.10765	0.13400	0.03920
	<.0001		0.0262	<.0001	<.0001	0.0262
	3219	3219	3219	3125	3125	3219

Table 4.4 – Correlation matrix continued.

Covaraite Matrix: Entire US

The CORR Procedure

	BA_20	BA_50	pop_99	popdens99	URBAN	URBAN_ INSIDEURBANAREAS	URBAN_ URBANCLUSTER	RURAL_POP
TREAT_99	0.03379	0.04322	0.00662	-0.00478	0.00252	0.00074	0.03600	0.06364
	0.0553	0.0142	0.7107	0.7895	0.8863	0.9663	0.0411	0.0003
	3219	3219	3141	3125	3219	3219	3219	3219
POINT_X	0.44793	0.36708	0.00397	0.11948	-0.01115	-0.00799	-0.06515	0.26907
	<.0001	<.0001	0.8244	<.0001	0.5331	0.6552	0.0003	<.0001
	3125	3125	3125	3125	3125	3125	3125	3125
POINT_Y	-0.25208	-0.21762	-0.00165	0.00959	-0.00058	-0.00084	0.00524	-0.04828
	<.0001	<.0001	0.9267	0.5922	0.9740	0.9625	0.7695	0.0070
	3125	3125	3125	3125	3125	3125	3125	3125
TREAT_99	0.03379	0.04322	0.00662	-0.00478	0.00252	0.00074	0.03600	0.06364
	0.0553	0.0142	0.7107	0.7895	0.8863	0.9663	0.0411	0.0003
	3219	3219	3141	3125	3219	3219	3219	3219

	MEAN_DTR	seas_ homes	MIG_ 1999_ 1998	INCOME	hgwy_dist	rest_ areas	wal_sams	CMPG_SUM	PICNIC_ SUM
TREAT_99	-0.01600	-0.00400	0.00503	0.03349	0.02889	-0.00886	0.01195	-0.01386	-0.01322
	0.3641	0.8204	0.7781	0.0576	0.1069	0.6155	0.5041	0.4318	0.4535
	3219	3219	3141	3217	3116	3219	3125	3219	3219
POINT_X	-0.20641	0.07661	-0.02032	0.08492	-0.04653	-0.05465	0.02935	-0.04141	0.00902
	<.0001	<.0001	0.2561	<.0001	0.0094	0.0022	0.1010	0.0206	0.6143
	3125	3125	3125	3124	3116	3125	3125	3125	3125
POINT_Y	0.03879	-0.01101	-0.04973	0.22567	0.07249	0.08400	-0.10228	-0.05053	-0.03580
	0.0301	0.5383	0.0054	<.0001	<.0001	<.0001	<.0001	0.0047	0.0454
	3125	3125	3125	3124	3116	3125	3125	3125	3125
TREAT_99	-0.01600	-0.00400	0.00503	0.03349	0.02889	-0.00886	0.01195	-0.01386	-0.01322
	0.3641	0.8204	0.7781	0.0576	0.1069	0.6155	0.5041	0.4318	0.4535
	3219	3219	3141	3217	3116	3219	3125	3219	3219

	WOOD	wood_pop	wood_area	DIST_1999	sawmill	ACRE_ CHRIS_ FARM	NURSERIES	nurse_ acreopen	countyarea_ acres
TREAT_99	0.03474	0.01348	0.01734	-0.07555	0.03261	0.03903	0.02404	0.00336	-0.00540
	0.0487	0.4445	0.3327	<.0001	0.0643	0.0268	0.1726	0.8489	0.7629
	3219	3219	3125	3125	3219	3219	3219	3219	3125
POINT_X	-0.16856	-0.20458	0.21461	-0.83521	0.27631	0.10784	0.03185	-0.02829	-0.28066
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0750	0.1138	<.0001
	3125	3125	3125	3125	3125	3125	3125	3125	3125

Table 4.4 – Correlation matrix continued.

Covaraite Matrix: Entire US

The CORR Procedure

	WOOD	wood_pop	wood_area	DIST_1999	sawmill	ACRE_ CHRIS_ FARM	NURSERIES	nurse_ acreopen	countyarea_ acres
POINT_Y	0.25479	0.28366	0.06777	0.21821	0.11160	0.17191	0.06140	-0.00777	0.27240
	<.0001	<.0001	0.0001	<.0001	<.0001	<.0001	0.0006	0.6643	<.0001
	3125	3125	3125	3125	3125	3125	3125	3125	3125
TREAT_99	0.03474	0.01348	0.01734	-0.07555	0.03261	0.03903	0.02404	0.00336	-0.00540
	0.0487	0.4445	0.3327	<.0001	0.0643	0.0268	0.1726	0.8489	0.7629
	3219	3219	3125	3125	3219	3219	3219	3219	3125

	park_area	airports	TREAT_99	POINT_X	POINT_Y	TREAT_99
TREAT_99	-0.00516	0.03920	1.00000	0.03655	0.02507	1.00000
	0.7696	0.0262		0.0411	0.1612	
	3219	3219	3219	3125	3125	3219
POINT_X	-0.16510	-0.10765	0.03655	1.00000	-0.36532	0.03655
	<.0001	<.0001	0.0411		<.0001	0.0411
	3125	3125	3125	3125	3125	3125
POINT_Y	0.23278	0.13400	0.02507	-0.36532	1.00000	0.02507
	<.0001	<.0001	0.1612	<.0001		0.1612
	3125	3125	3125	3125	3125	3125
TREAT_99	-0.00516	0.03920	1.00000	0.03655	0.02507	1.00000
	0.7696	0.0262		0.0411	0.1612	
	3219	3219	3219	3125	3125	3219

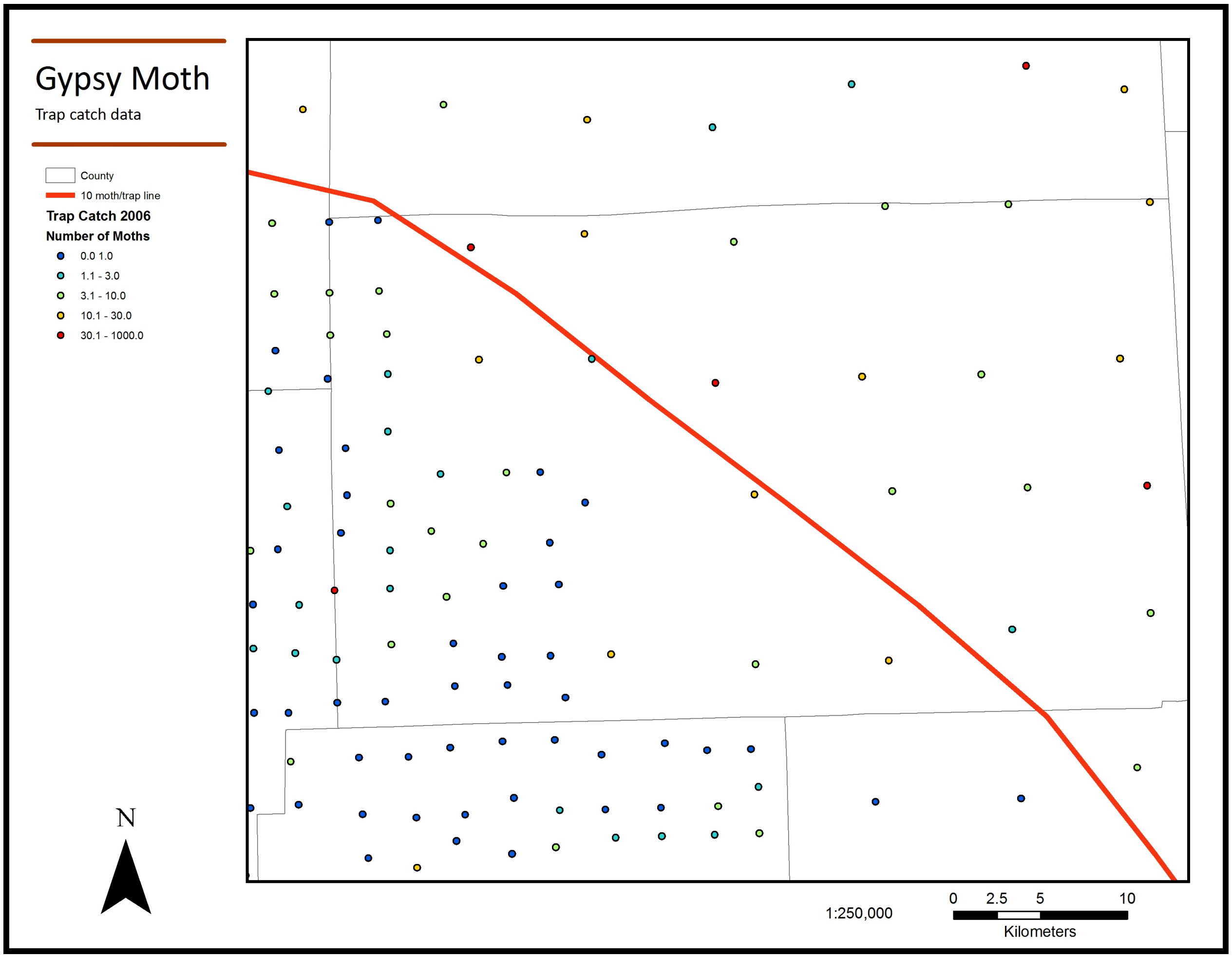


Figure 4.1 – Gypsy moth trapping grid. Traps dorsal to the 10 moth/trap line have a lower density than those distal the 10 moth/trap line.

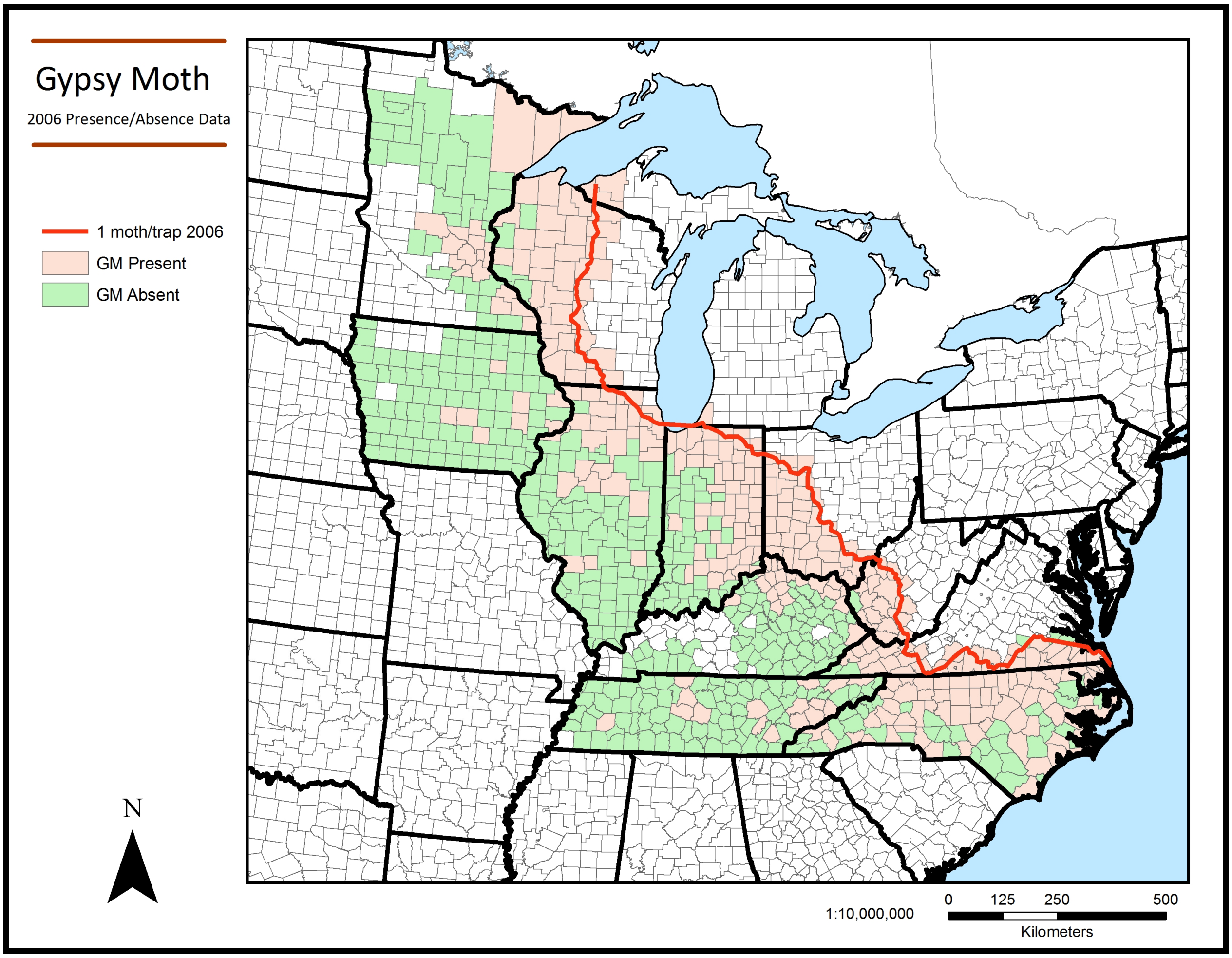


Figure 4.2 – An example of the dependent variable for the logistic regression, presence/absence of gypsy moth in a county.

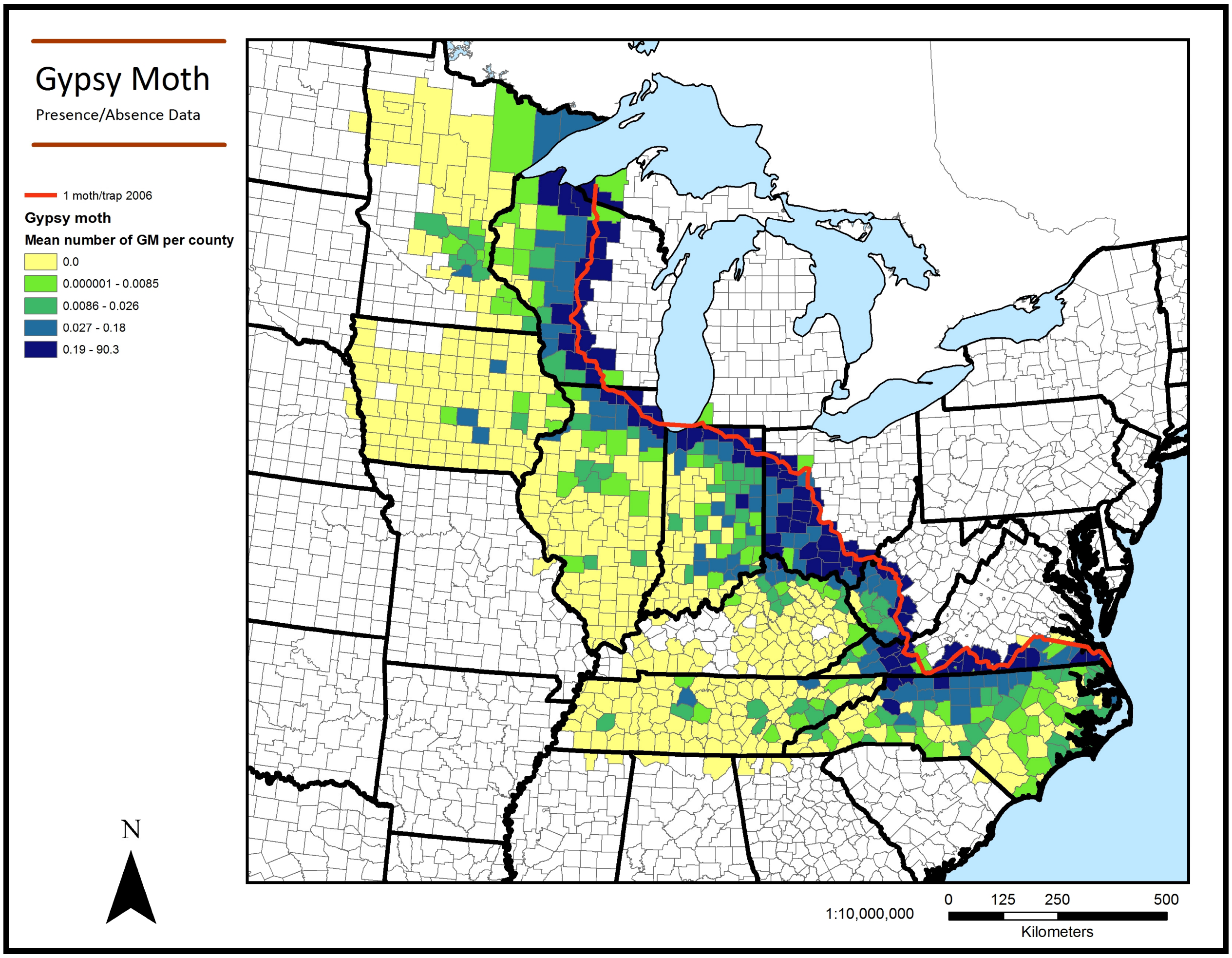


Figure 4.3 – An example of the dependent variable for the general linear model, the mean number of gypsy moths per trap.

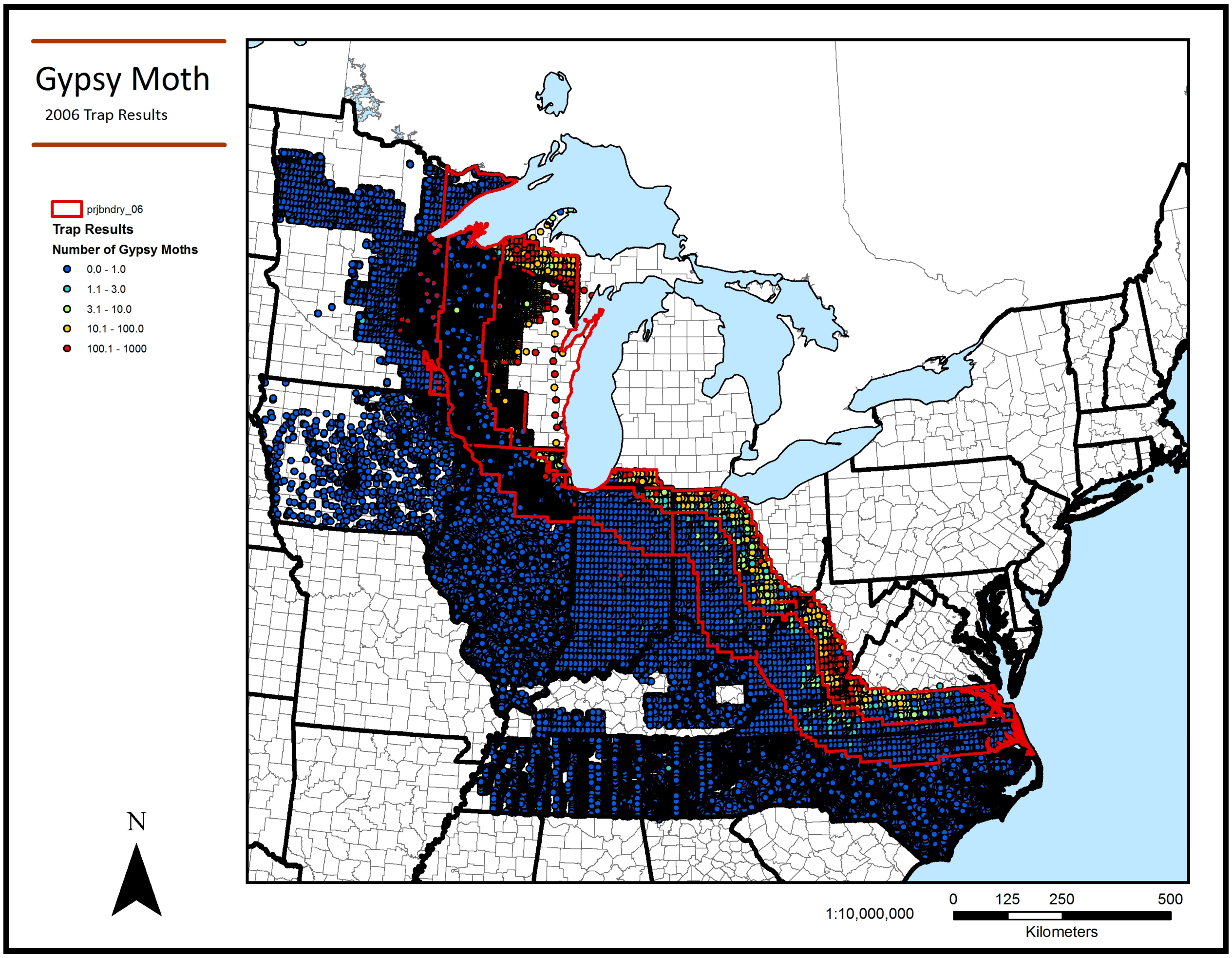


Figure 4.4 – Gypsy moth trap records depict the invasion front of gypsy moth and the spatial extent of data the analysis. The STS trapping zone lies within the red lines.

CHAPTER 5 - RESULTS

5.1 Introduction

This chapter begins with a brief discussion of the general linear model (Section 5.2). Then the logistic regression results for the sector models including variables representing each category of factors suggested by the conceptual framework are discussed (Section 5.3). Next, the results of diagnostic tests for spatial autocorrelation, including Moran's I for presence/absence and mean trap count (Section 5.4) and semivariograms for mean trap count (Section 5.5) are presented. Finally, the estimation results for the empirical specifications are presented (Section 5.6). The implications of these results for invasion theory and policy are discussed in the next chapter.

5.2 General linear model

The general linear model failed to control for spatial auto correlation (Figure 5.10 – 5.18, Appendix 5), therefore, it was dropped from the analysis and it will no longer be discussed. All models results that follow are logistic regression models.

5.3 Factor categories from the conceptual framework

For each year of the study, I estimated six logistic regressions, each with all of the potential variables representing one of the factors in the conceptual framework. Results of logistic regression models with and without a selection are presented here. Statistical significance for the empirical model without selection was determined at the $\alpha < 0.10$ level. Statistical significance for the stepwise selection of the factor category models was determined at the $\alpha < 0.05$ level. Though significance for the no selection model was determined at $\alpha < 0.10$, the benefit of this model is that the *P values* for all variables in the model can be examined. In contrast, the stepwise selection model removes insignificant variables, Thus, statistical significance is unknown for the insignificant variables.

The visitor attraction model yielded an average pseudo R square value of 0.039 (Table 5.1, Appendix 5). This category of variables was not a very strong predictor of gypsy moth presence. The stepwise selection process only produced statistically significant models for three years. Based on the estimation results without selection, I identified retirement county and seasonal homes as being the best predictors within this category, with statistical significance for six and seven years, respectively, out of a total of nine years. These variables are included in the final empirical model.

The source infestation category yielded an average pseudo R square value of 0.364 (Table 5.2, Appendix 5). This model category performed strongly, predominantly because of lagged distance to the nearest outbreak. When lagged four years, this variable was significant for six years; when lagged by only one year, this variable was significant in eight years. Due to multi-collinearity among the variables, only the distance to outbreak lagged by one year is selected for the empirical model.

The biophysical model category yielded an average pseudo R square value of 0.361 (Table 5.3, Appendix 5). This model category also performed strongly indicating the spatial dependence of the invasion process. County centroids (latitude and longitude) were significant each year, while county area was significant in five years. After controlling for these other variables, susceptible basal area is never significant. All four variables are included in the empirical model. As described below, county centroids are key to controlling for spatial autocorrelation by capturing unobserved spatially correlated factors. County area is indicative of the number of traps in a county.[6] Previous studies have found susceptible basal area to be significant (Sharov et al. 1999) when examining whether winter temperature or susceptible basal area are significant at predicting spread of gypsy moth but insignificant (Whitmire and Tobin 2006) when examining the persistence of low density colonies of gypsy moth. It is included here because susceptible basal area is theoretically one of the key biological determinants of the range of an invasive species.

The diffusion model yielded an average pseudo R square value of 0.521 (Table 5.4, Appendix 5). This model category was the strongest predictor of gypsy moth presence, distance to the 1 moth/trap line and lagged positive traps were both highly significant (*P values* < 0.0001) in all nine years. Both are included in the empirical model.

The demand for goods model yielded a mean pseudo R square of 0.162 (Table 5.5, Appendix 5) for the specification with all population variables and 0.145 for the specification with the gravity variable replacing all population variables (Table 5.6, Appendix 5). The gravity

[6] The number of traps may also be affected by local knowledge about the likelihood of catching gypsy moths, and thus is not included in the model due to endogeneity concerns. Its mean correlation with county area is 0.36.

variable was significant in five years while at least one of the population variables was significant seven out of nine years. Urban population inside urban areas was significant most often. However, this variable's coefficient is negative in some years because of multi-collinearity with other population variables (e.g. collinearity with urban population at .9988 with *P value* < 0.0001). In the model with population variables, income and household wood were significant six and seven years respectively. In the model with the gravity variable replacing the population variables, gravity and income were significant in six years, and households using wood was significant in eight years. The gravity variable, income, and household wood are included in the empirical model.

The transportation model yielded an average pseudo R square value of 0.097 (Table 5.7, Appendix 5). This model category was not a strong predictor of gypsy moth presence but may be important for capturing processes far from the generally infested zone. The number of big box retailers was significant in four years, airports were significant in six years, and highway distance was significant in seven years. Highway distance and airports are included in the empirical model.

5.4 Moran's I

Moran's I analyses generally suggested positive spatial autocorrelation in the two dependent variables: presence/absence in a county and the mean number of gypsy moths per trap (Table 5.8, Appendix 5). The mean Z scores for the entire temporal extent of the study were 31.04 and 7.83, respectively. In 2002 and 2004, the Z scores for the mean number of gypsy moths were less than 2, indicating no spatial autocorrelation. These two years, however, still yielded positive spatial autocorrelation using presence/absence data. Differences in spatial autocorrelation between presence/absence and mean observed values were predominantly a product of the reduced statistical power and resolution of the presence/absence data.

5.5 Semivariograms

The semivariograms of mean trap count tended to be anisotropic in the 0°, 45°, and 90° directions depending on the year (Figure 5.1 - 5.9, A, Appendix 5). The semivariogram range indicated a general spatial dependence up to approximately 200km.

I attempted to control spatial autocorrelation by progressively including variables: 1) X and Y coordinates, 2) X and Y coordinate squared and cubed, and 3) empirical model variables. Spatial autocorrelation was analyzed using the residual values from these models in the semivariogram. Including the X and Y coordinates in the logistic regression model reduces the value of the sill each year, generally by an order of magnitude for spatial dependence (Figures 5.1 - 5.9, B, Appendix 5). The next logistic regression model included X and Y coordinates and the coordinates' squared and cubed values, also reduced the sill when compared to the observational values. However, the sill and range were not reduced any more than the previous model (Figures 5.1 - 5.9, C, Appendix 5). When X and Y coordinates are included in the empirical model, the residuals are isotropic or at least display less anisotropy (Figures 5.1 - 5.9, D, Appendix 5).

5.6 Empirical models

Two empirical models were specified; 1) the empirical model, which includes at least one variable from each category described above, and 2) the empirical sector model, which includes additional variables from several of the categories. These empirical specifications were estimated using logistic regression without a selection and the coefficients are evaluated at $\alpha < 0.10$ significance level. Because there is still multicollinearity within and across categories, stepwise selection logit models are also estimated ($\alpha < 0.05$). Below I recap the variables used in the empirical model, define the variables used in the empirical sector model, and provide the results of both models.

The final empirical model included the following variables: 1) seasonal homes, 2) retirement counties, 3) the gravity variable, 4) income, 5) household wood, 6) airports, 7) highway distance, 8) lagged distance to nearest outbreak (1 year), 9) X coordinate, 10) Y coordinate, 11) susceptible basal area, 12) distance to 1 moth/trap line, and 13) lagged positive trap.

The empirical sector model included the above variables and 1) campgrounds, 2) nurseries, 3) sawmills, 4) Christmas trees, 5) rest areas, 6) Wal-Mart and Sam's Club, 7) county area, and 8) treatment. This model included additional variables to examine the effect of industries in quarantine that were not significant in their model categories.

The empirical model had a mean pseudo R square value of 0.565 (Table 5.9, Appendix 5). The stepwise selection procedure identified different sets of variables as significant in different years of the study. Every variable was statistically significant in at least one analysis year, as follows: gravity (significant two years, with negative coefficients), seasonal homes (significant one year, with positive coefficient), retirement county (significant one year, with negative coefficient), income (significant eight years, with positive coefficients), household wood (significant six years, with positive coefficients), airports (significant two years, with positive coefficient), highway distance (significant two years, with positive coefficients), lagged distance to nearest outbreak, 1 year (significant four years, with negative coefficients), susceptible basal area (significant one year, with positive coefficient), county latitude (significant five years, with positive coefficients), county longitude (significant four years, with positive coefficients), distance to the 1 moth/trap line (significant six years, with negative coefficient), and lagged positive trap (significant nine years, with positive coefficients).

The empirical sector model category resulted in an average pseudo R square value of 0.568 (Table 5.10, Appendix 5). Significant variables were gravity (significant three years, with negative coefficients), campground sum (significant four years, with both positive and negative coefficients), retirement county (significant 2 years, with negative coefficients), income (significant four years, with positive coefficients), nurseries (significant one year, with positive coefficient), sawmills (significant three years, with negative coefficients), household wood (significant six years, with positive coefficients), Christmas trees (significant two years, with positive coefficients), Wal-mart and Sam's club (significant two years, with positive coefficients), airports (significant two years, with positive coefficients), lagged distance to nearest outbreak, 1 year (significant four years, with negative coefficients), susceptible basal area (significant two years, with both positive and negative coefficients), county latitude (significant six years, with positive coefficients), county longitude (significant six years, with positive coefficients), distance to 1 moth/trap line (significant six years, with negative coefficients), lagged positive trap (significant nine years, with positive coefficients), treatment (significant one year, with positive coefficient).

The coefficients estimated by the logistic regression can be interpreted as the effect of a one unit change in the explanatory variables on the log odds of observing at least one moth in a trap in a given county and a given year. For example, the coefficient on retirement county indicates the difference in the log odds of a positive trap in a retirement county vs. other counties. These coefficients were converted to odds ratios (Table 5.11, Appendix 5). Odds ratios greater than one indicate that with the one unit increase in the independent this variable would be a stronger predictor of the dependent variable, while an odds ratio less than one would indicate the opposite, a weaker predictor of the dependent variable (SAS/STAT)[7]. The mean odds ratio of lagged positive trap and household wood over nine years has the largest odds ratio estimates, 4.47 and 3.45 respectively. A county were gypsy moth was present in the previous year is 4.47 times more likely to have gypsy moth in the next year. A one unit increase in household wood (1,000 more homes using wood as a primary heating source) in a county would make it 3.45 times more likely to have gypsy moth. The statistical significance of household wood in the no selection model ranged from a P value of 0.014 to 0.047, while lagged positive trap always had a *P value* < 0.0005. Comparatively household wood has a weaker *P value* than lagged positive trap, but still has a stronger *P value* than many of the other variables.

[7] Odds ratios in which the 95 C.I. include one suggest no significant change in the log odds in the response given a change in the explanatory variables.

Moran's I analysis of the Pearson's Chi-squared residuals from the empirical model indicate a reasonable job controlling for spatial autocorrelation (Table 5.12, Appendix 5). The Z score dropped below 2 in four years of the analysis and was substantially reduced in the other years. In addition, the semivariogram of the Pearson's Chi-squared residuals (Figures 5.1 – 5.9 D, Appendix 5) had substantially reduced variance, and in general, lacked a distinct sill and range when compared to the variance of the mean trap counts. Spatial autocorrelation in the empirical model has been controlled using the logistic regression model that includes variables accounting for the spatial dependence of the process (e.g. distance to the 1 moth/trap line and county latitude and longitude). The implications of my findings will be discussed in the next section with particular attention paid to quarantine policy, the movement of firewood, and stratified dispersal.

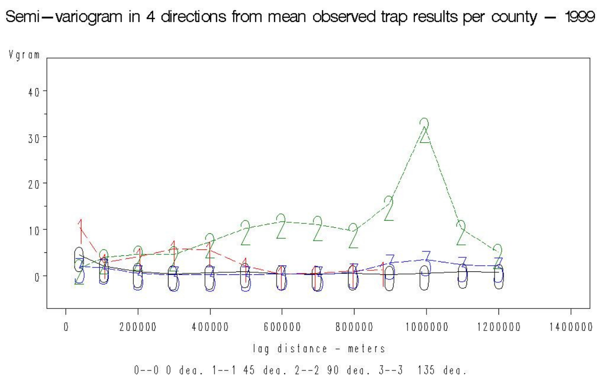
5.7 References

SAS/STAT 9.1 User's Guide (2004). SAS Institute, Cary, NC, USA. p. 5136.

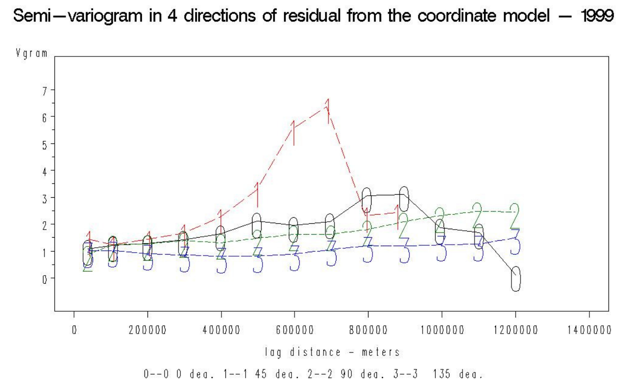
Sharov, A.A., Pjanowski, B.C., Liebhold, A.M., and Gage, S.H. (1999). What affects the rate of gypsy moth (Lepodoptera: Lymantriidae) spread: winter temperature or forest susceptibility? *Agricultural and Forest Entomology*, 1, 37-45.

Whitmire, S.L. and Tobin, P.C. (2006). Persistence of invading gypsy moth populations in the United States. *Oecologia*. 147;2 ,p230-237.

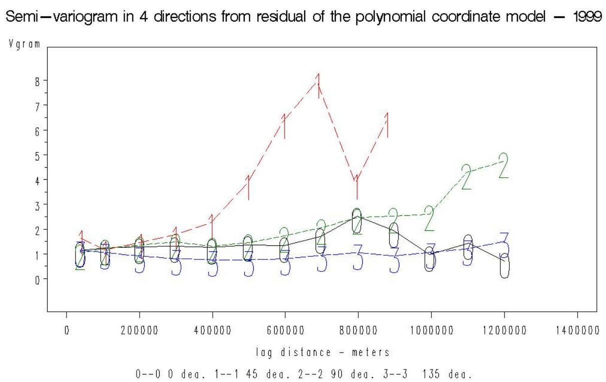
Appendix 5



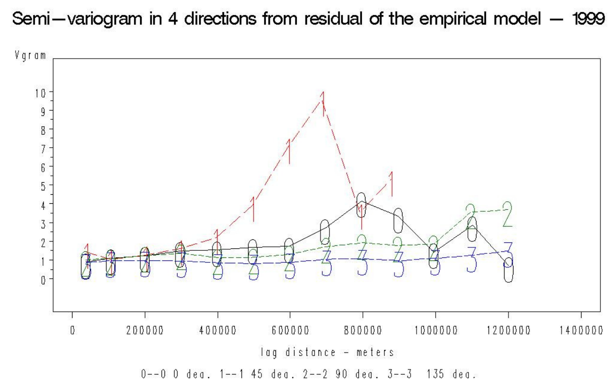
A.



B.

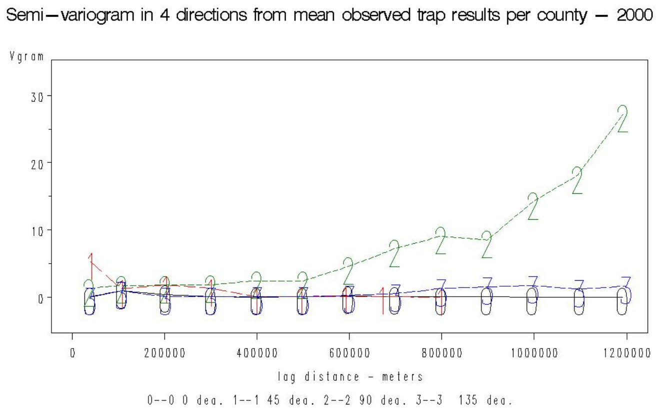


C.

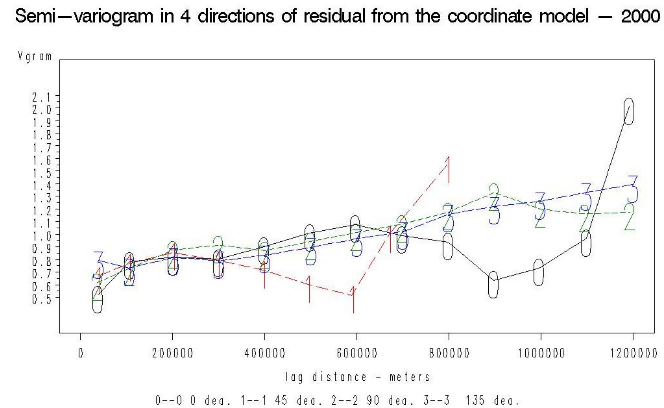


D.

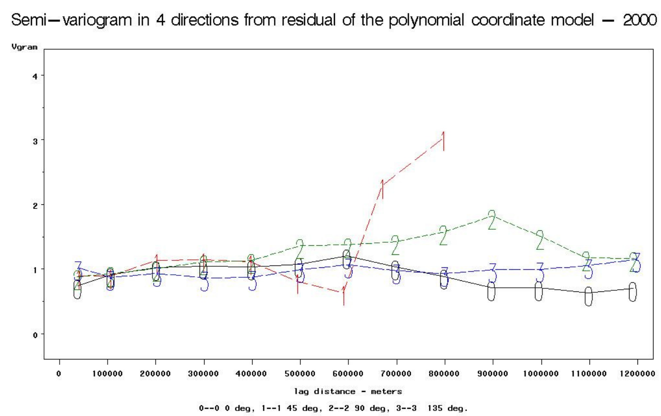
Figure 5.1 – The semivariogram results in 1999 using logistic regression for A) the observational values, B) the residuals from the X and Y coordinate model using, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. These semivariograms display spatial autocorrelation in the 90° direction (A) that is moderately controlled for by all other models (B,C,D). Note that the polynomial model (C) does not do any better than the coordinate model (B) controlling for spatial autocorrelation.



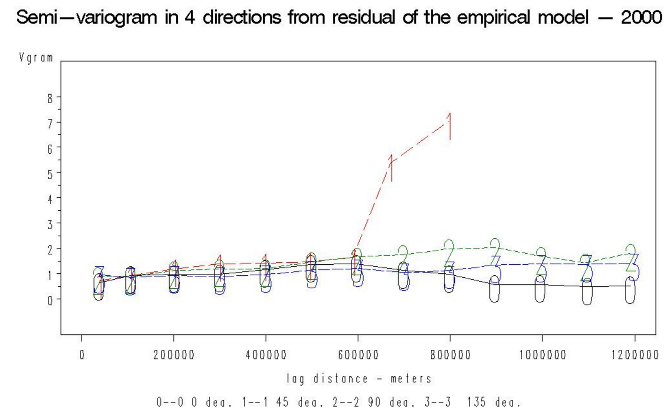
A.



B.

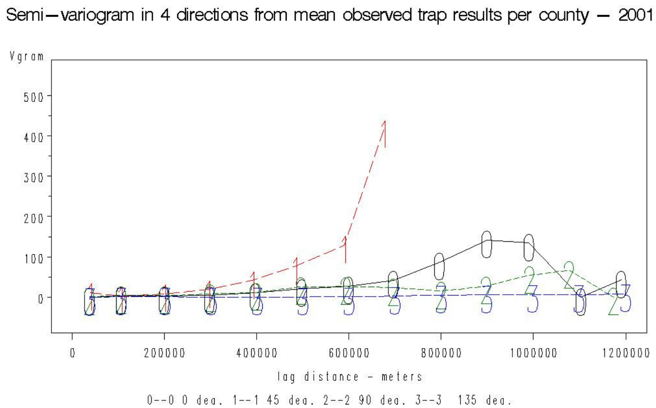


C.

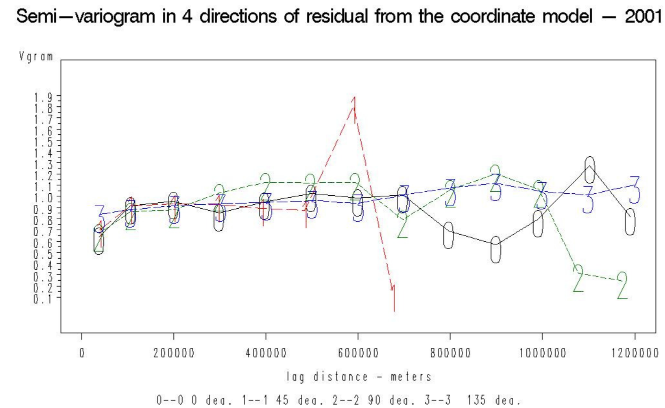


D.

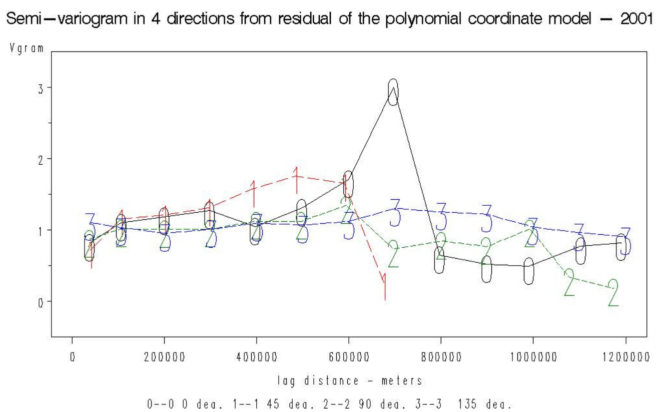
Figure 5.2 – The semivariogram results in 2000 using logistic regression for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. These semivariograms display spatial autocorrelation in the 90° direction (A) that is moderately controlled for by other models (B,C,D). Note that the polynomial model (C) does not do any better than the coordinate model (B) controlling for spatial autocorrelation.



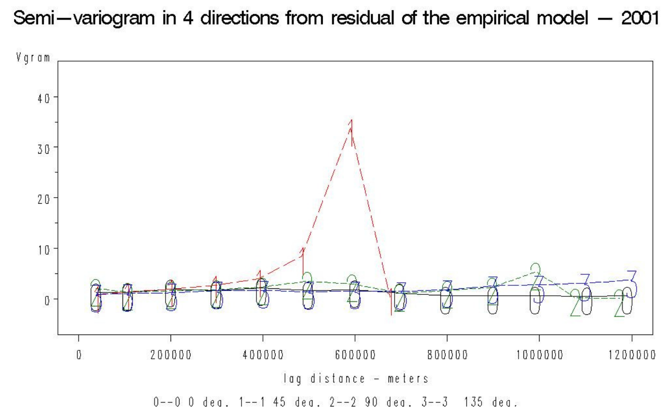
A.



B.

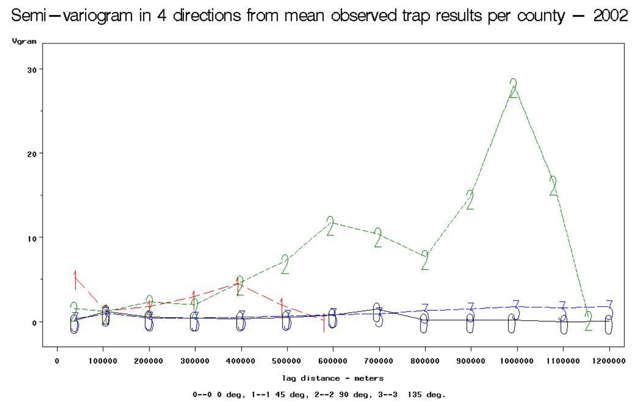


C.

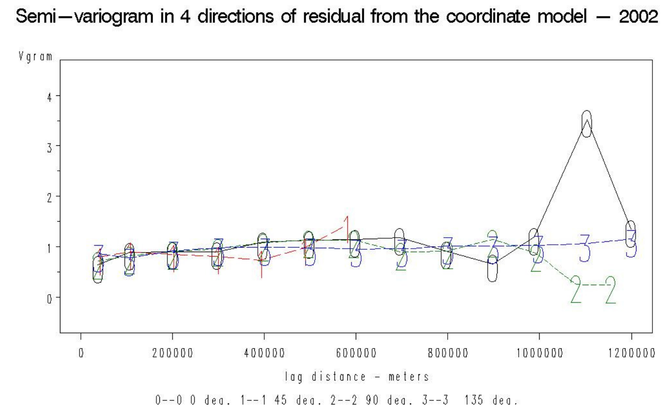


D.

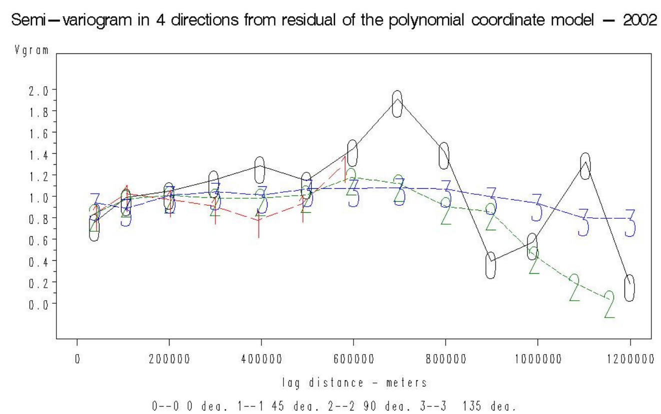
Figure 5.3 – The semivariogram results in 2001 using logistic regression for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. These semivariograms display spatial autocorrelation in the 45° direction (A) that is controlled for by the polynomial coordinate model that results in spatial autocorrelation in the north-south direction (C). Note that the polynomial model (C) does not do any better than the coordinate model (B) controlling for spatial autocorrelation.



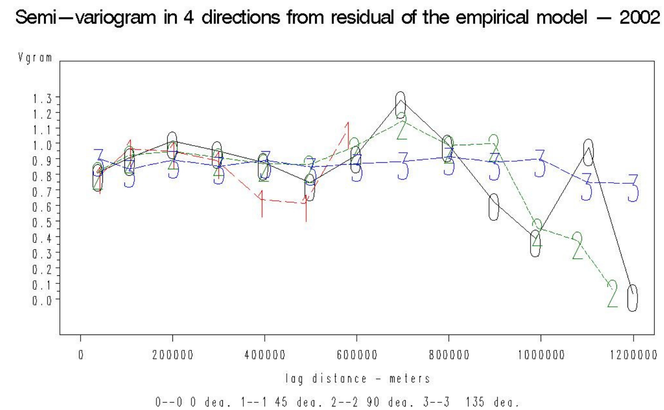
A.



B.

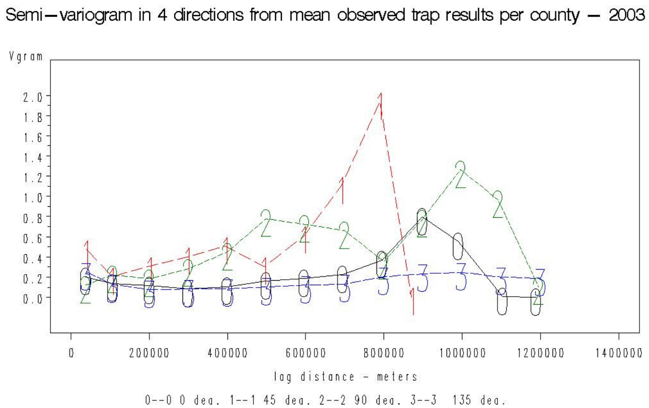


C.

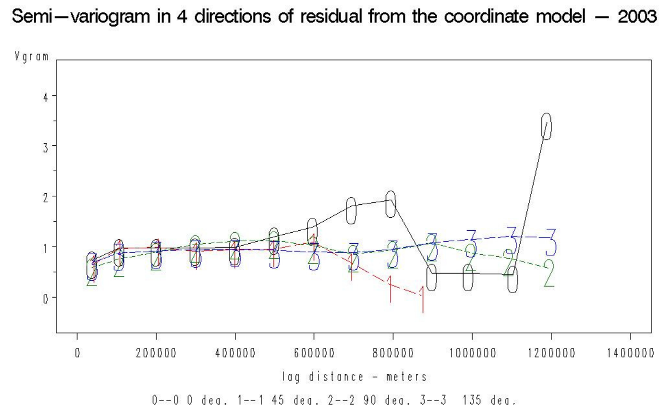


D.

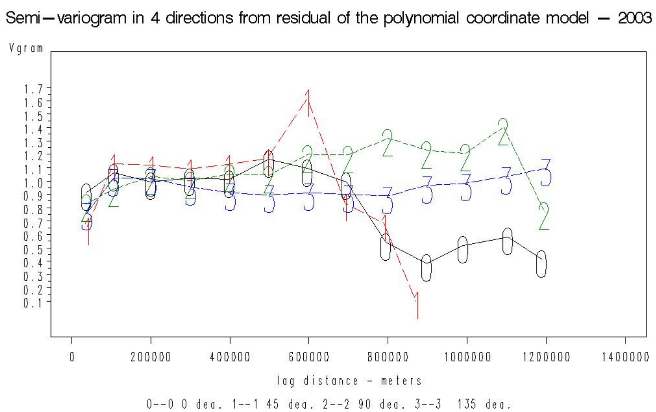
Figure 5.4 – The semivariogram results in 2002 using logistic regression for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. These semivariograms display autocorrelation in the 90° direction (A) that is controlled for in all other models (B,C,D). Note that the polynomial model (C) does not do any better than the coordinate model (B) controlling for spatial autocorrelation.



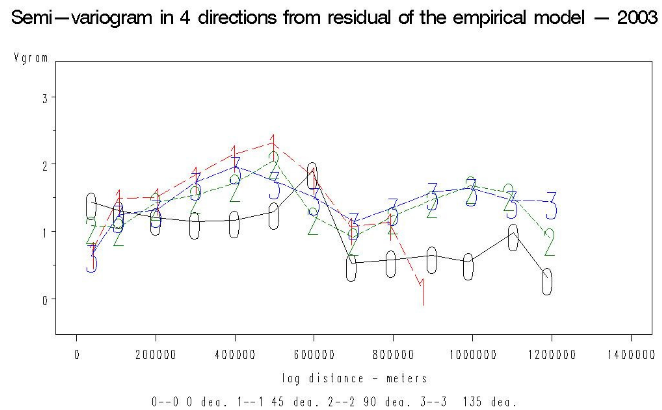
A.



B.



C.



D.

Figure 5.5 – The semivariogram results in 2003 using logistic regression for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. These semivariograms display spatial autocorrelation on the east-west and 45° degree direction. All models (B,C, and D) do a good job controlling for spatial autocorrelation. Note that the polynomial model (C) does not do any better than the coordinate model (B) controlling for spatial autocorrelation.

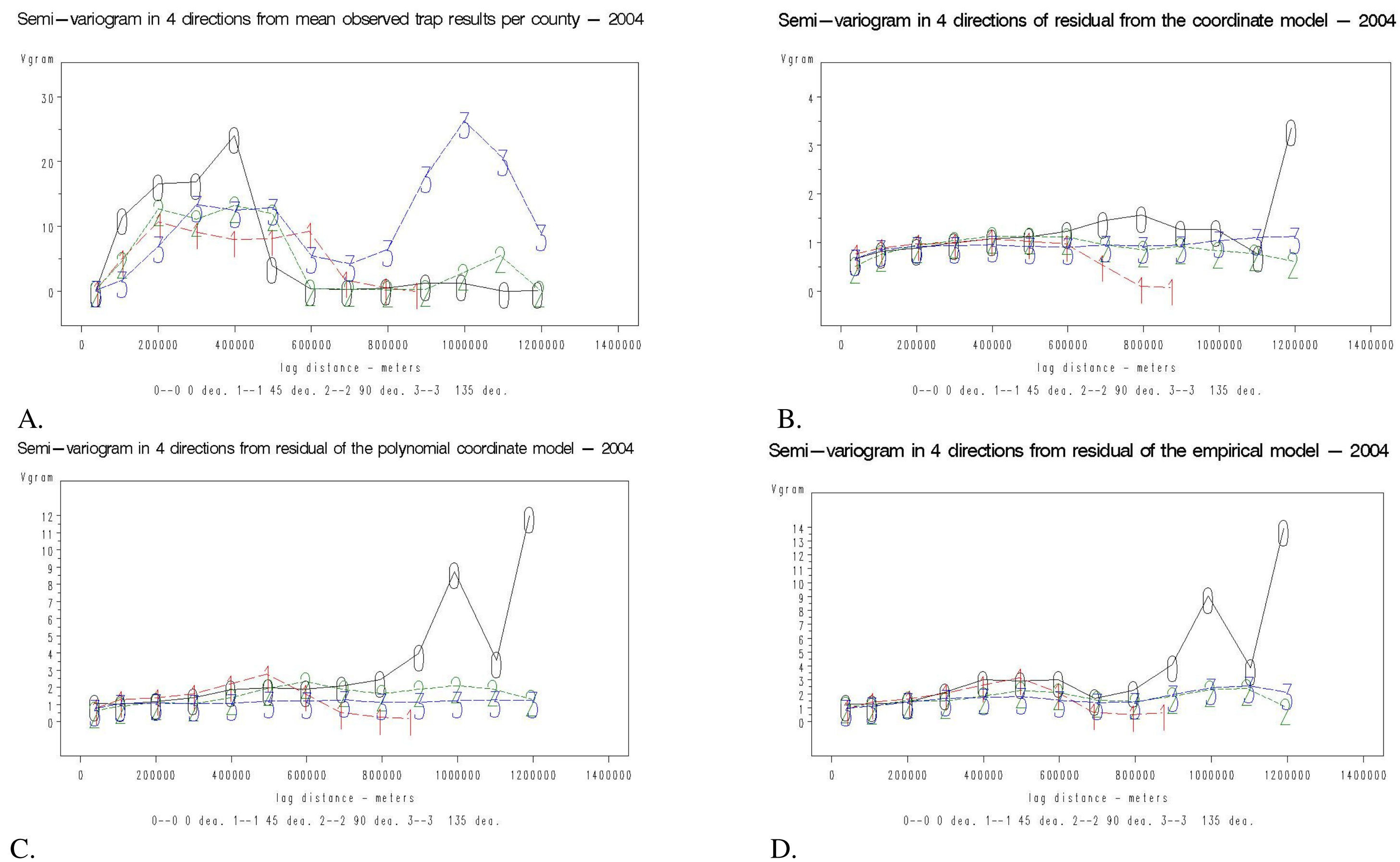


Figure 5.6 – The semivariogram results in 2004 using logistic regression for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. These semivariograms display spatial autocorrelation in all directions. All models (B, C, and D) do a relatively good job controlling for spatial autocorrelation. Note that the polynomial model (C) does not do any better than the coordinate model (B) controlling for spatial autocorrelation.

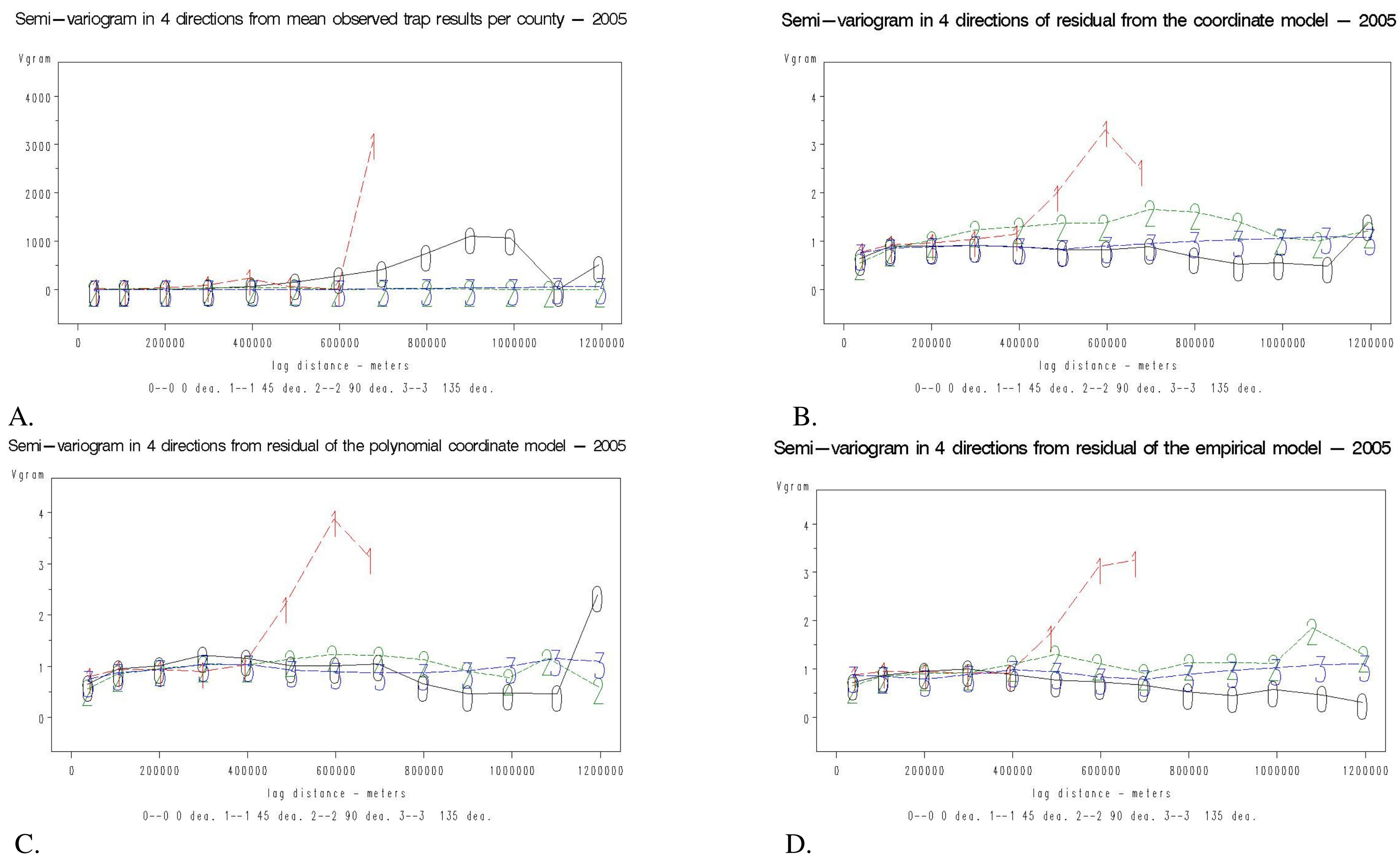
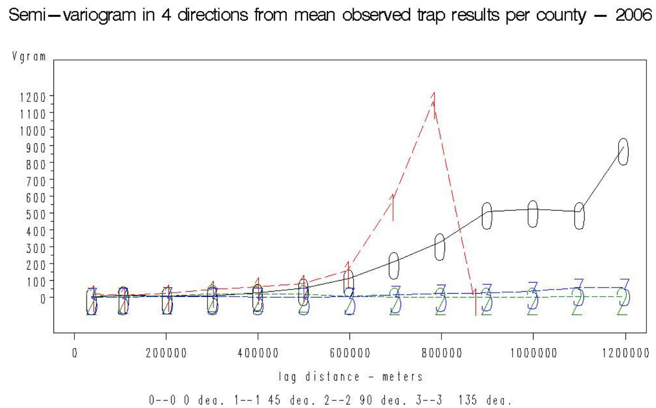
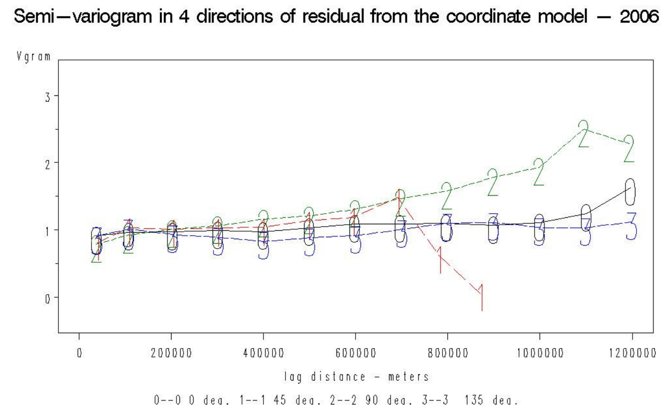


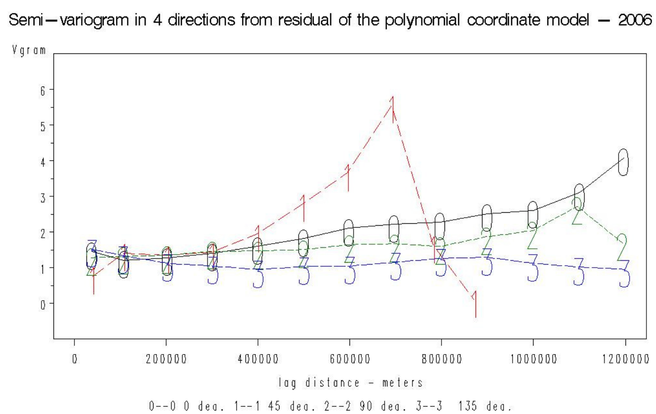
Figure 5.7 – The semivariogram results in 2005 using logistic regression for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. These semivariograms display spatial autocorrelation on the 45° direction. All models (B, C, and D) do a good job controlling spatial autocorrelation. Note the polynomial model (C) does not control for spatial auto correlation any better than the X and Y coordinate model (B).



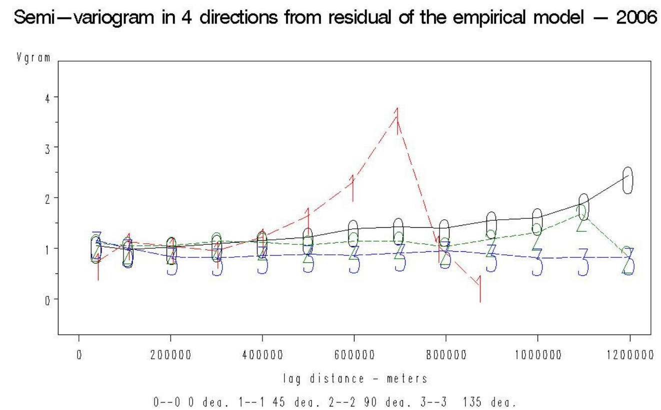
A.



B.

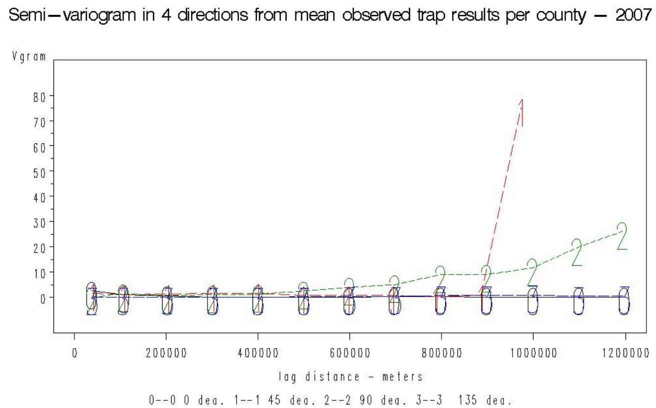


C.

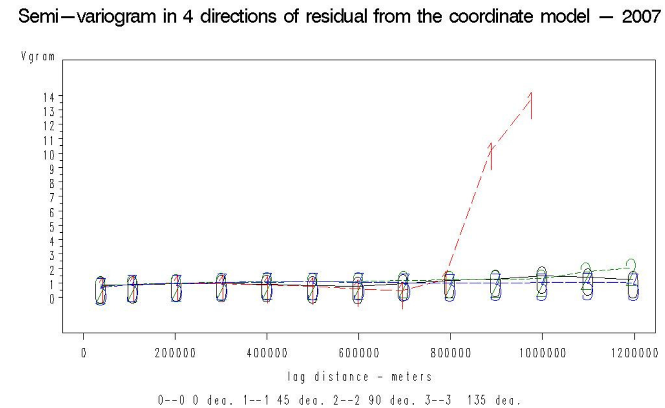


D.

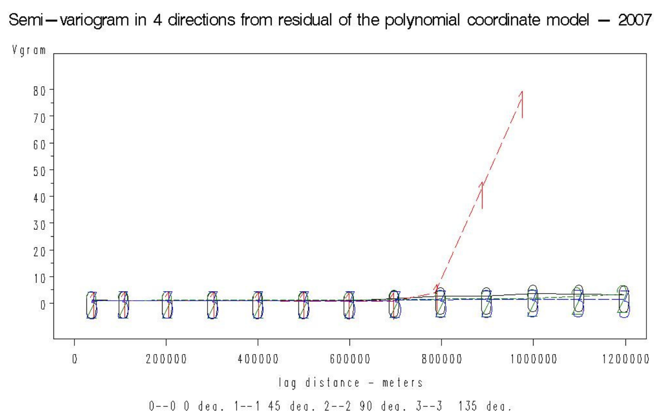
Figure 5.8 – The semivariogram results in 2006 using logistic regression for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomial, and D) the residuals from the empirical model. These semivariogram display spatial autocorrelation in the 180° and 45° directions (A) that is somewhat controlled the other models (B,C,D). Note the polynomial model (C) does not control for spatial auto correlation any better than the X and Y coordinate model (B).



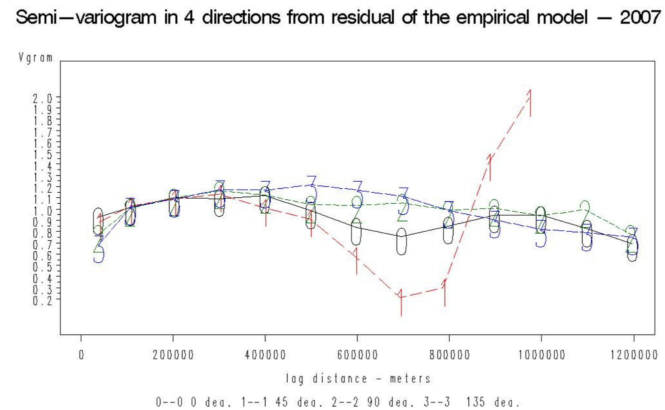
A.



B.

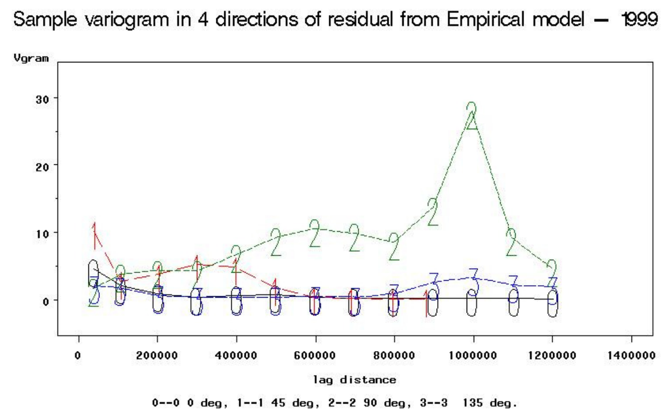


C.

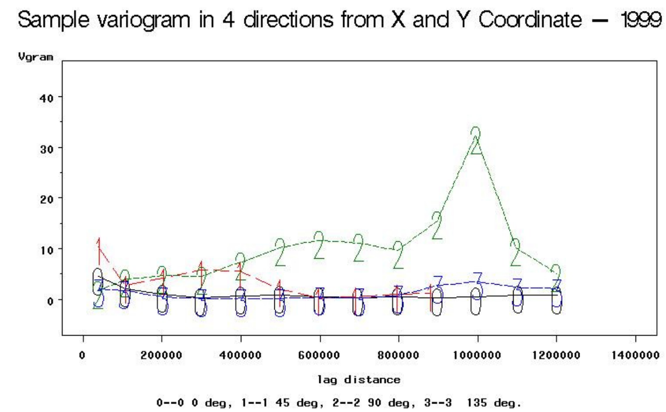


D.

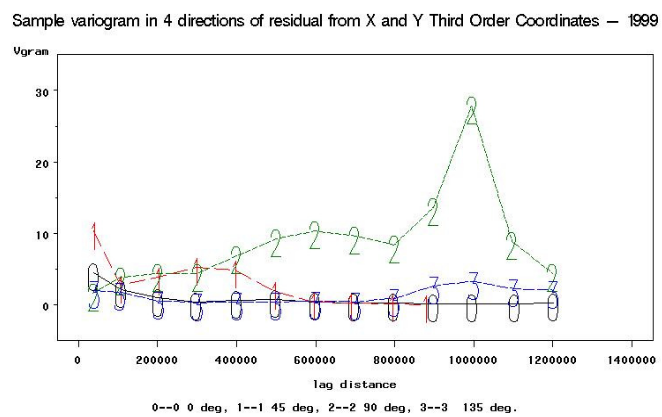
Figure 5.9 – The semivariogram results in 2007 using logistic regression for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. These semivariogram display spatial autocorrelation on the 90° and 45° direction (A) that is not controlled for by any other models (B,CD). Note the polynomial model (C) does not control for spatial auto correlation any better than the X and Y coordinate model (B).



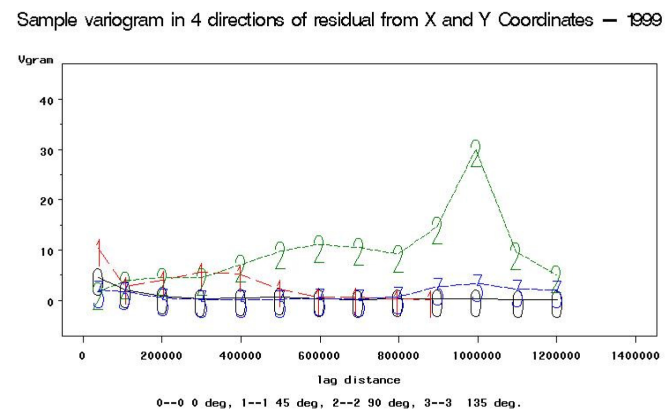
A.



B.



C.



D.

Figure 5.10 – The semivariogram results in 1999 using the general linear model for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. The general linear model fails to reduce the level of spatial autocorrelation.

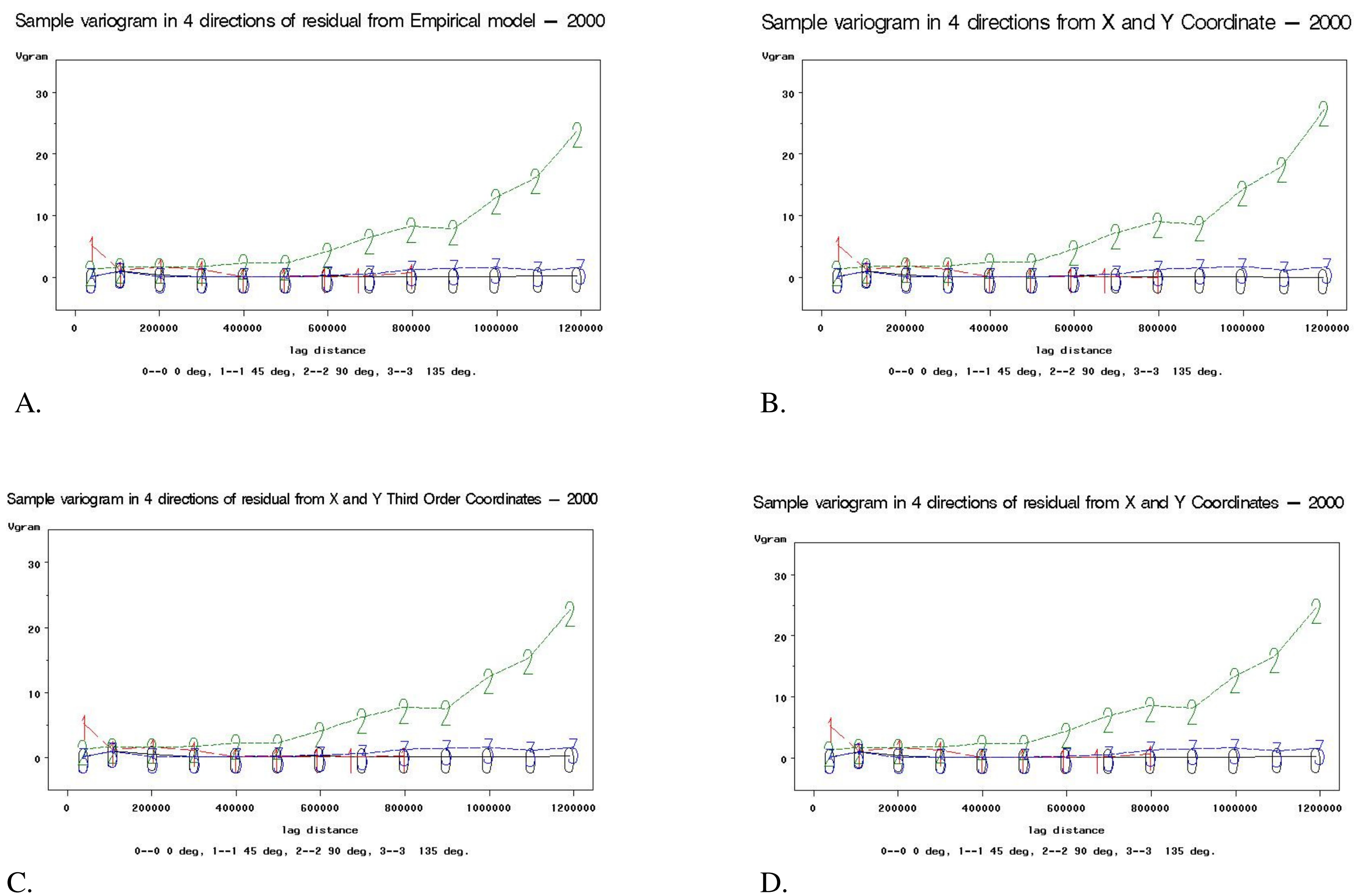
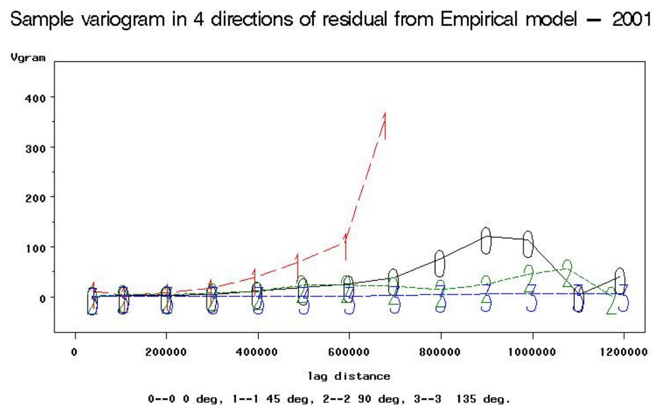
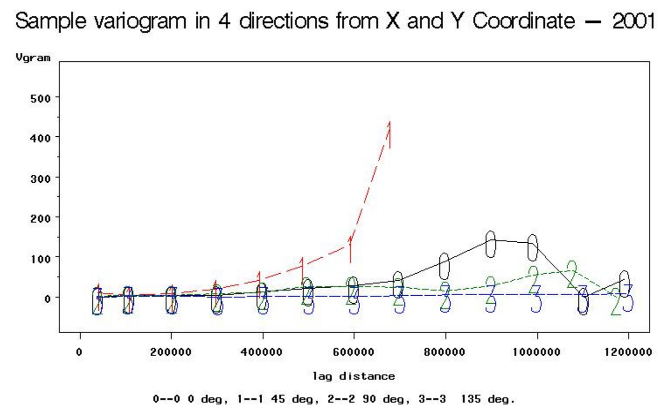


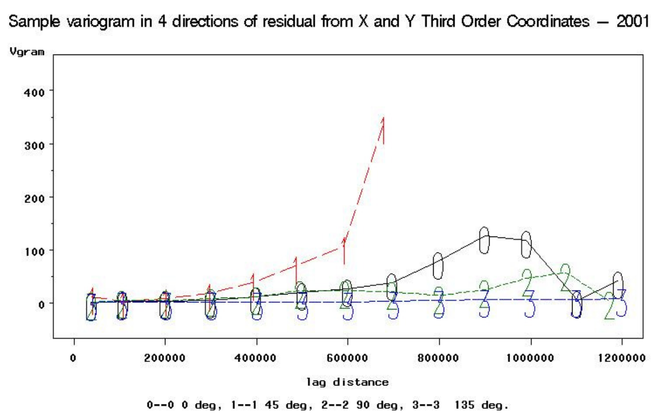
Figure 5.11 – The semivariogram results in 2000 using the general linear model for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. The general linear model fails to reduce the level of spatial autocorrelation.



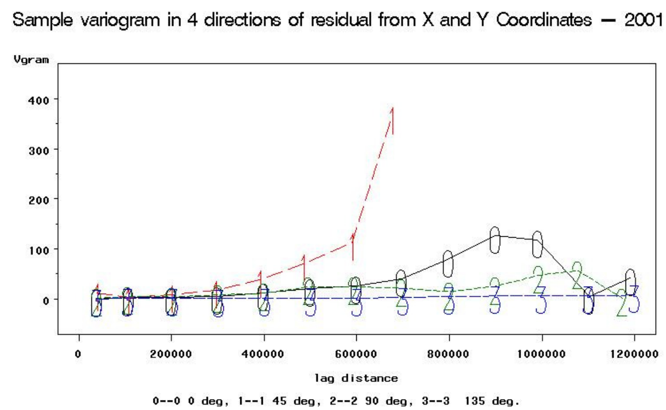
A.



B.



C.



D.

Figure 5.12 – The semivariogram results in 2001 using the general linear model for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. The general linear model fails to reduce the level of spatial autocorrelation.

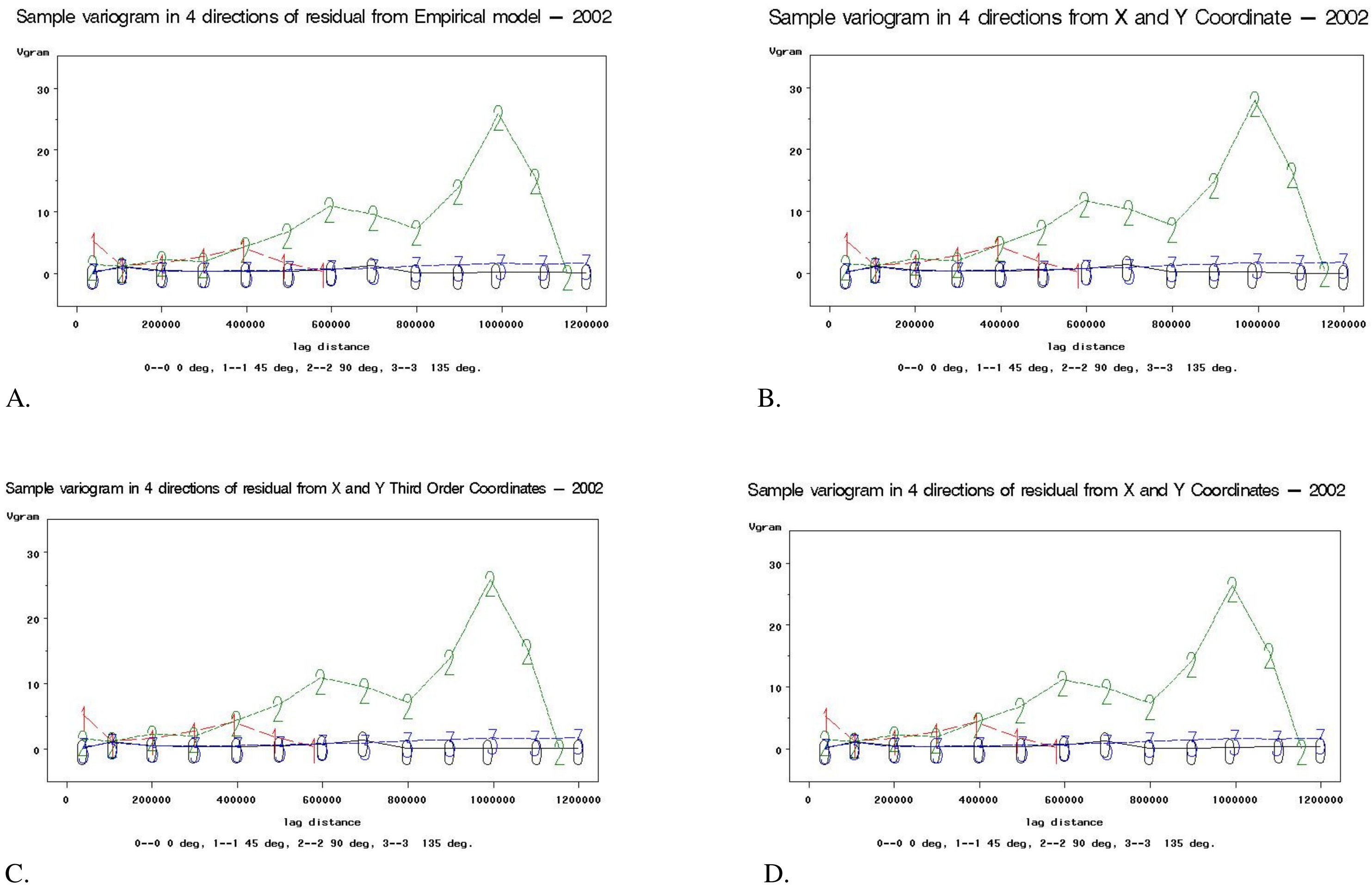
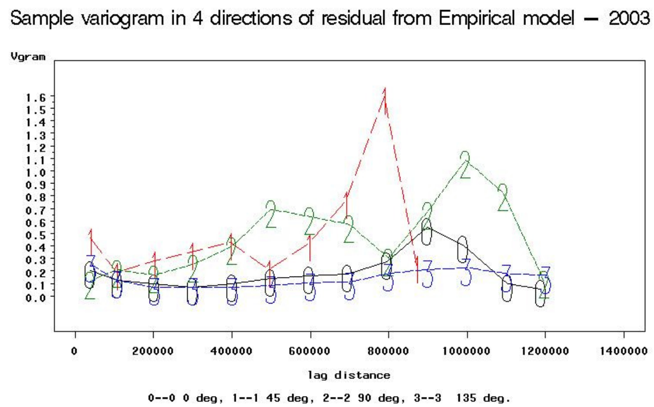
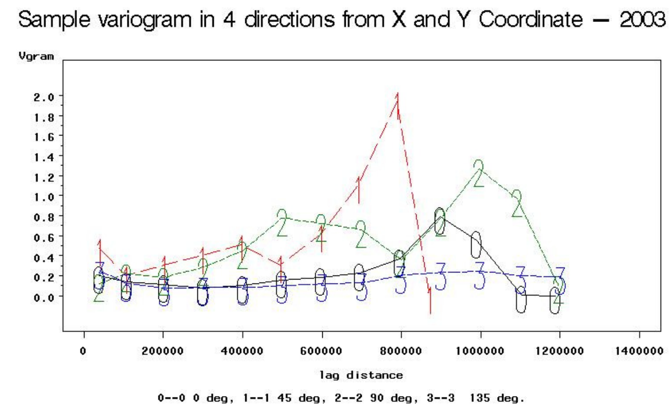


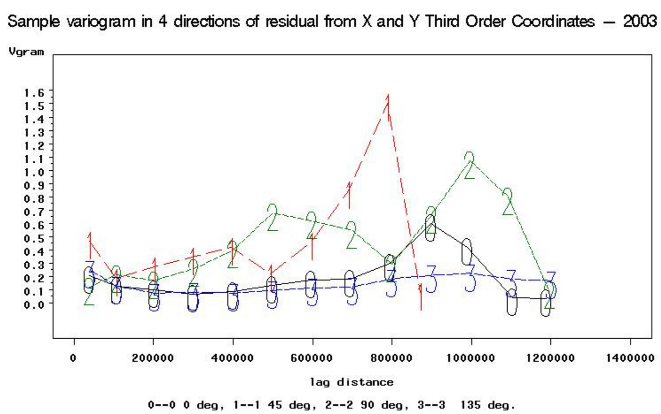
Figure 5.13– The semivariogram results in 2002 using the general linear model for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. The general linear model fails to reduce the level of spatial autocorrelation.



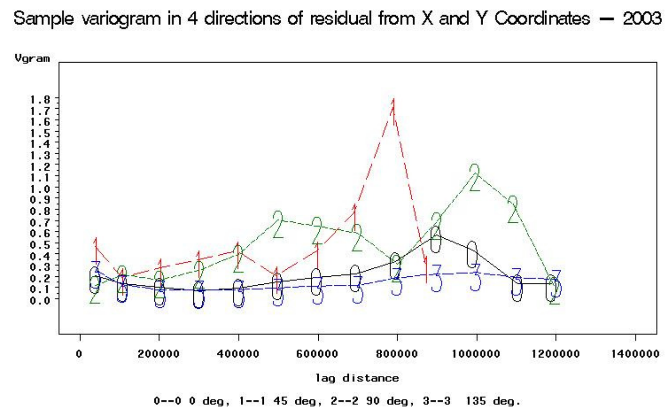
A.



B.

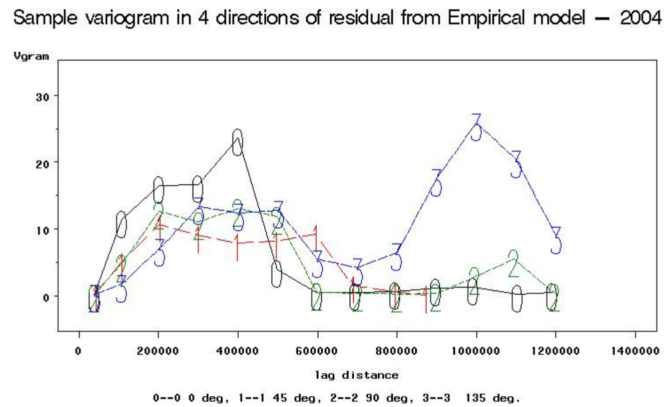


C.

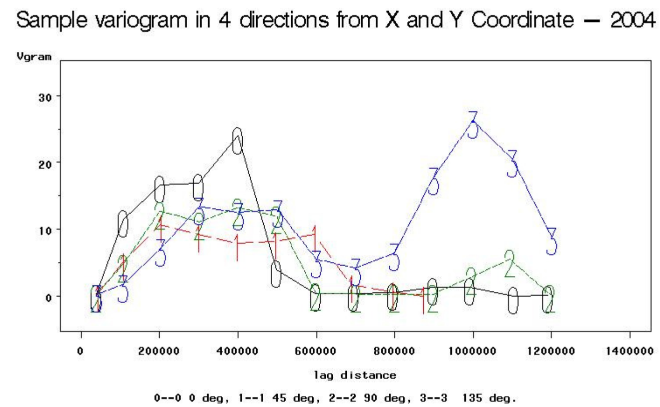


D.

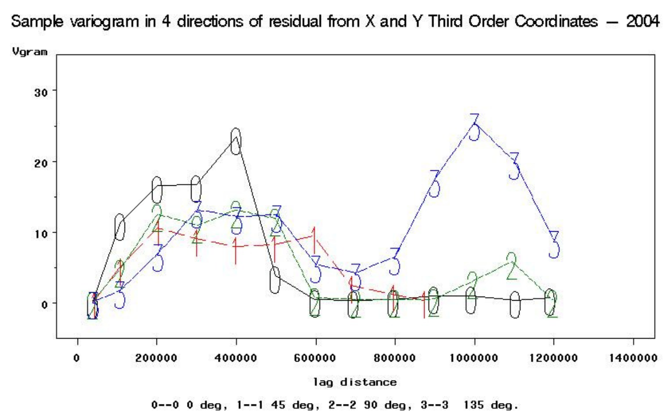
Figure 5.14 – The semivariogram results in 2003 using the general linear model for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. The general linear model fails to reduce the level of spatial autocorrelation.



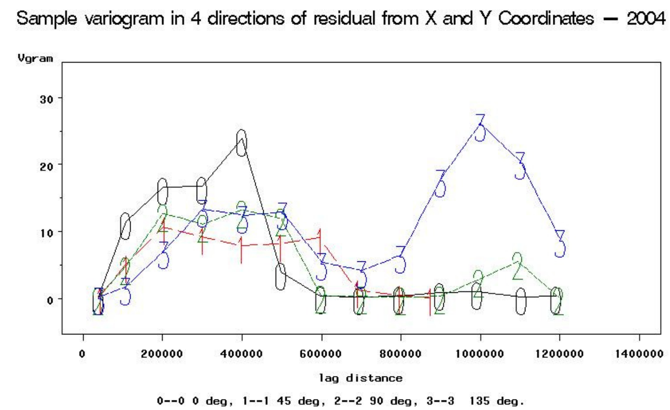
A.



B.



C.



D.

Figure 5.15 – The semivariogram results in 2004 using the general linear model for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. The general linear model fails to reduce the level of spatial autocorrelation.

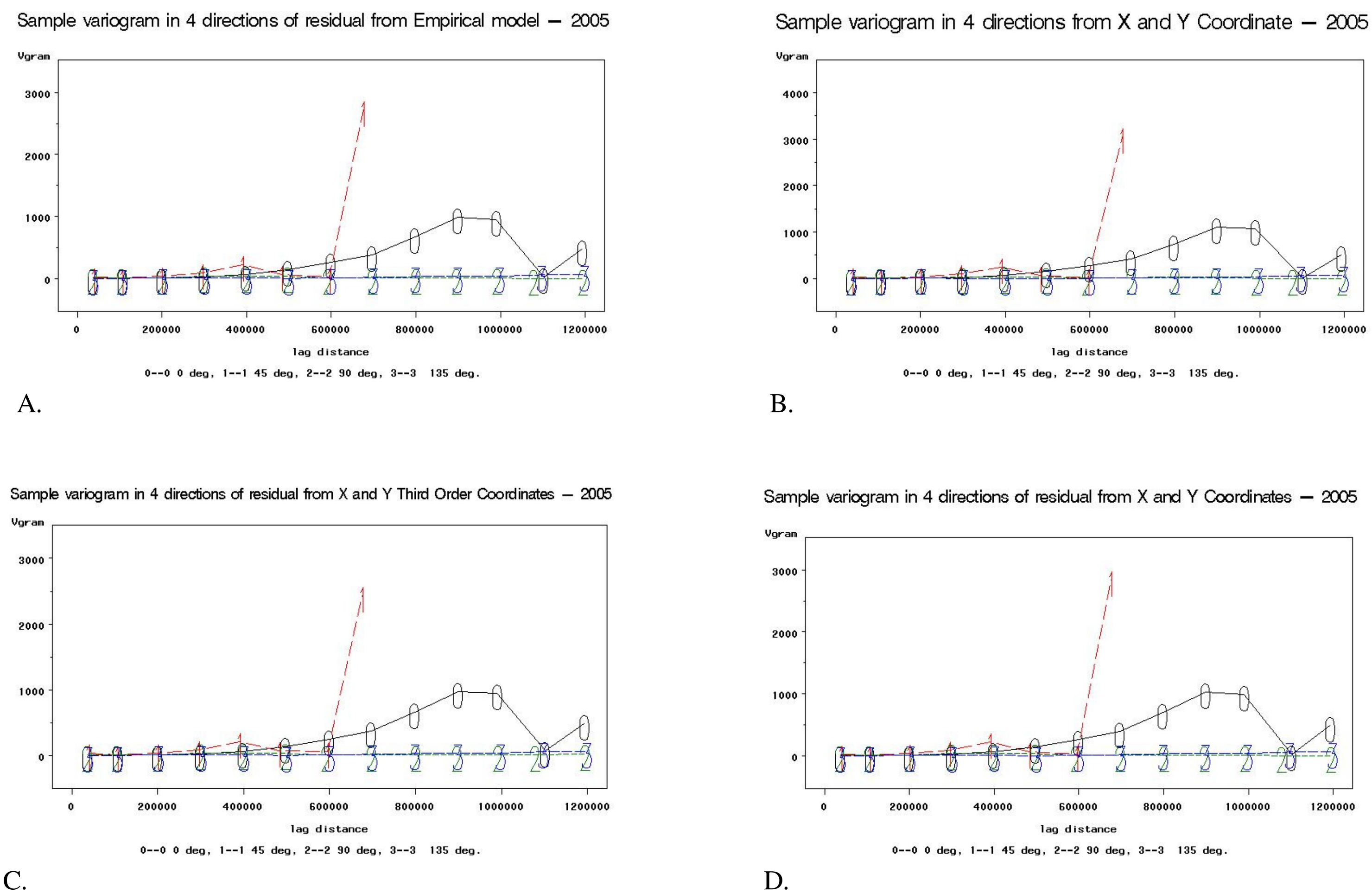
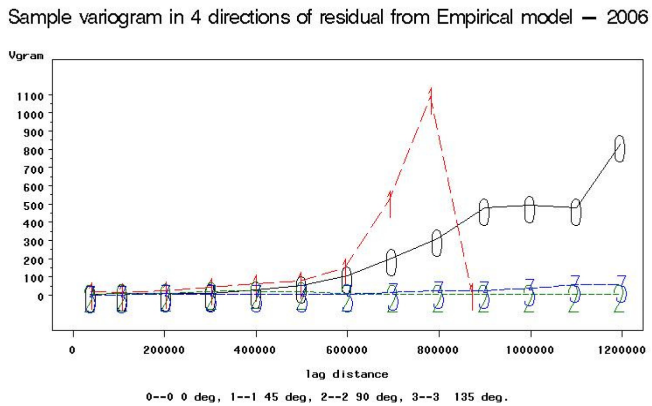
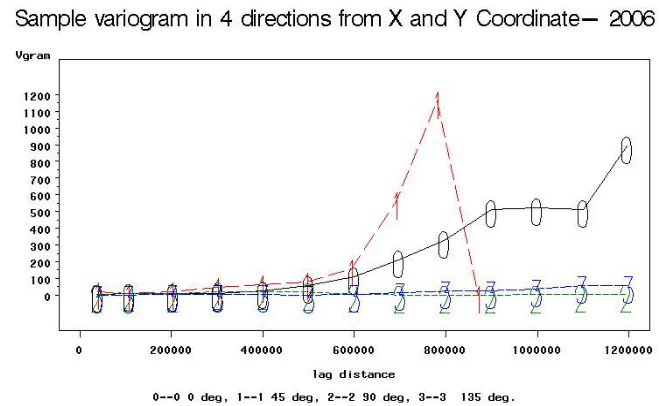


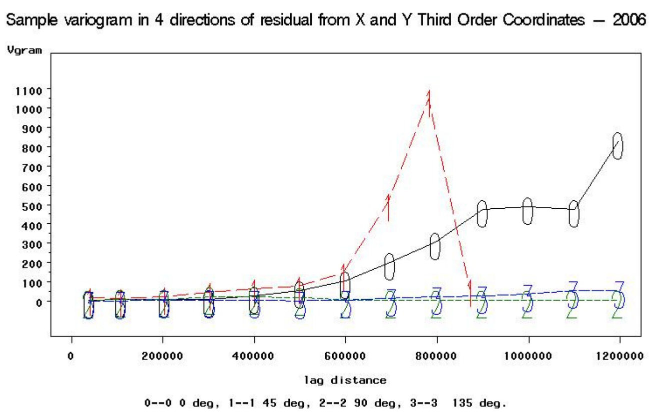
Figure 5.16 – The semivariogram results in 2005 using the general linear model for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. The general linear model fails to reduce the level of spatial autocorrelation.



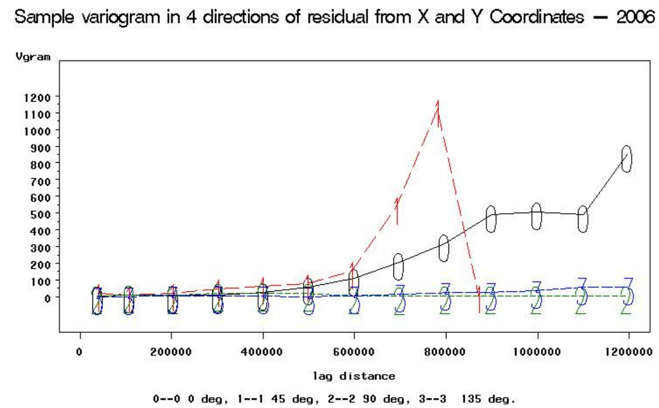
A.



B.



C.



D.

Figure 5.17 – The semivariogram results in 2006 using the general linear model for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model. The general linear model fails to reduce the level of spatial autocorrelation.

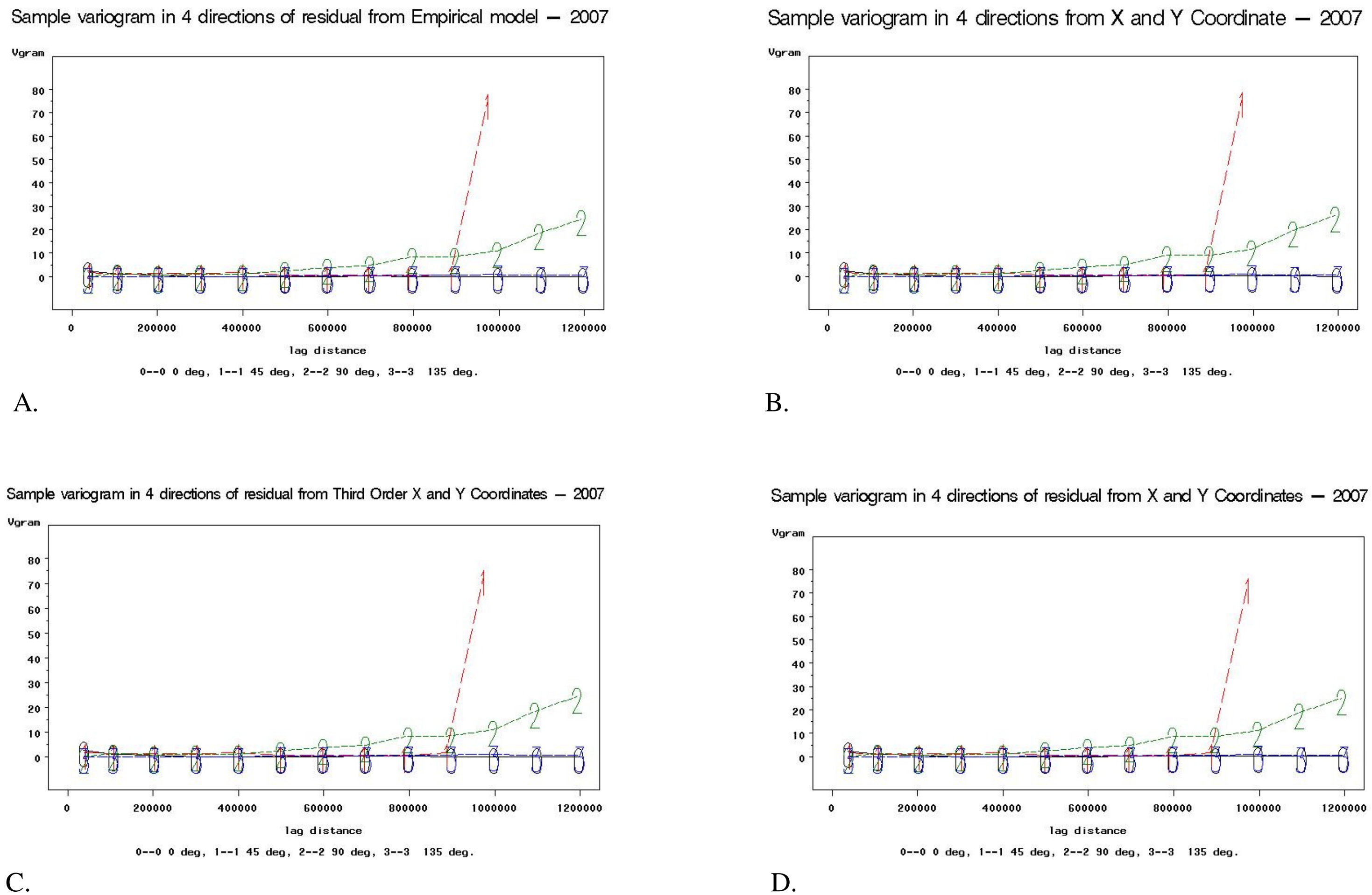


Figure 5.18 – The semivariogram results in 2007 using the general linear model for A) the observational values, B) the residuals from the X and Y coordinate model, C) the residuals from the X and Y coordinate model including a second and third order polynomials, and D) the residuals from the empirical model.. The general linear model fails to reduce the level of spatial autocorrelation.

Table 5.1 – Logistic regression results for the visitor attraction model category. Parameter estimate are from the no selection model. Highlighted values indicate significance at $\alpha < 0.10$.

Variable	1999 Estimate	1999 Pr > ChiSq	2000 Estimate	2000 Pr > ChiSq	2001 Estimate	2001 Pr > ChiSq	2002 Estimate	2002 Pr > ChiSq	2003 Estimate	2003 Pr > ChiSq
Intercept	-1.5769	0.0026	-0.2555	0.5036	0.1838	0.5626	-0.3823	0.2480	-0.2418	0.5001
Recreational Counties	-0.2356	0.4034	-0.0014	0.9957	-0.1334	0.5092	-0.2729	0.2090	-0.1213	0.6087
Park Area	0.0000	0.6613	0.0000	0.9982	0.0000	0.5783	0.0000	0.9207	0.0000	0.4368
Campgrounds	-6.3099	0.0343	-1.9562	0.3194	2.1612	0.7152	1.5715	0.7746	-1.1409	0.6466
Picnic Areas	0.0209	0.0340	0.0052	0.3053	0.0046	0.6442	0.0068	0.4565	0.0137	0.1003
Seasonal Homes	0.3196	0.0175	0.2133	0.0209	0.0501	0.3650	0.0986	0.1098	0.1394	0.0569
Retirement Counties	-0.6210	0.0087	-0.5759	0.0079	-0.1142	0.5678	-0.2866	0.1589	-0.2883	0.1663
Migration	0.0107	0.9782	-0.0930	0.9488	-0.2227	0.5703	-0.2923	0.4910	-0.0333	0.8881
Mead Distance to Nearest Road	0.0034	0.0020	-0.0005	0.4266	-0.0008	0.1627	-0.0005	0.3834	0.0003	0.6203

Variable	2004 Estimate	2004 Pr > ChiSq	2005 Estimate	2005 Pr > ChiSq	2006 Estimate	2006 Pr > ChiSq	2007 Estimate	2007 Pr > ChiSq
Intercept	-0.3896	0.2286	-0.2418	0.5607	-0.9383	0.0111	-0.5590	0.1160
Recreational Counties	0.1079	0.6093	-0.3443	0.1752	-0.1949	0.4124	-0.1751	0.4445
Park Area	0.0000	0.9238	0.0000	0.3820	0.0000	0.0175	0.0000	0.3366
Campgrounds	-3.1159	0.1728	0.6960	0.8399	-3.5261	0.0699	-3.5436	0.0263
Picnic Areas	0.0095	0.1142	-0.0038	0.5782	0.0042	0.2637	0.0081	0.0408
Seasonal Homes	0.0964	0.0936	0.1661	0.0530	0.3321	0.0003	0.2155	0.0081
Retirement Counties	-0.3618	0.0772	-0.2022	0.3393	-0.4240	0.0157	-0.4201	0.0083
Migration	-0.1334	0.6142	-0.1244	0.6952	0.3470	0.0925	0.2171	0.4057
Mead Distance to Nearest Road	0.0000	0.9886	0.0006	0.3730	-0.0006	0.3049	0.0000	0.9888

No Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.1025	0.0321	0.0217	0.0327	0.0267

No Selection	2004	2005	2006	2007	Mean Pseudo R square
N (Gypsy moth present)	290	164	441	371	
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.0201	0.0221	0.0625	0.0353	0.0395

Stepwise Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.0481	na	na	na	na

Stepwise Selection	2004	2005	2006	2007	Mean Pseudo R square
N (Gypsy moth present)	290	164	441	371	
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	na	na	0.0486	0.0212	0.0393

Table 5.2 – Logistic regression results for the source infestation model category. Parameter estimate are from the no selection model. Highlighted values indicate significance at $\alpha < 0.10$.

	1999		2000		2001		2002		2003	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	4.5787	<.0001	4.6548	<.0001	3.1987	<.0001	4.0906	<.0001	4.0168	<.0001
Size of outbreak within 100km (1 year)	0	.	0.000072	0.9056	3.19E-06	0.6925	6.01E-09	0.295	-1.71E-09	0.8971
Size of outbreak within 100km (4 year)	-1.70E-09	0.4034	-0.00002	0.9512	-8.14E-08	0.3974	0	.	0.000121	0.9079
Lagged Distance to Outbreak (1 year)	-0.0113	<.0001	-0.0111	<.0001	-0.00275	0.0564	-0.00333	0.0245	-0.00673	<.0001
Lagged Distance to Outbreak (4 year)	0.00185	0.2586	-0.001	0.1964	-0.00547	0.0004	-0.00825	<.0001	-0.0045	<.0001

	2004		2005		2006		2007	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	3.2774	<.0001	2.3818	<.0001	3.5058	<.0001	3.438	<.0001
Size of outbreak within 100km (1 year)	-5.21E-09	0.2382	6.73E-07	0.2439	-1.03E-10	0.9873	-8.58E-08	0.1154
Size of outbreak within 100km (4 year)	-5.65E-09	0.3196	-5.70E-10	0.725	6.32E-06	0.8073	-7.88E-09	0.183
Lagged Distance to Outbreak (1 year)	-0.0111	<.0001	-0.00233	0.0079	-0.00538	<.0001	-0.00011	0.6301
Lagged Distance to Outbreak (4 year)	-0.00045	0.6089	-0.00495	<.0001	-0.00693	<.0001	-0.00882	<.0001

No Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.3784	0.4342	0.3652	0.4449	0.432

No Selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.4005	0.2057	0.4889	0.3843	0.393

Stepwise Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.3737	0.4302	0.3536	0.4281	0.3755

Stepwise Selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.3929	0.1676	0.3733	0.3794	0.364

Table 5.3 – Logistic regression results for biophysical model category. Parameter estimate are from the no selection model. Highlighted values indicate significance at $\alpha < 0.10$.

	1999		2000		2001		2002		2003	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	-33.33	<.0001	-24.25	<.0001	-25.99	<.0001	-29.60	<.0001	-28.51	<.0001
Susceptible Basal Area	-0.34	0.69	-0.67	0.30	0.03	0.96	-0.29	0.68	-0.01	0.99
County Latitude	7.0E-06	<.0001	5.5E-06	<.0001	7.1E-06	<.0001	7.4E-06	<.0001	6.9E-06	<.0001
County Longitude	8.1E-06	<.0001	6.1E-06	<.0001	6.5E-06	<.0001	7.3E-06	<.0001	7.1E-06	<.0001
County Area	1.8E-06	0.057	-1.2E-06	0.014	-1.4E-06	0.008	-1.1E-06	0.017	-8.5E-07	0.05

	2004.00		2005.00		2006.00		2007.00	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	-24.35	<.0001	-18.60	<.0001	-22.68	<.0001	-19.32	<.0001
Susceptible Basal Area	0.30	0.66	1.14	0.10	-2.00	0.00	-2.46	<.0001
County Latitude	5.5E-06	<.0001	3.9E-06	<.0001	6.6E-06	<.0001	6.3E-06	<.0001
County Longitude	5.9E-06	<.0001	4.5E-06	<.0001	5.6E-06	<.0001	4.9E-06	<.0001
County Area	-5.7E-07	0.159	9.8E-08	0.876	5.1E-07	0.353	1.5E-06	0.02

No Selection	1999.00	2000.00	2001.00	2002.00	2003.00
N (Gypsy moth present)	297.00	302.00	261.00	244.00	221.00
N (Gypsy moth absent)	202.00	232.00	204.00	228.00	307.00
Pseudo R-square	0.45	0.31	0.32	0.34	0.42

No Selection	2004.00	2005.00	2006.00	2007.00	
N (Gypsy moth present)	290.00	164.00	441.00	371.00	Mean Pseudo R square
N (Gypsy moth absent)	269.00	268.00	328.00	419.00	
Pseudo R-square	0.33	0.16	0.46	0.46	0.36

Stepwise Selection	1999.00	2000.00	2001.00	2002.00	2003.00
N (Gypsy moth present)	297.00	302.00	261.00	244.00	221.00
N (Gypsy moth absent)	202.00	232.00	204.00	228.00	307.00
Pseudo R-square	0.45	0.31	0.32	0.34	0.42

Stepwise Selection	2004.00	2005.00	2006.00	2007.00	
N (Gypsy moth present)	290.00	164.00	441.00	371.00	Mean Pseudo R square
N (Gypsy moth absent)	269.00	268.00	328.00	419.00	
Pseudo R-square	0.33	0.16	0.46	0.46	0.36

Table 5.4 – Logistic regression results for the diffusion model category. Parameter estimate are from the no selection model. Highlighted values indicate significance at $\alpha < 0.10$.

	1999		2000		2001		2002		2003	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	1.9052	<.0001	1.829	<.0001	1.6951	<.0001	1.5431	<.0001	2.5876	<.0001
Distance to 1 moth/trap line	-0.00875	<.0001	-0.00987	<.0001	-0.011	<.0001	-0.0127	<.0001	-0.0132	<.0001
Lagged Positive Trap	0.5438	<.0001	0.4467	0.0001	0.5214	<.0001	0.8713	<.0001	0.878	<.0001
Treatment	0.00359	0.1874	0.0374	0.6653	0.0319	0.6596	0.00373	0.2037	0.000071	0.6456

	2004		2005		2006		2007	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	0.9176	<.0001	0.8614	0.0001	1.7173	<.0001	1.9295	<.0001
Distance to 1 moth/trap line	-0.00871	<.0001	-0.00445	0.0004	-0.00883	<.0001	-0.00824	<.0001
Lagged Positive Trap	0.6805	<.0001	1.0038	<.0001	1.0728	<.0001	0.6217	<.0001
Treatment	0.016	0.0942	0.00018	0.1282	0.00448	0.2353	0.0018	0.241

No Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.4529	0.4859	0.5116	0.6119	0.6037

No Selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.5387	0.4075	0.6378	0.5127	0.529

Stepwise Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.442	0.4621	0.4905	0.592	0.6031

Stepwise Selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.4846	0.3972	0.6307	0.5039	0.512

Table 5.5 – Logistic regression results for the demand for goods model category without the gravity variables. Parameter estimate are from the no selection model. Highlighted values indicate significance at $\alpha < 0.10$.

Variable	1999 Estimate	1999 Pr > ChiSq	2000 Estimate	2000 Pr > ChiSq	2001 Estimate	2001 Pr > ChiSq	2002 Estimate	2002 Pr > ChiSq	2003 Estimate	2003 Pr > ChiSq
Intercept	-2.188	0.002	-1.576	0.013	-0.755	0.302	-1.196	0.098	-3.086	<.0001
Population	0.225	0.510	0.169	0.246	-0.293	0.603	-0.430	0.221	-0.101	0.738
Population Density	0.240	0.621	-0.093	0.797	0.686	0.148	0.144	0.689	0.350	0.401
Urban Population	-0.089	0.801	-0.181	0.309	0.529	0.360	0.746	0.049	0.259	0.416
Urban Population Inside Urban Areas	-0.117	0.235	0.053	0.551	-0.222	0.026	-0.242	0.014	-0.135	0.187
Cluster of Urban Population	0.000	.	0.000	.	0.000	.	0.000	.	0.000	.
Rural Population	0.019	0.957	0.000	.	0.338	0.574	0.324	0.413	0.424	0.235
Income	0.041	0.013	0.027	0.065	0.000	0.997	0.010	0.540	0.052	0.003
Nurseries	-0.005	0.731	-0.010	0.482	0.007	0.578	-0.013	0.247	-0.016	0.245
Acres of Open Nurseries	0.000	0.286	0.000	0.539	0.000	0.452	0.000	0.127	0.001	0.064
Sawmills	-0.066	0.176	-0.128	0.004	-0.042	0.356	-0.033	0.453	-0.061	0.216
Household Wood	0.859	0.174	0.914	0.104	0.728	0.164	1.486	0.005	1.387	0.026
Household Wood/Population	2.925	0.006	1.531	0.177	0.365	0.743	0.127	0.906	1.684	0.155
Household Wood/Area	-0.026	0.085	0.001	0.747	-0.005	0.333	-0.002	0.728	-0.004	0.564
Christmas Tree Farms	0.010	0.211	0.032	0.049	0.022	0.039	0.008	0.178	0.003	0.642

Variable	2004.000 Estimate	2004.000 Pr > ChiSq	2005.000 Estimate	2005.000 Pr > ChiSq	2006.000 Estimate	2006.000 Pr > ChiSq	2007.000 Estimate	2007.000 Pr > ChiSq
Intercept	-3.333	<.0001	-1.271	0.089	-4.717	<.0001	-2.635	<.0001
Population	-0.033	0.878	-0.093	0.492	-0.091	0.485	-0.085	0.303
Population Density	-0.678	0.168	0.005	0.990	1.326	0.004	0.394	0.300
Urban Population	0.486	0.042	0.382	0.032	0.139	0.382	0.283	0.017
Urban Population Inside Urban Areas	-0.369	0.000	-0.256	0.018	-0.038	0.657	-0.188	0.027
Cluster of Urban Population	0.000	.	0.000	.	0.000	.	0.000	.
Rural Population	0.072	0.795	0.029	0.893	0.751	0.000	0.429	0.007
Income	0.047	0.006	0.017	0.316	0.067	<.0001	0.045	0.001
Nurseries	-0.019	0.121	0.002	0.857	-0.029	0.005	-0.012	0.188
Acres of Open Nurseries	0.000	0.228	0.000	0.173	0.000	0.186	0.000	0.117
Sawmills	-0.023	0.619	0.046	0.377	-0.135	0.001	-0.071	0.053
Household Wood	1.177	0.023	1.207	0.037	1.458	0.005	0.875	0.069
Household Wood/Population	2.310	0.034	0.868	0.478	2.364	0.018	0.638	0.486
Household Wood/Area	0.009	0.099	-0.004	0.511	-0.005	0.358	-0.001	0.813
Christmas Tree Farms	0.013	0.068	-0.001	0.872	0.013	0.019	0.006	0.189

No selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.1579	0.1428	0.1333	0.1573	0.2299

No selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.2607	0.1631	0.3158	0.1711	0.192

Stepwise Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.0999	0.0893	0.1177	0.1338	0.2055

Stepwise Selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.2424	0.1133	0.2993	0.1584	0.162

Table 5.6 – Logistic regression results for the demand for goods model category with the gravity variable. Parameter estimate are from the no selection model. Highlighted values indicate significance at $\alpha < 0.10$.

	1999		2000		2001		2002		2003	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	-1.642	0.010	-1.388	0.012	-0.559	0.369	-1.272	0.036	-2.483	0.000
Gravity	0.043	0.105	0.045	0.057	0.090	0.005	0.083	0.001	0.056	0.036
Income	0.037	0.025	0.024	0.090	0.000	0.993	0.013	0.396	0.049	0.003
Nurseries	0.003	0.835	0.002	0.867	0.006	0.596	-0.012	0.242	-0.005	0.694
Acres of Open Nurseries	0.000	0.226	0.000	0.520	0.000	0.424	0.000	0.129	0.001	0.064
Sawmills	-0.061	0.218	-0.124	0.005	-0.038	0.401	-0.037	0.399	-0.062	0.217
Household Wood	1.653	0.003	1.266	0.006	0.994	0.022	1.563	0.000	2.421	<.0001
Household Wood/Population	1.868	0.040	1.204	0.176	-0.259	0.775	-0.001	0.999	0.035	0.971
Household Wood/Area	-0.036	0.012	-0.004	0.278	-0.007	0.091	-0.009	0.210	-0.011	0.359
Christmas Tree Farms	0.007	0.357	0.029	0.067	0.023	0.034	0.008	0.187	-0.001	0.847

	2004		2005		2006		2007	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	-3.213	<.0001	-1.235	0.049	-3.643	<.0001	-1.958	<.0001
Gravity	0.040	0.037	0.036	0.104	0.119	<.0001	0.038	0.056
Income	0.054	0.000	0.019	0.209	0.061	<.0001	0.043	0.000
Nurseries	-0.003	0.741	0.004	0.736	-0.017	0.060	-0.007	0.428
Acres of Open Nurseries	0.000	0.276	0.000	0.202	0.000	0.196	0.000	0.206
Sawmills	-0.023	0.615	0.039	0.444	-0.128	0.001	-0.067	0.068
Household Wood	1.668	<.0001	1.394	0.005	3.013	<.0001	1.952	<.0001
Household Wood/Population	1.176	0.173	0.468	0.625	-0.215	0.793	-1.392	0.066
Household Wood/Area	0.000	0.922	-0.010	0.210	-0.008	0.136	-0.003	0.492
Christmas Tree Farms	0.008	0.266	-0.001	0.833	0.009	0.090	0.005	0.335

No selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.1463	0.1337	0.1272	0.1288	0.2151

No selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.2074	0.1393	0.2872	0.1483	0.170

Stepwise Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.0999	0.0893	0.0994	0.1018	0.1955

Stepwise Selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.194	0.1133	0.2766	0.1322	0.145

Table 5.7 – Logistic regression results for the transportation model category. Parameter estimate are from the no selection model. Highlighted values indicate significance at $\alpha < 0.10$.

	1999		2000		2001		2002		2003	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	-0.9163	0.0004	-0.4992	0.035	-0.367	0.1368	-0.6832	0.006	-0.7099	0.0052
Rest Areas	0.0839	0.4943	0.2065	0.0947	0.0443	0.7084	0.0686	0.5646	-0.0282	0.8078
Walmart and Sams Clubs	-0.1085	0.2052	0.035	0.6803	0.131	0.111	0.2763	0.0055	0.2027	0.0461
Railroad Distance	-0.00156	0.5184	-0.00146	0.488	-0.00066	0.7519	-0.00014	0.9482	-0.0012	0.6006
Airports	0.0949	0.0028	0.0513	0.0553	0.0199	0.428	0.0454	0.0877	0.0476	0.0937
Highway Distance	0.00461	0.0007	0.00222	0.0701	0.00135	0.2686	0.000479	0.6899	0.00295	0.0196

	2004		2005		2006		2007	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	-1.3897	<.0001	-0.597	0.0331	-1.3991	<.0001	-0.8059	<.0001
Rest Areas	0.0895	0.4126	-0.00645	0.9596	0.1014	0.2933	-0.0151	0.8731
Walmart and Sams Clubs	0.0259	0.7507	0.0383	0.6738	0.1787	0.0273	-0.0264	0.7059
Railroad Distance	-0.00094	0.6726	0.00286	0.2779	-0.00383	0.0422	0.0012	0.5454
Airports	0.0861	0.0018	0.0185	0.5062	0.0785	0.0015	0.0951	0.0003
Highway Distance	0.00349	0.0029	0.00302	0.0248	0.00362	0.0009	0.00223	0.0318

No Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.1268	0.0725	0.0467	0.1065	0.1114

No Selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.1428	0.0776	0.1487	0.0959	0.103

Stepwise Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.1198	0.0641	0.0373	0.1048	0.1034

Stepwise Selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.1407	0.0676	0.1402	0.0951	0.097

Table 5.1 – Moran's I statistic.

	Mean observational values				Presence/absence			
Year	Moran's Index	Expected Index	Variance	Z Score	Moran's Index	Expected Index	Variance	Z Score
1999	0.042332	-0.002008	0.000013	12.3754	0.124131	-0.002008	0.000022	26.67816
2000	0.034986	-0.001876	0.000008	13.28076	-0.020167	-0.001876	0.000016	28.298809
2001	0.023671	-0.002155	0.000017	6.296688	0.111723	-0.002155	0.000027	21.818027
2002	0.002977	-0.002123	0.000008	1.846918	0.108621	-0.002123	0.000023	22.945647
2003	0.040553	-0.001898	0.000011	12.75156	0.134376	-0.001898	0.000018	32.297717
2004	-0.000814	-0.001792	0.000001	0.928071	0.115351	-0.001792	0.000015	29.920556
2005	0.005055	-0.00232	0.000006	11.86262	0.063385	-0.00232	0.000031	11.862618
2006	0.008252	-0.001302	0.000002	3.136071	0.136349	-0.001302	0.000006	54.388018
2007	0.005291	-0.001267	0.000001	8.021316	0.123992	-0.001267	0.000006	51.215196

Table 5.9 – Logistic regression results for the empirical model. Parameter estimates are from the no selection model. Highlighted values indicate significance at $\alpha < 0.10$.

Variable	1999 Estimate	1999 Pr > ChiSq	2000 Estimate	2000 Pr > ChiSq	2001 Estimate	2001 Pr > ChiSq	2002 Estimate	2002 Pr > ChiSq	2003 Estimate	2003 Pr > ChiSq
Intercept	-20.5179	<.0001	5.9199	0.1791	-7.9472	0.0774	-8.5402	0.1255	-0.5044	0.9294
Gravity	-0.0787	0.0011	-0.0005	0.9808	0.0265	0.2722	0.0174	0.4636	-0.0298	0.2268
Seasonal Homes	0.1937	0.0717	0.1256	0.2161	-0.0345	0.6694	0.0745	0.3618	0.0688	0.3457
Retire	-0.5775	0.0540	-0.7073	0.0108	0.4178	0.1424	-0.1199	0.6773	-0.0179	0.9525
Income	0.0702	0.0028	0.0322	0.0905	0.0510	0.0183	0.0262	0.2550	0.1058	0.0001
Household wood	-0.3150	0.4642	0.8148	0.0327	1.2566	0.0036	1.1935	0.0108	0.5997	0.2196
Airports	-0.0021	0.9572	-0.0052	0.8796	0.0165	0.6232	0.0580	0.1401	-0.0104	0.7983
Highway Distance	0.0038	0.0167	-0.0004	0.7878	-0.0006	0.7042	-0.0002	0.8966	0.0021	0.2029
Lagged Distance to Outbreak (1 year)	-0.0069	0.0631	-0.0071	0.0010	0.0051	0.1460	0.0004	0.8893	0.0036	0.3463
Susceptible Basal Area	0.6524	0.5482	0.2612	0.7973	0.3103	0.7787	-0.0784	0.9467	2.6598	0.0317
County Latitude	0.0000	0.0002	0.0000	0.1384	0.0000	0.0026	0.0000	0.1046	0.0000	0.5124
County Longitude	0.0000	<.0001	0.0000	0.3077	0.0000	0.1646	0.0000	0.1435	0.0000	0.7233
Distance to 1 moth/trap line	0.0019	0.6486	-0.0068	0.0046	-0.0163	0.0003	-0.0139	0.0001	-0.0202	0.0005
Lagged positive trap	0.4601	0.0005	0.4779	0.0005	0.6812	<.0001	0.8233	<.0001	0.8250	<.0001

Variable	2004 Estimate	2004 Pr > ChiSq	2005 Estimate	2005 Pr > ChiSq	2006 Estimate	2006 Pr > ChiSq	2007 Estimate	2007 Pr > ChiSq
Intercept	-9.446	0.0149	-8.7716	0.0399	-0.6999	0.8532	-7.6183	0.0539
Gravity	-0.0209	0.2986	-0.00228	0.9081	0.0318	0.1697	-0.0528	0.0093
Seasonal Homes	0.0348	0.626	0.0184	0.8386	0.0178	0.8471	-0.0318	0.6938
Retire	-0.1468	0.5882	0.1272	0.6484	0.0954	0.6912	-0.2234	0.2651
Income	0.0378	0.0508	0.0359	0.0744	0.068	0.0004	0.0393	0.017
Household wood	0.9902	0.014	0.8402	0.0399	0.8371	0.0474	0.1397	0.6919
Airports	0.0692	0.0523	0.0308	0.3868	0.0105	0.779	0.1015	0.0025
Highway Distance	0.000914	0.4989	0.000699	0.649	0.00449	0.0023	0.000823	0.5113
Lagged Distance to Outbreak (1 year)	-0.0069	0.0039	-0.00461	0.0038	-0.00191	0.1167	-0.00035	0.7117
Susceptible Basal Area	-0.3534	0.7318	1.3086	0.2011	-1.5172	0.1352	-0.8799	0.286
County Latitude	1.70E-06	0.042	2.62E-06	0.0211	2.56E-07	0.7585	2.46E-06	0.0037
County Longitude	2.18E-06	0.0158	1.97E-06	0.0452	-1.61E-07	0.8556	1.86E-06	0.0606
Distance to 1 moth/trap line	-0.00287	0.2924	0.000584	0.8312	-0.00858	<.0001	-0.00539	0.0007
Lagged positive trap	0.636	<.0001	0.9788	<.0001	0.9849	<.0001	0.4958	<.0001

No Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.56	0.5241	0.5656	0.6353	0.67

No Selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.573	0.4715	0.6947	0.5525	0.583

Stepwise Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.5173	0.5049	0.5545	0.6253	0.6612

Stepwise Selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.5485	0.4436	0.686	0.5432	0.565

Table 5.10 – Logistic regression results for the empirical sector model. Parameter estimates are the from no selection model. Highlighted values indicate significance at $\alpha < 0.10$.

	1999		2000		2001		2002		2003	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	-23.8713	0.0002	0.0985	0.9844	-14.8731	0.0037	-14.7159	0.0252	-1.3235	0.8474
Gravity	-0.1047	0.0019	-0.0274	0.2777	0.005	0.8821	-0.0531	0.1031	-0.0577	0.0832
Campgrounds	1.0867	0.5326	5.6169	0.0046	7.3008	0.1542	10.4014	0.0761	6.6985	0.007
Seasonal Homes	0.1628	0.1054	0.1652	0.1161	-0.0659	0.4631	0.0226	0.7828	0.0489	0.5259
Retire	-0.6148	0.0412	-0.7904	0.0081	0.3571	0.2255	-0.0763	0.8097	0.0183	0.9532
Income	0.0512	0.0459	0.0157	0.4924	0.0291	0.2362	-0.00577	0.8348	0.0976	0.0012
Nurseries	0.0229	0.1139	-0.00316	0.8289	0.00236	0.8673	-0.0131	0.3366	-0.00099	0.9521
Sawmills	-0.0737	0.2409	-0.1396	0.0214	-0.1175	0.0791	-0.125	0.078	-0.1137	0.1191
Household wood	-0.4528	0.3971	1.4163	0.008	1.6588	0.0048	2.0674	0.0011	1.1623	0.0652
Christmas trees	0.000847	0.9225	0.0137	0.4952	0.0237	0.0735	0.00709	0.3398	-0.001	0.8937
Walmart Sams	0.0693	0.5959	0.1241	0.3664	0.0452	0.7268	0.5981	0.0014	0.1937	0.2002
Airports	-0.0159	0.7037	-0.00654	0.8583	0.0285	0.4375	0.0541	0.2093	-0.026	0.525
Highway Distance	0.0031	0.1475	0.000503	0.7994	0.000263	0.8939	-0.00206	0.3418	0.000391	0.8722
Lagged Distance to Outbreak (1 year)	-0.00564	0.1421	-0.00745	0.0012	0.00393	0.2875	0.00123	0.6734	0.00576	0.1666
Susceptible Basal Area	0.6001	0.5917	0.2641	0.8013	0.0813	0.9443	-1.0754	0.4093	2.651	0.0401
County Latitude	4.84E-06	0.0002	-2.87E-07	0.7823	4.37E-06	<.0001	3.27E-06	0.0167	-6.79E-07	0.6672
County Longitude	5.79E-06	0.0001	4.99E-07	0.6664	3.38E-06	0.0087	3.79E-06	0.0168	-1.90E-07	0.9086
County area	-4.06E-07	0.7246	-1.50E-06	0.0398	-1.24E-06	0.106	-1.33E-06	0.1319	3.82E-07	0.7515
Distance to 1 moth/trap line	0.00123	0.7765	-0.00478	0.0634	-0.0124	0.0104	-0.0125	0.0016	-0.023	0.0003
Lagged positive trap	0.4142	0.0026	0.4315	0.0021	0.6688	<.0001	0.8625	<.0001	0.7948	<.0001
Treatment	0.00312	0.2789	0.0279	0.6958	0.0291	0.6571	0.00299	0.3622	0.000051	0.6969

	2004		2005		2006		2007	
Variable	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	-16.3276	0.0004	-10.4587	0.0371	-0.7333	0.8729	-8.997	0.0281
Gravity	-0.0405	0.1487	-0.0221	0.3889	0.00427	0.8807	-0.0769	0.0043
Campgrounds	-0.3097	0.8627	-4.7863	0.0623	2.4247	0.1302	1.4191	0.2667
Seasonal Homes	0.0136	0.856	0.0318	0.7297	0.0493	0.6366	-0.0267	0.7624
Retire	-0.2485	0.4023	0.1101	0.7	0.0643	0.7973	-0.2243	0.2737
Income	0.0306	0.161	0.0127	0.584	0.0737	0.0007	0.0383	0.031
Nurseries	-0.0149	0.2446	0.0265	0.0724	-0.0155	0.2328	0.0126	0.1332
Sawmills	-0.0775	0.2277	0.0523	0.4203	-0.0818	0.1812	0.00638	0.8986
Household wood	1.3899	0.0073	0.3662	0.4785	1.4059	0.0128	-0.1784	0.6842
Christmas trees	0.016	0.0509	-0.00929	0.2165	-0.0005	0.9393	-0.00837	0.1256
Walmart Sams	0.1607	0.2298	0.0745	0.5749	0.2419	0.0847	0.1969	0.0824
Airports	0.0621	0.0949	0.0262	0.4896	-0.00838	0.8303	0.0692	0.0564
Highway Distance	0.00178	0.3388	-0.00107	0.631	0.00339	0.1031	-0.00159	0.3437
Lagged Distance to Outbreak (1 year)	-0.00712	0.0089	-0.00547	0.0014	-0.00248	0.0611	-0.00081	0.4384
Susceptible Basal Area	-0.653	0.5631	1.0765	0.3148	-1.8444	0.0803	-1.0275	0.2269
County Latitude	3.00E-06	0.0016	3.26E-06	0.0075	4.62E-07	0.6267	2.56E-06	0.0042
County Longitude	3.80E-06	0.0005	2.65E-06	0.0247	-1.38E-07	0.9014	2.31E-06	0.0264
County area	-1.10E-06	0.2359	8.28E-09	0.9943	2.07E-07	0.8596	1.09E-06	0.3406
Distance to 1 moth/trap line	0.000446	0.8841	0.00192	0.5108	-0.00813	0.0001	-0.00533	0.0014
Lagged positive trap	0.5982	<.0001	0.9865	<.0001	0.9396	<.0001	0.4443	0.0002
Treatment	0.0155	0.0838	0.000152	0.1571	0.00534	0.2127	0.0015	0.323

Table 5.10 continued – Logistic regression results for the empirical sector model continued.

No Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.5776	0.5695	0.6023	0.6781	0.6867

No Selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.6303	0.4949	0.7111	0.5719	0.614

Stepwise Selection	1999	2000	2001	2002	2003
N (Gypsy moth present)	297	302	261	244	221
N (Gypsy moth absent)	202	232	204	228	307
Pseudo R-square	0.5242	0.5255	0.5428	0.6383	0.6706

Stepwise Selection	2004	2005	2006	2007	
N (Gypsy moth present)	290	164	441	371	Mean Pseudo R square
N (Gypsy moth absent)	269	268	328	419	
Pseudo R-square	0.5485	0.4509	0.6892	0.5262	0.568

Table 5.11 – Odds ratio estimates and Hosmer Lemeshow test for the empirical model.

	1999			2000			2001		
	Point Estimate	95% Wald L.C.L.	95% Wald U.C.L	Point Estimate	95% Wald L.C.L.	95% Wald U.C.L	Point Estimate	95% Wald L.C.L.	95% Wald U.C.L
Parameter									
Gravity	0.931	0.893	0.971						
Seasonal Homes	1.188	1.013	1.393						
Retire				0.337	0.147	0.771			
Income	1.054	1.021	1.089				1.062	1.026	1.101
Household wood				2.591	1.449	4.636	3.377	1.706	6.684
Airports									
Highway Distance	1.004	1.002	1.006						
Lagged Distance to Outbreak (1 year)	0.99	0.987	0.992	0.989	0.987	0.992	1.009	1.003	1.014
Susceptible Basal Area									
County Latitude							1	1	1
County Longitude									
Distance to 1 moth/trap line							0.98	0.973	0.986
Lagged positive trap	2.861	1.774	4.615	3.039	1.903	4.853	3.939	2.211	7.016

	2002			2003			2004		
	Point Estimate	95% Wald L.C.L.	95% Wald U.C.L	Point Estimate	95% Wald L.C.L.	95% Wald U.C.L	Point Estimate	95% Wald L.C.L.	95% Wald U.C.L
Parameter									
Gravity									
Seasonal Homes									
Retire									
Income				1.129	1.078	1.183	1.057	1.027	1.088
Household wood	3.066	1.553	6.053				5.145	2.786	9.5
Airports	1.091	1.031	1.154						
Highway Distance				1.003	1	1.005			
Lagged Distance to Outbreak (1 year)									
Susceptible Basal Area				32.365	5.403	193.883			
County Latitude									
County Longitude									
Distance to 1 moth/trap line	0.985	0.982	0.989	0.986	0.982	0.989	0.989	0.987	0.992
Lagged positive trap	5.398	3.09	9.432	4.971	2.807	8.803	3.756	2.208	6.39

	2005			2006			2007		
	Point Estimate	95% Wald L.C.L.	95% Wald U.C.L	Point Estimate	95% Wald L.C.L.	95% Wald U.C.L	Point Estimate	95% Wald L.C.L.	95% Wald U.C.L
Parameter									
Gravity							0.953	0.917	0.99
Seasonal Homes									
Retire									
Income				1.082	1.051	1.115	1.05	1.019	1.081
Household wood	3.067	1.626	5.785						
Airports							1.121	1.06	1.185
Highway Distance				1.006	1.003	1.008			
Lagged Distance to Outbreak (1 year)	0.997	0.995	0.999	0.997	0.995	0.999			
Susceptible Basal Area									
County Latitude							1	1	1
County Longitude									
Distance to 1 moth/trap line	0.996	0.994	0.999	0.992	0.989	0.995	0.992	0.99	0.994
Lagged positive trap	7.341	4.293	12.554	7.147	4.305	11.864	2.687	1.726	4.182

Partition for the Hosmer and Lemeshow Test	1999	2000	2001	2002	2003	2004	2005	2006	2007
	0.356	<.001	0.2456	0.0457	0.0808	<.0001	0.0217	0.5634	0.5468

Table 5.12 – Moran's I analysis of the residuals from the mean observational values, presence/absence, and residuals from the empirical model.

	Pearson residuals from empirical model			
Year	Moran's Index	Expected Index	Variance	Z Score
1999	0.014255	-0.002008	0.000021	3.521103
2000	0.022929	-0.001876	0.000016	6.195834
2001	0.003049	-0.002155	0.000021	1.132753
2002	0.005192	-0.002123	0.000023	1.527387
2003	0.001486	-0.001898	0.000016	0.855928
2004	0.01289	-0.001792	0.000014	3.884501
2005	0.008196	-0.00232	0.00003	1.909541
2006	0.006736	-0.001302	0.000006	3.276173
2007	0.027896	-0.001267	0.000006	11.97413

CHAPTER 6 - DISCUSSION

6.1 Introduction

The estimation results presented in the previous chapter have implications for both invasive species policy – specifically quarantines – and invasion theory. I begin by discussing the gypsy moth quarantine (section 6.2) and then focus on one particularly important regulated item: firewood (section 6.3). Next I relate my findings to invasion theory (section 6.4) and conclude by identifying areas for future research (section 6.5).

6.2 Gypsy moth quarantine policy

The gypsy moth quarantine requires that logs, posts, pulpwood, bark, bark products, nursery stock, Christmas trees, outdoor household articles, and mobile homes be inspected and certified before being moved from inside to outside of the quarantine area (Animal and Plant Health Inspection Service, 2006). Thus, the quarantine directly affects the nursery, Christmas tree, timber, firewood, and moving sectors.

The initial set of explanatory variables considered in this analysis included proxies for each of these major sectors regulated by gypsy moth quarantine. Not all were found to be significant predictors of gypsy moth presence. Migration, a proxy for the moving industry, was only significant one year in the visitor attraction model. Christmas trees and nurseries were only significant in the empirical sector model twice and once, respectively. In addition, the estimated coefficients were always positive. This was contrary to expectations, which were that an increase in the number of Christmas tree farms and nurseries in a county in the uninfested zone would make it less likely that county residents would purchase Christmas trees and nursery stock from within the infested zone, thereby decreasing the probability of gypsy moth introduction. The general lack of statistical significance could indicate that the quarantine has been effective at preventing the spread of gypsy moth on nursery stock, Christmas trees, and in the moving industry. Alternatively, the coefficients may be insignificant because we have only weak proxies for flows of these regulated goods.

In contrast, the presence of sawmills in a county is a more direct proxy for the demand for logs, which could potentially be imported from the infested area. However, the coefficient on sawmills was statistically significant in only three years, and these significant coefficients were always negative, contrary to expectations. This evidence supports the decision to reject the hypothesis that sawmills increase the probability of gypsy moth introduction. This could be because logs are unlikely vectors, but most logs are transported with bark on and therefore could harbor egg masses. Thus, it seems likely that the quarantine has been effectively applied to the timber industry. The negative sign of the significant coefficients suggests that there may even be enforcement or information spill-overs.

The firewood sector cannot be proxied by number of business establishments or farms, but rather is represented by the number of households that depend on wood as a primary source of energy. Household wood use was a good predictor of gypsy moth presence, with a coefficient nearly as large as lagged positive trap in many years. For example, in ####, if the number of households using wood as a primary energy source in a county had been one unit (1,000 households) greater, then that county would have been approximately 3.45 times more likely to have a gypsy moth introduction. An alternative explanation for this result is some omitted variable, such as forest cover, that both attracts gypsy moth and encourages households to rely on wood. However, the correlation between susceptible basal area (20%) and household wood (0.12 *P value* <0.0001) is small and in any case, this variable was already included in the model.

Thus, the most likely interpretation of this finding is that counties dependent on wood as a primary source of residential energy are more likely to be invaded by gypsy moth than counties not dependent on wood energy. The role of firewood in the spread of invasive forest pests more generally and the implications for bioenergy are discussed in the next section.

6.3 Firewood

The consistently large and statistically significant effect of the use of firewood on the probability of observing gypsy moth in uninfested counties in the transition zone suggests that policy-makers concerned with controlling the spread of gypsy moth should consider measures targeted at firewood. These might include increased monitoring in areas where there is known high demand for firewood, better enforcement of existing regulations on the movement of firewood, expanded efforts to educate and raise public awareness about the potential costs and ways to prevent transport of invasive species on firewood, or certification programs that would identify "clean" firewood. Firewood is a low-value product, likely to be sold relatively close to where it is harvested. Thus efforts to prevent the spread of gypsy moth could be focused in the transition zone, where households in uninfested counties are likely to procure firewood from infested counties. However, firewood is believed to be a vector facilitating the spread of other invasive species as well.

Perhaps the greatest concern currently is the spread of emerald ash borer by firewood. . Emerald ash borer has been estimated to cost Ohio between $0.8 and $3.4 billion in landscape value losses alone (Sydnor et al. 2007). When expanded to the whole nation, impacts on forested areas from the emerald ash borer have the potential to cost upwards of $300 billion (Nowak 2003). The spread and impact of emerald ash borer on North America native ash is already so devastating that the USDA Natural Resource Conservation Service is beginning to implement a program to collect and store ash tree seeds under the National Ash Tree Seed Collection Initiative. In the event that the ash tree is completely lost from its native range, this seed bank would ensure the species' survival ex situ (United States Department of Agriculture – Natural Resource Conservation Service 2009).

Because of its wood boring behavior, emerald ash borer has the potential to be more difficult to detect in firewood than gypsy moth, for which life stages are visible externally. It can also survive at least two seasons in infested wood if the proper treatment is not applied[8] (i.e. heat treatment or properly storing firewood) (Petrice and Haack 2006, 2007). Thus, emerald ash borer has the potential to spread rapidly via infested firewood (BenDor et al. 2006).

Consequently, the emerald ash borer has been the motivation for firewood "blitzes" by state agencies in quarantined areas (Brewer, 2004, 2005, 2006). In 2004, 2005, and 2006

[8] A vast number of infested trees are being removed because of emerald ash borer infestations. As a result infested wood cannot be destroyed quickly enough. A number of recent publications (e.g. Petrice and Haack 2006; 2007, McCullough et al. 2007) attempt to solve this problem by proposing alternative treatments to ensure infested wood does not harbor live emerald ash borer life stages. Until the problem is solved infested wood can still disperse emerald ash borer when transported and waiting to be destroyed.

firewood blitzes in Ohio intercepted 7, 9, and 34 loads of firewood from searches of approximately 2,500, 3,100, and 3,800 vehicles respectively. Signs of emerald ash borer were found in 2004 and 2006. The blitzes only occurred over one weekend during the summer. Based on this, I estimated that 6 loads of firewood would be infested with emerald ash borer per checkpoint over the summer, including only weekends of May through September (i.e. a conservative estimate of firewood movement over 22 weekends)[9].

Programs to limit the risk of introducing invasive species through the transport of firewood will become even more important if climate change policy encourages the use of wood energy, either by taxing fossil fuels or directly encouraging use of bioenergy. Paradoxically, wood potentially polluted with invasive species could be a prime source for clean, renewable energy. Richter et al. (2009) argue that advanced wood combustion projects could more than double the current amount of wood used as energy in the United States. Consumer demand for clean, renewable and affordable energy could have unintended consequences if proposed community-based advanced wood combustion projects do not effectively manage shipments of wood to ensure they do not harbor invasive species.

6.4 Invasion theory

This research quantitatively links the residential use of wood energy to the presence of gypsy moth in the transition zone, adding to the empirical evidence that people are often

[9] While the movement of firewood may be a conservative estimate the checkpoints could have been placed in areas known to be popular for recreators, etc. Thus, there would be more firewood moved in these areas when compared to other parts of the state. However, if the checkpoints are random, Ohio could see thousands of loads of firewood with emerald ash borer moved on a given summer weekend.

dispersers of invasive species (Hodkinson and Thompson 1997, Mack et al. 2000, Suarez et al. 2001). It also provides evidence of a specific mechanism for the long distance component of stratified dispersal. Gypsy moth positive counties in the transition zone were an average of 97 km in front of the 1 moth/trap line, much further than the estimated 2.5 km per year that gypsy moth likely disperses through natural mechanisms (Liebhold et al. 1992). The occurrence and persistence of these isolated gypsy moth colonies effectively increases spread rates as the colonies coalesce with each other and the generally infested area (Tobin et al. 2004).

Over the past decade, gypsy moth has been spreading at much faster rates in the northern portion of the population front (Table 6.1, Appendix 6 http://da.ento.vt.edu/spread/sprdArea0.html). The northern transition zone also has a higher proportion of households that use firewood (Figure 6.1, Appendix 6), suggesting that there may be some relationship between the two. It could be that gypsy moth are more likely to be introduced via firewood, and then more likely to persist and coalesce due to a weak Allee effect[10] (Tobin et al. 2007a) or atmospheric transport mechanisms (Tobin and Blackburn 2008).

Mack et al. (2000) suggested that research priorities include 1) the epidemiology of invasive species, 2) comprehensive and collective economic assessment of invasive species, and 3) community awareness of invasive species. The empirical results of this thesis emphasize the critical importance of anthropogenic factors for understanding the epidemiology of invasive species and suggest that these could be incorporated into risk assessment mapping exercises (e.g. Carey et al. 1991 and Tobin 2007c). Anthropogenic factors could also be incorporated into the

[10] Low population densities of isolated colonies in the northern part of the invasion front display a weak Allee effect. These populations have a better chance of persisting when compared to the same population in other parts of the range because northern populations have a lower Allee threshold (Tobin et al 2007a).

concepts of “species invasiveness”, which reflects characteristics of species that influence its establishment and spread and “habitat invasibility”, which reflects characteristics of a habitat that affect its susceptibility to invasion. Although we highlight the significance of household use of wood energy as a predictor of the presence of gypsy moth, clearly this needs to be considered in combination with many other factors to assess habitat invasibility and species invasiveness. The findings of this thesis also directly support the call for research on community awareness, highlighting the need for better understanding of how to effectively engage the public in reducing the risk of spreading invasive species by transporting infested firewood.

6.5 Future research

This section dicusses four promising avenues of future research that could build upon the modeling strategy developed in this thesis: 1) additional explanatory variables, 2) analysis at varying spatial extents and scales, 3) alternative statistical models to control for spatial autocorrelation and better exploit the panel nature of the data, and 4) different species as the dependent variable.

Due to constraints on data availability, a number of potentially relevant variables were not included in this analysis (Table 6.2, Appendix 6). For example there is a strong visual correlation between scenic roadways and gypsy moth presence (Figure 6.2, Appendix 6). An unexplained outbreak of gypsy moth in the Arkansas Ozarks in the early 1990s (Liebhold and Tobin 2006) could perhaps be related to a nearby scenic roadway. The adjacency of forests to urban areas should also be explored, because counties that are both urban and close to forests could be more likely to receive gypsy moth in the transition zone. The relationship between acorn masting events and the white footed mouse could also impact not only populations of gypsy moth in the generally infested zone (Elkinton et al. 1996) but also their dispersal potential. On a similar note, the relationship between *entomophaga maimaiga* and precipitation not only impacts gypsy moth populations in the generally infested zone (Hajeck et al. 1990), but could also impact their dispersal potential.

Increasing the spatial extent of the analysis could identify additional layers of stratified dispersal. For example, household use of wood energy is probably not a significant driver of gypsy moth dispersal to the west coast, which would be too far from population front to receive firewood from the infested zone. At this extent, household moves from the generally infested zone are likely to be a stronger predictor (Lippitt et al. 2008). The analysis could also be conducted at a finer scale, since the exact location of gypsy moth traps is known. Some of the explanatory variables considered in this analysis could be constructed for specific points and some may be available at the census tract level but others are likely only available at the county level. Alternative statistical models should better exploit the panel and time series nature of the data and control for spatial autocorrelation. Additional analysis could be based on autoregressive time series model, fixed effects model, or random effects model.

Another way to build on this analysis is to estimate similar models of the spread of other invasive species to test whether and which anthropogenic factors are robust predictors of different species. Subject to data availability, it would be interesting to consider any of the forest pests that threaten the health of our natural resources and economy. Sirex (*Sirex noctilio*:

Fabricius) and emerald ash borer (*Agrilus planipennis:* Fairmaire) are successfully established and currently spreading (Hoebeke et al. 2005, Poland and McCullough 2006). Asian long horn beetle (*Anoplophora glabripennis*: Motschulsky) has been introduced (Cavey et al. 1998) and is currently established in New Jersey, New York, and Massachusetts. Hemlock woolly adelgid (*Adelges tsugae*: Annand) is currently devastating hemlocks along the east coast (Koch et al. 2006). These species have different ranges, different mechanisms and rates of natural spread, and different management plans. The interesting question across these different scenarios is, what role do we, people, play in the dispersal of invasive species and what does this imply for quarantine and other invasive species policy.

6.5 References

Animal and Plant Health Inspection Service (2006). The Gypsy Moth Program Manual. United States Department of Agriculture.

BenDor, T.K., Metcalf, S.S., Fontenot, L.E., Sangunett, B., and Hannon, B. (2006). Modeling the spread of the Emerald Ash Borer. *Ecological Modelling 197*, p221-236.

Brewer, M. (2004, September 7). Blitz Results in Confiscation of Seven Loads of Firewood. [Press release]. Ohio Department of Agriculture.

Brewer, M. (2005, June 1). Blitz Results in Confiscation of Infested Ash Tree Firewood. [Press release]. Ohio Department of Agriculture.

Brewer, M. (2006, June 9). Blitz Results in Confiscation of Infested Ash Tree Firewood. [Press release]. Ohio Department of Agriculture.

Carey, J.R. (1991). Establishment of the Mediterranean Fruit Fly in California. *Science*. 253;5026, p1369-1373.

Cavey, J.F., Hoebeke, E.R., Passoa, S., and Lingafelter S.W. (1998). A new exotic threat to North American hardwood forests: An Asian longhorned beetle, Anoplophora glabripennis (molschulsky) (Coleoptera: Cerambycidae). I. Larval description and diagnosis. *Proceedings of the Entomology Society of Washington*. 100;2, p373-381.

Elkinton, J.S., Healy, W.M., Buonaccorsi, J.P., Boettner, G.H., HAzzard, A.M., Smith, H.R., Liebhold, A.M. (1996). Interactions among gypsy moths, white-footed mice, and acorns. *Ecology*. 77:8 p2332-2342

Hajeck, A.E., Carruthers, R.I., and Soper, R.S. (1990). Temperature and Moisture Relations of Sporulation and Germination by Entomophaga-maimaiga (zygomycetes, entomophthhoraceae), a Fungal Pathogen of Lymantria-dispar (Lepidoptera, lymantriidae). *Environmental Entomology*. 19:1 p85-90.

Hodkinson, D.J. and Thompson, K. (1997). Plant dispersal: the role of man. *Journal of Applied Ecology*. 34;6, p1484-1496

Hoebeke, E.R., Haugen, D.A., and Haack, R.A. (2005). Sirex noctilio: Discovery of a Palearctic Siricid Woodwasp in New York. *Newsletter of the Michigan Entomological Society*. 50:1&2, p24-25

Koch, F.H., Cheshire, H.M., and Devine, H.A. (2006). Landscape-scale prediction of hemlock wolly adelgid, Adelges tsugae (Homoptera: Adalgidae), infestation in the southern Appalachian Mountains. *Environmental Entomology*. 35:5 p1313-1323.

Lippitt C.D., Rogan, J., Toledano, J., Sangermano, F., Eastman, R.J., Mastro, V., Sawyer, A. (2008). Incorporating anthropogenic variables into species distribution model to map gypsy moth risk. *Ecological Modelling*. 210;3, p339-350.

Liebhold, A.M., Halverson, J.A., and Elmes, G.A. (1992). Gypsy moth invasion in North America: a quantitative analysis. *Journal of Biogeography*. 19;5, p513-520.

Liebhold, A.M. and Tobin, P.C. (2006). Growth of newly established alien populations: Comparison of North American gypsy moth colonies with invasion theory. *Population Ecology*. 48;4, p253-262.

Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, H., Clout, M., Bazzaz, F.A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*. 10;3, p689-710.

McCullough, D.G., Poland, T.M., Cappaert, D., Clark, E.L., Fraser, I., Mastro, V., Smith, S., and Pell, C. (2007) Effects of Chipping, Grinding, and Heat on Survival of Emerald Ash Borer, *Agrilus planipennis* (Coleoptera: Buprestidae), in Chips. *Journal of Economic Entomolgy*. 100;4, p1304-1315

Nowak, D. (2003) Summary. *New threats to North American forests*. Global Institute for Sustainable Forestry, Yale University, New Haven, Connecticut. p13-16.

Petrice, T.R. and Haack, R.A. (2006). Effects of Cutting Date, Outdoor Storage Conditions, and Splitting on Survival of *Agrilus planipennis* (*Coleoptera: Buprestidae*) in Firewood Logs. *Journal of Economic Entomology*. 99;3 p790-796.

Petrice, T.R. and Haack, R.A. (2007). Can emerald ash borer, *Agrulus planipennis* (*Coleoptera: Buprestidau),* emerge from logs two summers after infested trees are cut? *The Great Lakes Entomologist*. 40;1&2 p92-95.

Pimentel, D., Zuniga, R., and Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, 52;3, p273-288.

Poland, M.T. and McCullough, D.G. (2006). Emerald Ash Borer: Invasion of the Urban Forest and the Threat to North America's Ash Resources. *Journal of Forestry*. 10;3, p118-124

Richter Jr., D. deB., Jenkins, D. H., Karakash, J.T., Knight, J., McCreery, L.R., Nemestothy,K.P. (2009). Wood Energy in America. *Science*. 323 p1432-1433.

Suarez, A.V., Holway, D.A., and Case, T.J. (2001). Patterns of spread in biological invasions dominated by long-distance jump dispersal: Insights from Argentine ants. *Proceedings from the National Academy of Sciences of the United States of America*. 98;3, p1095-1100.

Sydnor, D.T., Bumgardner, M., and Todd, A. (2007). The Potential Economic Impacts of Emerald Ash Borer (Agrilus planipennis) on Ohio, U.S., Communities. *Arboriculture and Urban Forestry*. 33;1 p48-54

Tobin, P.C., Sharov, A.A., Liebhold, A.A., Leonard, D.S., Roberts, E.A., and Learn, M.R. (2004). Management of the Gypsy Moth through a Decision Algorithm under the STS Project. *American Entomologist*, 50:4, p 200-209.

Tobin, P.C., Whitmire, S.L., Johnson., D.M. Bjornstad, O.N., and Liebhold, A.M. (2007b). Invasion speed is affected by geographical variation in the strength of Allee effects. *Ecology Letters.* 10, 36-43

Tobin, P.C. (2007c). Space-time patterns during the during the establishment of non-indigenous species. *Population Ecology*. 49, p257-263.

Tobin, P.C. and Blackburn, L.M. (2008). Long-Distance Dispersal of the Gypsy Moth (Lepidoptera: Lymantriidae) Facilitated its Initial Invasion of Wisconsin. *Environmental Entomology.* 37;1, 87-93

United States Department of Agriculture – Natural Resource Conservation Service (2009). February 3rd, 2009. http://www.mi.nrcs.usda.gov/programs/pmc.html

Appendix 6

Table 6.1 – Nine year spread rates of gypsy moth across three regions and the entire transition front. Data from the STS decision support system (http://da.ento.vt.edu/spread/sprdArea0.html)
.

	WI and MI		OH, IN, IL		NC, VA, WV		All states	
	Spread rate	3 year average	Spread rate	3 year average	Spread rate	3 year average	Spread rate	3 year average
1999	46.45	26.23	6.57	6.57	13.21	3.63	19.22	11.07
2000	6.77	18.04	-28.78	-11.1	16.6	11.56	-4	6.04
2001	14.96	22.72	11.12	-3.69	16.1	15.31	13.79	9.67
2002	33.42	18.38	-13.27	-10.31	-14.55	6.05	-0.41	3.12
2003	35.97	28.11	0.75	-0.46	7.45	3	12.56	8.64
2004	-6.56	20.94	-9.71	-7.41	-12.67	-6.59	-9.78	0.78
2005	-12.03	5.79	8.54	-0.13	14.27	3.01	4.81	2.53
2006	8.06	-3.51	4.87	1.23	8.52	3.37	6.73	0.58
2007	21.04	5.69	5.43	6.28	9.43	10.74	10.37	7.30
9 year mean	16.453	15.821	-1.609	-2.113	6.484	5.564	13.9	10.33

Table 6.2 – Additional data and sources that would be applicable to future research in this area.

Data	Source
Change of Address	USPS
Drivers License	State DMV
Housing Growth	SILVSIS Lab, University of Wisconsin
Wildland Urban Interface	SILVSIS Lab, University of Wisconsin
Parkways and Scenic Roads	National Atlas
Mast Production	Probably only exists locally
Predators	Nature Serve
Traffic	Bureau of Transportation
Commodity Flow	Bureau of Transportation
National Household Travel Survey	Department of Transportation

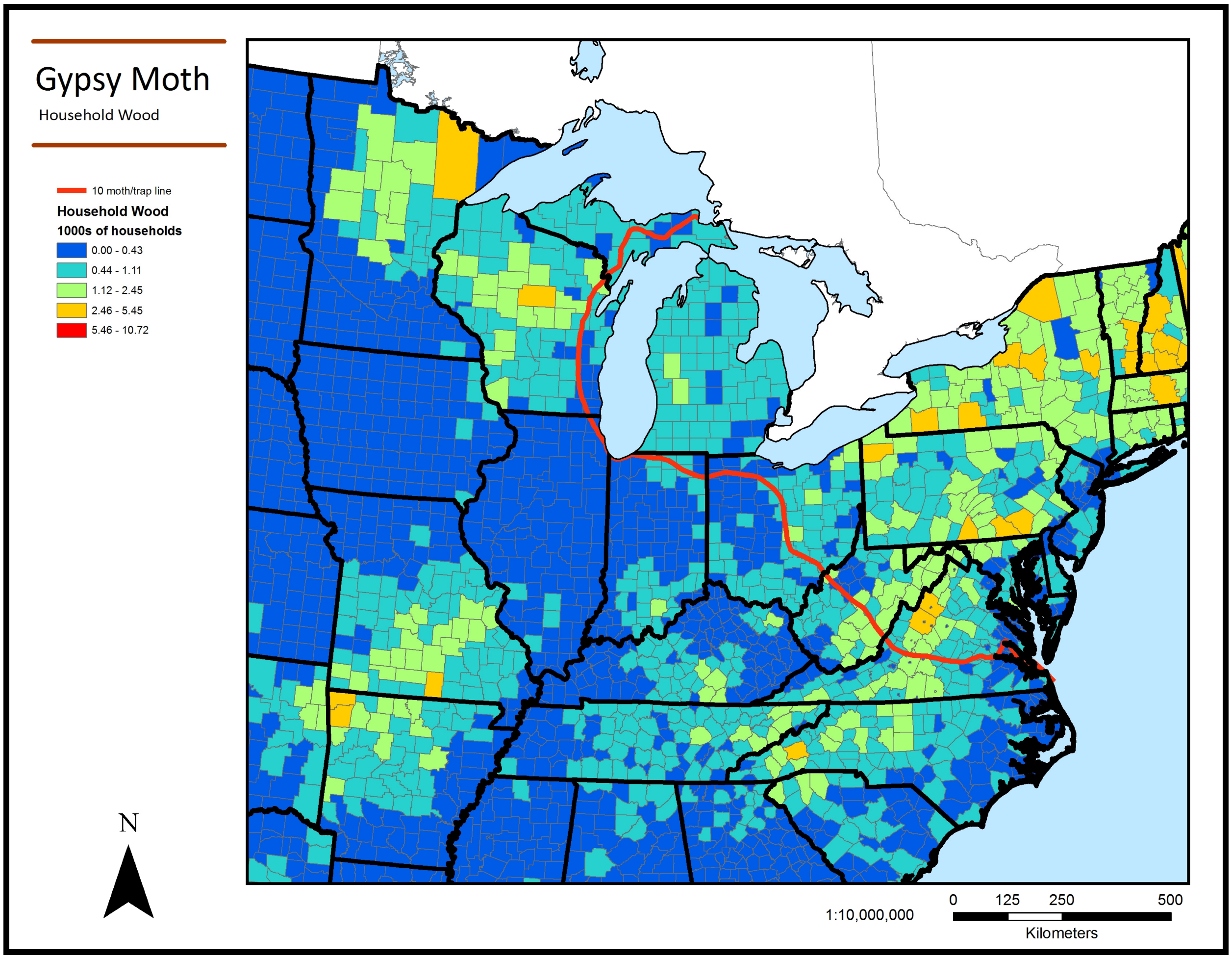


Figure 6.1 – Household wood.

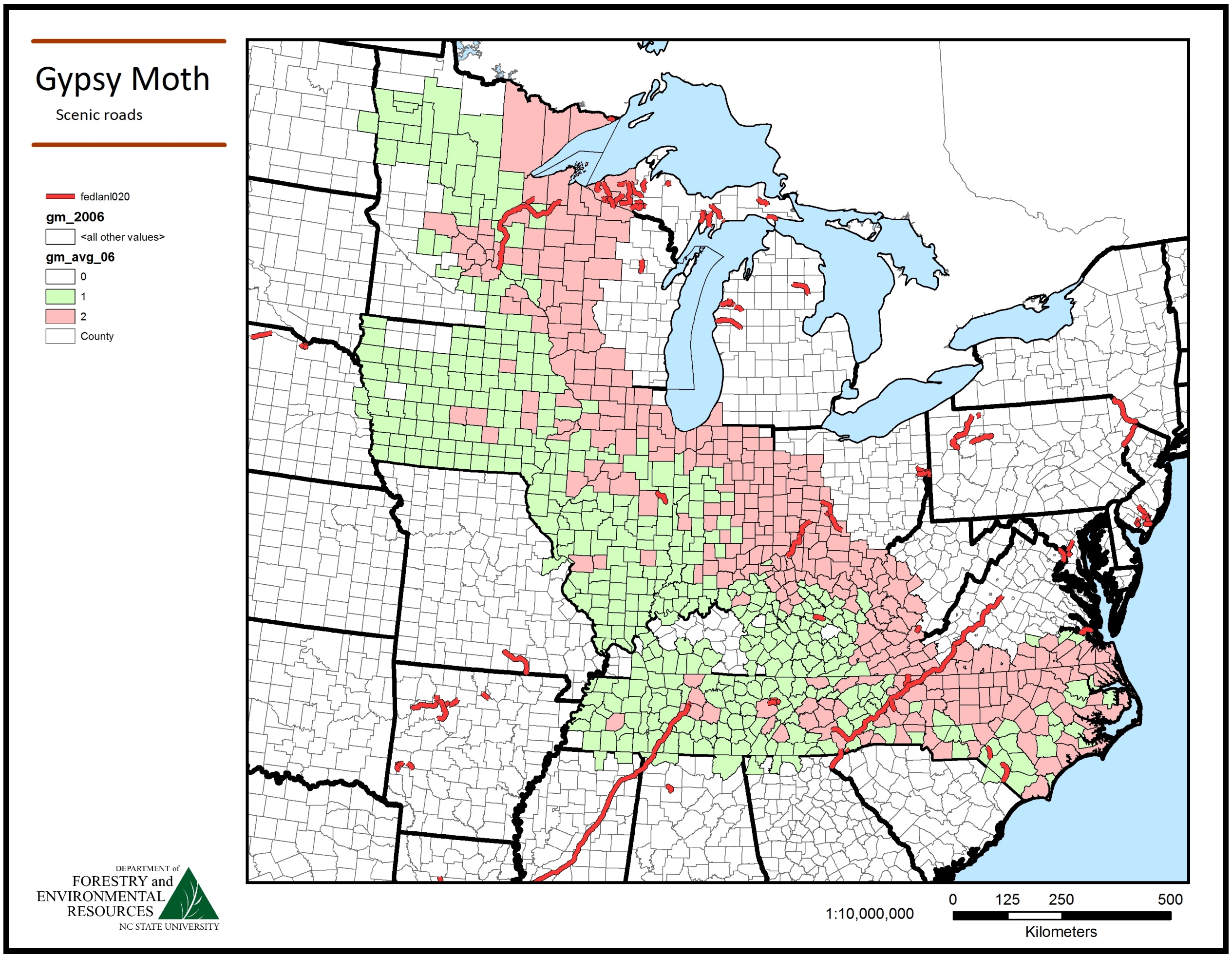


Figure 6.2 – Scenic roads and gypsy moth presence absence data.