#### AN ABSTRACT OF THE DISSERTATION OF

<u>Tzeng Yih Lam</u> for the degree of <u>Doctor of Philosophy</u> in <u>Forest Science</u> presented on <u>May 14, 2010.</u> Title: <u>Exploration of Statistical Methods for Synthesizing the Effects of Variable-</u><u>Retention Harvesting on Multiple Taxa</u>.

Abstract approved:

#### Douglas A. Maguire

Variable-retention harvesting was proposed to reduce loss of biodiversity and ecosystem processes associated with late-seral Douglas-fir (*Pseudotsuga menziesii*) forests in the Pacific Northwest. The Demonstration of Ecosystem Management Options experiment was established to test this hypothesis. Analysis presents various challenges to drawing statistical inferences about treatment effects. This dissertation explored novel statistical methods for understanding the response of multiple forest taxa to variable-retention harvesting.

Excessive zero counts are common among terrestrial small mammal species that are captured infrequently. Zero-inflated and hurdle models are appealing tools for analyzing these data. A simulation was performed to understand the properties and robustness of these models. When true mean abundance was low, the estimated parameters from these models were highly unstable. Goodness of fit criteria could not discern among the processes generating the data.

The Poisson and negative binomial Generalized Linear Models (GLMs) were fitted to four small mammal species with different rates of capture. Predictors included several variables representing vegetation structure. These models and overdispersed Poisson were then specified as Generalized Linear Mixed Models (GLMMs) to account for nesting and blocking in the experimental design. The fitted GLMs indicated that predictors were not consistent among models for the infrequently captured species. Differences in estimated coefficients between GLMs and GLMMs were noticeable. The overdispersed Poisson GLMM was suggested to be most suitable.

Structural Equation Modeling (SEM) is suitable for modeling interactions of many cause-and-effect relationships in forest ecosystems. SEM was applied to understand overstory-understory relationships of late-seral herb species under mature forest conditions and immediately after variable-retention harvesting. In undisturbed forests, light attenuation, belowground competition and stand age were the primary drivers of late-seral herb cover. After variable-retention harvesting, microclimatic stresses were inferred to primarily affect late-seral species diversity and composition. Logging debris had little discernible effect on the change in the late-seral herb community.

The explored statistical models complement conventional methods for studying the effects of variable-retention harvesting. These models address

distributional issues of response data and provide further insight into the complex processes driving managed forest ecosystems. Future analyses should apply a suite of statistical models to gain different perspectives. ©Copyright by Tzeng Yih Lam May 14, 2010 All Rights Reserved Exploration of Statistical Methods for Synthesizing the Effects of Variable-Retention Harvesting on Multiple Taxa

> by Tzeng Yih Lam

### A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Presented May 14, 2010 Commencement June 2010 Doctor of Philosophy dissertation of Tzeng Yih Lam presented on May 14, 2010.

APPROVED:

Major Professor, representing Forest Science

Head of the Department of Forest Ecosystems and Society

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Tzeng Yih Lam, Author

#### **ACKNOWLEDGEMENTS**

The undertaking of a doctoral dissertation could not have been accomplished without the help and tolerance from many people who are central to my life during these past four years. Some offered guidance and encouragement throughout the research and when the progress was stagnant, and some merely being tolerant and stayed out of my way when things were tough.

I offer the most sincere gratitude to my major professor, Doug Maguire. One might be asked, "What is the greatest privilege a major professor can give to his or her student?" I would say, "Freedom." Doug has given me much liberty for developing my own research, working independently, and involving in activities unrelated to my dissertation. It is the freedom to explore both within and without the bounds of my dissertation and not to worry about funding that I am most grateful for. The four years have been remarkable working for Doug.

My research would not have gone far without my committee members and the funding organization. I offer my sincere gratitude to Paul Anderson, Matthew Betts, Alix Gitelman and SueAnn Bottoms particularly for their time in meeting a demanding schedule and program as well as providing feedback, comments and advices. Furthermore, the gratitude is also extended to the Hayes family for graciously providing the Edmund Hayes Endowment for Silviculture Alternatives, which is the major source of support for me to complete the research.

The journey would have been dreadful without constant love and support from my family and friends at home. Without fully understanding what a doctoral degree

entails, they put up my odd-behavior-of-throwing-a-fit-without-reason. I knew they sometimes feared for my sanity but they never voiced it. I sincerely thank them for their understanding and tolerance, especially my parents, sister and my wife Yumei.

There are many to whom I owe my gratitude: John Kershaw from University of New Brunswick, Canada for introducing me to Doug and his neverending challenges and encouragement; Manuela Huso for many hours of discussion; Doug Mainwaring for his enlightening jokes or jests; Andrew Bluhm, Bianca Eskelson, Sean Garber, Yohan Lee, Aaron Wieskittel and for being fine colleagues and friends; and Angel Hu and YuTsai Wang for taking care of me and continue to do so. The list goes on as there are many who have affected every bits of my life, and what you did will always remain in my heart. Thank you the reader for reading to the end of my acknowledgement list and knows the people to whom I forever indebted.

Finally, I thank the many contributors to the Demonstration of Ecosystem Management Options (DEMO) study for allowing me to use the data in my research. The DEMO study is a joint effort of the USDA Forest Service Region 6 and Pacific Northwest Research Station. Research partners include the University of Washington, Oregon State University, University of Oregon, Gifford Pinchot and Umpqua National Forests, and the Washington State Department of Natural Resources. Funds were provided by the USDA Forest Service, PNW Research Station (PNW-93–0455, PNW-97-9021-1-CA, and PNW-01-CA-11261993-091).

## CONTRIBUTION OF AUTHORS

Manuela Huso provided extensive comments and professional expertise with the design and writing of Chapter 2.

## TABLE OF CONTENTS

	Ē	<u>age</u>
1	CHAPTER 1: INTRODUCTION	1
2	CHAPTER 2: MODELING ABUNDANCE OF INFREQUENT SPECIES: A SIMULATION STUDY FOR CLARIFYING THEORETICAL AND ECOLOGICAL INTERPRETATION OF ZERO-MODIFIED MODELS	6
	2.1 INTRODUCTION	7
	2.2 MODEL SPECIFICATION AND COMPARISON	12
	2.3 METHODS	20
	2.4 RESULTS	24
	2.5 DISCUSSION	32
	2.6 CONCLUSIONS	42
3	CHAPTER 3: DETECTING SMALL MAMMAL RESPONSES TO VARIABL RETENTION HARVESTING WITH STATISTICAL MODELS THAT ACCOMMODATE EXCESSIVE ZEROS AND HIERARCHICAL	E-
	SAMPLING	56
	3.1 INTRODUCTION	57
	3.2 MATERIALS	63
	3.3 METHODS	67
	3.4 RESULTS	88
	3.5 DISCUSSION	97
	3.6 CONCLUSIONS	110

# TABLE OF CONTENTS (Continued)

4 CHAPTER 4: STRUCTURAL EQUATION MODELING AND ITS	
APPLICATIONS TO OVERSTORY-UNDERSTORY RELATIONSHIP	' IN
MATURE DOUGLAS-FIR (Pseudotsuga menziesii (Mirb.) Franco)	
FORESTS AND TO LATE-SERAL HERBACEOUS SPECIES IMMED	IATE
RESPONSES TO VARIABLE-RETENTION HARVESTING IN THE	
PACIFIC NORTHWEST	130
4.1 INTRODUCTION	131
	125
4.2 STRUCTURAL EQUATION MODELING (SEM)	135
4 3 OVERSTORY-UNDERSTORY RELATIONSHIPS	148
	1 10
4.4 VARIABLE-RETENTION HARVESTING	169
4.5 CONCLUSIONS	197
	•••
5 CHAPTER 5: CONCLUSIONS	238
5 1 FUTURE RESEARCH	240
J.1 I UI UKE KESEAKUII	240
BIBLIOGRAPHY	

## LIST OF FIGURES

<u>Figure</u> <u>Page</u>
2.1 Relative frequency distributions implied by the 27 dgps. Parameters of the 49
2.2 Percent estimation error (%ERROR) in $\hat{\lambda}$ for the six models fitted to data 50
2.3 Estimated amount of zero-inflation $\hat{p}$ from data generated by 36 dgps
2.4 Results from 1000 simulations by two data generating processes, one
2.5 Percent estimation error (%ERROR) in $\hat{\pi}$ from the six models fitted to
2.6 Difference in AICc ( $\Delta$ AICc) between fitted models and the reference
2.7 The mean and 2.5 and 97.5 percentiles of p-values from $\chi^2$ goodness of fit 55
3.1 Locations of the six DEMO blocks in western Oregon and Washington; 122
3.2 Relative frequency histogram for captures of: (A) <i>Peromyscus maniculatus</i> 123
3.3 Relative frequency distribution for observed and predicted counts from the 124
3.4 Posterior coefficient estimates and 95% credible intervals of CWDVOL 125
3.5 Posterior coefficient estimates and 95% credible intervals of (A) HERB 126
3.6 Posterior standard deviation of random error $\tilde{\sigma}_h$ for each of the 7 estimated 127
3.7 Posterior standard deviation of random block effect $\tilde{\tau}_h$ for each of the 7 128
3.8 Posterior random block effect estimates $\tilde{d}_{hi}$ for all 7 estimated coefficients 129
4.1 A hypothetical path model for abundance of a small mammal species in
4.2 A hypothetical measurement model for species diversity. The circle depicts 218
4.3 A SR model for the overstory-understory relationship in mature Douglas 219
4.4 An equivalent SR model for the overstory-understory relationship in
4.5 Bivariate relationships and Pearson correlations between late-seral herb

# LIST OF FIGURES (Continued)

<u>Figure</u> <u>Page</u>
4.6 Bivariate relationships and Pearson correlations between understory cover 222
4.7 Bivariate relationships and Pearson correlations between fine litter cover 223
4.8 The final fitted SR model with unstandardized parameter estimates. The 224
4.9 The fitted equivalent SR model with unstandardized parameter estimates 225
4.10 Fitted equivalent SR model with the lower limit (0.025 quantile) of 226
4.11 Fitted equivalent SR model with the upper limit (0.975 quantile) of 227
4.12 A SR model for immediate post-harvest responses of late-seral herbs to 228
4.13 Bivariate relationships and Pearson correlations ( <i>r</i> ) between difference in pre- and post-harvest species diversity
4.14 Bivariate relationships and Pearson correlations ( <i>r</i> ) between difference in pre- and post-harvest species diversity
4.15 The final 15% retention SR model with unstandardized parameter 231
4.16 The final 40% retention SR model with unstandardized parameter 232
4.17 The alternative 40% retention SR model after removal of the HARVEST 233
4.18 Bivariate relationships and Pearson correlations ( <i>r</i> ) between: (A) pre 234
4.19 The reduced final 15% retention SR model after removal of the composite 235
4.20 The reduced final 40% retention SR model after removal of the composite 236
4.21 The reduced alternative 40% retention SR model after removal of

## LIST OF TABLES

<u>Table</u> <u>Page</u>
2.1 The 27 dgps (data generating processes) as defined by amount of zero
2.2 The range between the 2.5 and 97.5 percentiles (95W) of %ERROR in the 46
2.3 The mean and 95W (range between 2.5 and 97.5 percentiles) of estimated 47
2.4 The range between the 2.5 and 97.5 percentiles (95W) of the difference in 48
3.1 Estimated coefficients associated with each predictor (and corresponding standard errors) from the ZIP, ZINB,
3.2 Estimated coefficients associated with each predictor (and corresponding standard errors) from the ZIP, ZINB,
3.3 Estimated coefficients associated with each predictor (and corresponding standard errors) from the ZIP, ZINB,
3.4 Estimated coefficients associated with each predictor (and corresponding standard errors) from the ZIP, ZINB,
3.5 Comparison of GLMs and GLMMs for <i>Peromyscus maniculatus</i> (PEMA) in the DEMO study in 1999. Coefficient
3.6 Comparison of GLMs and GLMMs for <i>Clethrionomys gapperi</i> (CLGA) in the DEMO study in 1999. Coefficient
3.7 Comparison of GLMs and GLMMs for <i>Neurotrichus gibsii</i> (NEGI) in the DEMO study in 1999. Coefficient
3.8 Comparison of GLMs and GLMMs for <i>Peromyscus keeni</i> (PEKE) in the DEMO study in 1999. Coefficient
3.9 Relationship between habitat structures and <i>Peromyscus maniculatus</i>
4.1 Variance of observed variables and measurement error as average percent 200
4.2 Estimated direct effects on endogenous latent variables in the overstory 201
4.3 Estimated combined indirect effects on endogenous latent variables in the 202

# LIST OF TABLES (Continued)

<u>Table</u> Page
4.4 Estimated total effects on endogenous latent variables in the overstory 203
4.5 Estimated direct effects on the endogenous latent variables in the final 204
4.6 Estimated combined indirect effects on the endogenous latent variables in 205
4.7 Estimated total effects on the endogenous latent variables in the final 206
4.8 Harvest method and other activities in each experimental block (adapted from Halpern and McKenzie 2001)
4.9 The 11 observed variables used in modeling and their corresponding units, definitions, minimum, mean and
4.10 Estimated direct effects on the endogenous latent variables, difference in 209
4.11 Estimated combined indirect effects on the endogenous latent variables
4.12 Estimated total effects on the endogenous latent variables difference in 212
4.13 Estimated direct effects on the endogenous latent variables, difference in 213
4.14 Estimated combined indirect effects on the endogenous latent variables 215
4.15 Estimated total effects on the endogenous latent variables difference in 216

# EXPLORATION OF STATISTICAL METHODS FOR SYNTHESIZING THE EFFECTS OF VARIABLE-RETENTION HARVESTING ON MULTIPLE TAXA

#### 1 CHAPTER 1: INTRODUCTION

Douglas-fir (Pseudotsuga menziesii (Mirb.) Franco) forests in the Pacific Northwest, USA, have traditionally been managed under even-age silvicultural systems, most commonly the clearcutting system; however, public concern over dramatic reductions in old-growth forests and potential declines in forest biodiversity have fueled renewed interest in the feasibility of other silvicultural systems. Shelterwood with reserves (Matthews 1989) and innovations such as variableretention harvesting (Franklin et al. 1997) have been proposed as a means to meet diverse forest management objectives, particularly on public lands. These methods share a common working hypothesis that the retained structures and associated heterogeneity in stand structure could in the short-term maintain more taxa and ecological processes characteristic of mature forests, and in the long-term accelerate their recovery (Lindenmayer and Franklin 2002, Maguire et al. 2007). In particular, the proposed variable-retention system garners appeal among the public and professionals for its flexibility in accommodating any level of stand retention and a wide array of structural conditions, covering but not limited to the full range of traditional silvicultural systems like clearcutting and single-tree selection (Lindenmayer and Franklin 2002). Variable-retention harvesting is currently mandated on federal forestland in the Pacific Northwest; i.e., at least 15% of live trees in a harvest unit must be retained, with  $\geq$  70% of this retention in aggregates of 0.2– 1.0 ha, and  $\leq$  30% in either dispersed individual trees or smaller groups (< 0.2 ha) (USDA and USDI 1994, Tuchman et al. 1996).

The Demonstration of Ecosystem Management Options (DEMO) study (Aubry et al. 1999), a large-scale operational research experiment implemented in western Oregon and Washington, USA, was initiated in 1994 to feed the need for scientific evidence to support the mandated guidelines (Maguire et al. 2007). This study looked at the effects of variable-retention harvesting on various aspects of biodiversity, microclimate, and human perceptions. The experiment was implemented under a randomized complete block design (RCBD) with subsampling. Two blocks were established in Oregon on the Umpqua National Forest and four in Washington, with three on Gifford Pinchot National Forest and one on the state-owned Capitol Forest. In each block, six harvest treatments were established and defined by level (percentage of basal area) and/or spatial pattern (dispersed vs. aggregated) of retained trees. To date, numerous results from the DEMO experiment have been documented, e.g., changes in forest structure (Maguire et al. 2007), intensity of harvesting disturbances (Halpern and McKenzie 2001), and responses of forest understories (Halpern et al. 2005), terrestrial small mammals (Gitzen et al. 2007) and ectomycorrhizal fungus (Luoma et al. 2004). A main goal is also to synthesize the information across taxa and across ecological processes to assess the comprehensive viability of variable-retention harvesting as a silvicultural system, with a primary objective of ameliorating effects of timber production on forest biodiversity.

Synthesis of the diverse information generated by all the individual DEMO studies presents some challenging statistical issues which could be loosely classified into three groups: study design, response data, and analytical methods. Like other large-scale field experiments, the DEMO study has only a few independent experimental units with systematically arranged sampling units. The observations within an experimental unit are not independent because the sampling units are nested within the larger unit, and the experimental units may be spatially autocorrelated with neighboring units within a block. Hence, the nesting and blocking structures should be properly accounted for valid inferences on the sampling unit-level. The response data for some taxa may have distributional characteristics that require special consideration. For example, the count data of terrestrial small mammals – especially those from infrequently captured species – are dominated by zero counts. Conventional statistical methods such as multiple regression based on normal distribution theory may not be sufficiently flexible and robust to violations of their assumptions. Alternative modeling approaches would be preferable over exclusion of infrequent species from analysis.

The statistical method commonly applied to data from designed experiments is Analysis of Variance (ANOVA) which is regarded as an appropriate method for causal inferences from controlled experiments (Grace et al. 2009). Other methods such as multiple regression, ordination, and classification are exploratory in nature and are most suitable for observational studies. Irrespective of their strength in causal inference, variance and covariance are often the basic statistics behind these statistical methods. For example, the least square estimates of multiple regression coefficients,  $\hat{\beta} = (X'X)^{-1} X'Y$ , are based on the covariance of X and Y and the variance of X (Weisberg 2005), where X is a matrix of observed values of predictors centered at their means and Y is a vector of observed values of a response variable centered at its mean. Ecosystem processes involve complex interactions of many cause-and-effect relationships. Many hypotheses have been offered on relationships among taxa and processes, but the inherent multivariate nature of complex ecosystems frustrates traditional statistical testing. Perhaps statistical methods blending confirmatory and exploratory analyses and allowing tests of explicit hypotheses and models would be more apt for research on silvicultural control of biodiversity.

Many statistical methods commonly applied in other disciplines such as psychometrics and econometrics are relatively unfamiliar to the natural sciences. These methods include zero-inflated statistical models, hurdle models, and Structural Equation Modeling (SEM). In retrospect, studies across disciplines share some commonalities, e.g., some elements in a study cannot be fully manipulated and can only be observed. This condition suggests that methods must be transferable and capable of addressing the analytical challenges mentioned above. The overall goal of this dissertation was to explore the potential of specialized statistical methods for meeting the challenges associated with synthesizing responses of multiple forest taxa to variable-retention harvesting.

Chapter 2 attempts to gain insight into the potential utility and limitation of zero-inflated and hurdle models in modeling a wide variety of distributions that represent counts of infrequently captured species. This objective was pursued by conducting a carefully crafted simulation study and by subsequent ecological interpretation of model assumptions and performance. The simulation study quantified the bias and precision of parameter estimates from data simulated with known processes, and evaluated the efficacy of statistical fit criteria to infer the data generating process. Chapter 3 builds on the simulation study by modeling the association between habitat features and number of captures of two abundant and two infrequent forest floor small mammal species in the DEMO study. The nesting and blocking structures of the DEMO experiment were accounted for in the context of Bayesian Hierarchical Models. A number of empirical comparisons were made among alternative statistical models, including zero-inflated and hurdle models, Generalized Linear Models (GLM; McCullagh and Nelder 1989), and Generalized Linear Mixed Models (GLMM). Chapter 4 presents the concepts and framework of SEM, with the objective of introducing the methodology in sufficient detail to benefit researchers working in the natural sciences and having limited knowledge of SEM. SEM is then illustrated by an application to overstory-understory relationships in Douglas-fir forests, both in an undisturbed condition and after variable-retention harvesting.

## 2 CHAPTER 2: MODELING ABUNDANCE OF INFREQUENT SPECIES: A SIMULATION STUDY FOR CLARIFYING THEORETICAL AND ECOLOGICAL INTERPRETATION OF ZERO-MODIFIED MODELS

Tzeng Yih Lam

Manuela Huso

Douglas A. Maguire

#### **2.1 INTRODUCTION**

Modeling the relationship between habitat features and species presence or abundance provides the fundamental ecological understanding required for maintaining the biodiversity of managed forests. In most studies, modeling efforts rely on linear models that assume a Gaussian distribution or, in some cases, on Generalized Linear Models that assume conformity to another distribution in the exponential family (GLM; McCullagh and Nelder 1989), for example, the Poisson or negative binomial distribution. These models are adequate for relatively abundant species but less satisfactory for infrequent ones. As a consequence, statistical models for the latter species are typically not attempted. Among 15 studies of forest floor small mammals that we reviewed, the number of species captured ranged from 5 to 20, but on average 42% of the species were dropped from analysis (e.g., Cole et al. 1998, Sullivan et al. 2005, Waldien et al. 2006). The two main justifications for dropping species were preponderance of zero observations and model lack of fit. However, habitat associations of rare species are critical to effective management of forest biodiversity because these species are most prone to extinction (Meffe and Carroll 1997).

Zero-inflated (Lambert 1992) and hurdle (Mullahy 1986) models, loosely grouped as zero-modified models (ZMMs), provide a set of potential solutions to the analytical problem encountered in statistical analysis of infrequently captured species. These models explicitly recognize overdispersion caused by a greater number of zero counts than expected under the assumption of a Poisson distribution, and also can account for overdispersion attributable to unexplained heterogeneity in data (Zorn 1998). In short, ZMMs are ideal for modeling infrequent species in habitats where they yield data dominated by zeros and occasional large counts. Although zeroinflated and hurdle models are similar in application, they imply different data generating processes. The zero-inflated model assumes a dual-state latent structure: an observation belongs to either a perfect or an imperfect state (Lambert 1992). An observation in the perfect state can only have a zero count of an event whereas in the imperfect state its count can be any value – including zero. The hurdle model assumes a hurdle latent structure: an observation either crosses a hurdle or not, where the hurdle is generally set at zero (Mullahy 1986). Hence an observation that does not cross the hurdle has zero count. Once crossed, its count is constrained to be greater than zero (positive counts).

A number of ecological studies have evaluated the performances of ZMMs. Early work by Welsh et al. (1996) to model Leadbeater's possum abundance suggested that the hurdle model was preferable for its simple interpretation and ease of model fitting. Cunningham and Lindenmayer (2005) broadened the discussion to types of rarity and effects of low frequency of occurrence on ZMM parameters. Potts and Elith (2006) concluded that the hurdle model provided the best fit to data collected on a rare Australian plant species, *Leionema ralstonii*. For benthic macroinvertebrates community, Gray (2005) found that the negative binomial model was sufficient based on Akaike's Information Criterion (AIC; Burnham and Anderson 2002) and model predictability. Looking at the impact of livestock grazing on woodland birds in subtropical Australia, Martin et al. (2005) obtained mixed results for four bird species; i.e., the zero-inflated model had the best fit for three of the species (brown thornbill, superb fairy-wren and rufous whistler), while the negative binomial model proved best for the fourth (noisy miner). In an extensive study of 20 datasets containing 1672 different taxa, Warton (2005) compared models by AIC and concluded that the negative binomial model fit best for most count data and that the hurdle model either could not be fitted or performed poorly. Sileshi (2008) studied six soil animals in agroforestry and woodland environments in eastern Zambia and found that the negative binomial model performed best for 10 out of 12 cases and a zero-inflated model for the other two. Most of these studies concluded that the Poisson model was inadequate with respect to both goodness of fit and predictions on independent datasets.

A shared complication among the above comparative studies was that the underlying data generating process (dgp) of the field data was unknown and probably more complicated than implied by any of the theoretical statistical models. Specific models may therefore be identified as best by some goodness of fit criteria for one or a combination of two very different reasons; i.e., they simulate the underlying dgp or they are very flexible for fitting any distribution of data. Another problem is that it is impossible to know the extent of accuracy and precision of parameter estimates without knowing the dgp. A final complication is the impact of covariates on the relation between possible conditional distributions and the marginal distribution of target species. The relation between conditional and marginal distributions is difficult to visualize, particularly in data from field studies where observations are not equally frequent at different levels of the covariates. Count data appearing to have excessive zeros in the marginal count frequency distribution could in fact be negative binomialor Poisson-distributed after conditioning on covariates. This situation may characterize the systems described by Sileshi (2008). Furthermore the immense number of potential covariates in most ecological studies complicates evaluation of the dgp because a large number of alternative models are possible and each will have a different impact on apparent lack of fit and statistical properties of both conditional and marginal distributions. Simulation studies with known dgp and no complicating influence of covariates can therefore provide important guidelines for interpreting results from fitting alternative statistical models to data generated by more complicated and largely unknown ecological processes.

Some simulation studies have explored the properties of zero-inflated and hurdle models individually. Lambert (1992) included a small simulation study with one covariate and varying sample sizes in her theoretical development of the zeroinflated model. Hall and Zhang (2004) evaluated a zero-inflated model fitted to simulated data that included a random effect for clustering. Min and Agresti (2005) studied the effect of zero inflation and deflation on zero-inflated and hurdle models using the same simulation structure as Lambert (1992). For comparing ZMMs to other GLM specifications, many studies outside field of ecology work with empirical data (e.g., Gao and Khoshgoftaar 2007, Hall 2000, Rose et al. 2006). To our knowledge no one has rigorously compared the performance of zero-inflated, hurdle and other GLM models on data simulated by a variety of known dgps.

Lord et al. (2005, 2007) contended that choosing between zero-inflated models and other GLM specifications should be based on theoretical principles of the process that generated the event at specific temporal and spatial scales. They further downplayed the role of statistical fit in choosing the most appropriate model. If the dgp is different between models, choosing one over another based on statistical fit assumes that the fit criteria reflect conformity to the underlying dgp. Gray (2005) briefly related the dual-state process of zero-inflated models to ecological niche, but very little other work in the literature attempts to address any ecological interpretation of the assumptions and latent structures characterizing zero-inflated and hurdle models.

Cunningham and Lindenmayer (2005) noted that zero-inflated and hurdle models expanded the number of species that can be subjected to rigorous statistical analyses compared to those available for abundant taxa. The goal of our study was to gain insight into the potential utility and limitation of ZMMs through a carefully crafted simulation study, accompanied by ecological interpretation of model assumptions and performance. Specific objectives included: (1) to quantify the accuracy and precision of parameter estimates from data simulated with known dgps, (2) to evaluate the efficacy of statistical fit criteria commonly used in model selection, and (3) to provide ecological interpretation of the assumptions and latent structures underlying zero-inflated and hurdle models.

#### 2.2 MODEL SPECIFICATION AND COMPARISON

The six models considered in this study include the following: Poisson GLM (POIS), negative binomial GLM (NB), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), hurdle Poisson (HPOIS) and hurdle negative binomial (HNB). Increasing interest in these models has resulted in several excellent references for statistical properties, model specifications, and extensions (see Cameron and Trivedi 1998, Greene 1994, King 1989, Winkelmann 2008); therefore, only the most relevant model properties are presented below. In all models, *Y* is the count of an event,  $\lambda$  is the mean of a Poisson distribution, and  $\theta$  is the overdispersion parameter of a negative binomial distribution.

#### 2.2.1 Poisson GLM (POIS)

The Poisson GLM (POIS) is a benchmark model widely used for modeling count data. The Poisson distribution is,

$$\Pr(Y = y) = \frac{e^{-\lambda} \lambda^{y}}{y!}$$
(2.1)

where y = 0, 1, 2, ... and  $\lambda > 0$ . The expectation and variance are,

$$E(Y) = Var(Y) = \lambda \tag{2.2}$$

indicating equidispersion (Winkelmann 2008); i.e., the expectation equals the variance.

#### 2.2.2 Negative Binomial GLM (NB)

The restrictive equidispersion assumption of POIS may be violated when heterogeneity in counts is more than expected, resulting in overdispersion (larger variance) relative to the Poisson distribution. Following Venables and Ripley (2002), a random variable *Z*, assuming a gamma distribution with parameter  $\theta$ , is added to  $\lambda$ ,

$$Y \mid Z \sim Poisson(\lambda Z)$$
$$\theta Z \sim gamma(\theta)$$

This results in a gamma-Poisson mixture distribution, commonly known as negative binomial distribution,

$$\Pr(Y = y) = \frac{\Gamma(y + \theta)}{\Gamma(\theta) y!} \frac{\lambda^{y} \theta^{\theta}}{(\lambda + \theta)^{y + \theta}}$$
(2.3)

where  $y = 0, 1, 2, ..., \lambda > 0$  and  $\theta > 0$ . The expectation and variance are,

$$E(Y) = \lambda \tag{2.4}$$

$$Var(Y) = \lambda + \left(\frac{1}{\theta}\right)\lambda^2$$
(2.5)

Equation (2.5) shows that the model allows overdispersion relative to the Poisson model, with the variance exceeding the expectation by the amount  $\lambda^2/\theta$ .

#### 2.2.3 Zero-Inflated Models (ZI)

The concept of a zero-inflated model (ZI) originated from Rider (1961) and Cohen (1963) and was later applied by Johnson and Kotz (1969) to data with excessive zeros. Lambert (1992) expanded their work and formulated the concept under the GLM framework. The model is also known as the 'zero-altered probability model' or 'count model with added zeros'; however, we have adopted the terminology presented by Lambert (1992). In the context of finite mixture model (McLachlan and Peel 2000), the ZI model is a mixture of a point mass at zero and a count distribution (either Poisson or negative binomial), and has one latent structure (Baughman 2007). As mentioned earlier, the latent structure is a dual-state process: a perfect state observing only zero count and an imperfect state with any realization including zero. Let p be the probability of remaining in the perfect state or the amount of zero-inflation. The zero-inflated Poisson distribution (ZIP) is,

$$\Pr(Y = y) = \begin{cases} p + (1 - p)e^{-\lambda}, & y = 0\\ (1 - p)\frac{e^{-\lambda}\lambda^{y}}{y!}, & y > 0 \end{cases}$$
(2.6)

where  $p \in (0,1)$  and  $\lambda > 0$ . The expectation and variance are,

$$E(Y) = (1-p)\lambda \tag{2.7}$$

$$Var(Y) = (1-p)\lambda + (1-p)p\lambda^{2}$$
(2.8)

The ZIP formulation clearly shows that Pr(Y = 0) is a mixture of two sources of zero – from the perfect state *p* and the imperfect state  $(1-p)e^{-\lambda}$ . As *p* approaches zero, ZIP theoretically reduces to POIS (eqn. 2.6). The expected value of *Y* is always smaller than that for POIS by magnitude of 1-p due to zero-inflation (eqn. 2.7). The distribution is inherently zero-driven overdispersion – overdispersion caused by a greater number of zero counts than expected under the Poisson distribution (Zorn 1998). The amount of zero-driven overdispersion is  $(1-p)p\lambda^2$ .

The zero-inflated negative binomial distribution (ZINB) is,

$$\Pr(Y = y) = \begin{cases} p + (1-p)\frac{\theta^{\theta}}{(\lambda+\theta)^{\theta}}, & y = 0\\ (1-p)\frac{\Gamma(y+\theta)}{\Gamma(\theta)y!}\frac{\lambda^{y}\theta^{\theta}}{(\lambda+\theta)^{y+\theta}}, & y > 0 \end{cases}$$
(2.9)

where  $p \in (0,1)$ ,  $\lambda > 0$  and  $\theta > 0$ . The expectation and variance are,

$$E(Y) = (1-p)\lambda \tag{2.10}$$

$$Var(Y) = (1-p)\lambda + (1-p)\left(p + \frac{1}{\theta}\right)\lambda^2$$
(2.11)

A notable feature of these ZI models is that the expectation of ZIP and ZINB are identical, but the variance is not. ZINB has both sources of overdispersion (eqn. 2.11); i.e., zero-driven overdispersion as well as overdispersion captured by  $\theta$  that could result from added variability from omitted covariates (Minami et al. 2007), clustered Poisson process and inter-subject variability (McCullagh and Nelder 1989), or temporal dependency between observations (Rose et al. 2006) (hereafter referred to as overdispersion). Zorn (1998) stressed the importance of distinguishing between overdispersion driven by zero-inflation and that driven by these other sources. As is the case for ZIP, ZINB theoretically reduces to NB as *p* approaches zero.

#### 2.2.4 Hurdle Model (H)

Mullahy (1986) expanded the hurdle model (H) originally proposed by Cragg (1971). As mentioned earlier, the H model has a latent structure represented by the "hurdle" at zero; i.e., the count is zero before the hurdle is crossed and positive afterwards. As a finite mixture model, the H model is a mixture of a point mass at zero and a truncated-at-zero count distribution (either Poisson or negative binomial). Baughman (2007) argued that from a dual-state process perspective, the H model would have two latent structures; one is the hurdle and another is the imperfect state. The latter is implicit because every observation must be in the imperfect state given that they all have the inherent "potential" of crossing the hurdle. In fact, it is the probability of crossing the hurdle,  $\pi$ , that is the major feature of interest in this statistical model, implying that all observed zeros belong to the imperfect state. However, this apparent difference from ZI model is not fully recognized in the ecological literature.

The hurdle Poisson distribution (HPOIS) is,

$$\Pr(Y = y) = \begin{cases} 1 - \pi, & y = 0\\ \frac{\pi}{1 - e^{-\lambda}} \frac{e^{-\lambda} \lambda^{y}}{y!}, & y > 0 \end{cases}$$
(2.12)

where  $\pi \in (0,1)$  and  $\lambda > 0$ . The expectation and variance,

$$E(Y) = \frac{\pi}{1 - e^{-\lambda}}\lambda \tag{2.13}$$

$$Var(Y) = \frac{\pi}{1 - e^{-\lambda}} \lambda + \frac{\pi}{1 - e^{-\lambda}} \left( 1 - \frac{\pi}{1 - e^{-\lambda}} \right) \lambda^2$$
(2.14)

The expectation and variance of HPOIS resemble those of ZIP in that the expectation is smaller than that of the POIS model and the overdispersion is zerodriven. If Pr(Y = 0) is equal in the HPOIS and ZIP models (i.e.,  $1-\pi = p+(1-p)e^{-\lambda}$ ), then HPOIS will reduce to ZIP. This implies that if  $\pi$  and p are not conditioned on covariates under the GLM framework, then both models are equivalent; otherwise, they will be distinct specifications (Rose et al. 2006).

The hurdle negative binomial distribution (HNB) is,

$$\Pr(Y = y) = \begin{cases} 1 - \pi, & y = 0\\ \frac{\pi}{1 - \frac{\theta^{\theta}}{(\lambda + \theta)^{\theta}}} \frac{\Gamma(y + \theta)}{\Gamma(\theta) y!} \frac{\lambda^{y} \theta^{\theta}}{(\lambda + \theta)^{y + \theta}}, & y > 0 \end{cases}$$
(2.15)

where  $p \in (0,1)$ ,  $\lambda > 0$  and  $\theta > 0$ . The expectation and variance are,

$$E(Y) = \frac{\pi}{1 - \frac{\theta^{\theta}}{\left(\lambda + \theta\right)^{\theta}}} \lambda$$
(2.16)

$$Var(Y) = \frac{\pi}{1 - \frac{\theta^{\theta}}{(\lambda + \theta)^{\theta}}} \lambda + \frac{\pi}{1 - \frac{\theta^{\theta}}{(\lambda + \theta)^{\theta}}} \left(1 - \frac{\pi}{1 - \frac{\theta^{\theta}}{(\lambda + \theta)^{\theta}}} + \frac{1}{\theta}\right) \lambda^{2}$$
(2.17)

Similar to ZINB, HNB accounts for both zero-driven overdispersion and

overdispersion in the data representing counts greater than zero.

#### 2.2.5 GLM Framework

Under the GLM framework, **B**, **G** and **K** are three sets of covariates;  $\beta$ ,  $\gamma$  and  $\xi$  are the sets of parameters corresponding to **B**, **G** and **K**, respectively;  $\lambda$ , **p** and  $\pi$  are the parameters defined as above where  $\lambda = (\lambda_1, ..., \lambda_n)^T$ ,  $\mathbf{p} = (p_1, ..., p_n)^T$ , and  $\pi = (\pi_1, ..., \pi_n)^T$  for *n* samples of *Y*. Conditioning the parameters  $\lambda$  on covariates **B** yields the following Poisson regression model with a log link function (McCullagh and Nelder 1989):

$$\log(\lambda) = \mathbf{B}\boldsymbol{\beta} \tag{2.18}$$

Assume the **G** and **K** are two sets of covariates that determine the conditional distributions of the zero-inflation and hurdle components in the ZI and H models, respectively, yielding the following binomial regression component with logit link function:

$$\operatorname{logit}(\mathbf{p}) = \log\left(\frac{\mathbf{p}}{1-\mathbf{p}}\right) = \mathbf{G}\boldsymbol{\gamma}$$
(2.19)

$$\operatorname{logit}(\boldsymbol{\pi}) = \log\left(\frac{\boldsymbol{\pi}}{1-\boldsymbol{\pi}}\right) = \mathbf{K}\boldsymbol{\xi}$$
(2.20)

The sets of covariates between **B**, **G** and **K** could be entirely identical, partially overlapping or mutually exclusive.

#### 2.2.6 Model Comparison

In many empirical studies of ZMMs (e.g., Sileshi 2008, Warton 2005), information criteria such as AIC and Bayesian Information Criterion (BIC; Burnham and Anderson 2002) are common goodness of fit criteria for model selection and comparison. Multiple models have been observed to give nearly identical values when fitted to a set of data (e.g., Gray 2005). Burnham and Anderson (2002) examined results from Lindsey and Jones (1998) to validate use of AIC and BIC in comparing models with differing distributional assumptions, provided that the complete probability distribution (including the normalizing constant) was used to compute loglikelihood of each model. However, the study from Lindsey and Jones (1998) was empirical rather than theoretical. Comparison between POIS, NB and ZMMs may be feasible because all models are commonly based on discrete distributions. However, log-likelihoods of different models can be fundamentally different in magnitude, in turn affecting model comparison with different likelihood functions. Regardless, a version of AIC corrected for small sample sizes, AICc, was selected as a criterion for comparing models and was computed as,

$$AICc = -2\log L + \frac{2h(h+1)}{n-h-1}$$
(2.21)

where *L* is the likelihood, *n* is sample size and *h* is number of model parameters.

Another major criterion for comparing models is the degree of correspondence between the observed frequency distribution of counts and the expected distribution of counts under the fitted statistical model. The  $\chi^2$  goodness of fit statistic is one common measure of this degree of correspondence. To evaluate the performance of alternative statistical models in this simulation study, a pseudo- $\chi^2$  statistic,  $X^2$ , was calculated as

$$X^{2} = \sum_{c=1}^{C} \frac{\left[\sum_{i=1}^{n} I(y_{i} = c) - \sum_{i=1}^{n} \Pr(y_{i} = c)\right]^{2}}{\sum_{i=1}^{n} \Pr(y_{i} = c)}$$
(2.22)

where  $I(y_i = c) = 1$  if y in the *i*<sup>th</sup> observation has count c and 0 otherwise,  $Pr(y_i = c)$  is the probability of the event  $\{y_i = c\}$  under the fitted model, *n* is the sample size and *C* is the maximum count (Affleck 2006, Fortin and DeBlois 2007). Similar measure was used by Lambert (1992) to visually assess model fit in developing the ZI models. However, Affleck (2006) noted that  $X^2$  as calculated was not  $\chi^2$  distributed because the observed count data were first fitted to obtain parameter estimates. Nonetheless, the p-value assuming a  $\chi^2$  distribution with *C*-1 degree of freedom was applied as a relative measure of the degree of correspondence between observed and predicted count frequency.

Despite the described weaknesses of both AICc and the  $\chi^2$  goodness of fit statistic for comparing models, these statistics presented an opportunity to understand their behavior under known conditions of simulation, allowing stronger assessment of their applicability.

#### **2.3 METHODS**

The six models were evaluated by simulating data from a known dgp and without any covariates (i.e., **B**, **G** and **K** for this special case were a column vector of ones, so only the intercept term was included in eqns. 2.18–2.20). Excluding covariates in the simulation study simplified the design and allowed more

straightforward interpretation of results. The simulation study was set under a factorial design with varying levels of four factors: (1) p – amount of zero-inflation as defined in ZI models at five levels (0.0, 0.25, 0.5, 0.75), (2)  $\lambda$  – mean of a Poisson distribution at three levels (0.3, 1.5, 5.0), (3) RATIO<sub>NB</sub> – amount of overdispersion indexed by the ratio of the variance to the expectation of a negative binomial distribution,  $Var(Y)/E(Y) = (\lambda + (1/\theta)\lambda^2)/\lambda$ , at three levels (1, 1.5, 3), and (4) SAMPLE – sample size or number of observations at five levels (25, 50, 75, 100, 250). This design resulted in a total of 180 combinations.

The selected levels of  $\lambda$  were based on estimated mean abundances of small mammal species captured in the DEMO study. The infrequent species generally had estimated mean abundance between 0.004 and 0.05, and common species were represented by  $\lambda = 1.5$ . The  $\lambda$  level of 5.0 was not observed in the DEMO study, but was chosen as reasonable upper limit for the purpose of studying model behavior. The selected levels of *p* covered the range between the two limits to this parameter, 0 and 1. The selected levels of RATIO<sub>NB</sub> represented 0, 50, and 100% overdispersion.

Deciding the levels for p and  $\lambda$  was straightforward; however, appropriate levels for  $\theta$  in the NB, ZINB and HNB models were less obvious. As ZINB and HNB were derivatives of NB, one approach was choosing  $\theta$  levels based on a negative binomial distribution and extending them to the ZINB and NB models. Thus, RATIO<sub>NB</sub>, a function of both  $\theta$  and  $\lambda$  as defined above, was selected as a measure of the variance relative to the expectation of a negative binomial distribution, thereby standardizing the degree of overdispersion relative to  $\lambda$ . For any  $\lambda$ , the overdispersion
parameter was  $\theta = \lambda/(\text{RATIO}_{\text{NB}}-1)$ . One possible flaw in this approach is that interpretation of the selected  $\theta$  values would be different between NB, ZINB and HNB. For a negative binomial distribution,  $\theta$  did not distinguish between zero-driven overdispersion and other sources of overdispersion, so the selected  $\theta$  values encompassed both types. For ZINB and HNB, the selected  $\theta$  values only described the non-zero overdispersion. Regardless, the measure RATIO<sub>NB</sub> was simply a means of specifying dgps that would result in different distributions, with the objective of quantifying accuracy and precision of parameter estimates.

A unique dgp was defined by each combination of p,  $\lambda$  and RATIO<sub>NB</sub> (relative  $\theta$ ); therefore, the simulation study consisted of 36 dgps with five sample sizes each. For brevity, Table 2.1 lists 27 dgps excluding those with p = 0.5. Associated with each dgp were known values of  $\theta$  and  $\pi$ , data generating distributions and total overdispersion (Table 2.1). Without overdispersion, i.e., RATIO<sub>NB</sub> = 1,  $\theta$  was undefined ( $\infty$ , Table 2.1). However, when RATIO<sub>NB</sub> > 1, the implied  $\theta$  was listed. By definition (eqns. 2.12 and 2.15),  $\pi = 1-\Pr(Y=0)$  whereby  $\Pr(Y=0)$  could be easily calculated for each dgp by knowing the corresponding probability distribution that generating distributions or dgps. As mentioned, the ZI model without covariates was equivalent to the H model, so either *zip* or *hpois* could serve as the data generating distributions under zero-inflation (p > 0) as could either *zinb* or *hnb*. Total overdispersion was defined as Var(Y)/E(Y) for any data generating distribution, e.g., the total overdispersion of a *zinb* was the ratio of eqn. 2.10 to eqn. 2.11,  $[(1-p)\lambda+(1-$   $p)(p+1/\theta)\lambda^2]/(1-p)\lambda$ , so included both zero-overdispersion and overdispersion from other sources. Total overdispersion for each of the 27 dgps was computed to help characterize the resulting distributions.

For each of the 180 combinations of dgp and sample size, 1000 sets of observations were simulated based on the data generating distribution in Table 2.1. Each set of observations, OBS, was fitted to all six statistical models: POIS, NB, ZIP, ZINB, HPOIS, and HNB. For each model, therefore, a given dgp resulted in 1000 corresponding sets of estimated parameter(s), AICcs, and p-values from the  $\chi^2$  Goodness of Fit Test. For each value of the estimated parameter(s), absolute estimation error (|Estimated–True|) or percent estimation error (%ERROR = 100×[Estimated–True]/True) were then calculated. Finally, absolute estimation error, %ERROR, AICc and the  $\chi^2$  p-value were summarized by computing a mean and the range between the 2.5% and 97.5% sample quantiles (95W).

Assessing the performance of a model fitted to data generated by a process other than its matching dgp was analogous to fitting alternative models to field data resulting from complex and unknown ecological processes. Estimation error under these conditions provides a measure of model accuracy, and addresses the question of how robust the alternative models are to assumptions about the underlying dgp. Absolute %ERROR less than 10% was arbitrarily set as the level below which error could be considered negligible. To effectively understand goodness-of-fit between models, the difference in AICc ( $\Delta$ AICc) between models is the most meaningful criterion. This difference was calculated with respect to the model fitted to data generated by its matching dgp, hereafter referred to as the reference model, under the assumption that this reference model should have the best fit. For example, if *zip* was the dgp, then the ZIP model was the reference model and  $\Delta$ AICc was the difference between a given AICc and the AICc of the ZIP model. Any  $\Delta$ AICc smaller than ± 5 units was arbitrarily assumed not to differ from the reference model.

The simulation study was performed with R open source statistical software (R Development Core Team 2009) with the package pscl contributed from Zeileis et al. (2008). Sample size was at first thought to affect the simulation outcome. However, preliminary analyses showed that increasing sample size decreased 95W and had only minor effects on estimation error for most parameters. For the sake of brevity, we therefore only present results from sample size of 100. This sample size should be sufficiently large to observe any statistical properties of a model, particularly in the absence of any covariates, and approximates the sample size of typical field studies. For some comparisons, results from p = 0.5 were also excluded because they predictably fell between those of p = 0.25 and p = 0.75.

#### 2.4 RESULTS

Relative frequency distributions implied by probability mass functions of the 27 dgps varied substantially (Figure 2.1). Most dgps implied distributions with a right skew when  $\lambda \le 1.5$ , and as expected, the relative frequency at Y = 0 increased with increasing *p*. A notable observation was that the distribution of dgps with zero-inflation (i.e., p > 0) became bimodal when  $\lambda = 5$ .

## 2.4.1 Mean Parameter $\lambda$

The average %ERROR in estimating mean of a Poisson distribution  $\lambda$  from the fitted POIS and NB models was identical across all the dgps because both models produced the same  $\hat{\lambda}$  (Figure 2.2). The estimation error was negligible for these two models fitted to data from any dgp that had no zero-inflation (Figure 2.2A,D,G); however, this error became increasingly negative as *p* increased (e.g., Figure 2.2A–C). In contrast, estimation error from fitted POIS and NB models did not change dramatically with increasing overdispersion (e.g. Figure 2.2B,E,H) or increasing  $\lambda$  (see horizontal lines for POIS and NB in any panel of Figure 2.2). Although percent estimation error, which was a relative measure, was similar across all levels of  $\lambda$ , the absolute impact would by definition be greater for higher levels of  $\lambda$ . At *p* = 0.75 and  $\lambda = 5$ , the average  $\hat{\lambda}$  was 1.25; however, the unimodal frequency distribution implied by a POIS model with fitted  $\hat{\lambda} = 1.25$  would be quite different from the bimodal frequency distribution of a dgp with *p* = 0.75 and  $\lambda = 5$  (Figure 2.1).

The average %ERROR of fitted ZI and H models converged on zero with increasing  $\lambda$  for any given dgp. Most  $\hat{\lambda}$  have large error at  $\lambda = 0.3$ , but all estimation errors were trivial (< 0.5%) at  $\lambda = 5$  (Figure 2.2). The fitted ZIP model overestimated  $\lambda$  for datasets generated by any dgp other than a *zip* process, and the error increased three-fold when overdispersion (RATIO<sub>NB</sub>) doubled from 1.5 to 3 (e.g., Figure 2.2E and H). For the most part, results from fitting the HPOIS model were identical to those from fitting a ZIP model. Both ZINB and HNB models adequately estimated the  $\lambda$  of a *nb* dgp only when  $\lambda \ge 1.5$  (Figure 2.2D and G), but underestimated that of a *zip* dgp (Figure 2.2B and C). For the *pois* dgp, the ZINB model overestimated the  $\lambda$  but HNB underestimated it (Figure 2.2A). When fitted to counts generated by the *zinb* dgp, the average %ERROR at  $\lambda = 0.3$  of the HNB model was smaller than that of the ZINB model and the estimation error from both models increased with increasing *p* and overdispersion index RATIO<sub>NB</sub> (Figure 2.2E,F,H,I).

The range between the 2.5 and 97.5 percentiles (95W) of %ERROR consistently decreased with increasing levels of  $\lambda$ ;  $\hat{\lambda}$  was most unstable at  $\lambda = 0.3$  with range up to 1000% for some dgps (Table 2.2). Likewise, the 95W consistently increased with increasing overdispersion for all the six fitted models. For fitted ZI and H models, the 95W generally decreased with increasing  $\lambda$  irrespective of *p* and RATIO<sub>NB</sub>, but this range in H models was as large as or larger than that in the ZI model (Table 2.2).

#### 2.4.2 Amount of Zero-Inflation p

Simulation results for the amount of zero-inflation p was best presented by comparing estimated and known values ( $\hat{p}$  vs. p) because %ERROR was undefined when p = 0 (Figure 2.3). For fitted ZIP and ZINB models with  $\lambda = 5$  and at a given level of overdispersion (RATIO<sub>NB</sub>), average absolute estimation error was small to negligible and the  $\hat{p}$  was least variable (narrowest 95W, Figure 2.3C,F,I); however, both error and variability generally increased with decreasing  $\lambda$ , reaching a maximum at  $\lambda = 0.3$  (Figure 2.3A,D,G). This trend suggested that ZI models could fit a bimodal count distribution (data from a dgp with large  $\lambda$  and p >> 0, Figure 2.1) without difficulty, but parameter estimates became unstable when the distribution was unimodal and extremely right-skewed.

When count distribution was not overdispersed (RATIO<sub>NB</sub> = 1),  $\hat{p}$  from the fitted ZIP model has on average positive estimation error when  $\lambda = 0.3$  and p = 0, but the error was increasingly negative as p increased (Figure 2.3A). With increasing overdispersion (RATIO<sub>NB</sub>), the estimation error was increasingly positive at a given level of  $\lambda$  (Figure 2.3D–I). The estimate  $\hat{\lambda}$  from the fitted ZIP model displayed a strong positive error in the presence of overdispersion (RATIO<sub>NB</sub> > 1; Figure 2.2D–I), implying that this model accounted for overdispersion at low  $\lambda$  by increasing the estimate  $\hat{\lambda}$  and assuming that more of the zero observations originated from the perfect state, as confirmed by the positive error in  $\hat{p}$  (Figure 2.3D,G).

For the fitted ZINB model, the average estimation error was small to negligible and negative for most dgp with  $\lambda \ge 1.5$  and p > 0 (Figure 2.3B,C,E,F,H,I). The observed error at  $\lambda = 0.3$  followed the trend of the fitted ZIP model; however, the average absolute estimation error for the fitted ZINB model was generally higher than for the fitted ZIP model when RATIO<sub>NB</sub> = 1 and lower when RATIO<sub>NB</sub> > 1 (Figure 2.3A,D,G). The 95W from the fitted ZINB model was generally as wide as or wider than that of the fitted ZIP model for any dgp with  $\lambda \le 1.5$ , indicating uncertainty in separating the contribution of the perfect and imperfect states to the number of zero observations. At a low Poisson mean ( $\lambda = 0.3$ ), in the presence of zero inflation (p = 0.25 or p = 0.75), and in absence of overdispersion (RATIO<sub>NB</sub> = 1), the fitted ZIP models displayed strong curvilinear relationships between  $\hat{\lambda}$  and  $\hat{p}$  (Figure 2.4A). Although the  $\hat{\lambda}$  and  $\hat{p}$  of ZI models had large error at  $\lambda = 0.3$ , the error in estimated marginal probability of observing zero Pr(Y = 0) was negligible (Figure 2.4B). The lack of error in this marginal probability was attributable to the compensating effect between these two parameter estimates  $\hat{\lambda}$  and  $\hat{p}$  (Figure 2.4A). The probability mass function of the ZIP distribution (eqn. 2.6) indicated that when  $\lambda$  was estimated to be small,  $e^{-\lambda}$  would be large and the corresponding estimate of p had to be small for the model to fit the simulated number of zeros well.

## 2.4.3 Probability of Crossing Hurdle $\pi$

The average %ERROR in the estimated probability of crossing the zero hurdle  $(\hat{\pi})$  was identical between the fitted HPOIS and HNB models, and was negligible for all dgps (Figure 2.5). The range between the 2.5 and 97.5 percentile (95W) of %ERROR in  $\hat{\pi}$  decreased with increasing  $\lambda$ , matching the wider variability in %ERROR of  $\hat{\lambda}$  for dgps with  $\lambda = 0.3$  (Table 2.2). The 95W increased with increasing amount of zero inflation (*p*) and increasing overdispersion (RATIO<sub>NB</sub>) (Figure 2.5).

## 2.4.4 Overdispersion Parameter $\theta$

The striking result from estimated overdispersion parameter ( $\hat{\theta}$ ) was that the three models (NB, ZINB and HNB) produced reasonable mean estimates and an

acceptably narrow 95W for very few dgps (Table 2.3). Surprisingly, the ZMMs performed worse than NB model (Table 2.3). Results in this case did vary by sample size, suggesting that ZMMs required a sample size  $\geq 250$  to find a reasonable estimate of  $\theta$ . In the absence of overdispersion (RATIO<sub>NB</sub> = 1),  $\hat{\theta}$  needed to be exceptionally large for the *Var*(*Y*) of a negative binomial distribution to approach the *E*(*Y*) of that distribution; however, the fitted NB model yielded  $\hat{\theta} = 1.8$  for the dgp with p = 0.25 and  $\lambda = 5$ , and  $\hat{\theta} = 0.3$  for the dgp with p = 0.75 and  $\lambda = 1.5$  (Table 2.3).

The fitted NB model produced negligible error with comparatively narrow 95W for three dgps under the highest level of overdispersion (RATIO<sub>NB</sub> = 3), specifically with  $\lambda \ge 1.5$  and p = 0 or 0.25 (Table 2.3). This result suggested that the NB model could estimate  $\theta$  more readily from data generated under the specification of large overdispersion and large  $\lambda$ .

The ZINB and HNB models severely overestimated  $\theta$ , except for data generated under large overdispersion (RATIO<sub>NB</sub> = 3), large Poisson mean ( $\lambda$  = 5) and moderate amount of zero-inflation (p = 0.25; Table 2.3). Another type of instability in estimation was indicated by a large mean  $\hat{\theta}$  combined with narrow 95W, as illustrated by fitting the ZINB and HNB models to data generated under large overdispersion (RATIO<sub>NB</sub> = 3), large Poisson mean ( $\lambda$  = 5), and high degree of zero-inflation (p = 0.75; Table 2.3). In this case, only a few  $\hat{\theta}$  out of 1000 estimates were exceptionally large and the rest were generally more reasonable. Under a larger sample size (SAMPLE = 250), negligible estimation error was obtained when estimating  $\theta$  from a dgp with RATIO = 3 and  $\lambda \ge 1.5$  (results not shown in Table 2.3).

## 2.4.5 Akaike Information Criteria AICc

Two general results were obtained from comparing fitted models by AICc: (1) multiple models provided almost identical fits to a single dgp, and (2) the POIS model provided the worst fit to datasets generated by most of the dgps (large positive differences with respect to reference models,  $\Delta AICc$ , in Figure 2.6). All six models fit equally well to data generated from *pois* dgp, with all average  $\Delta AICc < 5$  (Figure 2.6A). For data generated by the *nb* dgp, the NB, ZINB and HNB models consistently provided the best fit (Figure 2.6D and G); however, the other models performed adequately at medium overdispersion (RATIO<sub>NB</sub> = 1.5; Figure 2.6D) and poorly at large overdispersion (RATIO<sub>NB</sub> = 3; Figure 2.6G). The ZINB and HNB models provided the best fit to data generated under the assumption of zero-inflation (p > 0), regardless of the level of overdispersion in the dgp (Figure 2.6B,C,E,F,H,I). The performance of fitted ZIP and HPOIS models was very similar to that of the fitted ZINB and HNB models, with the exception of slightly poorer performance under moderate zero inflation (p = 0.25) and high overdispersion (RATIO<sub>NB</sub> = 3; Figure 2.6H).

The range between the 2.5 and 97.5 percentiles (95W) indicated that variability in  $\Delta$ AICc increased with increasing  $\lambda$ , RATIO<sub>NB</sub>, and *p* (Table 2.4). In the absence of zero inflation (*p* = 0), fitted ZINB and HNB models consistently had equal or

narrower 95W than the other three models (Table 2.4); further supporting the fact that the two models were flexible in fitting any dgp.

# 2.4.6 P-Value of $\chi^2$ Goodness of Fit Test

The fitted ZINB and HNB models could accurately reproduce count distributions summarizing the data from any dgp, as suggested by the average p-value > 0.78 and 2.5 percentile > 0.29 (Figure 2.7). As expected, the average p-value of POIS model was commonly below 0.05 when fitted to data from any dgp with large zero-inflation (p) and/or overdispersion (RATIO<sub>NB</sub>). Likewise, the range between the 2.5 and 97.5 percentiles (95W) of the p-value distributions was the widest for any dgp with  $\lambda = 0.3$  and p > 0 (Figure 2.7B,C,E,F,H,I). The fitted NB model performed poorly for data generated by the *zip* and *zinb* dgps with  $\lambda = 5$ . When fitted to the data from the *zip* dgp, this model yielded average p-values near 0.05 level (Figure 2.7B and C), and when fitted to the data from the *zinb* dgp, the 2.5 percentile for the p-values was likewise near or below 0.05 (Figure 2.7E,F,H,I). The average p-values for the fitted ZIP and HPOIS models were above the 0.05 level for all dgps, but it generally decreased with increasing RATIO<sub>NB</sub> (Figure 2.7). For these same fitted models, the 2.5 percentiles were generally below 0.05 under high overdispersion (RATIO<sub>NB</sub> = 3), large Poisson mean ( $\lambda \ge 1.5$ ) and low or absent zero inflation ( $p \le 0.25$ ; Figure 2.7G and H).

#### **2.5 DISCUSSION**

#### 2.5.1 Model Evaluation

In application, it would be close to impossible to discern the exact dgp that generated an empirical dataset. However, simulation allows assessment of how a model performs when fitted to data from its matching dgp and from alternative dgps. In the simulations, all six models generally recovered the known parameter values when fitted to data from their matching dgps, albeit with some exceptions. The most notable exception was the large sample size required for accurate and precise estimate of  $\theta$  by the NB, ZINB and HNB models. In addition, the ZINB and HNB models on average produced positively estimation error in  $\hat{\lambda}$  when the known Poisson parameter  $\lambda$  was small; i.e., 0.3. Lastly, the ZI models did not perform well in recovering the amount of zero-inflation *p* when  $\lambda$  was 0.3.

When the six models were fitted to data from alternative dgps in the simulation, the known parameter values were not recovered in most cases. The POIS model recovered  $\lambda$  if the underlying dgp was nb, but generally produced negative estimation error for other dgps. Similarly for the NB model,  $\lambda$  was recovered when the dgp was *pois* but not for others. Furthermore, the NB model recovered  $\theta$  for some of the *zinb* dgps, but the estimates generally had error or were unreasonable. The ZIP model produced positive error in  $\hat{\lambda}$  and  $\hat{p}$  for many dgps unless  $\lambda = 5$  or for a *pois* dgp with  $\lambda \ge 1.5$ . The ZINB model reasonably recovered  $\lambda$  and p, provided that an alternative dgp had  $\lambda \ge 1.5$ . Both HPOIS and HNB models behaved similarly to their

ZI counterparts in terms of  $\lambda$ , but consistently recovered  $\pi$  regardless of dgps. Finally, as mentioned in numerous instances, the ZINB and HNB models had difficulty estimating  $\theta$  accurately for the *nb* dgp.

Except for Warton (2005), no empirical studies appear available as a comparison to this simulation analysis, primarily because most of the published studies address conditional distributions of the six models rather than marginal distributions. However, most studies generally agree that the POIS model is inadequate for modeling the abundance of most species. Warton (2005) reported that not all taxa exhibited overdispersion, and that the AIC of his POIS model was higher than that for the NB model in only about 47% of the studied taxa. The current simulation showed that AIC could help infer if a set of counts originated from a *pois* dgp. When the POIS model fitted poorly to the data from alternative models, the data were probably not generated by a *pois* dgp; in this case, one could expect  $\lambda$  to be underestimated. Conversely, a comparable fit from a POIS model could either imply a *pois* dgp or *zip* dgp with small  $\lambda$ .

Most studies have compared the performances of the NB model to alternative ZMMs. This approach has been motivated by the fact that if a NB model can adequately represent data with apparently excessive zeros (marginally), or at least as well as ZMMs, the NB model may be more desirable from the perspective of parsimony. Empirical studies that include covariates often show mixed results in model evaluation but provide relatively little explanation (e.g., Gray 2005, Martin et al. 2005). The influence of covariates on evaluating alternative models likely

complicates identification of the effects from two alternative sources, assumptions about the statistical model versus effects of covariates on marginal distributions. In one exception, Warton (2005) found that the NB model was sufficient to account for the high frequency of zeros in marginal count data, and further claimed that most multivariate abundance data probably conformed to negative binomial distributions with small means. The current simulation showed that Warton's (2005) claims were plausible because the fit of the NB model was generally comparable to alternative ZMMs for any level of overdispersion and/or zero-inflation, as long as the Poisson mean  $\lambda$  was small to moderate in size. However, results from the current simulation also suggested that the claimed adequacy of *nb* dgp was difficult to substantiate because information criteria were unable to establish the underlying dgp of an empirical dataset with any certainty. Furthermore, if the unknown dgp is not truly *nb*, one could risk underestimating  $\lambda$  by fitting NB model to the data, even if this model provides the best fit.

Results from the current simulation study strongly suggested that information criteria should not be used to infer the underlying dgp of empirical data. For example, one noteworthy result of ZINB and HNB models was their ability to provide the best fit, as measured by AICc and  $\chi^2$  good of fit tests, for almost any dgp. This result highlights the flexibility of the ZINB model for representing data with overdispersion and/or zero-inflation (Rose et al. 2006). If the models are sufficiently flexible for modeling marginal count data, they can probably be expected to fit well to empirical data that include covariates and a range of conditional distributions. It is also

important to recognize that some of the difference in AIC may be attributable to differences in the relative magnitude of log-likelihoods of the various statistical models. Nonetheless, despite the close fit, estimated parameters had error and were highly unstable. One possible cause might be limitation to fitting the ZINB or HNB model to data with predominantly non-zero observations (Minami et al. 2007, Warton 2005). It also appeared that estimation of  $\theta$  required a larger sample size than may be practically feasible.

The main application of the simulation results is to model habitat associations of infrequently captured species without knowing how well any given dgp represents the underlying ecological processes. Empirical marginal counts of infrequent species are by definition dominated by zeros, indicating low marginal mean abundance. Occasional large counts might be observed under conditions that promote aggregation of individuals in the population, producing a count distribution that is generally unimodal and positively skewed with the possibility of long positive tail. Field data would typically be conditioned on a set of habitat variables, so the conditional distribution is often of more interest than marginal distribution. Minami et al. (2007) asserted that the ZIP model might be more appropriate for species that were rare and occurred in small group while ZINB model might be appropriate for species that occurred in large aggregations. The current simulation study found that caution had to be exercised when the true mean abundance was very small, because both models could produce highly unstable parameter estimates with large error describing the marginal distribution under their corresponding dgp, as well as under alternative but

precisely specified dgps. These results probably extend to conditional distributions that estimate low mean abundance at a given level of one or more covariates, but the effect of alternative covariates on estimating parameters of the statistical distribution, effects of covariates, and the interaction between the two is not clear. Perhaps only a carefully crafted simulation study that includes covariates can help address this issue.

## 2.5.2 Finite Mixture Models

As mentioned in an earlier section, ZI and H models are finite mixture models with different specifications of latent structures. However, there has been little comprehensive discussion about the subtle differences between the two models and their assumptions in the context of ecological modeling, particularly the H model. When modeling a species frequency distribution, we implicitly assume that the data conform to the latent structure of the chosen model. Any data editing influenced by this assumption will introduce some circularity into assessment of model suitability. For example, 'naughty naughts' (Austin and Meyers 1996) are observations obtained from outside the environmental range of a species. Martin et al. (2005) emphasized the need for thoughtful sampling design to avoid including these type of data or filtering datasets to exclude them (Elith and Burgman 2002). Ultimately, the final datasets that exclude observations outside the environmental range of the target species imply the imperfect state for all observed zeros; i.e., the species can occur at the sampling point but is absent for various unspecified reasons. As a result the data have an implicit latent structure consistent with a H model, and analyzing them with a ZI model would be inappropriate. In short, it is important to understand the data origin for careful matching of a statistical model to the latent structure.

Many distinguish the perfect and the imperfect states by the type of zero each represents. This distinction is intuitively reasonable but poses some potential problems in the form of conflicting or ambiguous definitions and terminologies. Most of literature on this subject consistently agrees that structural zeros arise from the perfect state. The types of zero associated with the imperfect state include accidental, stochastic and sampling zeros. Cunningham and Lindenmayer (2005) defined random or accidental zeros to arise where conditions are potentially suitable for the species but it was absent due to non-detection or unidentifiable factors. Sileshi (2008) defined sampling zeros as a result of observational errors. Added to this confusing list are true and false zeros as defined by Martin et al. (2005); i.e., true zeros originate from demographic processes, unsuitable habitat or failure to saturate suitable habitat by chance, whereas false zeros originate from observational errors or aspects of sampling design such as length and area and involve detection failure (Martin et al. 2005). It is unclear how these definitions relate to the dual-state of a ZI model.

When formulating the ZI model, Lambert (1992) focused on describing the dual-state rather than on the sources of zeros. From this perspective, the above confusion can be avoided by focusing on the ecological interpretation of the two states, rather than attempting to define the origin of zeros. The perfect state is defined as a set of conditions under which individuals of a species cannot exist. In short, the perfect state describes inhospitable habitats that are incapable of supporting the

population, whereas the imperfect state is at least hospitable. The modeling objective is then to determine the habitat attributes associated with both states. An example of conditions associated with the perfect state is the matrix of hostile habitats defined in island biogeography theory (MacArthur and Wilson 1967). Conversely, Hutchinson's (1985) definition of fundamental niche could be viewed as the imperfect state. One could further argue that the imperfect state by definition could include ecological trap and sink habitats because they could support some individuals at least temporarily, regardless of how maladapted the individuals might be (Dias 1996, Schlaepfer et al. 2002). In summary, this approach seems better suited for understanding the latent structures of ZI and H models.

From a finite mixture model perspective, there is no restriction on the level of mixing. For example, Gao and Khoshgoftaar (2007) chose to model software fault with H models at hurdle levels of one and two faults. Therefore, the above approach has another advantage, specifically that definition of the two states could change according to the practical issue being addressed or the assumed properties of the statistical model. For example, the mixing of a ZI model could involve one state with either zero or one counts and the other state with counts exceeding one. The former could be coined as the "near-perfect" state and the latter could remain as the usual imperfect state. Lambert (1992) subtly recognized the restrictive nature of a perfect state in which defects were extremely rare but not impossible. The near-perfect state could represent habitat conditions that support only a minute population of a species; e.g., at the boundaries of environmental range or a severe sink habitat. One could

even declare several states in a model such as the N-mixture model (Royle 2004, Royle et al. 2005) according to ecological theories of a species abundance and distribution.

## 2.5.3 Rarity, Spatial Scale and Time

Cunningham and Lindenmayer (2005), modifying concepts from Cody (1986), Rabinowitz et al. (1986) and New (2000), provided an excellent general classification of species rarity based upon three factors: abundance of a species within a community (common or rare), habitat specificity (general or specialized) and geographic range (widespread or restricted). In ecological data, spatial scale influences analysis and interpretation and consequentially the choice of an appropriate model. Thus, the most appropriate model for habitat association of the categorized rare species will depend on the spatial scale of the sampling design in a study. The two primary elements of spatial scale are grain and extent (Bowman et al. 2001, Wiens 1989). Grain is defined as minimum space between sampling points and, hence, the smallest possible resolution for detecting spatial patterns. Extent is total area covered by a sampling grid.

The abundance of a species within its community will likely influence the choice between a Poisson or a negative binomial distribution, depending on its detectability. There is a higher chance of observing large counts for a species that is common and easily detectable within its community. As suggested by Minami et al. (2007), a negative binomial distribution would be more appropriate when the species

is common and aggregated within its community, whereas a Poisson distribution would be more appropriate when the species is less common and random in spatial distribution.

When the extent of a study is smaller than the known geographical range of a species, one could assume that all observations originate from the imperfect state. In this case, a H model would conform to the collected data. On the contrary, if the study approximately covers the geographical range of the species, or it is two to five times larger than the suitable habitat patches in the landscape (Mayer and Cameron 2003), the possibility of including observations where the species cannot occur becomes substantial. Therefore, a ZI model seems appropriate in this situation. Ultimately, neither a widespread or restricted geographical range of a species predetermines whether a H or ZI model is appropriate; rather, it largely depends on the extent of the study relative to the spatial distribution of suitable habitat.

Mayer and Cameron (2003) suggested that the sampling scale should be two to five times smaller than the inherent grain of habitat characteristics for the targeted species, even though this target could be difficult to achieve. However, the relationship between habitat specificity, grain and model choice is not straightforward. For a generalist species, we suggest that grain size inherent to the area of study plays a minor role as the species could occupy a range of habitats; hence the data are more likely to originate from the imperfect state, theoretically favoring a H model. For a specialist species, the relationship between the three would depends on the spatial distribution of suitable habitats. Without further knowledge, a conservative assumption is that the data are a mix of perfect and imperfect states regardless of grain size, theoretically favoring a ZI model. A small grain size and high frequency of suitable habitat would likely result in majority of the samples in the imperfect state, but a few of them could also fall into unsuitable habitat.

The above discussion implies a static view of the relationship between model choice and spatial scale of both sampling and habitat distribution. It is also possible that time interacts with model choice and spatial scale, particularly if population size fluctuates or animal activity varies by season. For a simple example, the current state of a forested area in an early successional stage (e.g., regenerated forest stand) might not support a certain group of species that is closely related to old-growth forest conditions, so the area would be in the perfect state. Over time as the vegetation structure develops, the area would probably include unsuitable and suitable habitat for the target species, so the model choice at this stage could be a ZI model. As the forests continue to mature, more habitats become suitable for the species group, so a H model could then describe the later stage of succession. Thus, the timing of sampling for the target species during forest succession would be one factor when considering model choice.

The current simulation study advised against using goodness of fit statistics to infer underlying processes that generated an observed set of data. A consequence of misidentifying the unknown dgp is misinterpreting the role that a habitat predictor has on a species community. For example, the ZINB and HNB models are the most flexible based on the simulation results, but the underlying dgp may truly be *pois*.

Consequently, a predictor may be misjudged to significantly cause the habitat conditions to be either inhospitable or hospitable (zero-inflation component), when it should only have affected the abundance of a species under this true *pois* dgp. Therefore, it is recommended instead that the choice of the most appropriate statistical model be made to the fullest extent possible in the context of known ecological processes and the spatial and temporal scale of the study. Results from the simulation study also suggest the need for exercising caution when modeling the response of infrequent species to habitat variables using ZMMs; i.e., the ZMMs produced unstable parameter estimates when true abundance was small. Because parameters of distributions describing detections are typically conditioned on habitat predictors in most research and applications, it is likely that the estimated effects of the habitat variables would be adversely affected and incorrect inferences drawn.

#### **2.6 CONCLUSIONS**

This simulation study focused on modeling marginal distributions free from the influence of covariates. However, in most ecological applications, identifying the conditional distribution of species abundance on habitat conditions is the main goal (Warton 2005). The extent to which these simulation results extend to selecting models for identifying covariates is unclear. For example, it is uncertain how estimation error in  $\hat{\lambda}$  from ZINB model progresses from conditional distributions with low mean abundance to those with high abundance across levels of a covariate. Further simulation studies may be necessary to explore this issue, but the current simulation study has provided insights into behavior of alternative models under several factors defining a data generating process.

The ability of goodness of fit criteria to discern an underlying dgp was evaluated, and in most cases the dgp could not be inferred accurately. However, the use of goodness of fit criteria would depend on a study's objective. If the objective is purely prediction, then a model that maximizes statistical fit may be desirable. In the simulation study, the ZINB or HNB models are most robust under this objective. Often, however, the objective involves a mix of prediction and identification of underlying process that generate the data (Lord et al. 2005). If parts of the dgp and associated ecological processes are known, the "best" fit model may not be the best simply because it is sufficiently flexible for handling a wide range in frequency distributions because it may not necessarily reflect the underlying dgp (Minami et al. 2007). In this context, maximizing statistical fit is largely trivial (Lord et al. 2007). When modeling empirical data of unknown dgp, Lord et al. (2007) advised that choice of the model be based on parsimony and agreement with theoretical expectation.

When applying a model, one should understand its latent structure and assumptions. Although excess zeros were the motivation behind ZI and H models, Lambert (1992) and Mullahy (1986) were concerned with defining a dual-state process or a hurdle when developing these respective models and did not attempt to define the sources of zeros. Beyond issues already discussed above, Lord et al. (2007) voiced several questions about the underlying logic of a dual-state process. One concerns the boundary conditions delimiting the two states when modeling with ZI model. The

same covariate is usually used to model two states simultaneously; so a question arises about the level of the covariate that demarcates the boundary. Recognizing complications in interpreting the results of a ZI model, Welsh et al. (1996) advocated the H model for its simplicity in understanding the association between covariates and abundance. However, this argument may not be the best justification for avoiding the ZI model.

If one is uncomfortable with the latent structure and assumptions of either ZI or H model, other statistical tools suggested by Lord et al. (2005) could model the preponderance of zeros found in the count data for many species. One such tool is small area statistics (SAS) or small area estimator (SAE) that is often used in survey science, where small sample sizes must be dealt with frequently (Lord et al. 2005, Rao 2003). Another is the extreme value model (Lord et al. 2007, Cole 2001). These models require further theoretical development and assessment in applications, but could provide alternatives to modeling rare species distribution and gaining insights into their effective management.

Table 2.1 The 27 dgps (data generating processes) as defined by amount of zeroinflation *p*, mean of a Poisson distribution  $\lambda$  and amount of overdispersion in negative binomial distribution RATIO<sub>NB</sub>. Associated with each dgp were known overdispersion parameter  $\theta$ , probability of crossing the hurdle  $\pi$ , the data generating distribution, and implied total overdispersion. Each dgp generated a sample size of 25, 50, 75, 100, and 250 observations. For RATIO<sub>NB</sub> = 1,  $\theta$  was undefined ( $\theta = \infty$ ). Total overdispersion was defined as Var(Y)/E(Y) of the corresponding data generating distribution.

	р	λ	$\theta$	_	Data Generating	Total
KATIO <sub>NB</sub>				π	Distribution	Overdispersion
1.0	0	0.3	$\infty$	0.26	pois	1.00
1.0	0	1.5	$\infty$	0.77	pois	1.00
1.0	0	5	$\infty$	0.99	pois	1.00
1.0	0.25	0.3	$\infty$	0.19	zip/hpois	1.08
1.0	0.25	1.5	$\infty$	0.58	zip/hpois	1.38
1.0	0.25	5	$\infty$	0.74	zip/hpois	2.25
1.0	0.75	0.3	$\infty$	0.06	zip/hpois	1.23
1.0	0.75	1.5	$\infty$	0.19	zip/hpois	2.13
1.0	0.75	5	$\infty$	0.25	zip/hpois	4.75
1.5	0	0.3	0.6	0.22	nb	1.50
1.5	0	1.5	3	0.70	nb	1.50
1.5	0	5	10	0.98	nb	1.50
1.5	0.25	0.3	0.6	0.16	zinb/hnb	1.58
1.5	0.25	1.5	3	0.53	zinb/hnb	1.88
1.5	0.25	5	10	0.74	zinb/hnb	2.75
1.5	0.75	0.3	0.6	0.05	zinb/hnb	1.73
1.5	0.75	1.5	3	0.18	zinb/hnb	2.63
1.5	0.75	5	10	0.25	zinb/hnb	5.25
3.0	0	0.3	0.15	0.15	nb	3.00
3.0	0	1.5	0.75	0.56	nb	3.00
3.0	0	5	2.5	0.94	nb	3.00
3.0	0.25	0.3	0.15	0.11	zinb/hnb	3.08
3.0	0.25	1.5	0.75	0.42	zinb/hnb	3.38
3.0	0.25	5	2.5	0.70	zinb/hnb	4.25
3.0	0.75	0.3	0.15	0.04	zinb/hnb	3.23
3.0	0.75	1.5	0.75	0.14	zinb/hnb	4.13
3.0	0.75	5	2.5	0.23	zinb/hnb	6.75

Table 2.2 The range between the 2.5 and 97.5 percentiles (95W) of %ERROR in the estimated parameter  $\hat{\lambda}$  for the six fitted models and all 27 combinations of Poisson mean parameter ( $\lambda$ ), amount of zero-inflation (p), and amount of overdispersion (RATIO<sub>NB</sub>).

<b>RATIO</b> <sub>NB</sub>	р	λ	POIS	NB	ZIP	ZINB	HPOIS	HNB
1.0	0	0.3	70	70	130	119	180	189
1.0	0	1.5	31	31	37	36	45	48
1.0	0	5	17	17	17	16	17	17
1.0	0.25	0.3	67	67	165	163	215	213
1.0	0.25	1.5	33	33	48	54	48	58
1.0	0.25	5	23	23	21	21	21	21
1.0	0.75	0.3	37	37	287	285	294	291
1.0	0.75	1.5	23	23	88	114	88	116
1.0	0.75	5	19	19	37	38	37	38
1.5	0	0.3	83	83	349	285	361	355
1.5	0	1.5	39	39	52	47	52	73
1.5	0	5	22	22	22	23	22	22
1.5	0.25	0.3	80	80	383	338	389	388
1.5	0.25	1.5	38	38	63	75	63	87
1.5	0.25	5	24	24	24	25	24	25
1.5	0.75	0.3	43	43	613	525	620	531
1.5	0.75	1.5	25	25	115	145	115	165
1.5	0.75	5	21	21	44	45	44	45
3.0	0	0.3	120	120	750	608	750	665
3.0	0	1.5	54	54	96	82	96	158
3.0	0	5	31	31	32	32	32	34
3.0	0.25	0.3	107	107	859	727	859	770
3.0	0.25	1.5	49	49	102	117	102	174
3.0	0.25	5	31	31	35	37	35	37
3.0	0.75	0.3	57	58	1334	1091	1337	1094
3.0	0.75	1.5	32	32	193	213	193	231
3.0	0.75	5	23	23	61	67	61	67

Table 2.3 The mean and 95W (range between 2.5 and 97.5 percentiles) of estimated overdispersion parameter ( $\hat{\theta}$ ) for each of the six fitted models and each dgp (one of 27 combinations of Poisson mean parameter ( $\lambda$ ), amount of zero-inflation (p), and amount of overdispersion (RATIO<sub>NB</sub>)). The estimates in bold were not significantly different from the known  $\theta$  values.

RATIO <sub>NB</sub>	р	λ	$\theta$ -	NB		ZINB		HNB	
				Mean	95W	Mean	95W	Mean	95W
1	0	0.3	_	$3 \times 10^{3}$	$8 \times 10^{4}$	$2 \times 10^{4}$	$1 \times 10^{5}$	$7 \times 10^{3}$	$3 \times 10^{4}$
1	0	1.5	_	$1 \times 10^{4}$	$3 \times 10^{4}$	$7 \times 10^{4}$	$3 \times 10^{5}$	$3 \times 10^{4}$	$1 \times 10^{5}$
1	0	5	_	$4 \times 10^{4}$	$1 \times 10^{5}$	$5 \times 10^{8}$	$1 \times 10^{9}$	$7 \times 10^{4}$	$3 \times 10^{5}$
1	0.25	0.3	_	$2 \times 10^{3}$	$6 \times 10^{3}$	$1 \times 10^{4}$	$5 \times 10^{4}$	$6 \times 10^{3}$	$3 \times 10^{4}$
1	0.25	1.5	_	$2 \times 10^{2}$	23.7	$3 \times 10^{4}$	$1 \times 10^{5}$	$3 \times 10^{4}$	$1 \times 10^{5}$
1	0.25	5	_	1.8	2.5	$8 \times 10^4$	$5 \times 10^{5}$	$1 \times 10^{9}$	$4 \times 10^{6}$
1	0.75	0.3	_	$5 \times 10^{2}$	$2 \times 10^{3}$	$7 \times 10^{3}$	$2 \times 10^{4}$	$6 \times 10^{3}$	$3 \times 10^{4}$
1	0.75	1.5	_	0.3	0.4	$2 \times 10^{4}$	$7 \times 10^{4}$	$4 \times 10^{4}$	$1 \times 10^{5}$
1	0.75	5	_	$3 \times 10^{4}$	$1 \times 10^{5}$	$7 \times 10^{4}$	$3 \times 10^{5}$	$9 \times 10^{4}$	$4 \times 10^{5}$
1.5	0	0.3	0.6	$1 \times 10^{2}$	$3 \times 10^{3}$	$5 \times 10^{3}$	$3 \times 10^{4}$	$7 \times 10^{3}$	$4 \times 10^4$
1.5	0	1.5	3	71.3	14.4	$3 \times 10^{3}$	$4 \times 10^4$	$3 \times 10^{3}$	$4 \times 10^4$
1.5	0	5	10	$4 \times 10^{2}$	45.2	$6 \times 10^{7}$	$7 \times 10^{7}$	$1 \times 10^{3}$	47.9
1.5	0.25	0.3	0.6	76.2	$1 \times 10^{3}$	$4 \times 10^{3}$	$3 \times 10^{4}$	$7 \times 10^{3}$	$4 \times 10^4$
1.5	0.25	1.5	3	1.3	1.9	$3 \times 10^{3}$	$4 \times 10^4$	$5 \times 10^{3}$	$5 \times 10^{4}$
1.5	0.25	5	10	1.4	1.5	$2 \times 10^{3}$	$6 \times 10^{2}$	$2 \times 10^{6}$	$1 \times 10^{5}$
1.5	0.75	0.3	0.6	$2 \times 10^{2}$	$2 \times 10^{3}$	$6 \times 10^{3}$	$2 \times 10^{4}$	$1 \times 10^{4}$	$5 \times 10^{4}$
1.5	0.75	1.5	3	0.2	0.3	$9 \times 10^{3}$	$6 \times 10^{4}$	$2 \times 10^{4}$	$1 \times 10^{5}$
1.5	0.75	5	10	$1 \times 10^{3}$	$8 \times 10^4$	$1 \times 10^{4}$	$2 \times 10^{5}$	$2 \times 10^{5}$	$2 \times 10^{5}$
3	0	0.3	0.15	3.3	0.4	$2 \times 10^{3}$	$2 \times 10^{4}$	$4 \times 10^{3}$	$4 \times 10^4$
3	0	1.5	0.75	0.8	0.9	90.5	3.6	$1 \times 10^{7}$	4.2
3	0	5	2.5	2.7	2.6	$7 \times 10^{9}$	3.1	2.8	3.4
3	0.25	0.3	0.15	8.5	0.5	$3 \times 10^{3}$	$2 \times 10^{4}$	$6 \times 10^{3}$	$5 \times 10^{4}$
3	0.25	1.5	0.75	0.5	0.5	52.3	6.6	$1 \times 10^{2}$	6.9
3	0.25	5	2.5	0.8	0.7	2.9	4.7	2.9	4.7
3	0.75	0.3	0.15	$2 \times 10^{2}$	$1 \times 10^{3}$	$7 \times 10^{3}$	$4 \times 10^4$	$2 \times 10^{4}$	$1 \times 10^{5}$
3	0.75	1.5	0.75	16.9	0.2	$2 \times 10^{3}$	$3 \times 10^{4}$	$5 \times 10^{3}$	$6 \times 10^{4}$
3	0.75	5	2.5	$4 \times 10^{4}$	$4 \times 10^{4}$	$4 \times 10^{2}$	21.3	$3 \times 10^{5}$	24.1

Table 2.4 The range between the 2.5 and 97.5 percentiles (95W) of the difference in AICc ( $\Delta$ AICc) between any given model and the reference model for all 27 combinations of Poisson mean parameter ( $\lambda$ ), amount of zero-inflation (p), and amount of overdispersion (RATIO<sub>NB</sub>). The 95W of  $\Delta$ AICc for a reference model was zero by definition.

RATIO <sub>NB</sub>	р	λ	POIS	NB	ZIP	ZINB	HPOIS	HNB
1.0	0	0.3	0	3	3	4	5	6
1.0	0	1.5	0	4	4	4	5	7
1.0	0	5	0	3	5	5	5	6
1.0	0.25	0.3	6	2	0	1	0	4
1.0	0.25	1.5	21	12	0	3	0	3
1.0	0.25	5	94	42	0	4	0	4
1.0	0.75	0.3	10	2	0	1	0	1
1.0	0.75	1.5	49	13	0	2	0	2
1.0	0.75	5	128	286	0	3	0	3
1.5	0	0.3	25	0	8	3	8	3
1.5	0	1.5	27	0	18	4	18	5
1.5	0	5	26	0	25	25	25	4
1.5	0.25	0.3	28	3	5	0	5	0
1.5	0.25	1.5	44	8	13	0	13	0
1.5	0.25	5	99	36	21	0	21	0
1.5	0.75	0.3	26	2	3	0	3	0
1.5	0.75	1.5	71	9	8	0	8	0
1.5	0.75	5	153	303	12	0	12	0
3.0	0	0.3	93	0	26	4	26	4
3.0	0	1.5	105	0	58	4	58	5
3.0	0	5	106	0	90	4	90	5
3.0	0.25	0.3	95	4	19	0	19	0
3.0	0.25	1.5	111	7	44	0	44	0
3.0	0.25	5	156	25	79	0	79	0
3.0	0.75	0.3	74	4	8	0	8	0
3.0	0.75	1.5	135	8	25	0	25	0
3.0	0.75	5	217	329	45	0	45	0



Figure 2.1 Relative frequency distributions implied by the 27 dgps. Parameters of the dgps were specified as a  $3^3$  factorial of: Poisson mean  $\lambda$  (0.3, 1.5, 5); amount of zero-inflation *p* (0, 0.25, 0.75); and amount of overdispersion RATIO<sub>NB</sub> (1, 1.5, 3). Each dgp corresponded to one of the following data generating distributions (Table 2.1): *pois*, *nb*, *zip*, *zinb*, *hpois* or *hnb*.



Figure 2.2 Percent estimation error (%ERROR) in  $\hat{\lambda}$  for the six models fitted to data generated by 27 dgps. Parameters of the dgps were specified as a 3<sup>3</sup> factorial of: Poisson mean  $\lambda$  (0.3, 1.5, 5); amount of zero-inflation p (0, 0.25, 0.75); and amount of overdispersion RATIO<sub>NB</sub> (1, 1.5, 3). The corresponding data generating distribution for each panel is shown in brackets, and the six fitted models included POIS ( $\circ$ ), NB ( $\Delta$ ), ZIP (+), ZINB (×), HPOIS ( $\Box$ ), HNB ( $\nabla$ ). The solid and dashed gray horizontal lines depict 0% and  $\pm$  10% ERROR.



Figure 2.3 Estimated amount of zero-inflation  $\hat{p}$  from data generated by 36 dgps. Parameters of the dgps were specified as a 3<sup>3</sup> factorial of: Poisson mean  $\lambda$  (0.3, 1.5, 5), amount of zero-inflation p (0, 0.25, 0.5, 0.75), and amount of overdispersion RATIO<sub>NB</sub> (1, 1.5, 3). The two fitted models were ZIP (+) and ZINB (×). The dashed and solid bars around the mean value indicate the 2.5 and 97.5 percentiles from separate fits of ZIP and ZINB, respectively, to 1000 sets of simulated data. Degree of deviation from the 1:1 line in gray indicates ERROR.



Figure 2.4 Results from 1000 simulations by two data generating processes, one assuming *zip* with  $\lambda = 0.3$ , p = 0.25 and RATIO<sub>NB</sub> = 1 ( $\circ$ ) and the other assuming *zip* with  $\lambda = 0.3$ , p = 0.75 and RATIO<sub>NB</sub> = 1 ( $\times$ ): (a) trend in estimated amount of zero-inflation parameter  $\hat{p}$  over estimated Poisson parameter  $\hat{\lambda}$  fitted to the ZIP models; and (b) estimated mean and 2.5 and 97.5 percentiles of marginal probability of a zero observation (Pr(Y = 0)) under each ZIP model (estimated and actual values of Pr(Y = 0) are virtually identical).



Figure 2.5 Percent estimation error (%ERROR) in  $\hat{\pi}$  from the six models fitted to data generated by 27 dgps. Parameters of the dgps were specified as a 3<sup>3</sup> factorial of: Poisson mean  $\lambda$  (0.3, 1.5, 5), amount of zero-inflation p (0, 0.25, 0.75), and amount of overdispersion RATIO<sub>NB</sub> (1, 1.5, 3). The corresponding data generating distribution for each panel is shown in brackets. The six fitted models included POIS ( $\circ$ ), NB ( $\Delta$ ), ZIP (+), ZINB (×), HPOIS ( $\Box$ ), HNB ( $\nabla$ ). The solid and dashed gray horizontal lines depict 0% and  $\pm$  10% error.



Figure 2.6 Difference in AICc ( $\Delta$ AICc) between fitted models and the reference model fitted to data generated by 27 dgps. Parameters of the dgps were specified as a 3<sup>3</sup> factorial of: Poisson mean  $\lambda$  (0.3, 1.5, 5); amount of zero-inflation p (0, 0.25, 0.75), and amount of overdispersion RATIO<sub>NB</sub> (1, 1.5, 3). The corresponding data generating distribution for each panel is shown in brackets. The six fitted models included POIS ( $\circ$ ), NB ( $\Delta$ ), ZIP (+), ZINB (×), HPOIS ( $\Box$ ), HNB ( $\nabla$ ). The solid and dashed gray horizontal lines depict 0 and ±5 unit difference in  $\Delta$ AICc.



Figure 2.7 The mean and 2.5 and 97.5 percentiles of p-values from  $\chi^2$  goodness of fit tests for each of six models fitted to data from27 dgps. Parameters of the dgps were specified as a 3<sup>3</sup> factorial of: Poisson mean  $\lambda$  (0.3, 1.5, 5); amount of zero-inflation p (0, 0.25, 0.75); and amount of overdispersion RATIO<sub>NB</sub> (1, 1.5, 3). The corresponding data generating distribution for each panel is shown in brackets. The six fitted models included POIS ( $\circ$ , black), NB ( $\Delta$ , gray), ZIP (+, red), ZINB (×, orange), HPOIS ( $\Box$ , blue), HNB ( $\nabla$ , green). The corresponding colored solid lines depict 2.5 and 97.2 percentiles of the p-value for each of the six models. The dashed gray horizontal line depicts p-value = 0.05.

## 3 CHAPTER 3: DETECTING SMALL MAMMAL RESPONSES TO VARIABLE-RETENTION HARVESTING WITH STATISTICAL MODELS THAT ACCOMMODATE EXCESSIVE ZEROS AND HIERARCHICAL SAMPLING

Tzeng Yih Lam

Douglas A. Maguire

#### **3.1 INTRODUCTION**

Silvicultural systems such as shelterwood with reserves and group selection (Matthews 1989) have been considered viable options for managing Douglas-fir (Pseudotsuga menziesii (Mirb.) Franco) forests on public lands in the Pacific Northwest, USA, particularly where management objectives call for a combination of timber production and other ecosystem services. The resulting complex stand structures are viewed as more amenable to achieving other management objectives such as conservation of biodiversity and maintenance of aesthetic quality. Variableretention systems, currently mandated on federal forestland, have a specific goal of providing refugia to late-seral or closed-canopy organisms ("lifeboating") and promoting more rapid recolonization of regenerated stands by these late-seral species (Franklin et al. 1997). To achieve this goal, two basic elements of residual forest structure – level and pattern of retention – must be manipulated in an appropriate manner (Aubry et al. 2009). However to establish successful refugia, knowledge about the habitat elements that targeted taxa perceive and the spatial scale at which they respond are essential.

One commonly targeted taxonomic group in the Pacific Northwest forests is the community of forest floor small mammals. These small rodents play an important role as predators, consumers, and prey (Martin and McComb 2002, Sullivan and Sullivan 2001), so contribute to both species and functional diversity of the forests in this region (Carey and Johnson 1995). Individual species are variably associated with
coarse woody debris, litter layers, herbaceous vegetation, shrubs and fungal fruiting bodies (Carey and Johnson 1995, Ure and Maser 1982). These habitat variables indicate quantity and quality of cover and food provided by vegetation comprising a forest stand. Waldien et al. (2006) demonstrated that Townsend's chipmunk (*Tamias townsendii*) selectively chose microhabitats with downed wood as it moved through a stand. This type of knowledge allows silviculturists to design silvicultural treatments that produce or protect specific habitat structures during regeneration harvests. Rigorous modeling of habitat associations becomes a key tool of forest management to successfully balance timber harvesting with maintenance of forest biodiversity.

Studies on the effects of silvicultural treatments on forest floor small mammal communities share three common characteristics: (1) only a few species dominate the total number of captures; (2) a majority of the species are captured infrequently, and (3) sampling of experimental units is typically systematic and often involves a hierarchical structure. The first two characteristics are quite natural (Cunningham and Lindenmayer 2005) and are caused by differences in both population abundance and detection probability. The population sizes of species contributing to the small mammal community inherently cover a wide range. The issue of detection probability is ubiquitous but often under-appreciated (Tyre et al. 2003). When detection probability is variable and unknown, species abundance or species presence/absence estimated by number of captures may be inaccurate and biased (Wenger and Freeman 2008). This would in turn affect inference on treatment effects because a detected change in capture may indicate a shift in detection probability, a shift in abundance, or

a combination of these and other factors. However, live trapping and capturerecapture techniques are required to estimate detection probability with a minimal level of confidence (MacKenzie et al. 2002); therefore, estimation of detection probabilities become problematic for studies like DEMO that use pitfall traps. Hence, it must often be assumed that changes in detection probability are negligible. Anderson (2001) summarized three factors affecting detection probability for a given species; i.e., observer bias, environmental conditions and species features that include abundance and behavior. The third characteristic is unavoidable especially in field experiments that strive for operational and ecological relevance (Maguire et al. 2007). These experiments generally have large experimental units with systematically arranged sampling units. The observations within an experimental unit are not independent because the sampling units are nested within the treatment unit, and experimental units may be spatially autocorrelated with neighboring units.

Unfortunately, a common approach to infrequently captured species is to exclude them from statistical analysis. Among 23 studies on small mammal responses to silvicultural treatments in the Pacific Northwest, a range of 2-20 species were captured but on average only 62% were analyzed (e.g. Klenner and Sullivan 2003, Suzuki and Hayes 2003). Some species were ignored because they were non-target species, or because they were transient species not normally associated with the sampled habitat, particularly when their home ranges were large compared to the size of experimental or sampling units (Carey and Johnson 1995). Another biological rationale to exclude some infrequent species is that the trapping technique may not be suitable for an unbiased estimate of abundance, although captures may still be a valuable index of relative abundance or relative activity in the context of a controlled experiment. From a statistical perspective, small counts are often precluded because they do not meet the normality assumption of conventional Analysis of Variance (ANOVA) or regression models. In one approach to address this issue, MacCracken (2005) grouped infrequently captured species based on similarity in life history to improve numerical stability in a Multivariate Analysis of Covariance (MANCOVA). Nonetheless, these rare species are particularly deserving of conservation efforts because they are most prone to extinction (Meffe and Carroll 1997), and capture frequency may still serve as a useful index of treatment response despite its unsuitability for estimating population size for some species. Potential statistical solutions to modeling treatment effects and habitat associations for these infrequent species include zero-inflated and hurdle models.

Lambert (1992) and Mullahy (1986) developed zero-inflated and hurdle models, respectively, for data characterized by excessive zero counts, relative to conventional Poisson and negative binomial models. Both models specifically account for overdispersion that is caused by greater-than-Poisson zero counts, and can account for overdispersion attributable to unexplained heterogeneity in data (Zorn 1998). Although similar in application, the assumption about the underlying data generating process (dgp) is different. Zero-inflated models assume a dual-state dgp: an observation is either in the perfect state, in which case its value must be zero, or in the imperfect state, in which case it can take any value including zero (Lambert 1992). Hurdle models assume a hurdle latent structure; i.e., the hurdle is set at zero, and an observation can only take greater-than-zero value conditional on the hurdle being crossed. Gray (2005) and Warton (2005) compared and evaluated these models in an ecological context.

To avoid autocorrelation between sampling units within experimental units, count data from individual sampling units are commonly summarized as a mean or total for the entire treatment unit (e.g. Steventon et al. 1998). Transformation of the summary statistics is sometimes necessary to meet assumptions of the statistical model such as normality of residuals (e.g. Suzuki and Hayes 2003). These steps are necessary for ANOVA or ANCOVA models that dominate most analyses of small mammal data. These models are appropriate when the primary objective is to test treatment effects on stand-level mean or total abundance. However, these tests at the level of experimental units address one spatial scale and might not be the most efficient use of information contained in individual sampling units. For example, microclimatic conditions likely influence the local distribution of southern red-backed vole (Clethrionomys gapperi) (Miller and Getz 1973). With data aggregated to treatment level, resolution is lost for detecting microhabitat attributes that drive responses to treatments. This type of data aggregation might contribute to contradictory results between studies with similar treatment categories (see Klenner and Sullivan 2003). Also, transformation of count data for application of analyses based on a normal distribution complicates interpretation of results and is ineffective for infrequent or rare species with many zero observations. Generalized Linear

Models (GLM; McCullagh and Nelder 1989) can accommodate count data and avert the interpretive problems associated with transformation. They can also be extended to multilevel models (Gelman and Hill 2007) when part of the objective includes detection of finer scale responses, because they avoid aggregation of data from individual sampling units.

Generalized Linear Mixed Models (GLMM) are GLMs in which random effects are included in addition to fixed effects of other covariates (Venables and Ripley 2002). In multilevel models, the random effects can help model the nested structure of subsamples, obviating the need for data aggregation. These random effects are also often effective for modeling the spatial autocorrelation between subsamples (Schabenberger and Gotway 2005). In the 23 small mammals studies mentioned above, the spacing for sampling grids commonly ranged from 10 to 30 m (e.g. Von Trebra et al. 1998). Subsamples among grid points at this spacing most likely are spatially autocorrelated, at least for many of the species with home ranges greater than 100-900 m2.

The goal of this study was to model the association between habitat features and the detection of forest floor small mammal species in the context of a large-scale silvicultural experiment. The experiment (Demonstration of Ecosystem Management Options or DEMO) was established to assess long term effects of variable-retention harvesting on plants, animals, ecological processes and public perceptions (Aubry et al. 1999, Halpern and Raphael 1999). One major set of response variables included the abundance and species composition of the forest floor small mammal community. The two specific objectives of this analysis were to make the following empirical comparisons in the process of modeling microhabitat associations of small mammal species: (1) relative performance of zero-inflated versus hurdle statistical models, and (2) relative performance of GLM versus GLMM. The analysis was carried out in two parts to meet the objectives. The first part of the analysis focused on fitting and interpreting various specifications of zero-inflated and hurdle models, collectively referred to as zero-modified models (ZMM). The second part of the analysis addressed specification and fitting of GLMMs.

### **3.2 MATERIALS**

#### **3.2.1 Experimental and Treatment Designs**

The DEMO study was designed as a large-scale operational experiment to test responses to variable-retention regeneration harvests in Douglas-fir forests of western Oregon and Washington, USA (Figure 3.1). The experiment was implemented under a randomized complete block design (RCBD) with subsampling. Two blocks were established in Oregon on the Umpqua National Forest (WF, DP) and four in Washington with three on Gifford Pinchot National Forest (PH, BU, LW) and one on the state-owned Capitol Forest (CF). McKenzie et al. (2000) and Halpern et al. (2005) described the physical environment, disturbance history, age, and understory dynamics of the study sites. In each block, six 13-ha experimental units (hereafter referred to as units) were established and treatments were randomly assigned to these units. The units were either square (360 m × 360 m) or slightly rectangular (320 m × 400 m). The six harvest treatments were defined by level (percentage of basal area) and/or spatial pattern (dispersed vs. aggregated) of retained trees as followed (Figure 3.1): (1) 100%: 100% retention (control); (2) 75%: 75% aggregated retention (three circular, 1ha patch cuts); (3) 40%D: 40% dispersed retention; (4) 40%A: 40% aggregated retention (five circular 1-ha residual aggregates); (5) 15%D: 15% dispersed retention; (6) 15%A: 15% aggregated retention (two circular 1-ha residual aggregates).

Harvesting systems varied by block depending on access and slope: skyline cable system for CF; ground-based system for WF and PH; and helicopter system for DP, BU and LW (Halpern and McKenzie 2001). Harvesting of all units within each block was completed within 3 to 7 months, and all blocks were logged between 1997 and 1998. Aubry et al. (1999) and Maguire et al. (2007) provided additional details on implementation of silvicultural treatments and harvesting prescriptions.

# **3.2.2 Small Mammal and Vegetation Sampling**

A permanent  $8 \times 8$  or  $7 \times 9$  sampling grid with 40 m spacing was installed in each 13-ha unit for vegetation and forest floor small mammal sampling, covering approximately 8 ha and yielding 63 or 64 sample points. The grid was buffered by 40 m between the outer grid points and the edge of the unit. Small mammals were trapped using a pitfall trap installed at each sample point and placed close to logs or other structures when possible (Corn and Bury 1990, Gitzen et al. 2007). Each trap was made of two No. 10 cans taped together to form a cylinder with diameter of 16 cm and depth of 35 cm, and was operated as death trap partially filled with water (Gitzen et al. 2007). The traps were opened continuously for about 28 days between late September and early November, and captured animals were collected and identified weekly (Gitzen et al. 2007). The response variable was the number of individuals of a species captured at each sample point.

Overstory and understory vegetation were studied concurrently with small mammal trapping, but were carried out only on a subset of sample points; 32-37 sample points depending on the treatment. Detailed sampling protocols are available from Halpern and McKenzie (2001) and Halpern et al. (2005). For understory vegetation, a cluster of 24 microplots ( $0.2 \times 0.5$  m) was used to observe percent cover of herbaceous and small shrub species (typically <1 m tall at maturity) at each sample point. In addition, percent cover of tall shrub species (typically >1 m tall at maturity) and understory coniferous trees (<5.0 cm dbh) was estimated on four 6-m line intercepts radiating out from the sampled grid point. Diameter of any down wood (stems  $\geq 10$  cm) was measured at the point of intersection with this same transect. Overstory trees were sampled with a set of nested circular plots: a 0.01-ha plot for trees with dbh  $\geq 5$  and <15 cm and a 0.04-ha plot for trees with dbh  $\geq 15$  cm (Maguire et al. 2007).

Mean plot attributes were computed for either the average of the 24 microplots (HERB) or the average of the four transects (CONIF, SHRUB, CWDVOL). Coarse woody debris volume per hectare for a transect was estimated by conventional techniques for line intersect sampling (de Vries 1986). A variable was also computed to describe the vertical complexity of overstory tree structure. At each sample point,

crown cross-sectional areas at 0.5-m height intervals were estimated for each tree (Dubrasich et al. 1997, Maguire et al. 2007). A crown-area profile was then generated by summing crown cross-sectional areas at each height interval. The Shannon diversity index (SHANNON, Staudhammer and LeMay 2001) was computed to represent the relative variability in total crown area among the 0.5-m height intervals. Higher SHANNON indices indicated greater vertical diversity or complexity, representing a more even distribution of total crown area among the 0.5-m height intervals. Lastly, stand density in terms of overstory trees per hectare (TPH, dbh  $\geq 5$ cm) at each sample point was estimated. The TPH was further divided by 100 (TPH\*) to facilitate model convergence. In summary, the six vegetation attributes for each sample point were: (1) mean percent herbaceous cover (HERB, %), (2) mean percent understory conifer cover (CONIF, %), (3) mean percent tall shrub cover (SHRUB, %), (4) mean coarse woody debris volume per hectare (CWDVOL,  $m^3/ha$ ), (5) overstory crown structural heterogeneity (SHANNON), and (6) stand density (TPH\*, trees/100ha).

Multiple pre- and post-treatment samplings were carried out, but this analysis was only based on data from the post-treatment sampling year 1999. For the purpose of habitat association modeling, sample points without vegetation information were discarded; hence only 32–37 sample points in each unit were used for analysis. Of the total 32 small mammal species captured in the DEMO study (Gitzen et al. 2007), four species were chosen for analysis, including two abundant and two infrequent species. The abundant species were the southern red-backed vole (*Clethrionomys gapperi*;

CLGA) and deer mouse (*Peromyscus maniculatus*; PEMA) with total captures of 378 and 322 respectively, and the infrequent species were the shrew-mole (*Neurotrichus gibsii*; NEGI) and Keen's deer mouse (*Peromyscus keeni*; PEKE) with total captures of 91 and 90 respectively. All species had at least five counts at some sample points, but the infrequent species had a higher proportion of traps on which 1-5 individuals were captured (Figure 3.2). The geographical range of CLGA and PEKE was restricted to Washington; therefore, their analyses were only based on the four blocks (BU, LW, PH and CF) located in that state. In short, 1181 sample points from six blocks were available for PEMA and NEGI, whereas 787 sample points from four blocks were available for CLGA and PEKE.

### **3.3 METHODS**

### 3.3.1 Zero-modified Models (ZMMs)

The six models considered in the first part of this study were: Poisson GLM (POIS), negative binomial GLM (NB), zero-inflated Poisson GLM (ZIP), zero-inflated negative binomial GLM (ZINB), hurdle Poisson GLM (HPOIS) and hurdle negative binomial GLM (HNB). The ZIP and HPOIS models were loosely grouped as the Poisson variation of zero modified models (ZMMs) and the ZINB and HNB models as the negative binomial variation of ZMMs. A brief description of model specifications are presented below, but Cameron and Trivedi (1998), Greene (1994) and Winkelmann (2008) provided excellent reviews of statistical properties, specifications and extensions. Throughout the current analysis, *Y* was defined as the total species

count of individuals captured at a sample point (hereafter referred to as trap),  $\lambda$  was the mean of the Poisson distribution, and  $\theta$  was the dispersion parameter of the negative binomial distribution.

The POIS model was the benchmark model because it was widely used for modeling count data. It was specified by the following Poisson probability mass function (pmf):

$$\Pr(Y = y) = \frac{e^{-\lambda} \lambda^{y}}{y!}$$
(3.1)

Under the POIS model, the variance and mean of the distribution were equal. The NB model was an extension of the POIS model accounting for overdispersion in count data, i.e., the case in which the variance exceeded the mean. It was specified as a mixture of the gamma probability density function (pdf) and Poisson pmf with parameter  $\theta$  capturing the overdispersion (Venables and Ripley 2002). The pmf for the gamma-Poisson distribution, more commonly known as the negative binomial distribution, was as followed:

$$\Pr(Y = y) = \frac{\Gamma(y + \theta)}{\Gamma(\theta) y!} \frac{\lambda^{y} \theta^{\theta}}{(\lambda + \theta)^{y + \theta}}$$
(3.2)

The negative binomial distribution had  $E(Y) = \lambda$  and  $Var(Y) = \lambda + (\lambda^2/\theta)$ . A smaller value of  $\theta$  would increase Var(Y); hence greater overdispersion.

To avoid confusion with other literature, the zero-inflated (ZI) and hurdle (H) models followed the specifications and terminologies of Lambert (1992) and Mullahy (1986), respectively, in the current analysis. Baughman (2007) viewed both models as

finite mixture models (McLachlan and Peel 2000) with different latent structures. For generality, define  $f_{Poisson}(y)$  as the Poisson pmf (eqn. 3.1) and  $f_{NegBin}(y)$  as the negative binomial pdf (eqn. 3.2), where *y* was the observed value of *Y*.

The ZI model was a mixture of a point mass at zero and a count distribution and had one latent structure (Baughman 2007). The latent structure was a dual-state process in which y = 0 was the perfect state and  $y \ge 0$  was the imperfect state. The dual-state process was governed by p, the probability of an observation in the perfect state or sometimes referred as the amount of zero-inflation. Hence the model consisted of a zero-inflation component and a count component.

The zero-inflated Poisson pmf was,

$$\Pr(Y = y) = \begin{cases} p + (1-p) f_{Poisson}(0), & \text{if } y = 0\\ (1-p) f_{Poisson}(y), & \text{if } y > 0 \end{cases}$$
(3.3)

The zero-inflated negative binomial pmf was,

$$\Pr(Y = y) = \begin{cases} p + (1 - p) f_{NegBin}(0), & \text{if } y = 0\\ (1 - p) f_{NegBin}(y), & \text{if } y > 0 \end{cases}$$
(3.4)

The H model was a mixture of a point mass at zero and a truncated-at-zero count distribution, and had two latent structures as described by Baughman (2007). The two latent structures were the hurdle and the imperfect state. An observation had strictly positive count (y > 0) when the hurdle was crossed; otherwise y = 0. The ability to cross the hurdle was measured by  $\pi$ , defined as the probability of crossing the hurdle by Mullahy (1986). Hence the model consisted of a hurdle component and a truncated-count component. The second latent structure was a natural extension

because every observation must be in the imperfect state to have the implicit "potential" of crossing the hurdle.

The hurdle Poisson pmf was,

$$\Pr(Y = y) = \begin{cases} 1 - \pi, & \text{if } y = 0\\ \frac{\pi}{1 - f_{Poisson}(0)} f_{Poisson}(y), & \text{if } y > 0 \end{cases}$$
(3.5)

The hurdle negative binomial pmf was,

$$\Pr(Y = y) = \begin{cases} 1 - \pi, & \text{if } y = 0\\ \frac{\pi}{1 - f_{NegBin}(0)} f_{NegBin}(y), & \text{if } y > 0 \end{cases}$$
(3.6)

The six pmfs were specified in the GLM context (McCullagh and Nelder 1989) by conditioning the parameter(s) of a pmf on a set of predictors. Let **B** be a  $n \times p$  matrix of predictors for the count component (ZI model) and the truncated-count component (H model), and **A** be a  $n \times q$  matrix of predictors for the zero-inflation component (ZI model) and the hurdle component (H model), where *n* was the number of observations, and *p* and *q* were the number of covariates for the count and zero-inflation components, respectively. Furthermore, let **β** and **a** be vectors of coefficients associated with **B** and **A**, respectively. Then the relationships between parameters and predictors through link functions were,

$$\log(\lambda) = \mathbf{B}\boldsymbol{\beta} \tag{3.7}$$

for both the count component and the truncated-count component. For the zeroinflation component and the hurdle component,

$$\operatorname{logit}(\mathbf{p}) = \log\left(\frac{\mathbf{p}}{1-\mathbf{p}}\right) = \mathbf{A}\boldsymbol{\alpha}$$
(3.8)

$$\operatorname{logit}(\boldsymbol{\pi}) = \log\left(\frac{\boldsymbol{\pi}}{1-\boldsymbol{\pi}}\right) = \mathbf{A}\boldsymbol{\alpha}$$
(3.9)

In this study, we assumed  $\mathbf{B} = \mathbf{A}$ , and the set of vegetation predictors for all components were HERB, CONIF, SHRUB, CWDVOL, SHANNON and TPH\*. Furthermore, there was slight variation in the number of trap nights (*tn*) for each trap, i.e., the number of nights that a trap was opened. It might be reasonable to assume that the number of individual captured was proportional to the number of trap nights. To correct for this effect, log(*tn*) was entered into **B** as an offset with its coefficient equal to 1 (McCullagh and Nelder 1989). Hence expanding eqns. (3.7), (3.8) and (3.9),

$$\log(\lambda_i) = \beta_{1i} + \beta_{2i}HERB_i + \beta_{3i}CONIF_i + \beta_{4i}SHRUB_i + \beta_{5i}CWDVOL_i + \beta_{6i}SHANNON_i + \beta_{7i}TPH_i^* + \log(tn_i)$$
(3.10)

$$logit(p_i) = \alpha_{1i} + \alpha_{2i}HERB_i + \alpha_{3i}CONIF_i + \alpha_{4i}SHRUB_i + \alpha_{5i}CWDVOL_i + \alpha_{6i}SHANNON_i + \alpha_{7i}TPH_i^* + log(tn_i)$$
(3.11)

$$logit(\pi_i) = \alpha_{1i} + \alpha_{2i} HERB_i + \alpha_{3i} CONIF_i + \alpha_{4i} SHRUB_i + \alpha_{5i} CWDVOL_i + \alpha_{6i} SHANNON_i + \alpha_{7i} TPH_i^* + log(tn_i)$$
(3.12)

where i = 1, ..., n.

We also assumed the observations were independent, so ignored any possible spatial correlations between traps. The tenability of this assumption ultimately depended on the specific covariates included in the model; therefore, we disregarded it to simplify model comparisons and interpretations, and in recognition of the fact that significance of predictors was not the primary interest in this part of the analysis. The corrected Akaike Information Criteria (AICc, Burnham and Anderson 2004) was used as a measure for model comparison. The data from the four species were analyzed with the pscl package (Zeileis et al. 2008) available from the open source statistical software R (R Development Core Team 2009).

### **3.3.2 Generalized Linear Mixed Models (GLMMs)**

Most statistical models assumed independence of observations. Nested subjects tended to be more similar due to proximity and similarities in the physical environment, biological community, and other factors on a spatial scale corresponding to the locations of nested subjects; hence their observations were not likely independent unless covariates inducing the spatial variability in influential conditions were included in the model. A multilevel model assumes observations to be independent after conditioning on random effects of each level in the nested structure (Congdon 2005). The primary motivation for eliminating autocorrelation was to have valid estimates for the standard errors of parameter estimates; otherwise, inferences might be incorrect if these different sources of variability were not taken into account and residual errors were autocorrelated (Congdon 2005).

In this study, the GLMMs were formulated in the context of Bayesian Hierarchical Models (BHMs) despite possible issues of non-conjugacy and complex Monte Carlo Markov Chain (MCMC) techniques (Congdon 2006). The method of BHMs was a sequential specification of probability distributions for a series of parameters. Given observed data *x*, the likelihood function of the observations with parameters  $\xi$  was  $L(x|\xi)$ . In Bayesian analysis, the parameters themselves were random variables and thus follow a probability distribution  $f(\xi)$  or commonly known as prior distribution (Congdon 2003). A prior distribution amounted to modeling assumptions and hypotheses about the nature of the parameters without considering the data (Congdon 2006). The parameters  $\xi$  might in turn depend on other parameters  $\psi$  (hyperparameters); therefore, the conditional prior distribution of  $\xi$  was  $f(\xi|\psi)$ , and the hyperparameters themselves had prior distribution  $f(\psi)$ . In summary, the posterior distribution of parameters and hyperparameters  $f(\xi, \psi|x)$ , i.e., updated knowledge on the prior beliefs of  $f(\xi|\psi)$  and  $f(\psi)$  with observed data, was (Congdon 2003),

$$f(\xi, \psi | x) \propto L(x | \xi) f(\xi | \psi) f(\psi)$$
(3.13)

When the posterior distribution had the same distributional form as the prior distribution, the prior distribution was said to be conjugate with the likelihood function and the advantage was analytical tractability (Congdon 2006). However in general application, particularly GLMMs, one often encountered non-conjugacy whereby the posterior distribution had no closed-formed expression; e.g., using normally distributed errors in the log link for count data (Congdon 2006). With the advent of sampling-based estimation methods such as MCMC techniques, one could relax the conjugacy restriction by choosing non-conjugate prior distribution. The sampling method would provide estimates of distributional characteristics relating to the parameters (Smith and Gelfand 1992). When specifying complex random effects for multilevel discrete data, more complicated estimation methods might be considered, e.g., hybrid Gibbs-Metropolis sampling (Browne and Drapper 2000). Other possible approaches to the data analysis included penalized quasi-likelihood and iterative generalized least squares, but BHM had several advantages over the alternatives. The primary advantages of the BHM were its ability to accommodate both the small sample size within clusters and the small number of higher level units, and its incorporation of all sources of uncertainty in estimating random effects (Congdon 2006).

The three models considered in the second part of this study were the Poisson GLMM (POIS-GLMM), overdispersed Poisson GLMM (overPOIS-GLMM) and negative binomial GLMM (NB-GLMM). Two random effects were considered: (1) a random effect for the nesting structure of sample points within experimental units, and (2) a random block effect consistent with the RCBD of the experiment. The layout of the GLMM followed the construct of the RCBD in Littell et al. (2006) and BHM in Congdon (2005).

The response variable and predictor variables for the models were defined as followed:

```
y_{ijk} = Observed count of individuals for a species at k-th trap in j-th treatment of i-th block
```

 $HERB_{ijk}$  = Mean percent herb cover at k-th trap in j-th treatment of i-th block

- $CONIF_{ijk}$  = Mean percent understory conifer cover at *k*-th trap in *j*-th treatment of *i*-th block
- $SHRUB_{ijk}$  = Mean percent tall shrub cover at *k*-th trap in *j*-th treatment of *i*-th block
- $CWDVOL_{ijk}$  = Mean coarse woody debris volume per ha at *k*-th trap in *j*-th treatment of *i*-th block (m<sup>3</sup>/ha)
- $SHANNON_{ijk}$  = Shannon index of heterogeneity in vertical structure at *k*-th trap in *j*-th treatment of *i*-th block

$$TPH_{ijk}^*$$
 = Tree density at k-th trap in j-th treatment of i-th block (trees/100-ha)

As in the previous section, the expectation of the trap-level count at *k*-th trap in *j*-th treatment of *i*-th block ( $\lambda_{ijk}$ ) was conditioned on a set of predictors in the GLM context and corrected for the variation in trap nights. The model formulation of the POIS-GLMM on the trap level was therefore as followed:

$$y_{ijk} \sim Poisson(\lambda_{ijk}) \tag{3.14}$$

$$\log(\lambda_{ijk}) = b_{1ij} + b_{2ij}HERB_{ijk} + b_{3ij}CONIF_{ijk} + b_{4ij}SHRUB_{ijk} + b_{5ij}CWDVOL_{ijk} + b_{6ij}SHANNON_{ijk} + b_{7ij}TPH^*_{ijk} + \log(tn_{ijk})$$
(3.15)

where

 $\lambda_{ijk}$  = Mean count of a species

- $b_{hij} = h$ -th random coefficient for *j*-th treatment of *i*-th block
- $tn_{ijk}$  = Number of trap nights at *k*-th trap in *j*-th treatment of *i*-th block

Designation of  $b_{hij}$  highlighted the fact that coefficients were random in a twolevel model. On the unit-level, each coefficient was partitioned into an overall mean, a random block effect and a random error as followed:

$$b_{1ij} = \mu_{1} + d_{1i} + e_{1ij}$$

$$b_{2ij} = \mu_{2} + d_{2i} + e_{2ij}$$

$$b_{3ij} = \mu_{3} + d_{3i} + e_{3ij}$$

$$b_{4ij} = \mu_{4} + d_{4i} + e_{4ij}$$

$$b_{5ij} = \mu_{5} + d_{5i} + e_{5ij}$$

$$b_{6ij} = \mu_{6} + d_{6i} + e_{6ij}$$

$$b_{7ij} = \mu_{7} + d_{7i} + e_{7ij}$$
(3.16)

where

$$\mu_h$$
 = Overall mean for the *h*-th random coefficient for the corresponding covariate

 $d_{hi}$  = Random block effect for the *h*-th random coefficient of *i*-th block  $e_{hij}$  = Random error for the *h*-th random coefficient for *j*-th treatment of *i*-th block and h = 1,...,7 coefficients (an intercept plus six predictors); i = 1,...,m blocks (m = 6for PEMA and NEGI and m = 4 for CLGA and PEKE); j = 1,...,6 treatments; k =  $1,...,n_{ij}$  traps (where  $n_{ij}$  was 32–37 depending on the *j*-th treatment of *i*-th block). This model specification implied that an *ij* combination represented a specific experimental unit.

The random block effect  $d_{hi}$  on the *h*-th coefficient  $b_{hij}$  represented the average deviation of units in the *i*-th block from the overall mean for all blocks (Kuehl 2000).

The random error  $e_{hij}$  on the *h*-th coefficient represented the residual for *j*-th treatment in *i*-th block after accounting for the average block effect.

In matrix notation, let

$$\mathbf{B}_{ij} = \begin{bmatrix} b_{1ij} \\ b_{2ij} \\ b_{3ij} \\ b_{3ij} \\ b_{4ij} \\ b_{5ij} \\ b_{6ij} \\ b_{7ij} \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \\ \mu_7 \end{bmatrix}, \quad \mathbf{D}_i = \begin{bmatrix} d_{1i} \\ d_{2i} \\ d_{3i} \\ d_{4i} \\ d_{5i} \\ d_{6i} \\ d_{7i} \end{bmatrix}, \quad \mathbf{E}_{ij} = \begin{bmatrix} e_{1ij} \\ e_{2ij} \\ e_{3ij} \\ e_{4ij} \\ e_{5ij} \\ e_{6ij} \\ e_{7ij} \end{bmatrix}$$
(3.17)

Equation (3.16) could be written as  $\mathbf{B}_{ij} = \mathbf{U} + \mathbf{D}_i + \mathbf{E}_{ij}$ . Furthermore, the full model for all *ij* on the unit-level was,

$$\begin{bmatrix} \mathbf{B}_{11} \\ \vdots \\ \mathbf{B}_{16} \\ \mathbf{B}_{21} \\ \vdots \\ \mathbf{B}_{26} \\ \vdots \\ \mathbf{B}_{26} \\ \vdots \\ \mathbf{B}_{m6} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ \vdots \\ \mathbf{D}_{m6} \end{bmatrix} \mathbf{U}' + \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \mathbf{D}_{1} \\ \mathbf{D}_{2} \\ \vdots \\ \mathbf{D}_{m} \end{bmatrix} + \begin{bmatrix} \mathbf{E}_{11} \\ \vdots \\ \mathbf{E}_{21} \\ \vdots \\ \mathbf{E}_{26} \\ \vdots \\ \mathbf{E}_{m1} \\ \vdots \\ \mathbf{E}_{m6} \end{bmatrix}$$
(3.18)

The full vector  $\mathbf{B}_{ij} \forall ij$  in eqn. (3.18) was assumed to follow a prior distribution characterized as a multivariate normal (MVN),

$$\begin{bmatrix} \mathbf{B}_{11} \\ \vdots \\ \mathbf{B}_{16} \\ \mathbf{B}_{21} \\ \vdots \\ \mathbf{B}_{26} \\ \vdots \\ \mathbf{B}_{m1} \\ \vdots \\ \mathbf{B}_{m6} \end{bmatrix} \sim MVN \begin{pmatrix} \begin{bmatrix} \mathbf{U} \\ \vdots \\ \mathbf{U} \\ \end{bmatrix}$$

where **Z**, **G**, and **R** are defined below. To construct the variance-covariance matrix (**ZGZ'** + **R**) in eqn. (3.19), the variance-covariance matrices of the random error and the random block effect were required. Let  $\sigma_h^2$  be defined as the variance of the random error associated with the *h*-th coefficient, and  $\rho_{hh'}$  as the correlation of random errors between *h*-th and *h'*-th coefficients, where  $h \neq h'$ . The prior distribution for random error vector **E**<sub>ij</sub> (eqn. 3.17) was assumed to follow a multivariate normal distribution (MVN),

$$\mathbf{E}_{ij} \sim MVN(\mathbf{0}_{7\times 1}, \mathbf{\Sigma}_R) \tag{3.20}$$

where  $\mathbf{0}_{7\times 1}$  was a 7 × 1 vector of zeros and  $\Sigma_R$  was a 7 × 7 variance-covariance matrix of the random error vector  $\mathbf{E}_{ij}$ ,

$$\boldsymbol{\Sigma}_{R} = \begin{bmatrix} \sigma_{1}^{2} & \rho_{12}\sigma_{1}\sigma_{2} & \cdots & \rho_{17}\sigma_{1}\sigma_{7} \\ \rho_{12}\sigma_{1}\sigma_{2} & \sigma_{2}^{2} & \cdots & \rho_{27}\sigma_{2}\sigma_{7} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{17}\sigma_{1}\sigma_{7} & \rho_{27}\sigma_{2}\sigma_{7} & \cdots & \sigma_{7}^{2} \end{bmatrix}$$
(3.21)

Furthermore, let  $\tau_h^2$  be defined as the variance of random block effect

associated with the *h*-th coefficient, and  $\phi_{hh'}$  as the correlation of random block effect between *h*-th and *h*'-th coefficients, where  $h \neq h'$ . The prior distribution for random block effect vector **D**<sub>*i*</sub> (eqn. 3.17) was also assumed multivariate normal,

$$\mathbf{D}_{i} \sim MVN\left(\mathbf{0}_{7\times 1}, \mathbf{\Sigma}_{D}\right) \tag{3.22}$$

where  $\Sigma_D$  was a 7 × 7 variance-covariance matrix of the random block effect vector  $\mathbf{D}_{i}$ ,

$$\boldsymbol{\Sigma}_{D} = \begin{bmatrix} \tau_{1}^{2} & \phi_{12}\tau_{1}\tau_{2} & \cdots & \phi_{17}\tau_{1}\tau_{7} \\ \phi_{12}\tau_{1}\tau_{2} & \tau_{2}^{2} & \cdots & \phi_{27}\tau_{2}\tau_{7} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{17}\tau_{1}\tau_{7} & \phi_{27}\tau_{2}\tau_{7} & \cdots & \tau_{7}^{2} \end{bmatrix}$$
(3.23)

For both  $\Sigma_R$  and  $\Sigma_D$ , we assumed nonzero correlation between any pairs of coefficients,  $\rho_{hh'}$  and  $\phi_{hh'}$ . This assumption was reasonable given that estimated coefficients generally were correlated to a certain extent in regression models.

The variance-covariance matrix  $\mathbb{Z}G\mathbb{Z}'+\mathbb{R}$  in eqn. (3.19) can be constructed with the variance-covariance matrices  $\Sigma_R$  and  $\Sigma_D$  as components. Based on eqn. (3.20), the full random error vector  $\mathbf{E}_{ij} \forall ij$  in eqn. (3.18) had a multivariate normal distribution as follows:

$$\begin{bmatrix} \mathbf{E}_{11} \\ \vdots \\ \mathbf{E}_{16} \\ \mathbf{E}_{21} \\ \vdots \\ \mathbf{E}_{26} \\ \vdots \\ \mathbf{E}_{m1} \\ \vdots \\ \mathbf{E}_{m6} \end{bmatrix} \sim MVN \begin{pmatrix} \begin{bmatrix} \mathbf{0}_{7\times 1} \\ \vdots \\ \mathbf{0}_{7\times 1} \\ \vdots \\$$

where  $\mathbf{R} = \mathbf{I}_{(m \cdot 6) \times (m \cdot 6)} \otimes \Sigma_R$  was the variance-covariance matrix of the full random error vector  $\mathbf{E}_{ij} \forall ij$  in eqn. (3.18) with dimension of  $(m \cdot 42) \times (m \cdot 42)$ , and  $\mathbf{I}_{(m \cdot 6) \times (m \cdot 6)}$  is a identity matrix with dimension of  $(m \cdot 6) \times (m \cdot 6)$ .

Based on eqn. (3.22), the full vector of random block effects  $\mathbf{D}_i \forall i$  in eqn. (3.18) followed a multivariate normal distribution as,

$$\begin{bmatrix} \mathbf{D}_{1} \\ \mathbf{D}_{2} \\ \vdots \\ \mathbf{D}_{m} \end{bmatrix} \sim MVN \begin{pmatrix} \begin{bmatrix} \mathbf{0}_{7\times 1} \\ \mathbf{0}_{7\times 1} \\ \vdots \\ \mathbf{0}_{7\times 1} \end{bmatrix}, \mathbf{G}$$
(3.25)

where  $\mathbf{G} = \mathbf{I}_{m \times m} \otimes \boldsymbol{\Sigma}_D$  was the variance-covariance matrix of the full random block effects  $\mathbf{D}_i \forall i$  in eqn. (3.18) with a dimension of  $(m \cdot 7) \times (m \cdot 7)$ , and  $\mathbf{I}_{m \times m}$  was a  $m \times m$ identity matrix. Furthermore, let  $\mathbf{Z}$  defined as the design matrix associated with the full random block effects  $\mathbf{D}_i \forall i$  in eqn. (3.18),

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$
(3.26)

Consequently,

$$\mathbf{Z}\mathbf{G}\mathbf{Z}' = \begin{bmatrix} \boldsymbol{\Sigma}_{D} & \cdots & \boldsymbol{\Sigma}_{D} & \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{0}_{7\times7} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{D} & \cdots & \boldsymbol{\Sigma}_{D} & \boldsymbol{0}_{7\times7} & \boldsymbol{\Sigma}_{D} & \cdots & \boldsymbol{D}_{7\times7} & \cdots & \boldsymbol{0}_{7\times7} \\ \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{0}_{7\times7} & \boldsymbol{\Sigma}_{D} & \cdots & \boldsymbol{\Sigma}_{D} & \cdots & \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{0}_{7\times7} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{0}_{7\times7} & \boldsymbol{\Sigma}_{D} & \cdots & \boldsymbol{\Sigma}_{D} & \cdots & \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{0}_{7\times7} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{0}_{7\times7} & \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{\Sigma}_{D} & \cdots & \boldsymbol{\Sigma}_{D} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{0}_{7\times7} & \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{0}_{7\times7} & \cdots & \boldsymbol{\Sigma}_{D} & \cdots & \boldsymbol{\Sigma}_{D} \end{bmatrix}$$

$$(3.27)$$

where **ZGZ'** has dimension of  $(m \cdot 42) \times (m \cdot 42)$ , identical to the dimension of **R**. Finally, the variance-covariance matrix for **B**<sub>*ij*</sub>  $\forall$ *ij* as specified in eqn. (3.19) was,

$$ZGZ' + R$$

$$= \begin{bmatrix} \Sigma_{D} + \Sigma_{R} & \cdots & \Sigma_{D} & \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \Sigma_{D} & \cdots & \Sigma_{D} + \Sigma_{R} & \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} \\ \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} & \Sigma_{D} + \Sigma_{R} & \cdots & \Sigma_{D} & \cdots & \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} & \Sigma_{D} & \cdots & \Sigma_{D} + \Sigma_{R} & \cdots & \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} & \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} & \mathbf{0}_{7\times7} & \cdots & \mathbf{0}_{7\times7} & \cdots & \Sigma_{D} + \Sigma_{R} & \cdots & \Sigma_{D} + \Sigma_{R} \end{bmatrix}$$
(3.28)

From eqn. (3.28), some important characteristics of the variance-covariance matrix of the coefficients are evident. The variance-covariance of h-th coefficients for a unit was,

$$\operatorname{var}(\mathbf{B}_{ij}) = \mathbf{\Sigma}_{D} + \mathbf{\Sigma}_{R} = \begin{bmatrix} \tau_{1}^{2} + \sigma_{1}^{2} & \phi_{12}\tau_{1}\tau_{2} + \rho_{12}\sigma_{1}\sigma_{2} & \cdots & \phi_{17}\tau_{1}\tau_{7} + \rho_{17}\sigma_{1}\sigma_{7} \\ \phi_{12}\tau_{1}\tau_{2} + \rho_{12}\sigma_{1}\sigma_{2} & \tau_{2}^{2} + \sigma_{2}^{2} & \cdots & \phi_{27}\tau_{2}\tau_{7} + \rho_{27}\sigma_{2}\sigma_{7} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{17}\tau_{1}\tau_{7} + \rho_{17}\sigma_{1}\sigma_{7} & \phi_{27}\tau_{2}\tau_{7} + \rho_{27}\sigma_{2}\sigma_{7} & \cdots & \tau_{7}^{2} + \sigma_{7}^{2} \end{bmatrix} (3.29)$$

which explicitly showed the contribution of two variance components; a component from variation on the block level ( $\Sigma_D$ ) and a component from variation among the units within a block ( $\Sigma_R$ ). The former measured the variation between block-level means of the *h*-th coefficient. The latter measured the extent of variability in the *h*-th coefficient among the units within a block. The correlations among coefficients for a given unit were interpreted similarly.

On the other hand, coefficients from two different units within the same block were correlated as,

$$\operatorname{cov}(\mathbf{B}_{ij}, \mathbf{B}_{ij'}) = \boldsymbol{\Sigma}_{D} = \begin{bmatrix} \tau_{1}^{2} & \phi_{12}\tau_{1}\tau_{2} & \cdots & \phi_{17}\tau_{1}\tau_{7} \\ \phi_{12}\tau_{1}\tau_{2} & \tau_{2}^{2} & \cdots & \phi_{27}\tau_{2}\tau_{7} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{17}\tau_{1}\tau_{7} & \phi_{27}\tau_{2}\tau_{7} & \cdots & \tau_{7}^{2} \end{bmatrix} \quad \text{where } j \neq j' \quad (3.30)$$

The purpose of blocking was for units within a block to be as homogeneous as possible (Kuehl 2000). As a result, units were more similar within a block than across blocks. Therefore, it was necessary to account for the correlation of coefficients from different units within a block, and this was properly captured by eqn. (3.30).

Lastly, it was assumed that covariances of coefficients among blocks were zero,

$$\operatorname{cov}(\mathbf{B}_{ij}, \mathbf{B}_{i'j'}) = \mathbf{0}_{7\times7} \qquad \text{where } i \neq i' \text{ for } \forall j, \forall j'$$
(3.31)

The hyperparameters of the variance-covariance matrices and the general mean were given non-informative prior distributions. A non-informative prior distribution reflects prior ignorance about the nature of the parameters, possibly due to lack of existing knowledge (Congdon 2006). Therefore, it is preferable when data will dominate the determination of the posterior estimates of the parameters (Banerjee et al. 2004). It is also commonly known as a flat prior distribution because it implies that a parameter is uniformly distributed between two limits such as  $-\infty$  and  $+\infty$  for a mean, or, between 0 and  $+\infty$  for a variance. For the general mean,

$$\mathbf{U} \sim MVN \begin{pmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, 1000000\mathbf{I}_{7\times7} \end{pmatrix}$$
(3.32)

where  $\mathbf{I}_{7\times7}$  is a 7 × 7 identity matrix.

For the variance-covariance matrices of the random error (eqn. 3.21) and random block effect (eqn. 3.23), an inverse-Wishart distribution was assumed,

$$\Sigma_{R} \sim inv - Wishart(\Omega, v) \tag{3.33}$$

$$\Sigma_{D} \sim inv - Wishart(\Omega, v) \tag{3.34}$$

where  $\Omega$  was a 7 × 7 identity matrix and v = 7. The inverse-Wishart distribution is the multivariate generalization of the inverse gamma distribution, and has commonly been used as a conjugate prior distribution for variance-covariance matrix (Congdon 2006). The latter use has been appealing because the distribution was defined on a real-valued positive definite matrix, which is an important characteristic of the variance-covariance matrix. The pdf of the inverse-Wishart distribution for  $\Sigma_R$  (or  $\Sigma_D$ ) was (Congdon 2006, Johnson and Wichern 2007),

$$f(\Sigma_{R} | \Omega, v) = \frac{|\Omega|^{(v/2)} |\Sigma_{R}|^{-(v+t+1)/2} \exp(-tr(\Omega\Sigma_{R}^{-1})/2)}{2^{vt/2} \pi^{t(t-1)/4} \prod_{i=0}^{t-1} \Gamma((v-i)/2)}$$
(3.35)

where  $\Omega$  was the scale parameter, *v* was the degrees of freedom, *t* was the *t* × *t* dimension of  $\Sigma_R$ , *tr*() was the trace of the matrix product and  $\Gamma(\cdot)$  was the gamma

function. As it was for  $\Sigma_R$  and  $\Sigma_D$ , the matrix  $\Omega$  is often an identity matrix with the same dimension of a variance-covariance matrix and *v* to equal the order of the matrix (Chib and Winkelmann 2001).

The overPOIS-GLMM extended the POIS-GLMM by including a trap-level random error term with the predictors in eqn. (3.15). A POIS model in general might be subjected to overdispersion because it did not have a variance parameter to capture the variation in the data (Gelman and Hill 2007). Therefore, the trap-level random error  $\varepsilon_{ijk}$  was added to directly model the overdispersion, which was meant to capture possibly additional level of variation beyond what was already accounted for by the multilevel modeling (Congdon 2005). Extending eqn. (3.15),

$$\log(\lambda_{ijk}) = b_{1ij} + b_{2ij}HERB_{ijk} + b_{3ij}CONIF_{ijk} + b_{4ij}SHRUB_{ijk} + b_{5ij}CWDVOL_{ijk} + b_{6ij}SHANNON_{ijk} + b_{7ij}TPH_{ijk}^* + \log(tn_{ijk}) + \varepsilon_{ijk}$$

$$(3.36)$$

The  $\varepsilon_{ijk}$  was assumed normally distributed with mean 0 and variance  $\sigma_{\varepsilon}^2$ . Thus, a higher  $\sigma_{\varepsilon}^2$  value would indicate greater overdispersion. A non-informative uniform prior distribution was given for the variance  $\sigma_{\varepsilon}^2$ ,

$$\sigma_{\varepsilon}^2 \sim U(0, 1000000) \tag{3.37}$$

Congdon (2005) cautioned that adding a random error term generally improved the model fit but may risk overparameterizing the model. However, the model converged satisfactorily for the data from all four species. As mentioned previously, the NB model was an extension of the POIS model to account for overdispersion in the data. Following Ntzoufras (2009) and Zuur et al. (2009), the NB-GLMM was specified as,

$$y_{ijk} \sim NegBin(p_{ijk}, \theta)$$
(3.38)

where  $p_{ijk} = \theta / (\theta + \lambda_{ijk})$  and *NegBin*() denotes a negative binomial distribution.

Subsequently,

$$\log(\lambda_{ijk}) = b_{1ij} + b_{2ij}HERB_{ijk} + b_{3ij}CONIF_{ijk} + b_{4ij}SHRUB_{ijk} + b_{5ij}CWDVOL_{ijk} + b_{6ij}SHANNON_{ijk} + b_{7ij}TPH_{ijk}^{*} + \log(tn_{ijk})$$
(3.39)

and a non-informative gamma prior distribution was assumed for the parameter  $\theta$ ,  $\theta \sim gamma(0.001, 0.001)$  (3.40)

As in the GLM, the parameter  $\theta$  captured overdispersion in the data not accounted for by the multilevel model. It would be ideal to have  $\theta$  parameterized for each unit, i.e.,  $\theta_{ij}$ . However, initial fitting of such a model failed to converge, probably due to overparameterization.

To facilitate comparison between the GLM and GLMM, posterior estimates of a given *h*-th coefficient were averaged over all units,

$$\overline{\tilde{b}}_{h} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{6} \tilde{b}_{hij}}{m \times 6}$$
(3.41)

and the associated 95% credible interval (analogous to confidence interval; Congdon 2006) was reported. The three GLMMs were specified and fitted using WinBUGS

(Lunn et al. 2000). Three disparate sets of initial values (chains) were used. The MCMC algorithm was run with 250,000 iterations and the initial 20,000 iterations were dropped. The convergence of the algorithm was judged from trace and history plots of the MCMC samples, the Brooks-Gelman-Rubin (BGR) statistic (Brooks and Gelman 1998), and the Monte Carlo (MC) errors of estimated parameters. The BGR statistic assessed convergence by comparing between-chain and within-chain variability (Ntzoufras 2009). A model was considered to have attained convergence when the BGR statistic was < 1.05. The MC errors measured the variation due to simulation for estimated parameters (Ntzoufras 2009). The parameters were estimated with high precision if the MC errors were relatively low compared to the corresponding estimated posterior standard deviations (Ntzoufras 2009).

The trace and history plots of the MCMC samples and the BGR statistics showed convergence for all three models and for all four species. The MC errors on average were less than 2% of the corresponding posterior standard deviations of the estimated parameters. Thinning of the MCMC samples was necessary for some parameters to reduce the autocorrelation between samples. Model comparison for each species was based on Deviance Information Criterion (DIC) of Spiegelhalter et al. (2002). The DIC was intended as a generalization of Akaike Information Criterion (AIC) that measured model complexity in hierarchical random effect models (Congdon 2006, Lunn et al. 2000); thus smaller values indicated better fitting models.

## **3.4 RESULTS**

#### **3.4.1 Zero-modified Models (ZMMs)**

Comparison among ZMMs showed two consistent results across species. First, estimates from the hurdle component were identical between HPOIS and HNB, but not for the zero-inflation component between ZIP and ZINB (Table 3.1–3.4). The explanation was straightforward. The H model conveniently maximized the log-likelihood functions of the hurdle and the truncated-count components separately (Mullahy 1986). Conversely, the ZI models did not have this flexibility due to the mixture specified in the pmf (Lambert 1992). When a predictor had significant coefficients in both the zero-inflation and the hurdle components, the signs of the two estimates were reversed (Table 3.1–3.4). This result was expected due to the contrasting definitions of *p* and  $\pi$ ; the zero-inflation component predicted the probability of an observation in the perfect state whereas the hurdle component predicted the probability of observing a positive count (Zeileis et al. 2008).

Predictors CONIF, CWDVOL and TPH\* were significantly correlated with the number of PEMA captures (Table 3.1 and 3.5). The POIS and NB models suggested that number of captures declined with an increase in each of these three variables. In the H models, however, these same predictors only discriminated between plots with and without any captures of this species; i.e., the predicted value of  $\pi$ , the probability of observing at least one capture, declined with increasing levels of CONIF, CWDVOL and TPH\*. Habitat variables were not significantly effective in estimating

the number of captures conditional on presence of PEMA ( $\geq 1$  capture). No relationship could be detected under the ZIP model between the habitat descriptors and number of captures. Conversely, the three predictors plus SHRUB significantly correlated with the count under the ZINB model, but the large estimated coefficients and standard errors in the zero-inflation component made the parameter estimates unreasonable and the model results in general unreliable.

In contrast to *Peromyscus maniculatus*, habitat variables were significantly correlated with both number of zero observations and conditional counts for *Clethrionomys gapperi* (CLGA; Table 3.2 and 3.6). Number of captures declined with increasing herb and small shrub cover (HERB) in the POIS and NB models. Similarly, the probability of crossing the zero hurdle ( $\pi$ ) declined with increasing HERB in the H models, and the probability of observing zero captures (p) increased with increasing HERB in the ZI models. All models predicted a lower number of captures with increasing SHRUB, and the H models also predicted a declining probability of crossing the zero hurdle with increasing SHRUB. The 'effect' of CONIF on *Clethrionomys gapperi* was generally positive under POIS, NB and ZINB models, increasing probability of crossing the zero hurdle, and decreasing probability of observing zero in ZIP model. CWDVOL was positively correlated with number of captures only for the POIS and ZIP models. Vertical complexity (SHANNON) was positively correlated with number of captures only in the POIS and NB models, but an increase in this habitat variable increased the probability of crossing the zero hurdle and reduced the probability of observing a zero in the ZIP model. Increasing stand

density (TPH\*) appeared to increase the count in the POIS, NB and HPOIS models as well as increased the probability of crossing the zero hurdle.

The predictors HERB and TPH\* had a consistently positive effect on *Neurotrichus gibsii* (NEGI) captures (Table 3.3 and 3.7). Increasing level of either predictors increased the number of captures in POIS and NB models, increased the probability of crossing the zero hurdle in the H models and decreasing the probability of observing a zero in the ZI models. Oddly, a conflicting result was the negative correlation of TPH\* with captures in the ZIP model. The ZIP model predicted increasing NEGI captures with increasing CONIF, and the HPOIS model predicted a decreased in the captures with increasing SHANNON. Otherwise, both predictors were unrelated to number of captures and other parameters.

Surprisingly, the ZINB model for *Peromyscus keeni* (PEKE) failed to converge (Table 3.4). Number of captures increased with increasing HERB in all models (Table 3.4 and 3.8), and likewise the probability of crossing the zero hurdle increased with increasing HERB. CONIF was positively correlated with number of captures, but only in the POIS and NB models. Increasing CWDVOL increased the probability of crossing the zero hurdle in the H models and the number of captures in the POIS and ZIP models. Oddly, the results from the ZIP model for SHANNON appeared contradictory; declining in both number of captures and probability of observing zeros with increasing SHANNON.

NEGI captures fitted to the NB model resulted in the greatest overdispersion relative to other species and models (0.08, Table 3.7). For a species, the

overdispersion was smaller when captures were fitted to either the ZINB or HNB models compared to the NB model (e.g. CLGA, Table 3.6); i.e., the predicted  $\hat{\theta}$  was larger for the former two models. However, The HNB model returned unreasonably small  $\hat{\theta}$  for NEGI and PEKE, which might indicate a poor fit of the model for infrequently captured species (Table 3.3 and 3.4).

The ranking of model goodness-of-fit based on AICc was slightly different between species (Table 3.1–3.8). The two best models for PEMA were the NB and ZINB models, the two best for both CLGA and NEGI were the ZINB and HNB models, and the two best for PEKE were the ZIP and NB models. The POIS model provided relatively poor fit for all species by having the highest AICc values (Table 3.5–3.8). Besides that, the POIS model also consistently under-predicted frequency of zeros and over-predicted frequency of ones (Figure 3.3). Although some slight variation was evident, all other models appeared to predict the marginal count frequency of all species fairly well (Figure 3.3). A surprising result was that the NB models predicted frequency of zero count as well as the ZMMs (Figure 3.3).

# 3.4.2 Generalized Linear Mixed Models (GLMMs)

Results between the GLMs and GLMMs were notably different (Table 3.5– 3.8). The 95% credible intervals for estimates of coefficients from the GLMMs were consistently wider than the corresponding 95% confidence intervals from the GLMs, regardless of species. As a result, some coefficients that were significant under GLM became non-significant under the corresponding GLMM, e.g., coefficients associated with TPH\* for PEMA, CLGA and NEGI (Table 3.5–3.7). Effects of predictor variables in the GLMMs were always greater than the effects of the same variable in the corresponding GLM, e.g., CONIF and CWDVOL for PEMA (Table 3.5). For the less frequent species, NEGI and PEKE, number of captures increased with increasing HERB in the GLM but decreased instead in the GLMM (Table 3.7 and 3.8).

In the more frequently captured species, overdispersion under the NB-GLMM was less than under the NB GLM, i.e.,  $\tilde{\theta}$  from the NB-GLMM was larger (Table 3.5 and 3.6). The relative overdispersion among species changed between the GLMs and the GLMMs. Overdispersion of the NB GLM fitted to the PEMA data was intermediate and comparable to PEKE, but overdispersion of the NB-GLMM fitted for this same species was the least and similar to CLGA. Conversely, overdispersion estimates were comparable between the GLMs and GLMMs for the two less frequent species (Table 3.7 and 3.8). The relative overdispersion among species as indicated by  $\tilde{\sigma}_{\varepsilon}$  from overPOIS-GLMM was in agreement with the NB-GLMM (Table 3.5–3.8).

The estimated effects of CWDVOL illustrated the variability in posterior estimates among units. The effect of CWDVOL on the number of PEMA captures was more strongly negative for the GLMM than for the GLM, and the effect on CLGA changed from a positive effect in the POIS GLM to a negative effect in the POIS-GLMM (Table 3.5 and 3.6). For NEGI and PEKE, CWDVOL had no significant effect on number of captures predicted by the GLMs, but was negatively correlated with number of captures in the GLMMs (Table 3.7 and 3.8). The cause of this behavior was revealed by examining posterior estimates for individual experimental units (Figure 3.4). CWDVOL exhibited a range from a negative to neutral effect on number of captures. For NEGI, almost half of the units exhibited strongly negative posterior estimates, although the 95% credible intervals were relatively wide (Figure 3.4C). Estimates for the CWDVOL coefficient varied widely among experimental units, so caution should be exercised when inferring general effects of CWDVOL ( $\overline{b}_5$ ) among different geographic locations.

Posterior estimates among units for other predictors were usually not as dichotomous as CWDVOL, as exemplified by the predictors HERB and SHANNON for CLGA (Figure 3.5). The effect of HERB on number of CLGA captures among units distributed regularly from marginally significant positive effect to strong negative effect with a majority closed to neutral effect (Figure 3.5A). Interestingly, the posterior estimates for 40%A treatment (Treatment 4) could be neutral, significantly positive or significantly negative depending on units ( $\tilde{b}_{2l4}$ , Figure 3.5A). On the other hand, the effects of SHANNON among units consistently concentrated on the positive side but with different levels of significance (Figure 3.5B). Despite a variety of effects or significance levels among units, the general effects of HERB ( $\vec{b}_2$ ) and SHANNON ( $\vec{b}_6$ ) on CLGA captures were significantly negative and positive, respectively (Table 3.6).
Differences in posterior estimates of coefficients were small among the alternative (POIS, NB and overPOIS) GLMMs for both PEMA and CLGA. However, this was not the case for the less frequent species, NEGI and PEKE. The number of NEGI captures declined significantly with increasing HERB only in the POIS-GLMM, and the posterior intercept from the overPOIS-GLMM for this same species was significantly more negative and imprecise than in the other models (Table 3.7). For PEKE, the posterior intercept was marginally different between the overPOIS-GLMM and the two other GLMMs (Table 3.8). The overPOIS-GLMM had the lowest DIC relative to the POIS and NB GLMMs for all species, and the overPOIS-GLMM improved the fit of the less frequent species particularly well. Surprisingly, the NB-GLMM had a poorer fit (higher DIC) than the POIS-GLMM for all species, in contrast to the superior performance (lower AICc) of the NB GLM relative to the POIS GLM.

The posterior standard deviation of random error  $\tilde{\sigma}_h$  (eqn. 3.21) had consistent patterns across species and models (Figure 3.6). Among the coefficients, the intercept had the largest posterior standard deviation  $\tilde{\sigma}_1$  implying that the estimates were highly variable among units, with highest variability in NEGI (Figure 3.6C). Furthermore, its wide 95% credible interval indicated imprecision but the precision seemed to improve going from the POIS-GLMM to other models; at least for species other than PEMA (Figure 3.6B–D). The posterior standard deviations of the coefficients for SHANNON ( $\tilde{\sigma}_6$ ) and TPH\* ( $\tilde{\sigma}_7$ ) were the second largest and were even comparable to that of intercept from the NB-GLMM and overPOIS-GLMM fitted to NEGI and PEKE captures (Figure 3.6C and D). Although the results are not shown, all posterior correlations between random errors of any pair of coefficients  $\tilde{\rho}_{hh'}$  (eqn. 3.21) were non-significant. However, some correlation was evident between intercept and the coefficient for SHANNON when the POIS-GLMM was fitted to NEGI captures with posterior estimate of -0.54 for  $\tilde{\rho}_{16}$  and 95% credible interval of (-0.95, 0.29).

Similar to the results for  $\tilde{\sigma}_h$ , the largest posterior standard deviation for the random block effect  $\tilde{\tau}_h$  (eqn. 3.23) was associated with the intercept, implying that mean intercept was more variable among blocks than the coefficients associated with habitat descriptors (Figure 3.7). The wide 95% credible intervals indicated imprecise estimates, but in this case model choice did not improve the precision except for PEKE (Figure 3.7D). The posterior standard deviations for coefficients associated with habitat variables were all comparable to one another (Figure 3.7). Despite some variation among blocks, the posterior block effects  $\tilde{d}_{hi}$  on all coefficients were not significantly different from zero across all blocks and species, as indicated by 95% credible intervals overlapping with zero (Figure 3.8). Although results are not shown, posterior correlations between random block effects of any pair of coefficients  $\tilde{\phi}_{hh}$ . (eqn. 3.23) were non-significant.

# 3.4.3 Small Mammal Responses to Habitat Structures

The overPOIS-GLMM with backward and forward eliminations in variable selection was used to finalize the four small mammal species responses to habitat

structures. Considering the results from previous sections, overPOIS-GLMM would be the best choice because the model properly accounted for the sources of variation in the data, i.e., nesting of traps, random block effects and overdispersion, and consistently had the lowest DIC among the alternatives. Although the results among GLMMs were similar for the more abundant species, they were not so for the less frequently captured species (Table 3.5–3.8). Therefore, overPOIS-GLMM would be a general model applicable to species with different capture frequencies.

The final results (Table 3.9) were in general agreement with those reported in Table 3.5–3.8, with minor exceptions for PEMA and NEGI. The predictor TPH\* was significant under variable selection in addition to CONIF and CWDVOL, and all three were negatively correlated with abundance of PEMA (Table 3.9). For NEGI, SHRUB was significant in addition to CWDVOL after eliminating other variables, and it had a negative effect on captures (Table 3.9). The correlation between CLGA captures and habitat variables HERB and CWDVOL was negative, but between CLGA captures and SHANNON was positive. Both HERB and CWDVOL were negatively associated with PEKE captures. Lastly, the DIC values for the final models of all four small mammal species (Table 3.9) were higher than those for overPOIS-GLMM without variable selection (Table 3.5–3.8), but they were still lower than the alternative POIS-GLMM and NB-GLMM (Table 3.5–3.8).

### **3.5 DISCUSSION**

Knowledge about the relationship between habitat structure and the presence and abundance of target species or species groups are required to actively conserve biodiversity in managed forests. This knowledge is built from sampling the target population and its habitat structure, and then applying appropriate statistical methods to infer relationships. Differences in detection probabilities among different species are caused by differences in population size and specific aspects of the sampling design that include spatial and temporal distribution and sampling intensity. Therefore, the choice of statistical models should consider the expected detection frequency and the characteristics of the sampling design. Analysis of the DEMO small mammal data assessed the relative performance of alternative statistical models for four species that differed in capture frequency but were sampled under the same sampling design. The zero-modified models appealed to the common problem of observing a high frequency of zeros and provided a theoretical basis for modeling them. Multilevel models addressed the hierarchical structure of the dataset so that sources of variances were properly quantified and inferences were correct. Other sources of variation such as random block effects were easily modeled under a Bayesian framework. In the end, the goal was to test the consistency of inferences across models and species and to identify the most robust statistical models for species with a wide range in detection frequency.

# 3.5.1 Zero-modified Models (ZMMs)

Any given statistical model for analyzing the relationship between number of captures and habitat attributes, particularly those influenced by silvicultural treatments, assumes the observed count of a species follows some statistical distribution. The ZI models assume that observations are made under two possible conditions, the first being inhospitable habitat with conditions that do not support the presence of the target species (the perfect state), and the second a tolerable habitat (the imperfect state) in which the number of captures may vary. The parameter *p* measures the probability of an observation being in the inhospitable habitat. A significant predictor of p indicates that the latter is conditional on that specific habitat feature. As mentioned earlier, the H models assume the imperfect state for all observations; i.e., all observations are made in tolerable habitat. The parameter  $\pi$  in the H models measures the probability that the target species is present; i.e., that it crosses the zero hurdle. A significant predictor of  $\pi$  implies that this probability is conditional on the habitat feature represented by that specific predictor. The same or a different set of predictors may be significantly correlated with the number of captures, conditional on the habitat being generally suitable for the species. In short, the ZI and H models imply different interpretation of the ecological process that generates the observed data: zeros under the ZI models are assumed to be observed in a mix of suitable and unsuitable habitats, while all zeros in the H models are implied to be observed in suitable habitats only.

Some insights may be gained by assessing the consistency of relationships between number of captures and habitat variables among the alternative statistical models, and whether this consistency varies between relatively frequent and relatively infrequent species. The models are linked by considering the POIS model as the parent model for the others. The NB model extends the POIS model by accounting for overdispersion that could result from added variability attributable to omitted covariates (Minami et al. 2007), a clustered Poisson process (McCullagh and Nelder 1989), inter-subject variability (McCullagh and Nelder 1989), or temporal dependency between observations (Rose et al. 2006). Nonetheless both models assume a parsimonious ecological process in that the expected number of captures is correlated with some predictors of habitat structure. The ZMMs attempt to elaborate on the simpler POIS and NB processes by accounting for a secondary process that yields a higher number of zero observations than expected. The same predictors or different predictors then describe both the POIS and NB expectations and the added theoretical structure for generating excess zeros. In essence, the ZMMs are extensions of the POIS and NB models, because the latter are subsumed in the former. This relationship is illustrated by the empirical results from four small mammal species in the DEMO study.

PEMA is a relatively abundant generalist species that inhabits a wide range of habitat types (Moses and Boutin 2001, Muzika et al. 2004). Sullivan et al. (2008) found the species to be unaffected by clearcut, single seed-tree, group seed-tree and patch cut treatments. The presumed generalist behavior of this species is in general agreement with the results, particularly with respect to the lack of significant habitat variables for predicting number of captures in the ZMMs. CONIF, CWDVOL and TPH\* were significant predictors of mean number of captures under the POIS and NB models; however, the H models indicated that the predictors could also help distinguish the habitat conditions in which the species was less likely to be present. In fact, the generalist nature of the species would theoretically cause it to conform to a H model rather than a ZI model, because the H model assumes the habitat, in which sampling is carried out, is hospitable to PEMA.

For CLGA, both ZI and H models are plausible and supported by the results. The choice of a model for best representing the underlying ecological process would depend on the assumptions about within-unit habitat variability. CLGA has been described as a good indicator of old-growth forest conditions (Sullivan et al. 2005), and does not respond negatively to light regeneration harvests such as single-tree selection treatments in boreal forests (Von Trebra et al. 1998). However, a drastic reduction in population size is observed in clearcuts (Klenner and Sullivan 2003) or when residual overstory basal area of a harvested stand falls below approximately 5.1 m<sup>2</sup>/ha (Medin 1986). Runciman and Sullivan (1996) suggested that the reduction was due to removal of vegetation structure and shifts in species composition that constituted preferred habitat preferences of the species. A ZI model is appropriate for the DEMO data because the treatments created some areas with inhospitable habitat conditions within a unit. This effect was particularly strong for the aggregated treatments (15%A or 40%A) because the cut area between undisturbed aggregates

created two very distinct habitats. On the other hand, a H model is more appropriate if the change in vegetation structure was not so severe as to eliminate the population, e.g., the 75% retention treatment. Regardless of which assumption holds, the response of CLGA captures to a given habitat variable is consistent across models: negative association with HERB and SHRUB but positive association with CONIF, CWDVOL, SHANNON and TPH\*. This consistency strengthens the evidence for the general relationship.

For the less frequent species NEGI and PEKE, the link across models is harder to interpret. Consistently for both species, the number of captures increased with increasing HERB in the POIS and NB models, but the ZMMs distinguished between HERB effects on presence and HERB effects on number of captures. Other habitat variables such as SHANNON lacked significant correlation with number of PEKE captures in the POIS and NB models, but this variable was a significant predictor for p in the ZIP model and  $\lambda$  in the HPOIS model. The predicted coefficients associated with SHANNON in the ZIP model for PEKE are difficult to explain. The simulations described by Lam et al. (2010; Chapter 2) showed that ZMMs estimated highly imprecise and positively error in parameter estimates for distributions generated by combinations of  $\lambda$  and p that would be typical of rare species. This tendency might extend to the situation here in which parameters  $\lambda$  and p are conditioned on predictors in the GLM context. A carefully crafted simulation study with predictors will probably be required to fully understand the effect of covariates that result in a suite of conditional distributions that include those with an inflated number of zeros at one end

of covariate range, grading into those with few zeros and an expectation much larger than zero at the other end of the covariate range. Raphael (1984) and Carey and Johnson (1995) suggested that NEGI and PEKE were more closely associated with older forests, but these species have rarely been studied. Nonetheless, ZMMs seem capable of identifying statistically significant relationships between infrequent species and habitat variables. This class of statistical models therefore may prove useful in other biodiversity studies that aim to link habitat features to relative abundance of infrequent species, although some caution should be exercised until these models are more fully understood.

The above discussion focuses on comparing various specifications of ZMMs. However, the POIS and NB models can still often be valid models. For example, the abundance of a given species might not depend on latent structures such as those implied by ZMMs, but rather reflect unexplained heterogeneity such as that introduced by omitting key covariates (Minami et al. 2007), including interaction with other small mammal species. When the precise mechanism that produces overdispersion is unknown, the NB model would be an appropriate model (McCullagh and Nelder 1989). In reality, the data generating process is never known. Goodness-of-fit criteria might narrow the number of feasible models, but selection of the model should also based on as much theoretical understanding of the underlying ecological process as possible (Lord et al. 2005).

### **3.5.2 Generalized Linear Mixed Models (GLMMs)**

The structure of the GLMMs resulted in individual regression models fitted to the data for each unit. This process properly accounted for the nesting structure of the traps. Because units were blocked by design, the unit-level coefficients accounted for the blocking in a fashion similar to an ANOVA model for RBCD that assumed blocks were random (Kuehl 2000). However, treatment effects were excluded from this analysis under the assumption that the habitat structural variables more directly reflected the effects of treatments on residual stand structure.

The structure of the GLMMs explained some of the differences in results from the GLMs. First, the GLMM properly accounted for the degrees of freedom during estimation. Coefficients were estimated from the 32–37 traps within treatment units instead of the 787–1181 traps that were pooled in the GLMs. This recognition of hierarchical sampling explained the fact that 95% credible intervals were wider than 95% confidence intervals; i.e., the smaller sample size resulted in less precise estimates. If vegetation structure had been measured at all grid points, sample size would have doubled to 63–64 and precision would have been improved. This expanded description of vegetation structure would have allowed inclusion of small mammal data at all grid points, and thus provided more information to detect trends within experimental units. Although these additional measurements may not have been necessary for detecting treatment effects on the level of experimental units, they would have the added benefit of providing stronger insight into mechanisms driving responses at within-unit level.

By fitting individual regressions for each unit, the GLMM structure appropriately captured variation in the coefficients among units. Data pooling in the GLMs could have masked the within-unit relationships between predictors and number of captures. As one example, the number of captures was positively correlated with a predictor when ignoring the units (data pooling), but the relationship became negative within units after block and unit effects were accounted for. In other words, the trend was positive across units but negative within units. These differences among levels in the hierarchy probably explain the difference between the average of coefficient estimates from the GLMMs and the estimates from the GLMs. A related benefit was the confidence gained by observing consistency in the association of a predictor with number of captures among units. Parameter estimates were expected to be generally consistent in direction but different in value among units because the levels of a predictor only reflect treatment effects and not the effect of other covariates distinguishing the individual experimental units. This was the case for most predictors as exemplified by SHANNON for CLGA (Figure 3.5B). Consistency in a covariate effect across units provides strong evidence for the general effect on number of captures. Conversely, inconsistency in a covariate effect, e.g., HERB for CLGA (Figure 3.5A), implies interaction between that covariate and unobserved characteristics of the units receiving the same treatment.

The use of BHM in the DEMO study has several advantages. One of them is improved precision in the coefficient estimates for each unit and facilitation of inference on the collection of units (Congdon 2006). This comes from the structure of BHM whereby parameters are drawn from the same probability density conditioned on hyperparameters. For example, the intercept  $b_{1ij}$  (eqn. 3.19) from each unit was drawn from a multivariate normal distribution with mean  $\mu_I$  and a complex variancecovariance matrix that described the relationship of  $b_{1ij}$  to other coefficients. Conceptually, information from all units collectively determines the probability density for the intercept parameter. The estimate for unit *ij* is then a form of weighted average between information from that unit and the pooled information. This structure is particularly efficient when units with fewer observations are included (Gelman and Hill 2007), although the within-unit sample size of 32–37 in DEMO did not differ excessively.

A related advantage of BHM is that it provides an estimate for species and units in which no captures were observed. The number of units with no PEMA captures was 6, for no CLGA captures was 2, for no NEGI captures was 17, and for no PEKE captures was 6. In classic regression, two approaches are typically taken. In the first, one common line is fitted to all sampling units across all experimental units, but the resulting estimates have biased standard errors due to the autocorrelation within experimental units. In the second, the line is fitted to averages for the experimental units and the resulting relationship is only applicable for estimating mean responses of individual units. In the BHM context, unique regression estimates are obtained for all units, even those with only zero observations in all sampling units; however, the resulting precision may be low. Units with no captures do contain relevant information, and they are shared to predict the hyperparameters of the probability densities. An excellent example was the CWDVOL, where some of the estimated negative effects of this variable came from units with no captures for the target species. In essence, no information is "lost" and inference is appropriate for the whole study.

The parameters  $\sigma_h$  and  $\tau_h$  measure variability in a coefficient among units and blocks, so increasing sample size does not affect this inherent variability (Gelman and Hill 2007). A high value of  $\sigma_h$  suggests that the marginal effect of a given habitat variable on number of captures is smaller in some units than in others. This behavior may result from units having a different range in the habitat variable, from an interaction between this variable and others, or from the effects of unobserved characteristics of units such as fine woody debris (Manning and Edge 2004), unit history or geographical location (Carey and Johnson 1995). Blocking of experimental units is motivated by expected heterogeneity among blocks, so  $\tau_h$ , as a source of variation in unit-level coefficients is expected. From a technical perspective,  $\sigma_h$  and  $\tau_h$ describe the variation of the population for which  $b_{hij}$  and  $d_{hi}$  are estimated (Gelman and Hill 2007). If the blocks and units are random draws from a given population,  $\sigma_h$ and  $\tau_h$  should represent the population variability and be applicable to new units not currently represented in the study (Gelman and Hill 2007). Hence the posterior  $\sigma_h$  and  $\tau_h$  are generally higher than what are suggested by the observed data.

As also noted by Congdon (2005), the overPOIS-GLMM had the best goodness-of-fit, especially for species having large overdispersion. The trap-level random error  $\varepsilon_{ijk}$  could be referred to as the unstructured random effect because it was intended to capture undefined variability in the data. This variability may come from omitted covariates, spatial and temporal autocorrelation, or even perhaps different data generating processes such as zero-inflation suggested by ZI models. Thus, the  $\varepsilon_{iik}$ could potentially be structured to account for spatial autocorrelation between traps in a unit with methods described by Schabenberger and Gotway (2005). However, Mantel and Moran's *I* tests for spatial autocorrelation in each unit indicated lack of spatial autocorrelation, with only one or two units showing weak evidence of autocorrelation. Hence, to avoid further complexity and overparameterization of the overPOIS-GLMM, direct modeling of spatial autocorrelation was not pursued. The highest overdispersion indicated by  $\sigma_{\varepsilon}^2$  occurred for NEGI captures, probably because this species was not observed on approximately 50% of the units. Parameter  $\varepsilon_{ijk}$  may have accounted for the large group of traps that failed to capture any individual, or the effects of unobserved covariates. Including the trap-level random error in the model for the infrequent species also caused the posterior intercept to differ significantly from that of the other GLMMs, but had little effect on coefficients associated with habitat variables. This result coupled with a better DIC suggested that the overPOIS-GLMM may be a more suitable model for the infrequent species.

Lastly, models documented in the literature differ between small mammal species mainly due to differences in frequency of captures and the statistical methods applied. The abundant species, PEMA and CLGA, are well studied in their natural habitats and their responses to silvicultural treatments are relatively well understood. A review of the literature indicated that the habitat associations of NEGI and PEKE have seldom been studied; NEGI was captured in eight studies but was only analyzed in four, and PEKE was captured in four studies but analyzed in only two. In addition to accounting for the different sources of variation imposed by the sampling and experimental design, the GLMMs accommodated the relative rarity of these species and identified some tentatively important habitat variables. Thus, these methods are promising alternatives for modeling habitat relationships in species that are observed over a wide range of capture frequency.

### **3.5.3 Small Mammal Responses to Habitat Structures**

Habitat variables CONIF, CWDVOL and TPH\* were negatively correlated with PEMA captures for the final overPOIS-GLMM. In a companion study, Gitzen et al. (2007) found that the capture rates of PEMA increased as retention level decreased; a result consistent with our analysis. The relationship between CWDVOL and number of PEMA captures, however, has been inconsistent among different studies; e.g., Craig et al. (2006) concluded that down wood was not a critical habitat factor, but Manning and Edge (2004) suggested otherwise. Nonetheless, our study suggested that CWDVOL might be an important predictor for PEMA abundance. The final model suggested that HERB and CWDVOL were negatively correlated with CLGA captures but that SHANNON was positively correlated. Gitzen et al. (2007) concluded that the number of CLGA captures was independent of the variable-retention treatments; partially supporting our analysis in that TPH\* was not a significant predictor for CLGA abundance. However, their analysis focused on unitlevel responses to unit-level attributes, in contrast to our analysis that addressed within-unit responses and to within-unit covariates. The contrasting results therefore suggest that CLGA responses differ between these two spatial scales. As noted, CLGA is considered more closely associated with old-growth forest conditions, which may be expressed in higher overstory vertical complexity. Sullivan et al. (2005) suggested that stumps, rotting logs and exposed roots could help provide a moist micro-climate that was preferable to CLGA, but DEMO results suggested that a similar effect may not necessarily be imposed by increasing CWDVOL.

Gitzen et al. (2007) found that the closed-canopy species NEGI showed inconsistent or no response to the variable-retention treatments. Suzuki and Hayes (2003) also found that thinning treatments did not affect the abundance of this species. Our analysis supported these findings in that stand density was not a significant predictor for NEGI captures. The final model described a negative correlation between SHRUB and CWDVOL and number of NEGI captures, with large overdispersion. Contrary to our study, Carey and Johnson (1995) predicted a positive stand-level relationship between CWDVOL and the population estimates for NEGI. Their small mammal trapping design was a  $10 \times 10$  grid with 40-m spacing, which was comparable to the DEMO study. However, the regression was carried out with stand-level mean habitat variables. Thus, the discrepancy between studies could be due to the different spatial scales in the analysis, whereby the relationship might be negative within-stand but positive across stands. Wilson and Carey (2000) suggested that this species was more directly related to the soil-litter interface than to understory vegetation, suggesting that the observed large overdispersion may have been due to variability from omitted covariates such as litter layer depth or soil conditions.

The final model for PEKE suggested that the number of captures was negatively correlated with HERB and CWDVOL. On the contrary, West (1991) found that PEKE captures on the stand-level were positively correlated with the number of large coniferous trees and the percent cover of well-decayed logs. The trap design used in his study was a  $6 \times 6$  grid with 15-m spacing in each of 45 sites on the western slopes of the southern Washington Cascade Range (West 1991). Because comparative studies were few for this species, Wilson and Carey (2000) acknowledged that the habitat determinants of PEKE abundance remained largely unknown. However, the more flexible modeling options explored in this study have the potential to advance our understanding of this relatively uncommon species in the future studies.

# **3.6 CONCLUSIONS**

Inferences are commonly made on models identified as the best of many alternatives, based on some type of goodness-of-fit. This approach is appropriate for a series of nested models, but is insufficient for choosing among models based on

different assumptions about the underlying data generating process (dgp) and distributional properties. Lam et al. (2010; Chapter 2) showed that goodness-of-fit criteria such as information criteria measured model flexibility for fitting data but did not allow definitive identification of the underlying dgp. Information criteria are also helpful for variable selection. A potentially useful approach is to first understand the ecological implications of zero inflation and interpreting the results from information provided by multiple models rather than selecting one model as the best. This process is actually common when working with nonlinear or mechanistic models. Nonlinear models have a relatively strong foundation in biological theory and principle, and each parameter or combinations of parameters have direct interpretation with regard to attributes such as the asymptote or inflection point of a logistic growth model (Schabenberger and Pierce 2001). This type of model requires identification of links between model assumptions, model behavior, and the processes underlying the phenomenon of interest. Lord et al. (2005, 2007) pointed out that without insights into underlying dgp, the resulting improper choice of a model could lead to counterintuitive results and inappropriate inferences.

The use of GLM advanced our understanding of treatment effects by conditioning the number of small mammal captures within a treatment unit on habitat variables, assuming that variation in these habitat variables were at least partly attributable to variable-retention treatments. The GLMMs refined this understanding by recognizing the relationship between captures and habitat variables at several levels, including individual treatment units and blocks. A main consequence of not accounting for the nested subsampling and random block effects in the GLMs was bias in variance of estimates and improper inference on statistical correlations. This consequence could very well extend to the ZMMs; thus, the GLMM inferences were considered more reliable, particularly where they deviated from those of the ZMMs. The development of zero-modified random effects models has started to appear due to their many potential applications (Hall and Zhang 2004). For example, Hall (2000) and Yau and Lee (2001) extended zero-inflated models to clustered data, and Min and Agresti (2005) developed a hurdle random effects model for repeated-measure count data. However, currently available specifications of zero-modified mixed effects models are not yet sufficiently intricate to properly account for nesting and blocking structures in the DEMO study. For example, Yau and Lee (2001) only allowed random effects on the intercept term. Specification of a zero-modified mixed effects model in a Bayesian framework is considerably more complex, but may become a future option as the theory and software for these models are more fully developed.

Table 3.1 Estimated coefficients associated with each predictor (and corresponding standard errors) from the ZIP, ZINB, HPOIS and HNB models for *Peromyscus maniculatus* (PEMA), one of the two most frequently captured small mammal species in the DEMO study in 1999. The estimates from count and truncated-count components (eqn. 3.7) are listed in the section labeled COUNT and the estimates from zero-inflation (eqn. 3.8) and hurdle (eqn. 3.9) components are listed in the section labeled ZI/H. Estimated overdispersion  $\hat{\theta}$  and AICc are listed last, and statistically significant estimates are shown in bold ( $\alpha = 0.05$ ).

Dradiators		COU	NT			ZI/	Ή	
Fieulciois	ZIP	ZINB	HPOIS	HNB	ZIP	ZINB	HPOIS	HNB
Intercept	-2.975	-3.805	-2.976	-3.391	0.198	-457.21	-0.635	-0.635
	(0.179)	(0.165)	(0.171)	(0.365)	(0.273)	(669.58)	(0.174)	(0.174)
HERB	-0.012	-0.006	-0.011	-0.011	-0.011	1.463	-0.001	-0.001
	(0.008)	(0.005)	(0.007)	(0.009)	(0.012)	(2.218)	(0.006)	(0.006)
CONIF	-0.047	-0.061	-0.068	-0.074	0.017	-17.89	-0.053	-0.053
	(0.038)	(0.022)	(0.050)	(0.056)	(0.047)	(27.03)	(0.022)	(0.022)
SHRUB	0.008	0.009	0.007	0.009	0.010	6.721	-0.002	-0.002
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(9.860)	(0.004)	(0.004)
CWDVOL	-0.0006	-0.0012	-0.0006	-0.0008	0.0010	0.0866	-0.0013	-0.0013
	(0.0006)	(0.0005)	(0.0006)	(0.0008)	(0.0008)	(0.1274)	(0.0005)	(0.0005)
SHANNON	-0.026	-0.087	-0.020	-0.024	0.083	-61.53	-0.083	-0.083
	(0.059)	(0.055)	(0.059)	(0.077)	(0.086)	(90.31)	(0.057)	(0.057)
TPH*	-0.060	-0.142	-0.052	-0.053	0.132	46.74	-0.165	-0.165
	(0.063)	(0.042)	(0.065)	(0.076)	(0.074)	(68.38)	(0.045)	(0.045)
$\hat{ heta}$	_	0.427	_	1.327				
AICc	1452.0	1431.8	1451.8	1446.2				

Table 3.2 Estimated coefficients associated with each predictor (and corresponding standard errors) from the ZIP, ZINB, HPOIS and HNB models for *Clethrionomys gapperi* (CLGA), one of the two most frequently captured small mammal species in the DEMO study in 1999. The estimates from count and truncated-count components (eqn. 3.7) are listed in the section labeled COUNT and the estimates from zero-inflation (eqn. 3.8) and hurdle (eqn. 3.9) components are listed in the section labeled ZI/H. Estimated overdispersion  $\hat{\theta}$  and AICc are listed last, and statistically significant estimates are shown in bold ( $\alpha = 0.05$ ).

Dradiatora		COU	NT			ZI/	Ή	
Fieulciois	ZIP	ZINB	HPOIS	HNB	ZIP	ZINB	HPOIS	HNB
Intercept	-3.696	-4.132	-3.625	-3.938	1.510	0.719	-2.329	-2.329
	(0.408)	(0.492)	(0.420)	(0.539)	(0.515)	(0.718)	(0.328)	(0.328)
HERB	0.001	0.001	-0.002	-0.004	0.037	0.059	-0.025	-0.025
	(0.005)	(0.007)	(0.006)	(0.008)	(0.010)	(0.017)	(0.007)	(0.007)
CONIF	0.005	0.023	0.002	0.005	-0.144	-0.164	0.083	0.083
	(0.009)	(0.011)	(0.008)	(0.013)	(0.042)	(0.084)	(0.019)	(0.019)
SHRUB	-0.014	-0.011	-0.016	-0.017	-0.003	-0.002	-0.009	-0.009
	(0.005)	(0.005)	(0.005)	(0.006)	(0.008)	(0.011)	(0.004)	(0.004)
CWDVOL	0.0004	0.0005	0.0003	0.0003	0.0001	0.0006	0.0004	0.0004
	(0.0002)	(0.0003)	(0.0002)	(0.0004)	(0.0005)	(0.0010)	(0.0004)	(0.0004)
SHANNON	0.157	0.165	0.156	0.159	-0.355	-0.347	0.435	0.435
	(0.111)	(0.132)	(0.114)	(0.139)	(0.146)	(0.196)	(0.095)	(0.095)
TPH*	0.019	0.010	0.032	0.034	-0.080	-0.326	0.057	0.057
	(0.014)	(0.016)	(0.014)	(0.020)	(0.045)	(0.156)	(0.023)	(0.023)
$\hat{ heta}$	—	1.570	_	1.661				
AICc	1283.5	1265.2	1287.3	1276.6				

Table 3.3 Estimated coefficients associated with each predictor (and corresponding standard errors) from the ZIP, ZINB, HPOIS and HNB models for *Neurotrichus gibsii* (NEGI), one of the two most infrequently captured small mammal species in the DEMO study in 1999. The estimates from count and truncated-count components (eqn. 3.7) are listed in the section labeled COUNT and the estimates from zero-inflation (eqn. 3.8) and hurdle (eqn. 3.9) components are listed in the section labeled ZI/H. Estimated overdispersion  $\hat{\theta}$  and AICc are listed last, and statistically significant estimates are shown in bold ( $\alpha = 0.05$ ).

Dradiatora		COU	NT			ZI/	Ή	
Fieulciois	ZIP	ZINB	HPOIS	HNB	ZIP	ZINB	HPOIS	HNB
Intercept	-2.343	-3.078	-2.180	-5.398	4.409	3.735	-4.356	-4.356
	(0.341)	(0.748)	(0.347)	(22.818)	(0.521)	(0.744)	(0.475)	(0.475)
HERB	0.003	0.008	-0.003	-0.002	-0.037	-0.036	0.035	0.035
	(0.008)	(0.011)	(0.009)	(0.015)	(0.010)	(0.013)	(0.008)	(0.008)
CONIF	0.035	0.052	0.018	0.027	0.039	0.060	-0.001	-0.001
	(0.014)	(0.027)	(0.016)	(0.040)	(0.023)	(0.033)	(0.014)	(0.014)
SHRUB	-0.011	-0.011	-0.008	-0.011	-0.002	-0.002	-0.007	-0.007
	(0.010)	(0.012)	(0.011)	(0.017)	(0.011)	(0.013)	(0.006)	(0.006)
CWDVOL	-0.0004	-0.0003	-0.0010	-0.001	-0.0010	-0.0009	0.0006	0.0006
	(0.0007)	(0.0009)	(0.0009)	(0.002)	(0.0009)	(0.0011)	(0.0006)	(0.0006)
SHANNON	-0.223	-0.269	-0.243	-0.304	-0.033	-0.032	0.015	0.015
	(0.124)	(0.221)	(0.122)	(0.296)	(0.165)	(0.218)	(0.135)	(0.135)
TPH*	-0.085	-0.096	-0.026	-0.042	-0.236	-0.329	0.093	0.093
	(0.038)	(0.049)	(0.044)	(0.098)	(0.079)	(0.113)	(0.029)	(0.029)
$\hat{ heta}$	-	0.646	_	0.022				
AICc	568.8	558.4	571.4	562.0				

Table 3.4 Estimated coefficients associated with each predictor (and corresponding standard errors) from the ZIP, ZINB, HPOIS and HNB models for *Peromyscus keeni* (PEKE), one of the two most infrequently captured small mammal species in the DEMO study in 1999. The estimates from count and truncated-count components (eqn. 3.7) are listed in the section labeled COUNT and the estimates from zero-inflation (eqn. 3.8) and hurdle (eqn. 3.9) components are listed in the section labeled ZI/H. Estimated overdispersion  $\hat{\theta}$  and AICc are listed last, and statistically significant estimates are shown in bold ( $\alpha = 0.05$ ).

Prodictors		COU	NT			ZI	Ή	
Fieulciois	ZIP	ZINB	HPOIS	HNB	ZIP	ZINB	HPOIS	HNB
Intercept	-3.967	NA	-3.482	-11.86	1.415	NA	-3.190	-3.190
	(0.574)		(0.524)	(97.55)	(0.934)		(0.373)	(0.373)
HERB	0.050	NA	0.033	0.052	0.056	NA	0.022	0.022
	(0.007)		(0.013)	(0.023)	(0.033)		(0.007)	(0.007)
CONIF	-0.016	NA	0.061	0.139	-0.186	NA	0.041	0.041
	(0.031)		(0.043)	(0.100)	(0.098)		(0.025)	(0.025)
SHRUB	0.015	NA	0.006	0.021	0.031	NA	0.002	0.002
	(0.006)		(0.008)	(0.014)	(0.018)		(0.005)	(0.005)
CWDVOL	0.0012	NA	-0.0017	-0.0031	0.0012	NA	0.0013	0.0013
	(0.0006)		(0.0014)	(0.0023)	(0.0017)		(0.0005)	(0.0005)
SHANNON	-0.983	NA	-0.528	-0.843	-1.775	NA	0.015	0.015
	(0.159)		(0.295)	(0.507)	(0.744)		(0.110)	(0.110)
TPH*	0.115	NA	-0.041	-0.127	0.436	NA	-0.058	-0.058
	(0.067)		(0.097)	(0.196)	(0.215)		(0.042)	(0.042)
$\hat{ heta}$	—	NA	_	0.0002				
AICc	557.7	NA	564.5	562.0				

Table 3.5 Comparison of GLMs and GLMMs for *Peromyscus maniculatus* (PEMA) in the DEMO study in 1999. Coefficient estimates and their 95% confidence intervals were reported for the GLMs. Average posterior coefficient estimates  $\overline{b}_h$  (eqn. 3.41) and their 95% credible intervals were reported for the GLMMs. Estimated and posterior overdispersion parameters  $\theta$  were reported for the NB GLM and -GLMM, respectively, in addition to the variance of the random error  $\sigma_{\varepsilon}^2$  (eqn. 3.37) for the overPOIS-GLMM, and AICc or DIC for all five models. Statistically significant estimates are shown in bold.

Covariatas	РО	IS	N	overPOIS	
Covariates	GLM	GLMM	GLM	GLMM	GLMM
Intercept	-3.751	-3.457	-3.782	-3.408	-3.741
	(-3.976, -3.527)	(-3.923, -3.007)	(-4.114, -3.450)	(-3.935, -2.906)	(-4.302, -3.200)
HERB	-0.004	-0.042	-0.005	-0.044	-0.043
	(-0.012, 0.004)	(-0.104, 0.013)	(-0.015, 0.006)	(-0.113, 0.014)	(-0.108, 0.015)
CONIF	-0.067	-0.309	-0.058	-0.299	-0.301
	(-0.107, -0.028)	(-0.526, -0.131)	(-0.100, -0.016)	(-0.534, -0.113)	(-0.529, -0.117)
SHRUB	0.001	-0.008	0.002	-0.006	-0.010
	(-0.004, 0.006)	(-0.095, 0.073)	(-0.005, 0.009)	(-0.095, 0.081)	(-0.100, 0.076)
CWDVOL	-0.0014	-0.046	-0.0014	-0.047	-0.046
	(-0.002, -0.0007)	(-0.075, -0.025)	(-0.002, -0.0004)	(-0.075, -0.025)	(-0.075, -0.025)
SHANNON	-0.063	-0.208	-0.066	-0.194	-0.195
	(-0.140, 0.013)	(-0.427, 0.003)	(-0.174, 0.042)	(-0.433, 0.034)	(-0.437, 0.046)
TPH*	-0.164	-0.148	-0.156	-0.153	-0.158
	(-0.235, -0.093)	(-0.370, 0.057)	(-0.240, -0.077)	(-0.392, 0.076)	(-0.396, 0.064)
heta	_	_	0.393	1.422	-
$\sigma^2_arepsilon$	_	_	-	-	0.866
AICc	1573.3	_	1441.3	-	-
DIC	_	1276.1	_	1310.8	1210.8

Table 3.6 Comparison of GLMs and GLMMs for *Clethrionomys gapperi* (CLGA) in the DEMO study in 1999. Coefficient estimates and their 95% confidence intervals were reported for the GLMs. Average posterior coefficient estimates  $\overline{b}_h$  (eqn. 3.41) and their 95% credible intervals were reported for the GLMMs. Estimated and posterior overdispersion parameters  $\theta$  were reported for the NB GLM and -GLMM, respectively, in addition to the variance of the random error  $\sigma_{\varepsilon}^2$  (eqn. 3.37) for the overPOIS-GLMM, and AICc or DIC for all five models. Statistically significant estimates are shown in bold.

Covariatas	PO	IS	N	overPOIS	
Covariates	GLM	GLMM	GLM	GLMM	GLMM
Intercept	-5.394	-5.891	-5.320	-5.769	-6.233
	(-5.874, -4.914)	(-6.809, -5.502)	(-5.866, -4.775)	(-6.815, -4.824)	(-7.273, -5.295)
HERB	-0.014	-0.071	-0.022	-0.079	-0.075
	(-0.022, -0.006)	(-0.122, -0.026)	(-0.033, -0.011)	(-0.134, -0.032)	(-0.130, -0.028)
CONIF	0.022	-0.036	0.044	-0.021	-0.024
	(0.009, 0.034)	(-0.210, 0.116)	(0.021, 0.067)	(-0.202, 0.140)	(-0.203, 0.141)
SHRUB	-0.014	-0.019	-0.013	-0.015	-0.018
	(-0.020, -0.008)	(-0.073, 0.027)	(-0.020, -0.005)	(-0.070, 0.034)	(-0.073, 0.030)
CWDVOL	0.0005	-0.031	0.0004	-0.032	-0.032
	(0.0001, 0.0008)	(-0.061, -0.011)	(-0.0002, 0.001)	(-0.061, -0.012)	(-0.060, -0.012)
SHANNON	0.440	0.445	0.419	0.434	0.452
	(0.308, 0.572)	(0.132, 0.762)	(0.263, 0.574)	(0.105, 0.795)	(0.124, 0.806)
TPH*	0.036	0.160	0.046	0.189	0.180
	(0.016, 0.056)	(-0.049, 0.376)	(0.014, 0.079)	(-0.043, 0.429)	(-0.051, 0.410)
heta	_	_	0.687	1.328	-
$\sigma^2_arepsilon$	_	_	-	-	0.895
AICc	1406.9	_	1288.2	-	-
DIC	_	1194.8	_	1215.6	1105.3

Table 3.7 Comparison of GLMs and GLMMs for *Neurotrichus gibsii* (NEGI) in the DEMO study in 1999. Coefficient estimates and their 95% confidence intervals were reported for the GLMs. Average posterior coefficient estimates  $\overline{b}_h$  (eqn. 3.41) and their 95% credible intervals were reported for the GLMMs. Estimated and posterior overdispersion parameters  $\theta$  were reported for the NB GLM and -GLMM, respectively, in addition to the variance of the random error  $\sigma_{\varepsilon}^2$  (eqn. 3.37) for the overPOIS-GLMM, and AICc or DIC for all five models. Statistically significant estimates are shown in bold.

Covariatas	PO	IS	N	overPOIS	
Covariates	GLM	GLMM	GLM	GLMM	GLMM
Intercept	-6.514	-6.142	-6.452	-6.221	-13.49
	(-7.097, -5.931)	(-7.940, -4.805)	(-7.239, -5.665)	(-8.411, -3.959)	(-19.52, -8.991)
HERB	0.028	-0.242	0.029	-0.177	-0.184
	(0.017, 0.039)	(-0.477, -0.078)	(0.010, 0.048)	(-0.471, 0.049)	(-0.461, 0.038)
CONIF	0.005	-0.136	0.012	-0.050	-0.116
	(-0.015, 0.024)	(-0.568, 0.220)	(-0.024, 0.048)	(-0.555, 0.416)	(-0.662, 0.356)
SHRUB	-0.009	-0.201	-0.010	-0.247	-0.240
	(-0.019, 0.002)	(-0.530, 0.063)	(-0.025, 0.005)	(-0.624, 0.069)	(-0.637, 0.087)
CWDVOL	0.0001	-0.204	0.00003	-0.212	-0.212
	(-0.001, 0.001)	(-0.334, -0.114)	(-0.002, 0.002)	(-0.351, -0.115)	(-0.347, -0.119)
SHANNON	-0.110	-0.005	-0.176	-0.216	0.174
	(-0.283, 0.063)	(-0.568, 0.627)	(-0.417, 0.065)	(-1.129, 0.672)	(-0.918, 1.319)
TPH*	0.076	-0.105	0.107	0.113	-0.011
	(0.034, 0.118)	(-0.604, 0.389)	(0.032, 0.182)	(-0.639, 0.894)	(-0.780, 0.740)
heta	_	_	0.080	0.082	-
$\sigma^2_{arepsilon}$	_	_	-	-	4.945
AICc	694.9	_	570.3	-	-
DIC	_	523.7	_	569.2	309.7

Table 3.8 Comparison of GLMs and GLMMs for *Peromyscus keeni* (PEKE) in the DEMO study in 1999. Coefficient estimates and their 95% confidence intervals were reported for the GLMs. Average posterior coefficient estimates  $\overline{b}_h$  (eqn. 3.41) and their 95% credible intervals were reported for the GLMMs. Estimated and posterior overdispersion parameters  $\theta$  were reported for the NB GLM and -GLMM, respectively, in addition to the variance of the random error  $\sigma_{\varepsilon}^2$  (eqn. 3.37) for the overPOIS-GLMM, and AICc or DIC for all five models. Statistically significant estimates are shown in bold.

Covariates	PO	IS	N	overPOIS	
Covariates	GLM	GLMM	GLM	GLMM	GLMM
Intercept	-6.310	-5.158	-6.151	-4.725	-7.038
	(-6.911, -5.705)	(-6.374, -4.165)	(-6.832, -5.470)	(-6.120, -3.295)	(-9.326, -5.291)
HERB	0.029	-0.144	0.025	-0.162	-0.166
	(0.018, 0.040)	(-0.282, -0.28)	(0.010, 0.039)	(-0.319, -0.022)	(-0.332, -0.028)
CONIF	0.053	-0.140	0.050	-0.058	-0.092
	(0.015, 0.090)	(-0.449, 0.130)	(0.001, 0.098)	(-0.415, 0.269)	(-0.456, 0.219)
SHRUB	0.001	-0.130	0.003	-0.141	-0.142
	(-0.007, 0.009)	(-0.297, 0.006)	(-0.007, 0.013)	(-0.342, 0.026)	(-0.337, 0.022)
CWDVOL	0.0009	-0.081	0.0008	-0.084	-0.084
	(0.0001, 0.002)	(-0.144, -0.048)	(-0.0002, 0.002)	(-0.147, -0.039)	(-0.148, -0.039)
SHANNON	-0.053	-0.077	-0.077	-0.035	0.029
	(-0.231, 0.126)	(-0.475, 0.315)	(-0.282, 0.128)	(-0.588, 0.511)	(-0.562, 0.621)
TPH*	-0.061	0.006	-0.057	0.048	0.029
	(-0.134, 0.012)	(-0.301, 0.305)	(-0.141, 0.027)	(-0.391, 0.491)	(-0.398, 0.457)
heta	_	_	0.320	0.199	-
$\sigma^2_{arepsilon}$	_	_	-	-	2.484
AICc	586.7	_	559.8	_	_
DIC	_	529.7	_	597.2	421.2

Table 3.9 Relationship between habitat structures and *Peromyscus maniculatus* (PEMA), *Clethrionomys gapperi* (CLGA), *Neurotrichus gibsii* (NEGI) and *Peromyscus keeni* (PEKE) abundance in the DEMO study in 1999. Average posterior coefficient estimates  $\overline{b}_h$  (eqn. 3.41) and their 95% credible intervals were reported for the overPOIS-GLMM and the selected significant habitat variables, in addition to the variance of the random error  $\sigma_{\varepsilon}^2$  (eqn. 3.37) and DIC.

Covariates	PEMA	CLGA	NEGI	PEKE
Intercept	-4.393	-6.435	-7.545	-5.821
-	(-4.804, -3.985)	(-7.600, -5.440)	(-9.231, -6.241)	(-6.909, -4.844)
HERB		-0.045		-0.146
	—	(-0.083, -0.011)	—	(-0.261, -0.053)
CONIF	-0.298			
	(-0.542, -0.115)	—	—	
SHRUB			-0.313	
			(-0.633, -0.086)	
CWDVOL	-0.045	-0.028	-0.204	-0.079
	(-0.074, -0.024)	(-0.054, -0.010)	(-0.329, -0.117)	(-0.140, -0.037)
SHANNON		0.645		
	—	(0.375, 0.991)	—	—
TPH*	-0.302			
	(-0.495, -0.145)	—	—	—
$\sigma^2_{arepsilon}$	0.818	0.821	2.448	1.429
DIC	1223.1	1114.6	379.3	470.8



Figure 3.1 Locations of the six DEMO blocks in western Oregon and Washington; DP: Dog Prairie, WF: Watson Falls, LW: Little White Salmon, PH: Paradise Hills, BU: Butte, and CF: Capitol Forest. The six harvest treatment designs are shown above the map. Solid gray represents uncut forest (100% and 75% treatments) and uncut 1-ha forest aggregates (15%A and 40%A treatments); black dots represented dispersed residual trees (15%D and 40%D treatments).



Figure 3.2 Relative frequency histogram for captures of: (A) *Peromyscus maniculatus* (PEMA); (b) *Clethrionomys gapperi* (CLGA); (C) *Neurotrichus gibsii* (NEGI); and *Peromyscus keeni* (PEKE).



Figure 3.3 Relative frequency distribution for observed and predicted counts from the POIS, NB, ZIP, ZINB, HPOIS and HNB models for: (A) *Peromyscus maniculatus* (PEMA); (b) *Clethrionomys gapperi* (CLGA); (C) *Neurotrichus gibsii* (NEGI); and *Peromyscus keeni* (PEKE).



Figure 3.4 Posterior coefficient estimates and 95% credible intervals of CWDVOL  $(\tilde{b}_{5ij}, \text{eqn. 3.11})$  from POIS-GLMM for all the units and the small mammal species in the study. The species were (A) *Peromyscus maniculatus* (PEMA); (b) *Clethrionomys gapperi* (CLGA); (C) *Neurotrichus gibsii* (NEGI); and *Peromyscus keeni* (PEKE). Sample size was 36 units for PEMA and NEGI and 24 units for CLGA and PEKE.



Figure 3.5 Posterior coefficient estimates and 95% credible intervals of (A) HERB  $(\tilde{b}_{2ij})$  and (B) SHANNON  $(\tilde{b}_{6ij})$  from POIS-GLMM for all the units for the species *Clethrionomys gapperi* (CLGA).



Figure 3.6 Posterior standard deviation of random error  $\tilde{\sigma}_h$  for each of the 7 estimated coefficients (eqn. 3.21) and 95% posterior intervals for the POIS-GLMM (circle and solid line), NB-GLMM (square and dotted line) and overPOIS-GLMM (triangle and dashed line) for each species: (A) *Peromyscus maniculatus* (PEMA); (b) *Clethrionomys gapperi* (CLGA); (C) *Neurotrichus gibsii* (NEGI); and *Peromyscus keeni* (PEKE).



Figure 3.7 Posterior standard deviation of random block effect  $\tilde{\tau}_h$  for each of the 7 estimated coefficients (eqn. 3.23) and 95% posterior intervals for the POIS-GLMM (circle and solid line), NB-GLMM (square and dotted line) and overPOIS-GLMM (triangle and dashed line) for each species: (A) *Peromyscus maniculatus* (PEMA); (b) *Clethrionomys gapperi* (CLGA); (C) *Neurotrichus gibsii* (NEGI); and *Peromyscus keeni* (PEKE).



Figure 3.8 Posterior random block effect estimates  $\tilde{d}_{hi}$  for all 7 estimated coefficients (eqn. 3.16) and 95% posterior intervals across blocks from POIS-GLMM for each small mammal species: (A) *Peromyscus maniculatus* (PEMA); (b) *Clethrionomys gapperi* (CLGA); (C) *Neurotrichus gibsii* (NEGI); and *Peromyscus keeni* (PEKE). Analyses were based on six blocks for PEMA and NEGI and four blocks for CLGA and PEKE.
4 CHAPTER 4: STRUCTURAL EQUATION MODELING AND ITS APPLICATIONS TO OVERSTORY-UNDERSTORY RELATIONSHIP IN MATURE DOUGLAS-FIR (*Pseudotsuga menziesii* (Mirb.) Franco) FORESTS AND TO LATE-SERAL HERBACEOUS SPECIES IMMEDIATE RESPONSES TO VARIABLE-RETENTION HARVESTING IN THE PACIFIC NORTHWEST

Tzeng Yih Lam

Douglas A. Maguire

### **4.1 INTRODUCTION**

Douglas-fir (Pseudotsuga menziesii (Mirb.) Franco) forests in the Pacific Northwest have traditionally been regeneration harvested under even-age silvicultural systems, most commonly the clearcutting system; however, public concern over dramatic reductions in old-growth forests and potential declines in forest biodiversity have fueled renewed interest in the feasibility of other silvicultural systems. Shelterwood with reserves (Matthews 1989) and innovations such as variableretention harvesting (Franklin et al. 1997) have been proposed as means to meet diverse forest management objectives particularly on public lands. These methods share a common working hypothesis that the retained structures and associated heterogeneity in stand structure could in the short-term maintain taxa and ecological processes characteristic of mature forests and in the long-term accelerate their recovery (Lindenmayer and Franklin 2002, Maguire et al. 2007). In their review of silvicultural methods that retain old-growth attributes, Bauhus et al. (2009) and Rosenvald and Lõhmus (2008) suggested that both stand and landscape dynamics under these systems are in dire need of further research. It is also evident that routine statistical methods may not be sufficient for understanding the mechanisms linking stand structure to biodiversity and for testing working hypotheses about the interactions of various taxa and ecological processes.

Ecosystem processes involve complex interactions of many cause-and-effect relationships. Research on silvicultural control of biodiversity should employ

statistical methods that test working hypotheses about these complex interactions. The strength of causal inference depends largely on a study design, particularly the principle of random assignment of treatments to experimental units as advocated by Fisher (1926). This randomization of treatments in a designed experiment is the foundation for inferring causality (Shipley 2000). Because Analysis of Variance (ANOVA) models are closely connected to experimental design, the general implication is that ANOVA is an appropriate method for causal inference (Grace et al. 2009). Multivariate Analysis of Variance (MANOVA) is an extension of univariate ANOVA to multivariate responses. These models may be appropriate for assessing net effects of experimental treatments; however, they could fail to uncover any information about underlying processes or response mechanisms (Grace et al. 2009).

Data reduction by categorization (classification) and synthetic continuous variables (ordination) are very efficient for analyzing multivariate ecological data (McCune et al. 2002). Examples of these methods include cluster analysis and nonmetric multidimensional scaling, respectively. Classification and ordination are conventionally exploratory and descriptive because their capability for understanding functional links is limited (Grace 2006). Some general idea of causal mechanisms usually drives data collection, but these statistical methods cannot explicitly incorporate *a priori* and complex inter-relationships between variables (Grace 2006). In spite of this limitation, the methods are useful for discovering underlying trends in data and for supplementing the process of model building and hypothesis formulation.

Structural equation modeling (SEM) is an alternative method for testing ecological processes. SEM is a collection of procedures that test hypothesized relationships among observed variables (Grace 2008, Schumacker and Lomax 2004). Complex interactions are first translated into a network of directional paths linking variables, and are then evaluated against multivariate data (Bollen 1989). These paths postulate direct effects, indirect effects and spurious associations due to common causes; hence, SEM is often related to causal modeling (Kenny 1979). It is philosophically a confirmatory data analysis, but its application extends to testing alternative *a priori* models or model-building (Jöreskog 1993), and SEM can therefore be regarded as blending confirmatory and exploratory analyses (Kline 2005). The key to successful SEM rests on the competence of a researcher to posit initial cause-andeffect models drawing from accumulated knowledge, prior experiences and published results.

SEM is an attractive modeling tool for testing our understanding of complex ecological processes. Comprehensive assessments by Shipley (2000), Pugesek et al. (2003) and Grace (2006) have brought SEM into the context of natural systems. Laughlin and Abella (2007) and Laughlin et al. (2007) applied SEM to an observational study of abiotic and biotic factors influencing plant community composition and species richness in a ponderosa pine (*Pinus ponderosa* P. & C. Lawson) forest ecosystem. Youngblood et al. (2009) studied the effects of experimental thinning and prescribed burning treatments on mortality of ponderosa pine with ANOVA and SEM. The cascading effects of fire intensity, surface fire severity and bole charring on mortality attributed to bark beetles and wood borers were well represented by SEM, but not by the ANOVA model. These studies demonstrate the flexibility of SEM for accommodating various study designs. With increasing recognition of the benefits from embedding existing knowledge in data analysis and the limitation of common statistical methods for large-scale ecological experiments (Miao et al. 2009), SEM may offer a promising alternative.

Overstory-understory relationships have been studied for decades with early efforts focused on forage production (e.g., Mitchell and Bartling 1991, Uresk and Severson 1989). More recent interest addresses conservation of functional groups such as late-seral herbaceous species under alternative silvicultural systems such as variable-retention harvesting (e.g., Halpern et al. 2005). Overstory tree cover (McKenzie et al. 2000), tree density (Bailey and Tappeiner 1998), shrub cover (Berger and Puettmann 2000) and soil litter (Dzwonko and Gawroński 2002) are among the variables found strongly associated with the richness and abundance of the understory herb community. These variables serve as surrogates for ecological factors such as light attenuation, throughfall precipitation, and soil water availability. The predominant approach for analysis has been classical univariate models and exploratory analyses, but many researchers have underscored the complex and multivariate nature of these inter-relationships (Nemati and Goetz 1995). A SEM approach to the overstory-understory relationship could apply existing knowledge in proposing functional links and could thereby test and expand our understanding of relationships and mechanisms of response to silvicultural treatments.

Many natural science studies that apply SEM provide only a brief account of the methodology (e.g., Johnstone et al. 2009). In one exception, Grace (2008) summarized the general aspects of applying SEM in observational ecology studies, and illustrated the technique with two examples. In another, Grace and Bollen (2006) provided details on a SEM framework with emphasis on a specific issue related to composite variables. In a third exception, Shipley (2009) described generalized multilevel SEM. We found that the existing publications on the methodology were either too brief or too detailed to benefit those with limited knowledge of SEM but having interest in its application. Hence, the goal of this paper was to bridge general and specific treatments of the methodology in the natural sciences. Two specific objectives were to: (1) systematically present the concepts and framework for SEM, and (2) illustrate the application of SEM to overstory-understory relationships in an experimental test of responses to variable-retention harvesting of Douglas-fir (*Pseudotsuga menziesii*) forests in the Pacific Northwest region of the United States.

# **4.2 STRUCTURAL EQUATION MODELING (SEM)**

The two central components of SEM are the path model and the measurement model. The path model or path analysis was developed by the geneticist Sewall Wright (1921, 1934) to quantify specific cause-and-effect relationships between observed variables through computation of path coefficients. Exploratory factor analysis, as introduced by Charles Spearman (1904), is usually considered the origin of the measurement model. The measurement model quantifies linkages between hypothetical constructs and observed variables through computation of factor loadings. Because SEM can specify any combination of these two components, a unifying and flexible mathematical framework, LISREL (LInear Structural RELations), was developed by Karl Jöreskog (1973) and adopted for modern application of SEM (Grace 2008). Hayduk (1987) and Kelloway (1998) described the LISREL system in detail. The following summary of SEM is primarily drawn from Grace (2006), Kline (2005) and Schumacker and Lomax (2004).

### 4.2.1 Path Model

Specification of a path model involves hypothesized cause-and-effect relationships between observed variables. These relationships are usually based on theoretical considerations or evidence from prior studies. However, certain conditions must be met for a variable to be designated as a cause versus an effect. Kenny (1979) proposed three: time precedence, functional relationship and nonspuriousness. First, for variable A to cause variable B, A has to precede B in time, so time precedence implies an asymmetric relationship between the two. Second, A and B should be functionally related because there is no causal relationship if they are independent. Third, if the relationship between A and B is spurious due to a common cause, it will disappear once the common cause is identified and represented in the model.

Consider a hypothetical model for representing the response of forest floor small mammal abundance to stand structural characteristics (Figure 4.1). In this scenario, overstory tree density and coarse wood debris are *exogenous* observed variables whereas understory tree cover, herbaceous cover and abundance of small mammal species are *endogenous* variables. An exogenous variable is always considered only a cause; its causes are generally unknown or not of interest, or it is measured directly in the application at hand; therefore, it is not represented in a model. An endogenous variable is an effect, but it may also be a cause to other endogenous variables. A single arrowhead is a path that represents a direct effect of the causal variable on a response variable, whereas a double arrowhead represents an unanalyzed association. Unanalyzed associations are customarily specified between exogenous variables because no hypothesis is included to explain why they covary. In general not all causal processes acting on a system are specified, and the model merely represents a simplification (Laughlin et al. 2007).

Once the model is fitted to data, the path coefficient (direct effect) of each path is estimated and interpreted similarly to a regression coefficient. An indirect effect of any causal variable is estimated as the product of direct effects, and a total effect is the sum of all direct and indirect effects. In the example, tree density has only indirect effects on abundance, mediated by understory cover, herb cover, or both. Thus, one of the indirect effects is the product of two path coefficients, one representing the effect of tree density on herb cover and the other the effect of herb cover on small mammal abundance. Coarse wood debris has both a direct and an indirect effect on abundance, the latter through herb cover as the mediating causal variable. Hence, its total effect on abundance is the sum of the indirect and direct effects. A cautionary note is warranted for interpretation of direct and indirect effects. As with multiple regression, an effect is interpreted as the change induced by fixing other variables in a model and changing only the subject variable. A direct effect would occur if all other variables in a model remained constant (Shipley 2000). In estimating an indirect effect, all other variables in a model are controlled except for the mediating variables in the path representing the indirect effect in question (Shipley 2000). These technicalities are important to be kept in mind for valid interpretation of the effects. However, one may question the justification for holding constant any one of a set of variables that are non-orthogonal but rather multicollinear in the modeling dataset, which is the typical conditions in observational variables measured in complex forest ecosystems.

Path coefficients can be expressed as standardized or unstandardized coefficients. Unstandardized coefficients are more intuitive because they represent direct effects of the causal variables on the response variable on a scale consistent with original units of measurements. Furthermore, it is important to present unstandardized results in a path model because significance tests on the coefficients are based on standard errors of the unstandardized solutions (Grace 2006). Standardized coefficients are computed by first standardizing all variables by subtracting their mean and dividing by their standard deviation, and then computing the coefficients. These coefficients are therefore expressed in units of standard deviations for the corresponding variables, and are interpreted as the expected change in the endogenous variable per unit change in an exogenous variable, with change in both variables

measured in units of their corresponding standard deviations. Standardized coefficients allow direct comparison of the magnitude of effects of two causal variables measured on different scales (Grace 2006).

As in linear regression, a basic assumption underlying the path model is linearity between variables. Nonlinear relationship can be accommodated by transformations or adding higher-order terms. However the latter usually complicate path model specification (Hayduk 1987). A second important assumption is that the exogenous variables are measured without error. Any measurement errors, including those attributable to data entry, field recording, or other causes (Hayduk 1987), lead to bias in estimated path coefficients (Bollen 1989).

# 4.2.2 Measurement Model

A common method for evaluating a measurement model is confirmatory factor analysis (CFA). CFA is a process of specifying the number and types of observed variables associated with one or more hypothetical constructs, and analyzing how well the observed variables measure the constructs. A hypothetical construct is a conceptual variable which cannot be directly measured. Conversely, an observed variable can be measured and is used to infer the construct. Two types of constructs are *latent* and *composite* variables. A latent variable is a cause of its corresponding observed variables, whereas a composite variable is a collective effect of the variables. For brevity, only latent variables are further discussed. Grace and Bollen (2008) provided more details about the theory and application of composite variables. Choice of the observed variables for a measurement model must consider the *validity* and *reliability* of the observed variables. Validity refers to the accuracy of an observed variable for representing the effect of a latent variable. Because a latent variable is multifaceted, the observed variables selected should measure these different facets, and ideally have no effects other than those through the latent variable (Thompson 2003). Reliability refers to the consistency in measurement of an observed variable or the amount of random measurement error. The idea is similar to estimating the precision of a measuring device by repeatedly measuring the same observation under similar conditions.

Consider a hypothetical measurement model for species diversity (Figure 4.2). Conventionally, a circle represents a latent variable and a rectangle represents an observed variable. Following Krebs (1999), the model postulates that species diversity is a multifaceted concept with three observed variables measuring its different facets. The total number of species (observed and unobserved), which could be estimated from a Jackknife estimator (Burnham and Overton 1979, Krebs 1999), measures the total species richness. Shannon diversity index measures the diversity and the evenness of a community. The Simpson evenness index measures only the evenness of a community, which is hypothesized to be measuring a different facet of diversity than the Shannon diversity index. The single arrowheads pointing from the latent variable to the observed variables are factor loadings. They represent the direct effects and are interpreted similarly to regression coefficients. The thicker arrowheads to left of the observed variables depict the random measurement errors. A general recommendation is to have three or more observed variables per latent variable to ensure model convergence, proper solutions and adequate accounting for measurement error. Observed variables should be both valid and reliable with respect to measurement of the hypothetical constructs. Hayduk (1987) commented that observed variables with greater than 40% measurement error were likely prone to estimation problems. Bollen (1989) provided further details on the use of latent variables.

### 4.2.3 Structural Regression (SR) Model

A structural regression (SR) model is a path model with hypothetical constructs. The goal is to take measurement errors of observed variables into account when evaluating a path model. A fully latent SR model has only constructs in the path model whereas a partially latent SR model is a mix of observed variables and constructs. The SR model combines principles of path and measurement models. In both path and measurement models, and hence in SR models, an important phase of analysis is model identification. Model identification is a property that determines whether the model allows for unique parameter estimates. The two basic conditions for identification are: (1) model degrees of freedom equal to or greater than zero ( $df_M$  $\geq 0$ ), and (2) a known scale for every latent variable. The total degrees of freedom for the model is v(v+1)/2, where v is the number of observed variables. This total corresponds to the number of variances and unique covariances in a variancecovariance matrix for v variables. Thus,  $df_M$  is the total degrees of freedom minus the number of estimated parameters. A *just-identified* model ( $df_M = 0$ ) will have unique parameter estimates, but an *overidentified* model ( $df_M > 0$ ) is desirable for modeling testing. An *underidentified* model ( $df_M < 0$ ) will not have unique solutions for all parameters. Scale is a property of each latent variable. Because a latent variable is not measurable, it must take on the scale or units of measure from one of its observed variables. A way to assign scale is by imposing a unit loading identification (ULI) constraint by fixing a factor loading of one observed variable at a value of 1.0.

A model that meets these conditions on degrees of freedom and scale does not guarantee identification. Bollen (1989) suggested a two-step rule for checking identification of a SR model. The first step is specifying a SR model as a CFA model – replacing all directional paths with double arrowheads (unanalyzed association). The resulting CFA model is identified if it meets the following sufficient requirements and assumptions: (1) at least two observed variables per latent variable, (2) independence between measurement errors and latent variables, and (3) independence between measurement errors and latent variables, and (3) independence between measurement errors. The second step for checking identification is to check *recursiveness* of the path model part of the SR model, ignoring any observed variables used to measure latent variables. A path model is identified if it meets the following requirements for recursiveness: (1) errors are uncorrelated, and (2) all causal effects are unidirectional. If models in both steps are identified, the whole SR model is identified, and model fitting can proceed.

Some SR models can fail the two-step rule and still be identified. One example is correlated measurement errors in a CFA model. Another common example is a nonrecursive path model, e.g., a model with feedback loops or correlated errors, which requires special rules for identification. A special case is a SR model that has one observed variable per latent variable and therefore requires *a priori* assignment of measurement errors. Last but not least, Kenny (1979) described the condition of *empirical underidentification* in which the effective degrees of freedom were reduced due to two highly correlated observed variables.

### 4.2.4 Model Estimation

The LISREL framework can be summarized into three matrix equations, two for the measurement model component and one for the path model component (Grace 2006). For the measurement model component,

$$\mathbf{x} = \mathbf{\Lambda}_{\mathbf{x}} \boldsymbol{\xi} + \boldsymbol{\delta} \tag{4.1}$$

$$\mathbf{y} = \mathbf{\Lambda}_{\mathbf{y}} \mathbf{\eta} + \mathbf{\epsilon} \tag{4.2}$$

where, **x** is a  $p \times 1$  vector of observed exogenous variables and it is a linear function of a  $j \times 1$  vector of exogenous latent variables  $\xi$  and a  $p \times 1$  vector of measurement error  $\delta$ .  $\Lambda_x$  is a  $p \times j$  matrix of factor loadings relating **x** to  $\xi$ . Similarly, **y** is a  $q \times 1$  vector of observed endogenous variables, **q** is a  $k \times 1$  vector of endogenous latent variables,  $\varepsilon$  is a  $q \times 1$  vector of measurement error for the endogenous variables, and  $\Lambda_y$  is a  $q \times k$  matrix of factor loadings relating **y** to **q**. Associated with eqns. (4.1) and (4.2) respectively are two variance-covariance matrices,  $\Theta_{\delta}$  and  $\Theta_{\varepsilon}$ . The matrix  $\Theta_{\delta}$  is a  $p \times p$  matrix of variances and covariances among measurement errors  $\varepsilon$ . For flexibility, LISREL describes the path model component as relationships among latent variables,

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{4.3}$$

where **B** is a  $k \times k$  matrix of path coefficients describing the relationships among endogenous latent variables,  $\Gamma$  is a  $k \times j$  matrix of path coefficients describing the linear effects of exogenous variables on endogenous variables, and  $\zeta$  is a  $k \times 1$  vector of errors of endogenous variables. Associated with eqn. (4.3) are two variance-covariance matrices:  $\Phi$  is a  $j \times j$  variance-covariance matrix of latent exogenous variables, and  $\Psi$  is a  $k \times k$  matrix of covariances among errors of endogenous variables.

With only these three equations, LISREL is a flexible mathematical framework that can accommodate any specification of a SEM model. SEM is also referred to as covariance structure modeling because the variance-covariance matrix is the basic statistic for modeling. Model fitting is based on a fitting function that minimizes the difference between the model-implied variance-covariance matrix  $\Sigma$  and the observed variance-covariance matrix S,

$$\min f(\mathbf{\Sigma}, \mathbf{S}) \tag{4.4}$$

where **S** is estimated from observed data,  $\Sigma$  is predicted from the causal and noncausal associations specified in the model, and  $f(\Sigma, S)$  is a generic function of the difference between  $\Sigma$  and **S** based on an estimation method that follows. As Shipley (2000) concisely stated, *causation implies correlation*; i.e., if there is a causal relationship between two variables, there must exist a systematic relationship between them. Hence, by specifying a set of theoretical causal paths, one can reconstruct the model-implied variance-covariance matrix  $\Sigma$  from total effects and unanalyzed associations. For the recursive path model with only observed variables, Wright (1960) proposed tracing rules that help to reconstruct  $\Sigma$ ; however, these rules were susceptible to errors, especially for complex models. Hayduk (1987) outlined a step-by-step formulation under the LISREL mathematical framework, specifying the following mathematical equation for  $\Sigma$ :

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Lambda}_{y} \mathbf{A} \left( \boldsymbol{\Gamma} \boldsymbol{\Phi} \boldsymbol{\Gamma}' + \boldsymbol{\Psi} \right) \mathbf{A}' \boldsymbol{\Lambda}_{y}' + \boldsymbol{\Theta}_{\varepsilon} & \boldsymbol{\Lambda}_{y} \mathbf{A} \boldsymbol{\Gamma} \boldsymbol{\Phi} \boldsymbol{\Lambda}_{x}' \\ \boldsymbol{\Lambda}_{x} \boldsymbol{\Phi} \boldsymbol{\Gamma}' \mathbf{A}' \boldsymbol{\Lambda}_{y}' & \boldsymbol{\Lambda}_{x} \boldsymbol{\Phi} \boldsymbol{\Lambda}_{x}' + \boldsymbol{\Theta}_{\delta} \end{bmatrix}$$
(4.5)

where  $\mathbf{A} = (\mathbf{I}-\mathbf{B})^{-1}$ . Note that in eqn. (4.5) the derivation of  $\Sigma$  does not involve the observed and latent exogenous and endogenous variables (i.e.,  $\mathbf{x}$ ,  $\mathbf{y}$ ,  $\xi$  and  $\eta$ ).

A common method in SEM for estimating parameters in  $\Sigma$  is maximum likelihood (ML). In ML estimation, the algorithm iteratively searches for a set of parameter values that maximizes the likelihood that the differences between **S** and  $\Sigma$ are due only to sampling error or, in other words, that minimizes the deviations between elements of **S** and  $\Sigma$  (Grace 2006). This minimization is accomplished by deriving a fitting function  $f(\Sigma, S)$  (eqn. 4.4) based on the logarithm of a likelihood ratio, where the ratio is the likelihood of a given fitted model to the likelihood of a perfectly fitting model. The maximum likelihood procedure requires the endogenous variables to follow a multivariate normal (MVN) distribution, and **S** to follow a Wishart distribution. Hayduk (1987) described the steps in the derivation and expressed the fitting function  $F_{ML}$  as,

$$F_{ML} = \log |\mathbf{\Sigma}| + \operatorname{tr}(\mathbf{S}\mathbf{\Sigma}^{-1}) - \log |\mathbf{S}| - \operatorname{tr}(\mathbf{S}\mathbf{S}^{-1})$$
(4.6)

where tr() refers to the trace of a matrix and  $\Sigma$  and S are defined as above. Proper application of eqn. (4.6) also requires that observations are independently and identically distributed, and that matrices  $\Sigma$  and S are positive definite (Hayduk 1987). After minimizing eqn. (4.6) through an iterative process of parameter estimation, the final results are the estimated variance-covariance matrix and path coefficients for the specified model.

### 4.2.5 Model Assessment

A multitude of indices and criteria are available to assess model fit. Kline (2005) and Schumacker and Lomax (2004) provided a comprehensive listing, but three basic fit statistics are summarized here. The first is the overall model chi-square test based on a test statistic that is a function of the fitting function as follows:

$$\chi_M^2 = (n-1)F_{ML}$$
(4.7)

where  $F_{ML}$  is computed from eqn. (4.6), *n* is sample size, and  $\chi^2_M$  follows a chi-square distribution with degree of freedom  $df_M$  as defined above. Subsequently, a p-value is estimated and evaluated against a significance-level.

The overall model chi-square test is only applicable for an overidentified model; i.e., when  $df_M > 0$ . The null hypothesis associated with the test is that there is no difference between model estimates and the data, and the alternative hypothesis is otherwise. Therefore, *failure to reject* the null hypothesis is the ultimate objective of the modeling process. Although it may seem to contradict common hypothesis

testing, this approach is consistent with the accept-support context where the null hypothesis represents a researcher's belief (Steiger and Fouladi 1997). Nonetheless, as with common hypothesis testing, failure to reject the fitted model does not prove the specified causal relationships in the model. One should be particularly aware of existing equivalent models; i.e., models that have different hypothesized causal relationships but fit the data equally well.

The second fit statistic to consider is the Root Mean Square Error of Approximation (RMSEA), which is parsimony-adjusted index that accounts for model complexity. The index approximates a noncentral chi-square distribution with the estimated noncentrality parameter as,

$$\hat{\delta}_M = \max\left(\chi_M^2 - df_M, 0\right) \tag{4.8}$$

where  $\chi_M^2$  is computed from eqn. (4.7) and  $df_M$  is defined above. The magnitude of  $\hat{\delta}_M$  reflects the degree of misspecification of the fitted model. The RMSEA is then defined as:

$$RMSEA = \sqrt{\frac{\hat{\delta}_{M}}{df_{M}(n-1)}}$$
(4.9)

Thus, RMSEA measures the degree of misspecification per model degree of freedom, adjusted for sample size. RMSEA also reflects the view that the fitted model is an approximation of reality, so that RMSEA measures the error of approximation (Raykov and Marcoulides 2000). Browne and Cudeck (1993) suggested that RMSEA

 $\leq 0.05$  indicates a close approximation or fit, a value between 0.05 and 0.08 indicates a reasonable approximation, and a value  $\geq 0.1$  suggests a poor fit.

The third index is the standardized root mean square residual (SRMR), which is relatively easy to compute. Both S and  $\Sigma$  are transformed into correlation matrices, and the residual matrix is the difference between the two. Hence the mean square of each residual is the SRMR. In general, SRMR less than 0.10 is considered a good fit of S as an approximation to  $\Sigma$ .

In general, statistical tests for the overall model fit and p-values of parameter estimates are less important in SEM than in univariate regression models. One reason for this difference is that SEM is mostly a full information method; i.e., all parameters are simultaneously estimated, so the significance of a parameter estimate should be viewed in the context of the whole model. Second, the confirmatory aspect of the model is weakened if model modification is based on the significance of estimates rather than the theory behind the model structure. Finally, SEM is still a large-sample technique, and hypothesis testing is generally affected by sample size.

## 4.3 OVERSTORY-UNDERSTORY RELATIONSHIPS

The understory layer directly or indirectly supports much of the flora and fauna diversity in the Pacific Northwest (Halpern and Spies 1995). Hence, the ecological processes influencing the understory plant community are fundamental to conserving biodiversity. Research into overstory-understory relationships in this region has primarily focused on silvicultural effects, and the periods of early stand succession and

stand closure (McKenzie et al. 2000). Few studies have attempted to model the relationship in mature forests or for late-seral herbaceous species that have high conservation value (for exceptions see McKenzie and Halpern 1999, McKenzie et al. 2000). Across all studies, a fundamental assumption is that understory vegetation is controlled to a large degree by overstory structure, or that there exists a time precedence of overstory as a causal factor over the understory as an effect. However, this simple assumption downplays a suite of direct and indirect effects on the understory layer. SEM is a statistical method capable of addressing the hierarchical structure of forest vegetation and the complex interactions among vegetation layers and plant species. This study applies SEM to quantify ecological mechanisms influencing the abundance of late-seral herb species in mature Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) forests of the Pacific Northwest.

### 4.3.1 Hypotheses and Processes

Application of SEM requires a set of well defined hypotheses generated from theoretical considerations or informal observation. In this study, the processes of interest are light attenuation through the canopy, competition for belowground resources, and effects of forest floor litter. The effects of these processes on the forest herb community have been widely examined in both experimental and observational studies. Barbier et al. (2008) presented a comprehensive review of these processes and suggested additional influential factors such as the production of phytotoxic compounds and net effects such as spatial distribution of water. The following four hypotheses were formulated to address the processes of interest.

**Hypothesis 1**. Overstory tree cover directly affects late-seral herb cover, with a secondary and indirect effect through its influence on mid-story shrub and tree cover.

Overstory effects on understory vegetation are most commonly assumed to be mediated by light interception and attenuation. Light is a major limiting factor and influences forest stand development and understory plant establishment (Jennings et al. 1999, Oliver and Larson 1990); however, it is also closely correlated with throughfall precipitation and soil moisture in many situations (Anderson et al. 1969, Barbier et al. 2008). Although it is not possible to control the amount of light that impinges on the forest canopy, light levels under the canopy can be controlled by manipulating the amount, spatial arrangement, morphology and species of vegetation that absorbs incident light in various forest layers (Lieffers et al. 1999). Cannell and Grace (1993) observed that, in general, understory leaf area was inversely proportional to overstory leaf area. Overstory tree cover in this analysis is applied as a reasonable surrogate for leaf area and the amount of light intercepted by the overstory canopy. Understory plant species have been shown to behave as ecological filters, particularly in regard to their influence on resource availability for tree seedlings (Maguire and Forman 1983, George and Bazzaz 1999). Light attenuation is probably one of the major mechanisms driving this process; for example, Bartemucci et al. (2006) found that the understory plant community acted as a filter by reducing light penetration to

the forest floor. In this analysis, understory shrub and tree cover served as a proxy for light attenuation by this middle layer of forest vegetation.

**Hypothesis 2**. Belowground competition from overstory trees directly affects lateseral herb cover, with a secondary and indirect effect through its influence on midstory shrub and tree cover.

Tree density is further hypothesized to represent the degree of belowground resource competition. Numerous studies document the effect of root competition between trees and understory vegetation (Barbier et al. 2008). Root trenching experiments have demonstrated the importance of belowground competition in driving overstory-understory relationships (Lindh et al. 2003, Riegel et al. 1992). The diversity, abundance and biomass of the shrub and herb community respond positively to trenching, most likely due to increased availability of soil water and nutrients. Direct measurement of coarse and fine root mass of species and individuals, and the implications for belowground resource competition, is considerably more difficult. Silvicultural studies such as those that reduce tree density by thinning (e.g., Bailey et al. 1998) provide only limited information on mechanisms because thinning simultaneously increases light levels reaching the forest floor and belowground resource availability. For this study, tree density is tested as a surrogate for belowground resource competition under the assumption that overstory and mid-story cover (Hypothesis 1) account for light attenuation and light availability at the forest floor.

**Hypothesis 3**. Cover of coarse and fine woody litter directly affects late-seral herb cover, with a secondary and indirect effect through its influence on mid-story shrubs and tree cover.

This hypothesis represents the contribution of forest floor litter to nutrient and soil moisture availability (Spies and Franklin 1991) or mechanical interference with understory plant establishment (Ellsworth et al. 2004). Prescott (2002) considered that the mass and nutrient content of litter produced by a stand were the best indicators of soil mineral availability, and that these litter attributes depended on overstory species composition. In coniferous forests, the soil tends to have lower pH, lower nutrient availability, and higher carbon/nitrogen ratio, all of which may be unfavorable to understory vascular species (Barbier et al. 2008, Saetre et al. 1997). Forest floor litter can also directly impede the growth and germination of shrubs and herbs by reducing light and temperature and preventing root contact with mineral soil (Ellsworth et al. 2004, Lindgren and Sullivan 2001).

Two different processes were also considered; i.e., both coarse and fine woody litter accumulation. Coarse woody litter is produced by tree mortality in response to a number of interrelated factors such as stand density, site quality, age class and disease intensity (Greenwood and Weisberg 2008). Overstory tree density and stand age were hypothesized as the main driving factors for tree mortality. As a stand develops and self-thinning proceeds, death of overstory trees contributes to the coarse woody litter pool (Tappeiner et al. 2007). Fine litter accumulation is proportional to rate of crown recession, and because higher stand density generally hastens crown recession (Garber et al. 2008), tree density and tree size were proposed as the main factors influencing fine woody litter accumulation. The contributions of tree density and overstory tree cover to fine woody litter accumulation were considered separately, with the latter perhaps representing distinct mechanisms such as crown abrasion during a storm. Lastly, we assumed that there were a myriad of unknown factors contributing to both coarse and fine woody litter production.

**Hypothesis 4**. Topographic aspect directly affects late-seral herb cover, with a secondary and indirect effect through mid-story shrubs and tree cover.

Environmental factors such as aspect, slope and elevation are useful surrogates for the spatial and temporal distribution of factors such as radiation and temperature (Stage and Salas 2007). Aspect affects the amount of incident solar radiation; thus, it could strongly influence the microclimatic conditions such as air and soil temperature (Fekedulegn et al. 2003). In the northern atmosphere, it is well known that southwest aspects are often the most severe sites for vegetation establishment and growth (e.g., Beers et al. 1966).

# 4.3.2 Data Collection

The data for modeling overstory-understory relationships came from the Demonstration of Ecosystem Management Options (DEMO) study (Aubry et al. 1999), a large-scale operational research experiment implemented in western Oregon and Washington, USA. This study looked at the effects of variable-retention harvesting on various aspects of biodiversity, microclimate, and human perceptions. For details on the experimental and treatment designs, refer to Chapter 3. The analysis presented here was based on pre-harvest data collected in 1994/1995 from mature Douglas-fir stands ranging in age from 65 to 170 years.

A permanent  $8 \times 8$  or  $7 \times 9$  grid of 40-m spacing was installed in each 13-ha experimental unit. The grid was buffered by 40 m along the edge of each unit. Overstory and understory vegetation were studied only on a subset of sample points: 32–37 sample points depending on the unit. Detailed sampling protocols are available from Halpern and McKenzie (2001) and Halpern et al. (2005). Percent cover of herbaceous species (typically  $\leq 1$  m tall at maturity) was recorded for each of 24 microplots  $(0.2 \times 0.5 \text{ m})$  clustered at each sample point. Percent cover of coarse and fine woody litter cover was also recorded for the same microplots. Percent cover of tall shrub species (typically  $\geq 1$  m tall at maturity) and of understory coniferous and hardwood trees (<5.0 cm dbh) was measured by the line intercept method on four 6-m long transects radiating from each sample point. At the 0-m and 6-m marks of each transect, a Moosehorn densiometer was used to estimate percent overstory tree cover. Overstory trees were sampled with a set of nested circular plots: 0.01 ha plot for trees with dbh >5 and <15 cm and 0.04 ha plot for trees with dbh >15 cm (Maguire et al. 2007). At each sample point, aspect was recorded as azimuth in the downhill direction.

In the DEMO study, 48 herb species were classified as late-seral herb species; i.e., a species that reached maximum abundance in old-growth forest conditions and were sensitive to canopy removal or disturbance (Halpern et al. 2005). The nine observed variables for modeling late-seral herb cover with SEM were: (1) mean percent late-seral herb cover (LSHERB, %), (2) mean percent tall shrub and understory coniferous and hardwood trees cover (UNDER, %), (3) mean percent coarse woody litter cover (CLITTER, %), (4) mean percent fine woody litter cover (FLITTER, %), (5) mean percent overstory tree cover (TREE, %), (6) tree density expressed as trees per hectare (TPH, number/ha), (7) quadratic mean diameter at breast height (dbh) of overstory trees (QMD, cm), (8) cosine-transformed aspect (ASPECT), and (9) stand age (AGE, year).

LSHERB, CLITTER and FLITTER were the average of 24 microplots, and UNDER and TREE were the average of four transects at each sample point. As recommended by Beers et al. (1966), aspect was cosine transformed with a predetermined phase shift of 45° resulting in a variable that would range from one at a northeast aspect and negative one at a southwest aspect (Stage and Salas 2007).

### 4.3.3 Structural Equation Modeling

The four hypotheses described above were translated into the SR model depicted in Figure 4.3. The SR model failed the two-step rules for model identification by Bollen (1989) because there was only one observed variable per latent variable, and the errors of the fine and coarse litter were correlated. The former problem was addressed by assuming that the observed variables were measured with errors and that the latent variables represented the true values. Hayduk (1987) suggested that the solution for this identification issue was to assign *a priori*  measurement errors and the ULI constraint. Measurement error expressed as the percent of variance of each observed variable was estimated as the average of three expert opinions (Table 4.1; Paul Anderson, Douglas Maguire and Douglas Mainwaring, Personal Communication). The amount of variance attributed to measurement error was computed as the product of this percentage and the observed variance, and was then entered into the SR model (Figure 4.3). Subsequently, the ULI constraint assigned a value of 1.0 for each factor loading (Figure 4.3). The lower (0.025 quantile) and upper (0.975 quantile) limits of percent of measurement error were also provided for a sensitivity analysis on the effects of measurement errors on parameter estimates (Table 4.1). The path model part of the SR model had a *bow-free pattern*; i.e., errors of fine and coarse woody litter were correlated but there was no direct effect between the two latent variables (Kline 2005). In practice, a path model with a bow-free pattern is considered a recursive model, and thus is identified (Kline 2005). Consequently, the whole SR model was identified.

The bivariate relationship between LSHERB and TPH was nonlinear; thus to satisfy the linearity assumption, TPH was transformed by the natural logarithm. Furthermore, the observed variance-covariance matrix was ill-scaled; i.e., the difference between the largest and smallest variances was extremely large (Kline 2005). This would cause a problem in the iterative model estimation process. Thus, ASPECT and log(TPH) were each multiplied by 10 to improve the properties of the observed variance-covariance matrix (Figure 4.3). Because the maximum likelihood method is generally scale-free and scale-invariant, the parameters estimated for the

transformed ASPECT and log(TPH) could be algebraically converted back to the original metric (Kline 2005).

Initial testing revealed that most observed variables were not univariate normally distributed due to moderate or extreme skewness and/or kurtosis. In one exception, QMD exhibited no significant kurtosis. SEM is sensitive to violation of the multivariate normality assumption, particularly kurtosis; therefore, proper adjustment was needed to obtain valid standard error estimates. Following the recommendations of Jöreskog et al. (2000), both observed and asymptotic variance-covariance matrices were fitted by Robust Maximum Likelihood (RML) to obtain standard error estimates that reflected the non-normality in the data.

As documented in Chapter 3, the DEMO experiment was a Randomized Complete Block Design (RCBD) with six blocks and six treatments, for a total of 36 experimental units. Because sample points were nested within each unit, observations within the unit were not independent. If the nesting structure of the data was ignored, the parameter estimates were unbiased but the associated standard errors might be underestimated (Grace et al. 2009). For brevity, this study ignored blocks but accounted for the nested study design. Following methods discussed by Asparouhov and Muthén (2006) and described in LISREL documentation (2005), standard errors were adjusted and the model chi-square statistic (eqn. 4.7) was scaled to reflect the nesting structure of the data.

An equivalent model was designed with the only change being a hypothesized effect of Understory Cover on Fine Litter (Figure 4.4). The rationale was that limited

light availability caused by an increase in both overstory cover and understory density would hasten understory crown recession, thus contributing to the fine litter pool. Both models were fitted with LISREL 8.8 (Jöreskog and Sörbom 2006), but the fit was poor for both models. The initial SR model (Figure 4.3) had a scaled- $\chi_M^2$  of 33.75,  $df_M = 9$  and p-value < 0.001, and the RMSEA was 0.049 with 95% confidence interval between 0.032 and 0.067. Similarly, the equivalent model (Figure 4.4) had a scaled- $\chi_M^2$  of 29.40,  $df_M = 9$  and p-value < 0.001, and the RMSEA was 0.049 with 95% confidence interval between 0.027 and 0.062.

Sets of modification indices for both models were provided by LISREL with suggested paths and unanalyzed associations to improve model fit. However, most of the suggestions could not be supported by theory. Also to be consistent among the two models, the two paths Stand Age  $\rightarrow$  Late-Seral Herb Cover and Aspect  $\rightarrow$  Fine Litter were plausible additions. The first suggested path was intuitively appealing because stand structure and microclimatic conditions would gradually become more favorable to the development of late-seral herb community through the course of stand development. With regard to the second suggested path, Sariyildiz et al. (2005) found that, in the northern hemisphere, fine litter deposited in stands on northern aspects had a higher decomposition rate than litter deposited on southern aspects, likely due to drier and hotter microclimatic conditions at the latter. The modified initial SR model (hereafter referred to as the base model) and the modified equivalent model therefore included both of these suggested paths. Only unstandardized solutions are presented

as path coefficients in Figures 4.3 and 4.4 and in subsequent path diagrams. However, to facilitate comparison between relative magnitudes of effects, the width of arrows in the diagrams corresponds to the values of standardized solutions; i.e., wider arrows indicated larger standardized coefficients and thus stronger relative effects.

# 4.3.4 Results

#### 4.3.4.1 Bivariate Relationships

The bivariate relationships between LSHERB and its six predictors were depicted in Figure 4.5. LSHERB was positively correlated with FLITTER, UNDER and ASPECT (p-values < 0.001) but negatively correlated with TREE and log(TPH) (p-values < 0.001). In contrast, the correlation between LSHERB and CLITTER was not significant (p-value = 0.221). UNDER was predicted by five observed variables and all correlations were significant (p-values < 0.001) (Figure 4.6). FLITTER and ASPECT were positively correlated with UNDER whereas CLITTER, TREE and log(TPH) were negatively correlated.

The predictors for FLITTER were TREE, log(TPH) and QMD (Figure 4.7), and all were significantly correlated with FLITTER (p-values < 0.001). TREE and QMD were positively correlated with FLITTER, and log(TPH) was negatively correlated. The two predictors for CLITTER were log(TPH) and AGE (Figure 4.7). Both predictors were significantly and positively correlated with CLITTER (p-values < 0.001 and = 0.027 respectively). In some cases, visual inspection of the bivariate relationships suggested the possibility of nonlinearity, e.g., between LSHERB and FLITTER or CLITTER (Figure 4.5). These data were also fitted with quadratic equations and a logarithmic transformation of the predictors, but they did not improve the fit based on model  $R^2$  and residual plots. Ultimately, bivariate relationships provided only a crude check on the multivariate linearity assumption because apparent nonlinearities may result in part from non-orthogonal distribution of observations across all predictors.

## 4.3.4.2 Base SR Model

The final base model converged to admissible solutions with the indices indicating relatively good overall model fit. For this model, the scaled- $\chi_M^2$  was 7.94 with  $df_M = 7$  and a p-value = 0.34. The RMSEA was 0.011 with 95% confidence interval between 0.0 and 0.038, indicating a close approximate fit. The SRMR was 0.81, which was higher than the recommended level. Inspection of the standardized residual variance-covariance matrix (standardized **S** – standardized  $\Sigma$ ) revealed that the model did not estimate covariances between TREE and CLITTER very well, and the same was observed for covariances TREE and FLITTER. The predicted path coefficients are presented in Figure 4.8.

Overstory tree cover directly and indirectly affected late-seral herb cover (Figure 4.8); however, the direct effect was stronger than indirect effects through Fine Litter and Understory Cover. A 1% increase in overstory tree cover directly decreased the late-seral herb cover by 0.19% (p-value = 0.01, Table 4.2). The indirect effects

were mediated by fine litter cover and/or understory cover, but the combined indirect effects were insignificant (p-value = 0.35, Table 4.3). However, the path Overstory Cover  $\rightarrow$  Understory Cover  $\rightarrow$  Late-Seral Herb Cover had a strong negative effect on late-seral herb cover (-0.12 = -0.64×0.18, Figure 4.8). Nevertheless, the total effect of overstory tree cover was significantly negative (-0.24, p-value < 0.001, Table 4.4).

Tree density effects on late-seral herb cover were primarily indirect and mediated by forest floor litter and/or understory cover (Figure 4.8). This conclusion was supported by an insignificant direct effect (p-value = 0.61, Table 4.2). The combined indirect effect suggested that a two-fold increase in tree density reduced late-seral herb cover by 1.97% (p-value =0.003, Table 4.3). Of all the indirect effects, the path with the largest estimated absolute effect was Tree Density  $\rightarrow$  Understory Cover  $\rightarrow$  Late-Seral Herb Cover (-0.15 = -0.82×0.18, Figure 4.8).

The effects of coarse and fine woody litter on late-seral herb cover were predominantly mediated by understory cover (Figure 4.8). As indicated in Table 4.2, the direct effects of coarse and fine litter were insignificant (p-values = 0.33 and 0.12 respectively). However, a 1% increase in coarse litter cover indirectly increased the late-seral herb cover by 0.17% (p-value = 0.046, Table 4.3), and a similar increase in fine litter cover indirectly increased late-seral herb cover by 0.24% (p-value = 0.005, Table 4.3). Of all the predictors of fine litter cover, average tree size (QMD) turned out to be insignificant and tree density (log(TPH)) had the strongest effect (Figure 4.8). Tree density had a negative effect on fine litter cover (-0.39, p-value = 0.004, Table 4.2) whereas overstory cover (0.12, p-value < 0.001, Table 4.2) and aspect

(0.12, p-value = 0.031, Table 4.2) had positive effects. Both tree density and stand age were significant predictors for coarse litter cover with respective positive effects of 0.16 (p-value = 0.002, Table 4.2) and 0.03 (p-value = 0.014, Table 4.2). Nonetheless, the fitted model could not explain much of the observed variation in the coarse and fine woody litter; the squared multiple correlations  $R^2$  were 0.03 and 0.05 respectively (Figure 4.8). Finally, fine and coarse litter covers were negatively correlated with large error covariance (-43.18; Figure 4.8) and a standardized error covariance of -0.725.

Aspect had direct and indirect effects on late-seral herb cover, with the latter mediated by fine litter and/or understory cover. Both effects were significantly positive (p-values < 0.02, Tables 4.2 and 4.3), with a total effect amounting to 0.73 (pvalue < 0.001, Table 4.4). Of the indirect effects, the pathway Aspect  $\rightarrow$  Understory Cover  $\rightarrow$  Late-Seral Herb Cover amounted to a positive effect of 0.25 (1.38×0.18). As aspect shifted from southwest towards northeast, the late-seral herb cover increased. As expected, stand age had a strong direct effect on late-seral herb cover, with a 1 year increase in stand age directly increasing cover by 0.16% (p-value < 0.001, Table 4.2). The model also predicted an indirect effect of stand age, which was mediated by coarse litter and understory cover; however this effect was marginally significant (p-value = 0.057, Table 4.3). With all the predicted direct and indirect effects, the fitted model explained approximately half of the observed variance in lateseral herb cover ( $R^2 = 0.49$ ; Figure 4.8).

# 4.3.4.3 Equivalent SR Model

The equivalent model converged to admissible solutions, and the indices indicated a slightly superior fit to the base model: the scaled- $\chi_M^2$  was 6.42 with  $df_M =$ 7 and a p-value = 0.49, the RMSEA < 0.001 with 95% confidence interval between 0.0 and 0.034. However, similar to the final base model, the SRMR was 0.82 and the standardized residuals for the covariances between TREE and CLITTER, and TREE and FLITTER were large. The predicted path coefficients are presented in Figure 4.9.

Most direct effects in the final equivalent model were comparable to the final base model with a few noticeable exceptions. Foremost was the hypothesized direct effect of understory cover on fine woody litter, which was significantly positive with 1% increase in understory cover directly increasing litter cover by 0.07% (p-value < 0.001, Table 4.5). Second, two paths became insignificant; i.e., Aspect  $\rightarrow$  Fine Litter and Coarse Litter  $\rightarrow$  Understory Cover. The former has p-value of 0.999 whereas the latter coefficient was -0.28 with p-value of 0.313 (Table 4.5). Contrary to the base model, average tree size had a significant effect on fine woody litter (-0.10, p-value = 0.007, Table 4.5). A final notable difference was that the model explained larger amount of the observed variance in fine woody litter cover ( $R^2 = 0.12$ ; Figure 4.9).

Another outcome of the equivalent model was the different set of indirect and total effects. An indirect effect of understory cover on late-seral herb cover mediated by fine woody litter was predicted to be insignificant (0.020, p-value = 0.165, Table 4.6). Despite that, the total effect of understory cover was significant suggesting the

direct effect had the largest contribution (0.196, p-value < 0.001, Table 4.7). The overstory tree cover, tree density and aspect indirectly affected fine woody debris (p-values = 0.022, 0.007 and 0.003 respectively, Table 4.6), and all effects were mediated by understory cover (Figure 4.9). However, the cascading effects of these three variables on late-seral herb cover through both understory cover and fine woody litter were small, e.g., the largest predicted effect among the three was Aspect  $\rightarrow$  Understory Cover  $\rightarrow$  Fine Litter  $\rightarrow$  Late-Seral Herb Cover (0.029 = 1.54×0.07×0.27, Figure 4.9). This indirect cascading effect contributed little to the total effect of aspect on late-seral herb cover (0.725, Table 4.7).

#### 4.3.4.4 Sensitivity Analysis on Measurement Errors

General results from the sensitivity analysis on measurement errors for the base and equivalent models were similar; hence, only those from the equivalent model were depicted in Figures 4.10 and 4.11. Parameter estimates from the model with the lower limit (0.025 quantile) of percent measurement error (Figure 4.10) were almost identical to that assuming average measurement error final model (Figure 4.9). Conversely, there were major differences in the results from the model with the upper limit (0.975 quantile) of percent measurement errors (Figure 4.11 vs. Figure 4.9). For example, estimated path coefficients for Tree Density  $\rightarrow$  Fine Litter and Tree Size  $\rightarrow$  Fine Litter indicated larger effects for the model with the upper limit (unstandardized coefficients presented in Figure 4.11 vs. Figure 4.9), although the direction of effects remained the same. In addition, these effects were stronger in standardized

coefficients as indicated by the wider arrows in Figure 4.11 versus Figure 4.9. Secondly, the  $R^2$  of all endogenous latent variables increased implying that the model with upper limit explained more of the observed variance in those variables.

## 4.3.5 Discussion

Both fitted SR models supported the hypothesis that light attenuation associated with increasing overstory tree cover negatively affects late-seral herb cover. McKenzie et al. (2000) found similar results in their analysis of late-seral herb cover in mature forests of the Pacific Northwest. At first glance, typically shade-tolerant late-seral species might be expected to increase rather than decrease with increasing overstory cover. However, the range of overstory cover was relatively narrow. The 25% quantile of TREE was 65.3%, indicating that the majority of sample points had dense overstory cover (Figure 4.5). Late-seral herbs probably cannot thrive under an extremely dense overstory canopy because light availability would be limiting (DeGranpré et al. 1993). Thus as overstory cover increased from 65% to 100%, lateseral herb cover would decline in response to the decrease in light availability. The model accounted for two negative effects that act concurrently on late-seral herb cover; i.e., a cascading effect of overstory-understory-herb and the direct effect of overstory tree cover. Mid-story shrubs and trees apparently act as filters and play a role beyond the direct effect of overstory cover on late-seral herb cover. In short, superimposed effects of forest vegetation layers determine the integrated effect of vertical stand structure on late-seral herbs.
The hypothesis of belowground resource competition was supported by both fitted models; i.e., decreasing tree density had a positive net effect on late-seral herb cover. In unthinned stands, McKenzie et al. (2000) also found that maximum percent cover of late-seral herb cover increased with decreasing stand density index (SDI). However, the effect of root competition from overstory trees on late-seral herbs has received scant attention; instead, most work has focused on the composite response of all understory species (e.g., Powell and Bork 2006). Lindh and Muir (2004) found that thinning dense 20-year-old Douglas-fir stands had a positive impact on the frequency of late-seral herbs 20 years later. However, their results may not immediately discriminate between release from belowground resource competition and the combination of other factors such as time and light availability. The fitted models suggested that the indirect effect through understory shrub and tree cover was stronger than the direct effect of tree density. This result might be expected if the general depth of root systems differ between layers of the aboveground vegetation. Lindh et al. (2003) suggested that shade-tolerant plants would invest resources into above-ground development at the expense of a limited root system. Hence, understory trees and shrubs may have denser and deeper root systems than late-seral herbs, and may be more likely to interact extensively with tree root systems. In areas with lower overstory density, understory shrubs and trees may establish by more readily exploiting available belowground resources than late-seral herbs. With a local increase in understory shrub and tree cover made possible by reduced below- and

above-ground competition from the overstory trees, late-seral herbs may gain from the more favorable microclimatic conditions created by the denser mid-story vegetation.

Although there was no strong evidence to reject either the base or equivalent model, the statistical tests favored the equivalent model over the base model and, hence, the alternative hypothesis that fine woody litter was the effect of understory shrub and tree cover rather than the cause. This is also substantiated by the larger explanatory power the fitted equivalent model had on the observed variance of fine litter cover. In mature Douglas-fir forests of the Pacific Northwest, forest floor litter may have a negative effect on late-seral herbs by physically obstructing growth and establishment (Lindgren and Sullivan 2001), although positive effects may also be possible. The equivalent model supported the postulate that coarse woody debris negatively affects late-seral herbs indirectly through understory shrubs and trees, possibly as physical obstruction to the latter, but the effect was not significant. Conversely, the base model indicated a significant positive effect of coarse wood on both understory cover and later seral herb cover. Both models did support the hypothesized role of tree mortality and crown recession in coarse and fine litter production, respectively. Furthermore, the equivalent model predicted a stronger contribution of average tree size to the fine litter production. In both models, coarse and fine woody litter did not have a significant direct effect on late-seral herbs, contrary to what was hypothesized. In addition, effects of other unknown processes controlling accumulation of coarse and fine wood are suggested by the low  $R^2$  and large correlation between the two. Further research into the interaction between forest floor litter accumulation and understory vegetation is required to improve our understanding about mechanisms proposed in Hypothesis 3.

Both fitted models confirmed the influence of aspect on vegetation establishment and growth. Conditions in Douglas-fir stands on northeast aspects are more favorable to both understory shrubs and trees and late-seral herbs than on southwest aspects. Gazol and Ibáñez (2009) found that a southern aspect was associated with higher shrub diversity in mixed acidophilus beech (Fagus sylvatica L.) and oak (Quercus robur L.) forests in northern Spain. Other studies such as Mitchell and Bartling (1991) and Gracia et al. (2007) found that herb production and shrub diversity were not related to changes in aspect for ponderosa pine forests in Colorado and Wyoming, USA or for mixed species stands of Scots pine (*Pinus sylvestris* L.) and mountain pine (Pinus uncinata Ramond ex DC. In Lam. & DC.) in northeastern Spain, respectively. Neither of these studies addressed late seral species, but their mid-story shrubs response differed from the DEMO results for the mix of shrubs and small trees. Some of the differences among these studies are probably attributable to the different forest types and climates, and the fact that many environmental factors change simultaneously with aspect.

Given enough time, natural stand dynamics gradually lead to stand conditions that are favorable to the late-seral herb community. Many aspects of the biotic and abiotic environment change with stand development, including changes in the chemical and structural properties of soil (Barbier et al. 2008, Dupouey et al. 2002). In some forest types (e.g., Whitney and Foster 1988), age may be a strong predictor of understory community structure. In the DEMO SR models, average age of overstory trees emerged as a strong predictor of late-seral herb cover for the mature Douglas-fir forests of the Pacific Northwest.

Including latent variables in SEM acknowledges that a certain degree of measurement error exists in the observed variables (Grace 2006). Sensitivity analysis made clear that the parameter estimates from the fitted models were affected by measurement errors to some degree. In general, path coefficients and the proportion of explained variance ( $R^2$ ) are biased downward if measurement errors are ignored (Grace 2006). In addition, sensitivity analysis raises a question about robustness of conclusions to the usual assumption in most statistical models (e.g., multiple regressions) that predictors are measured without error. If measurement errors are small, the effects on parameter estimates would not be strong. On the other hand, if measurement errors are large, it is necessary to account for these errors to avoid biased estimates. SEM is a flexible method that can account for measurement errors when necessary.

## 4.4 VARIABLE-RETENTION HARVESTING

Increasing public concern about possible adverse effects of clearcutting on the biodiversity and ecological processes associated with mature Douglas-fir forests in the Pacific Northwest has prompted research into the efficacy of variable-retention harvesting. The number and spatial distribution of retained trees in the variableretention harvesting system have been hypothesized to influence the ameliorating

effects on taxa that are sensitive to abrupt changes in stand conditions. Creating aggregates might bring greater benefits to late-seral herbaceous species than having the retained trees dispersed over the harvesting unit (Halpern et al. 2005), but leaving a larger number of dispersed residual trees may ameliorate effects of regeneration harvesting better than leaving fewer dispersed trees. Residual stand conditions exert immediate effects on the survival of taxa and the persistence of processes, but also have long term implications for the recolonization and recovery. Conditions immediately after harvesting are largely driven by retention level, distribution of residual trees, and direct disturbance from logging activities. However, separating the effects of these factors would be difficult due to inter-correlations and complex interactions inherent in operational settings (Halpern et al. 2005). Part of the challenge is to find statistical methods that explicitly address as many of these interactions as possible. To our knowledge, few studies have analytically examined the interactions of these factors and their effects on biodiversity, although many hypotheses have been proposed. This study applied SEM in an attempt to understand responses of late-seral herb species diversity and composition to variable-retention harvesting in the Pacific Northwest, with emphasis on early responses to overstory removal and logging disturbances.

### 4.4.1 Hypotheses

Two sets of hypotheses were formulated. The first set related to microclimatic stresses induced by overstory removal, e.g., increased solar radiation, increased

temperature, and decreased relative humidity at the forest floor. The second set addressed direct disturbance from logging activities, e.g., uprooting, displacement and burial. The target population was late-seral herbs, a group of plants that shows an affinity for deeply shaded microsites buffered by multi-layered canopies (Halpern and Spies 1995). These microsites are characterized by relatively cool and moist conditions during the dry and warm summer (Halpern and Spies 1995), as well as minimal fluctuation in temperature and humidity (Chen 1991). The late-seral herb population is likely to be sensitive to changes in microsite and direct physical disturbance (Halpern et al. 2005). Microclimatic stresses and logging activities may therefore risk local extirpation and affect the speed or success of future recovery. The specific hypotheses are:

**Hypothesis 1**. The reduction in overstory crown cover from variable-retention harvesting directly and adversely affects late-seral herb species diversity and composition, and imposes an indirect effect by its influence on mid-story shrub and tree cover.

Crown cover captures processes related to light transmittance and moderation of relative humidity and air and soil temperature under the forest canopy (Barbier et al. 2008, Sharpe et al. 1996). Late-seral herb species are adapted to conditions under dense canopy, but seem to vary in their tolerance of sudden exposure (Halpern et al. 2005). Sudden and heavy reductions in canopy cover from harvesting can cause abrupt changes in microsite conditions, to the extent that some species may not be able to physiologically endure the change and could at least temporarily eliminated from the site. Mid-story shrubs and trees may act as a buffer by ameliorating the effects of overstory reduction. However, understory trees may also experience initial stress analogous to thinning shock (Harrington and Reukema 1983), including foliar chlorosis, top dieback, growth reduction (Maguire et al. 2006) or mortality due to sudden exposure to solar radiation, high air temperatures, and high vapor pressure deficits. These effects on the understory tree layer may limit its ability to ameliorate overstory reduction and its effect on late-seral herb species. Hence, we proposed direct and indirect effects of the reduction in overstory canopy cover on late-seral herb species.

**Hypothesis 2**. Reduction in tree density from variable-retention harvesting indirectly affects late-seral herb species diversity and composition through its influence on mid-story shrub and tree cover.

A reduction in tree density from the regeneration harvests may induce a net positive effect on residual plants by releasing belowground resources that would otherwise be pre-empted by overstory trees (Lindh et al. 2003, Riegel et al. 1992). However, understory shrubs and trees are speculated to respond more rapidly and opportunistically to the reduced belowground competition than late-seral herbs. Understory woody plants have denser and more extensive root systems so are in a better position to rapidly access the newly available resources. Root growth of Douglas-fir advance regeneration accelerated in response to recent partial harvests and released belowground resources (Kneeshaw et al. 2002). A major effect from reduction in tree density on late-seral herbs may therefore be indirect. **Hypothesis 3**. Aspect directly affects late-seral herb species diversity and composition, with a secondary and indirect effect through mid-story shrubs and trees cover.

Aspect affects the amount of incident solar radiation; thus, it could strongly influence the microclimatic conditions such as air and soil temperature (Fekedulegn et al. 2003). In the northern atmosphere, it is well known that southwest aspects are often the most severe sites for vegetation establishment and growth (e.g., Beers et al. 1966). With opening of the canopy after harvesting, aspect probably plays a greater role in determining the community of late-seral herbs and understory shrubs and trees. **Hypothesis 4**. Slash, coarse woody debris and disturbed mineral soil resulting from harvesting directly affects late-seral herb species diversity and composition.

Deposition of slash, movement of coarse woody debris and exposure of mineral soil during variable-retention harvesting will directly influence survival, persistence and recovery of late-seral herbs through physical damage to the plants and change in forest floor and soil conditions. Harvesting activity can result in burial of herbs, resulting in mortality or impeded growth (Lindgren and Sullivan 2001). Effects of burial would be more severe in non-clonal versus clonal herbs because clonal herbs may move perennating structures into more favorable environments (Halpern et al. 2005). Exposed mineral soil, e.g., from displacement of organic matter or skid trails, is also an effect of local harvesting activities. Other disturbances associated with harvesting include uprooting and soil compaction. **Hypothesis 5**. Production of slash, movement of coarse woody debris and exposure of mineral soil increase with greater amounts of overstory removal and are influenced by the types of harvesting system.

Tree harvesting generates logging slash in the form of fine needles, twigs, branches, and tops, depending on whether whole trees or only merchantable stems are yarded. Even under whole-tree yarding, much of this material can be generated by damaging trees and other vegetation during harvesting. In general, the level of reduction in overstory crown cover and understory shrub and tree cover should indicate the level of potential slash production. Coarse woody debris can be generated by cull logs, stem breakage, or snag felling, so again greater overstory reduction has the potential for generating a greater amount of new coarse woody debris on forest floor. Finally, the quantity of disturbed mineral soil is also a direct effect of the level of overstory reduction.

The yarding methods and other harvesting activities on the DEMO units were described by Halpern and McKenzie (2001) and are summarized in Table 4.8. Two general harvesting systems were applied to yard material from the treatment units, a ground-based system using a shovel loader to yard individual logs (in WF and PH blocks, Chapter 3) and helicopter system in which whole trees were removed (in DP, BU and LW blocks, Chapter 3). Helicopter logging was expected to generate less slash, coarse woody debris and disturbed mineral soil because non-merchantable trees were not felled, branches and tops were yarded to the landing, and temporary roads were not constructed. In contrast, shovel logging was expected to leave branches and tops where they fell in the units, creating more slash and greater variability in slash production. The slash was piled and burned in only the WF block (Table 4.8).

In addition to the five hypotheses, we also hypothesized that error variances of late-seral herb species diversity and composition were correlated because other common factors not represented in the model were expected to affect both of these variables.

## 4.4.2 Data Collection

The vegetation sampling protocol for the DEMO study was described in Section 4.3.2. Sampling was carried out in 1994–1996 before harvest, and immediately after harvest in 1998–2000. Exploratory data analysis showed that the Capitol Forest (CF) block had low number of late-seral herb species before harvesting, reaching a maximum of four, probably because the stands were relatively young (~ 65 years old). The data from the CF block were therefore excluded from modeling. Furthermore, we limited the analysis to treatment combinations represented by two retention levels (15% and 40%) and two patterns of retention (Dispersed and Aggregated; refer to Chapter 3).

The main responses of interest to the study were species diversity and species composition of the late-seral herb community. The Shannon diversity index is a common index for measuring species diversity (Krebs 1999), and takes into account the number and evenness of species. The index was calculated for pre- and post-harvest data as follows:

$$SHANNON = -\sum_{i=1}^{spp} p_i \ln p_i$$
(4.10)

where  $p_i$  was the pre- or post-harvest relative frequency of late-seral herb species *i* at a sample point.

The proposed index for measuring species composition was the Renkonen similarity index (Krebs 1999). It measured the percent similarity in the relative abundance of species between pre- and post-harvest community at a given sample point as follows:

$$RENKONEN = 100\% \times \sum_{i=1}^{spp} \min\left(p_{pre,i}, p_{post,i}\right)$$
(4.11)

where  $p_{pre,i}$  and  $p_{post,i}$  were the relative frequency of late-seral herb species *i* at a sample point before and after harvest, respectively. The frequency of a late-seral herb species *i* at a given sample point was the number of the 24 microplots in which the species was detected ( $n_i$ ). The abundance of all late-seral herb species at a sample point was measured as the total of all frequencies  $N = \sum_{i=1}^{spp} n_i$ , and the relative frequency of species *i* at a given sample point was therefore measured as  $p_i = n_i/N$ .

Except for crown volume, the variables depicting vegetation structure were described in Section 4.3.2. Crown volume was estimated by first computing crown cross-sectional areas at 0.5-m height intervals for each sampled tree (Dubrasich et al. 1997, Maguire et al. 2007). All crown cross-sectional areas were then summed over all height intervals for all the trees on a given plot, and this total crown cross-sectional area served as a surrogate for total crown volume.

The information on ground conditions immediately after harvesting was collected along the four 6-m transects as described in Section 4.3.2. Percent cover of slash and disturbed mineral soil and volume of coarse woody debris were estimated, and slash depth was estimated at 12 points along each transect yielding a total of 48 measures of slash depth per sample point (Halpern and McKenzie 2001). Slash consisted of needles, leaves, twigs and branches <10 cm in diameter, disturbed mineral soil included skid trails and places where mineral soil had been exposed or deposited, and coarse woody debris was down wood with diameter  $\geq$ 10 cm (Halpern and McKenzie 2001).

In summary, the 11 observed variables for modeling changes in species diversity and composition of late-seral herb communities with SEM were: (1) difference in pre- and post-harvest Shannon diversity index (dSHANNON), (2) Renkonen similarity index (RENKONEN, %) (3) difference in pre- and post-harvest understory shrubs and trees cover (dUNDER, %), (4) difference in pre- and postharvest crown volume (dCROWNVOL, m<sup>2</sup>/ha), (5) difference in pre- and postharvest tree density (dTPH, trees/ha), (6) cosine-transformed aspect (ASPECT), (7) postharvest slash cover (SLASHCOV, %), (8) post-harvest slash depth (SLASHDEP, cm), (9) post-harvest coarse woody debris volume (CWDVOL, m<sup>3</sup>/ha), (10) post-harvest disturbed mineral soil cover (DISTSOIL, %), and (11) harvesting system (HARVEST, Table 4.8). These variables are summarized in Table 4.9 by their minima, means and maxima for 15% retention (i.e., 15%A and 15%D treatments combined) and 40% retention (i.e., 40%A and 40%D treatments combined). All differences were defined as the post-harvest value minus the pre-harvest value.

## 4.4.3 Structural Equation Modeling

The six hypotheses were translated into a SR model depicted in Figure 4.12. Similar to the model in Section 4.3.3, each latent variable had only one observed variable. Therefore, *a priori* measurement errors and ULI constraints were imposed on the observed variables. For flexibility, measurement error was expressed as a percentage of the observed variance for a given observed variable (Figure 4.12). The SR model was considered a partially latent model because the dichotomous variable HARVEST was measured without error. Some of the observed variables were transformed to avoid ill-scaled variance-covariance matrices (Figure 4.12).

Modeling the response to variable-retention harvesting entailed use of a composite variable, Logging Residuals (depicted as a hexagon in Figure 4.12). A composite represents a collection of causes (Grace and Bollen 2006), including in this case the effects from latent variables slash cover, slash depth and coarse wood. This construct allowed for a comprehensive test of the combined effects of logging residuals on late-seral herb species diversity and composition. As a composite variable, Logging Residuals was by definition endogenous and therefore contained error variance. A common practice is to assume zero for the error variance because a composite variable is completely determined by the latent variables that influence it (Grace and Bollen 2006). To avoid model underidentification, a ULI constraint was

imposed on the path Slash Depth  $\rightarrow$  Logging Residuals (Figure 4.12) to establish the scale of measurement (Grace and Bollen 2006), a practice that is similar to a measurement model.

Exploratory data analysis revealed that all endogenous observed variables had significant skewness and kurtosis. Additionally, sample points were nested within experimental units. To account for non-normality and non-independence of observations, a maximum likelihood estimation procedure that produced robust standard errors was applied and effects were tested by a chi-square statistic (Muthén and Muthén 2007). Standard errors were estimated using a 'sandwich estimator' (Muthén and Muthén 2007), and the chi-square statistic was asymptotically equivalent to the Yuan-Bentler test statistic (Yuan and Bentler 2000). The modeling and analysis were carried out with Mplus 5.21 (Muthén and Muthén 2007).

An elaboration on standard SEM modeling is multigroup SEM analysis (Grace 2006). The general procedure is to assign data to unique groups, and to fit a model to each group using data from all groups simultaneously (Grace 2006). The advantage of this approach is its ability to test hypotheses about whether parameter estimates differ among groups and to determine whether a common model adequately describes all groups (Grace 2006). In this study, the four treatments were the groups, and a multigroup SEM analysis was carried out with the proposed SR model (Figure 4.12). As in the previous section, only unstandardized solutions are presented. However, to facilitate comparison of magnitude of effects, the width of arrows in the path diagrams

corresponds to the standardized solutions; i.e., wider arrows indicated larger standardized coefficients and thus stronger relative effects.

# 4.4.4 Results

#### 4.4.4.1 Bivariate Relationships

The bivariate relationships between direct predictors and dSHANNON and RENKONEN were depicted for 15% retention (Figure 4.13) and 40% retention (Figure 4.14). Both dSHANNON and RENKONEN for 15% retention showed strong correlations with most direct predictors. Both response variables were positively correlated with dCROWNVOL and ASPECT but negatively and more weakly correlated with dUNDER and DISTSOIL (Figure 4.13). The trends in the relationships between four predictors and both dSHANNON and RENKONEN for 40% retention were similar to those of 15% retention albeit the correlations were weaker (Figure 4.14). In general, a bivariate linear relationship between a response and predictor appeared plausible for both retention levels, but the scatterplots indicated a large amount of variation around these linear trends. However, as mentioned earlier, bivariate relationships provide very limited insight into the multidimensional relationships inherent to a SEM.

### 4.4.4.2 SR Models Exploration

The multigroup SEM analysis on the proposed SR model (Figure 4.12) did not converge to any admissible solution. The reported problem was that there were not enough experimental units within each group to generate reliable parameter estimates; i.e., each treatment had only five experimental units. To circumvent this problem, the data were regrouped by retention level by pooling both the dispersed and aggregated retention treatments; i.e., data from 15%D and 15%A treatments were grouped into a single 15% treatment and data from 40%D and 40%A treatments were similarly grouped. The current mandate for federal forestland in the Pacific Northwest is to retain at least 15% of live trees in a harvest unit, to ensure that  $\geq$  70% of this retention is in aggregates of 0.2–1.0 ha, and to allow  $\leq$  30% in either dispersed smaller groups (< 0.2 ha) or as individual trees (USDA and USDI 1994, Tuchman et al. 1996). In practice, therefore, a gradient in aggregate size from 1-ha to individual trees may be applied, with the aggregated and dispersed DEMO treatments representing the extremes. Also, Halpern et al. (2005) found the vegetation to have stronger response to retention level than dispersion types.

A second attempt was then made at multigroup analysis of the regrouped data to compare response mechanisms in 15% versus 40% retention. However, the analysis still failed to generate any admissible solution, even though number of experimental units per group (10) seemed adequate. One likely explanation was that the SR model could not simultaneously fit the variance-covariance matrices of both groups. In other words, there might not be a common model for the two retention levels. The next step was to fit the proposed SR model individually to each group. The drawback was that the hypotheses for multigroup analysis could not be tested. Nonetheless, it helped to discover possible contrasting features between the groups. The proposed SR model failed to converge when fitted to 15% retention level data. Subsequent exploration suggested adding error covariances between endogenous latent variables. When error covariances between Slash Cover and Slash Depth, and between Slash Cover and Coarse Wood were added, the model converged to an admissible solution with the test statistics indicating good fit: the  $\chi^2_M$  = 25.83 with  $df_M$  = 21 and a p-value = 0.21, the RMSEA = 0.027 with 95% confidence interval between 0.0 and 0.057, and the SRMR = 0.033. The final SR model for 15% retention with predicted path coefficients and the two added error covariances is depicted in Figure 4.15.

The initially proposed SR model without added covariances converged to an admissible solution when fitted to 40% retention level data; however, the test statistics indicated poor fit: the  $\chi_M^2 = 179.40$  with  $df_M = 23$  and a p-value < 0.0001, and the RMSEA = 0.141 with 95% confidence interval between 0.122 and 0.160. The two added error covariances (Slash Cover and Slash Depth, Slash Cover and Coarse Wood) from the final 15% retention SR model were among the suggested modifications. To facilitate comparison of the 15% and 40% models, the two error covariances were added. The resulted model fit was improved, but the tests still indicated lack of fit: the  $\chi_M^2$  reduced to 68.84 with  $df_M = 21$  and a p-value < 0.0001, and the RMSEA was 0.081 with 95% confidence interval between 0.060 and 0.103. From the list of suggested modifications, three appeared theoretically plausible: two error covariances (Disturbed Soil and Slash Depth, Disturbed Soil and Coarse Wood)

and a direct path Tree Density  $\rightarrow$  Slash Cover. With these three additions, tests indicated adequate model fit: the  $\chi_M^2 = 25.27$  with  $df_M = 18$  and a p-value = 0.118, the RMSEA was 0.034 with 95% confidence interval between 0.0 and 0.063, and the SRMR was 0.038. The final SR model for 40% retention with predicted path coefficients, including the four additional error covariances and one additional direct path, is depicted in Figure 4.16.

Throughout the model building process for 40% retention, the effects of HARVEST on slash, coarse woody debris and disturbed mineral soil was consistently insignificant. As an alternative model, the HARVEST variable was removed, with other structures kept the same. The resulting model converged to an admissible solution with tests showing adequate fit: the  $\chi^2_M$  = 15.80 with  $df_M$  = 15 and a p-value = 0.395, the RMSEA was 0.012 with 95% confidence interval between 0.0 and 0.053, and the SRMR was 0.038. Because this alternative model and the 40% retention SR model with HARVEST were not nested, it would be improper to compare the fit between the two models. The alternative model with predicted path coefficients is depicted in Figure 4.17.

# 4.4.4.3 The Final 15% Retention SR Model

The final 15% retention SR model predicted that retaining higher crown volume after harvest increased late-seral herb species diversity and similarity in species composition before and after harvest (Figure 4.15). For each reduction of  $10,000 \text{ m}^2$ /ha in crown volume by harvesting, the Shannon diversity index declined by

1.21 units (p-value = 0.012, Table 4.10). Likewise, for this same reduction in overstory crown volume the similarity between pre- and post harvest late-seral species composition (RENKONEN) declined by 0.74% (p-value < 0.001, Table 4.10). A move from southwest to northeast aspects resulted in a smaller reduction in species diversity and greater community similarity before and after harvest (Figure 4.15). In contrast to direct effects of reduction in crown volume, indirect effects of reduction in tree density were insignificant for species diversity (0.063, p-value = 0.216, Table 4.11) and community similarity (0.002, p-value = 0.945, Table 4.11).

The negative effect of aspect on understory cover was contrary to expectation (Figure 4.15). More northeastern aspects implied a greater reduction in understory shrub and tree cover relative to more southwestern aspects (-1.391, p-value < 0.001, Table 4.10). The lower average initial understory cover on southwest aspects prior to harvest (Figure 4.18A) may have limited the maximum potential reduction in understory cover caused by the harvesting disturbances (Figure 4.18B). Conversely, northeast aspects had a greater range in the reduction of understory cover (Figure 4.18B) because the average initial understory cover before harvest was greater (Figure 4.18A).

Of all the harvesting disturbances to the forest floor, only disturbed mineral soil was found to have negative effect on any aspect of the late-seral herb community, and that effect was limited to only species diversity (Figure 4.15). The model predicted that 1% increase in the disturbed mineral soil was associated with a 0.27 unit reduction in diversity index of late-seral herbs (p-value = 0.004, Table 4.10). The

combined effect of logging residuals or forest floor disturbances were not significant for either species diversity (-0.130, p-value = 0.439, Table 4.10) or community similarity (-0.088, p-value = 0.515, Table 4.10). Moreover, the slash cover and coarse woody debris volume did not contribute significantly to the composite variable Logging Residuals (Figure 4.15).

All factors hypothesized to contribute to the production of slash, coarse woody debris and disturbed mineral soil were significant, with the exception of HARVEST (Figure 4.15, Table 4.10). As expected, higher retention of tree density, crown volume and understory cover were associated with less slash, coarse woody debris and disturbed mineral soil after harvest (Table 4.10). These factors explained a considerable amount of the observed variance of Slash Cover ( $R^2 = 0.57$ ) and Slash Depth ( $R^2 = 0.36$ ) but not Coarse Wood ( $R^2 = 0.12$ ) or Disturbed Soil ( $R^2 = 0.10$ ) (Figure 4.15). Helicopter vs. shovel yarding did not appear to affect slash cover and coarse wood significantly (Figure 4.15), but helicopter logging reduced disturbed mineral soil by 2.86% (-2.86 = -14.32/5; p-value = 0.037, Table 4.10), but increased the slash depth by 5.04 cm (5.04 = 25.22/5; p-value < 0.001, Table 4.10).

Consistent with model building process and the original hypotheses, the estimated error correlations between the following three sets of endogenous latent variables were large: 0.561 for Diversity and Composition, 0.769 for Slash Cover and Slash Depth and -0.233 for Slash Cover and Coarse Wood. The large error correlation between Diversity and Composition suggested that: 1) other factors not represented in the model may have affected both variables; and/or 2) this was a direct result of both variables being a function of number of species and relative abundance  $p_i$  (eqns. 4.10 and 4.11). This effect of other factors was further supported by the low explanatory power model for Diversity and Composition ( $R^2 = 0.29$  and 0.32 respectively, Figure 4.15).

The total indirect effects of factors on Diversity and Composition were insignificant (Table 4.11). However, the model predicted a specific indirect path, Tree Density  $\rightarrow$  Disturbed Mineral Soil  $\rightarrow$  Diversity, to be significant (0.052, p-value = 0.019, result not shown in Table 4.14), probably because the two direct paths were significant. Therefore, we would conclude that the significant total effects (Table 4.12) were mostly driven by significant direct effects.

### 4.4.4.4 The Final 40% Retention SR Model

Many of the results from the final 40% retention SR model were similar to those for 15% retention, but a striking difference was fewer significant direct effects (Figure 4.16 vs. Figure 4.15). Lower reductions in crown volume resulted in lower reductions in late-seral species diversity and community similarity (Figure 4.16). For a reduction of 10,000 m<sup>2</sup>/ha in the crown volume after harvesting, diversity (dSHANNON) decreased by 0.78 unit (p-value = 0.029, Table 4.13). Likewise, for the same crown volume reduction, community similarity (RENKONEN) declined by 0.51% (p-value = 0.002, Table 4.13). Contrary to results for 15% retention, aspect had no significant effects on species diversity or community similarity (p-value = 0.152 and 0.225 respectively, Table 4.13). On the other hand, changes in understory shrub and tree cover did influence community similarity (Figure 4.16), in spite of a weak bivariate correlation (r = -0.05, Figure 4.14). A reduction of 1% in understory cover was associated with a 0.12% decrease in community similarity of late-seral herbs (p-value < 0.001, Table 4.13).

As was found for 15% retention, northeastern aspects were associated with greater loss of understory shrub and tree cover (Figure 4.16). This result is best explained again by the greater initial shrub and tree cover on northern aspects (Figure 4.18C–D). Neither the composite variable Logging Residuals nor disturbed mineral soil significantly affected species diversity or composition (Figure 4.16, Table 4.13). In contrast to 15% retention, however, the model predicted a significant contribution of Slash Cover to Logging Residuals (-1.079, p-value = 0.022, Table 4.13).

The hypothesized factors represented in the final 40% retention model were significant contributors only to Slash Cover and Slash Depth (Figure 4.16). With any reduction in Tree Density, Crown Volume or Understory Cover by harvesting, both Slash Cover and Slash Depth increased (Table 4.13, Figure 4.16). For removal of every 10 trees/ha by harvesting, slash cover increased by 0.27% (p-value = 0.001, Table 4.13). In contrast to results for 15% retention, the yarding system did not have any significant effects on slash, coarse wood or disturbed mineral soil in 40% retention (Figure 4.16).

As depicted in the model, there were a number of significant error correlations between endogenous latent variables. The predicted error correlation between Diversity and Composition was 0.381, and the proportion of observed variance that was explained by the model was low for both Diversity and Composition ( $R^2 = 0.21$ and 0.13 respectively, Figure 4.16). Other factors not represented in the model were therefore concluded to affect the two responses. The predicted error correlations between Slash Cover and Slash Depth, Slash Cover and Coarse Wood, Disturbed Soil and Slash Depth, Disturbed Soil and Coarse Wood were 0.811, -0154, -0.154 and -0.170 respectively. As was obtained in the model for 15% retention, the 40% retention model had strong explanatory power for Slash Cover ( $R^2 = 0.67$ ) and Slash Depth ( $R^2 = 0.57$ ), but did poorly for estimating Coarse Wood ( $R^2 = 0.02$ ) and Disturbed Soil ( $R^2 = 0.07$ ) (Figure 4.16).

Although none of the total indirect effects on Diversity and Composition were significant (Table 4.14), the model predicted one significant indirect path, Aspect  $\rightarrow$  Understory Cover  $\rightarrow$  Composition (0.069, p-value = 0.025, result not shown in Table 4.14). The significant total effects in Table 4.15 were driven mostly by significant direct effects.

The alternative 40% retention model that eliminated the HARVEST variable (Figure 4.17) produced results almost identical to the previous 40% retention model (Figure 4.16). Two major differences were that the path Tree Density  $\rightarrow$  Disturbed Soil became significant (-0.21, p-value = 0.013), and the path Slash Cover  $\rightarrow$  Logging Residuals became insignificant (-1.18, p-value = 0.090). Another minor difference was that some paths were stronger in the alternative model than in the previous model, e.g., Crown Volume  $\rightarrow$  Slash Depth and Understory Cover  $\rightarrow$  Slash Depth (Figure 4.17 vs. Figure 4.16).

#### 4.4.4.5 Reduced 15% and 40% Retention SR Models

According to Grace and Bollen (2006), one issue with composite variables is the significance of the path with a ULI constraint; i.e., Slash Depth  $\rightarrow$  Logging Residuals. The path coefficient was set to 1.0 to establish the scale of measurement for Logging Residuals, ignoring whether the coefficient was significantly different from zero (Grace and Bollen 2006). One approach to evaluate the validity of this constraint was to reduce the model by omitting the composite variable Logging Residuals and specifying direct effects of Slash Cover, Slash Depth and Coarse Wood on Diversity and Composition. These reduced models for 15% and 40% retention are depicted in Figure 4.19 and Figure 4.20, respectively.

The reduced 15% retention model converged to an admissible solution and the tests indicated an adequate fit: the  $\chi_M^2 = 22.46$  with  $df_M = 19$  and a p-value = 0.262, the RMSEA was 0.024 with 95% confidence interval between 0.0 and 0.057, and the SRMR was 0.033. The significant path coefficients and explained variances were almost identical to the final 15% retention model (Figure 4.19 vs. Figure 4.15), primarily because the direct effects of slash, coarse woody debris and disturbed mineral soil were not significant, as was the case for the full model. Hence, logging residuals were concluded to have little detectable effect on late-seral herb species diversity and community similarity.

In contrast to 15% retention, the reduced 40% retention model did not adequately fit the observed variance-covariance matrix, although it did converge to admissible solution: the  $\chi_M^2 = 27.18$  with  $df_M = 16$  and a p-value = 0.039, the RMSEA was 0.045 with 95% confidence interval between 0.010 and 0.073, and the SRMR was 0.038. In addition to the model lack of fit, path coefficients Understory Cover  $\rightarrow$  Composition (p-value = 0.080) and Crown Volume  $\rightarrow$  Diversity (p-value = 0.070) became insignificant. However, the path Tree Density  $\rightarrow$  Slash Cover became stronger in the reduced model. Although Slash Cover contributed significantly to the Logging Residuals in the final model, neither Logging Residuals in the full model nor the slash, coarse woody debris or disturbed mineral soil in the reduced model had a significant direct effect on species diversity or community similarity.

The alternative 40% retention SR model with both the HARVEST variable and composite variable Logging Residuals removed (Figure 4.21) yielded an adequate fit: the  $\chi_M^2 = 16.01$  with  $df_M = 13$  and a p-value = 0.249, the RMSEA was 0.026 with 95% confidence interval between 0.0 and 0.062, and the SRMR was 0.038. All significant parameter estimates and proportions of explained variance were almost identical to the alternative 40% retention model with only HARVEST removed (Figure 4.17). The one exception was that the path Understory Cover  $\rightarrow$  Composition became insignificant in the reduced model (p-value = 0.074). Regardless, none of the 40% retention models detected a significant effect of slash, coarse wood or disturbed mineral soil on late-seral herb species diversity or community similarity.

# 4.4.5 Discussion

The process of building a model to explain the response of late-seral herbs to variable-retention harvests was not strictly confirmatory because it relied on a combination of suggested modification indices, theoretical support, and various ad hoc hypotheses. Instead, it was a mixture of exploratory and confirmatory analysis. Initial models were designed on the basis of well established hypotheses. The final models adequately fit the observed variance-covariance matrices, so the results could not support rejection of the hypothesized relationships. However, equivalent models may exist, and the final models need to be validated with independent datasets to establish consistency in responses across populations.

A major challenge in applying SEM to large-scale ecological experiments is accommodating the experimental design (Grace et al. 2009). A large-scale experiment generally includes blocking of experimental units and nesting of sampling units, as well as other features such as split-plot structures. If the structure imposed by these designs is not properly taken into account then estimates of standard errors will be biased (Grace et al. 2009). The DEMO analysis ignored the blocking nature of the experiment because it necessitated multilevel SEM, which is still under theoretical development. Another possibility is a Bayesian approach to SEM (Lee 2007). Regardless, the estimated standard errors and the significance tests of estimated coefficients in this analysis were likely conservative with respect to not rejecting the null hypothesis. Also, these large operational experiments inevitably have a relatively small number of independent experimental units because locating a larger set of sufficiently uniform units is logistically challenging and monitoring responses on them may be not be financially feasible. This problem is not unique to SEM, however, because statistical models such as ANOVA face similar issues. Perhaps the effects of small sample size are more noticeable with SEM because it is a multivariate statistical method. We resorted to grouping the DEMO data by retention level and ignoring spatial distribution to have a larger number of experimental units per group. This approach seemed reasonable because one of the primary mechanisms represented in the SEMs was the plot-level effect of overstory reduction on late-seral herbs. Overstory reduction effects are best regarded as a gradient from total overstory removal to no removal, a gradient that is well represented by the combined dispersed and aggregated treatments.

The results of all model specifications suggested that Hypothesis 1 could not be rejected, leading to a tentative conclusion that reductions in overstory crown volume from variable-retention harvesting had direct negative effects on late-seral herb species diversity and community similarity at both 15% and 40% retention levels, but no indirect effects through mid-story shrub and tree cover. Even in 40% retention, microclimatic conditions on forest floor were likely influenced by the harvest reductions in crown cover to the extent that late-seral herbs were affected. Late-seral herbs are expected to be relatively sensitive to a rapid decline in relative humidity, increase in temperature and other microclimatic conditions, and experience shock immediately after harvest. However, 15% retention had a greater effect of overstory reduction on late-seral herbs than 40% retention. Given comparable overstory reduction as measured by crown volume, species diversity and community similarity decreased much more under 15% retention than under 40%. This difference may be partly attributable to the range in overstory density reduction, which was -561,000 to 167,200 m<sup>2</sup>/ha for 15% retention and -443,000 to 131,000 m<sup>2</sup>/ha for 40% retention (Figures 4.13 and 4.14). The larger reduction signals a potentially more drastic change in microclimate (Heithecker and Halpern 2006) and hence a potentially more deleterious effect on late-seral herb diversity and lower degree of similarity between pre- and post-harvest communities. Similarly, a nonlinear response to overstory reduction would cause an apparent increase in effect even though the linear effect within the range covered by 40% retention might be indistinguishable.

In contrast to Hypothesis 1, Hypothesis 2 was rejected by all models regardless of retention level or variable specification. The hypothesized direct effect of reduction in tree density on understory shrub and tree cover was not significant, and its indirect effect on late-seral herb species diversity and community similarity likewise was not significant. If undamaged by overstory reduction, understories may respond positively and immediately by investing in root growth, perhaps because an increase in belowground resources is more stimulating than the possible mix of positive and negative effects of sudden exposure to light (Kneeshaw et al. 2002). Lindh et al. (2003) found that root trenched plots had greater shrub and herb cover 10 years after trenching treatments were implemented in Douglas-fir forests. A lag in understory response might explain the apparent lack of immediate effects from reductions in tree density. On the contrary, however, Riegel et al. (1992) implemented a root trenching experiment that included thinning in *Pinus ponderosa* forests in northeastern Oregon, and found that understory aboveground biomass increased by 53% immediately after harvest and by 94% a year later. They concluded that belowground resources competition was the primary controlling factor for understory production, rather than light availability.

Few have studied the effects of silvicultural treatments or silvicultural regimes on late-seral herb species diversity and composition. Lindh and Muir (2004) studied responses in 40-year-old Douglas-fir stands that had been thinned 20 years ago to an average residual density of 600 trees/ha. The thinned stands had a higher frequency of late-seral herbs and understory composition that was more similar to old-growth forests than to unthinned stands. Similarly, in 60-110-year-old Douglas-fir stands of western Oregon, Bailey et al. (1998) found that, 7–23 years after thinning, species richness of shade-tolerant forest herbs was similar to that of old-growth forests. In Sierran mixed conifer forests of northern California, Battles et al. (2001) found that the composition of late-seral plant species differed between silvicultural treatments. Group selection (openings of 0.6 ha) and single-tree selection had a greater proportion of late-seral versus early-seral plant species than clearcutting and shelterwood regimes (Battles et al. 2001). In red and eastern white pine forests of central Ontario, Quinby (2000) found that, immediately after shelterwood regeneration cuts ( $\sim 50\%$  basal area removal), microclimate changed to a drier and more light-intensive environment at forest floor, species richness and number of uncommon plant taxa declined, and plant

community composition shifted. In another analysis of the DEMO data, Halpern et al. (2005) found greater compositional change and loss of late-seral species when retention level was reduced to 15%. These studies all focused on responses to overstory reduction over both the short and long term. Results from the SEM analysis complement these other studies by identifying causal pathways, suggesting that microclimatic stresses probably outweigh increased availability of belowground resources in influencing late-seral species diversity and composition.

The hypothesis related to aspect was rejected for the 40% retention treatment but not for 15% retention. At 15% retention, overstory cover was reduced to a greater degree, so the topographic effect of aspect on the forest floor would be greater. Likewise, even an aggregate within 15% retention units would receive less protection from adjacent aggregates compared to the 40% retention units. Aspect would therefore have an increasing ameliorating (northeast) or intensifying (southwest) effect as retention level declined to 15% regardless of spatial distribution of trees.

Halpern et al. (2005) proposed that the significant decline of late-seral herbs species at lower retention levels may be attributable to increased physical disturbance and burial beneath logging slash. However, SEMs fitted at the plot level led to rejection of the hypothesis that slash and coarse wood affect late-seral herb species diversity and composition. On the other hand, disturbed mineral soil did have a significantly negative effect, but only in the 15% retention treatments. Evidence for rejecting Hypothesis 4 was therefore mixed, in spite of the mechanisms suggested by Halpern et al. (2005). When simultaneously testing multiple hypotheses in SEMs, other processes such as microclimatic stresses and topographic effects might be stronger and more critical in affecting late-seral herbs, washing out any burial effect. Disturbed mineral soil probably represents less the burial effect than direct disturbance of forest floor that otherwise may have hosted late-seral plants. In 15% retention, logging disturbances such as displacement, uprooting and soil compaction would undoubtedly be more intensive, so the models likely captured this effect.

Support for Hypothesis 5 depended on retention level and type of logging residuals. The predictors for production of slash were consistent for both retention levels, but the predictors for production of coarse wood and disturbed mineral soil were significant only for 15% retention, perhaps because logging intensity was greater at this level. Nevertheless, the SEM models indicated weak causal links between degree of overstory tree removal and the amount of coarse wood and disturbed mineral soil in both retention levels. The added error covariances during model building may be attributable to one or two possible causes: (1) external factors not represented in the model, and (2) direct effects between two types of logging residuals. The first explanation is more likely than the second. The direct path Tree Density  $\rightarrow$  Slash Cover suggested that for 40% retention level, the local reduction in tree density was a significant predictor of slash cover.

In another analysis of the DEMO data, Halpern and McKenzie (2001) concluded that yarding method and harvesting operations explained more of the variation in ground disturbance than level or pattern of retention. The SEM analysis supported this conclusion and Hypothesis 5, but only for slash depth and disturbed mineral soil in the 15% retention treatment. Harvesting method apparently plays a lesser role in 40% retention, so yarding system may not contribute to variation in the production of slash and hence to late-seral herbs diversity and community similarity. Within the 15% retention units, the lack of effects on slash cover and coarse wood, and the positive effect of helicopter logging on slash depth, were unexpected. Shovel yarding in the WF and PH blocks (Table 4.8) was expected to result in greater slash and coarse woody debris because tops and branches were left where the trees were felled and because non-merchantable trees were felled and left. However, the amount of slash cover and coarse wood volume were equally high among blocks. Slash depth at the helicopter-logged LW block was notably higher than at other blocks, which raised the possibility that the variable for helicopter logging (HARVEST) may be indicating other factors associated with the LW site.

### **4.5 CONCLUSIONS**

The aim of this paper was to provide a conceptual framework for SEM applied to forestry. The goal was to establish a basic understanding of principles, limitations and assumptions of SEM so that potential applications can be recognized. Many other excellent and more comprehensive sources are available for deeper insights into the methodology. Data from natural systems typically present analysis issues, such as non-normality in endogenous variables, nonlinear relationships, ill-scaled variancecovariance matrices and nested study designs. SEM continues to evolve and address special issues such as modeling binary, ordinal and categorical data (Muthén 1983, 1984), and accommodating interactions and curvilinear relationships between hypothetical constructs (Kenny and Judd 1984, Ping 1996). Recent years have seen the development of different model estimation procedures (see Schumacker and Lomax 2004), including a Bayesian approach to SEM (Lee 2007). The latter has been an attractive alternative for modeling hierarchical SEMs.

The overstory-understory relationship involves cascading effects of higher layers of forest vegetation on successively lower layers. The inter-relationships can be complex and difficult to test experimentally, but considerable theory is available to embed a set of working hypotheses within the conceptual framework. The overstoryunderstory model is therefore a fitting illustration of the application of SEM to understand vertical forest structure, and it builds on previous univariate models and exploratory analyses of forest composition. The modeling approach involved a mix of confirmatory and exploratory approaches, with the former specifying hypotheses and associated processes, and the latter imposing tentative structures to improve the model. Because Maximum Likelihood is a full-information method, all processes are considered simultaneously during model estimation. As a result, we gain insights into the processes behind the observed patterns; particularly the importance of light attenuation, belowground resource competition, microclimatic effects associated with aspect, and stand development over time. Most of these processes have direct effects on late-seral herb cover, and more have indirect effects mediated through understory vegetation. Although the final model fitted the data reasonably well, the results do not prove the causal relationships hypothesized in the model, and it is possible that

equivalent models exist that fit the data just was well. Finally, it is important to note that the effects were explained in terms of processes that we did not directly measure, e.g., light attenuation was represented by amount of overstory crown cover rather than a measure of amount of light. Hence, there should be a degree of reservation in inferring processes from effects.

The second SEM analysis explored potential causes of immediate responses of late-seral herb species diversity and composition to variable-retention harvesting. Previous analyses of the same data by Halpern and McKenzie (2001) and Halpern et al. (2005) focused on net effects of nominal treatments by applying ANOVA and speculating about causal mechanisms, e.g., burial by logging slash and microclimatic stresses from overstory reduction. SEM is a logical next step to formulate hypothesized mechanisms and simultaneously testing them with approaches such as multigroup analysis. Results from the SEMs suggested that microclimatic stresses likely played a more important role in affecting late-seral herbs than logging disturbances. By validating the developed models with data sources from other geographical regions or populations, we could evaluate the generality of the models or identify other processes that might be operating. This study has shown the value of not limiting analysis to any one approach to statistical modeling, and how different methods can complement one another to provide more comprehensive insights and understanding of the complex processes driving managed forest ecosystems.

Table 4.1 Variance of observed variables and measurement error as average percent variance of the variable, with corresponding lower (0.025 quantile) and upper (0.975 quantile) limits. The average percent measurement errors were obtained from three expert opinion surveys.

Observed	Variance	Measurement Error	Lower Limit	Upper Limit
Variable		(%)	(%)	(%)
LSHERB	307.68	8.3	5.7	19.3
UNDER	1034.79	8.3	5.7	17.7
FLITTER	82.15	8.3	5.7	12.7
CLITTER	52.74	8.3	5.7	12.7
TREE	265.50	10.7	8.3	23.3
log(TPH)×10	64.67	4.0	2.7	7.3
ASPECT×10	49.34	4.0	3.0	7.7
QMD	376.21	4.0	2.7	7.3
AGE	1680.37	10.0	5.0	15.0

Table 4.2 Estimated direct effects on endogenous latent variables in the overstoryunderstory SR model, along with corresponding standard errors and p-values (see Figure 4.8). Statistical tests were based on a *t*-distribution with 1181 degrees of freedom to adjust for the effects of the nesting structure on standard errors.

Path Model	Direct Effect	Standard Error	P-value
Late-Seral Herb Cover (LSHERB) =			
Overstory Cover (TREE)	-0.192	0.075	0.010
Stand Density (log(TPH)×10)	-0.085	0.167	0.610
Stand Age (AGE)	0.155	0.044	0.000
Aspect (ASPECT×10)	0.424	0.163	0.010
Fine Litter (FLITTER)	0.276	0.177	0.118
Coarse Litter (CLITTER)	0.209	0.213	0.326
Understory Cover (UNDER)	0.175	0.047	0.000
Fine Litter (FLITTER) = $(TDEE)$	0.110	0.022	0.000
Overstory Cover (TREE)	0.119	0.033	0.000
Mean DBH (QMD)	-0.071	0.046	0.127
Stand Density (log(TPH)×10)	-0.391	0.134	0.004
Aspect (ASPECT×10)	0.115	0.053	0.031
Coarse Litter (CLITTER) =			
Stand Density (log(TPH)×10)	0.164	0.053	0.002
Stand Age (AGE)	0.027	0.011	0.014
Understory Cover (LINDER) =			
Overstory Cover (TPEE)	0.638	0 165	0.000
Stand Density (log(TPH)×10)	-0.038	0.105	0.000
$A \operatorname{space} (A \operatorname{SDECT}_{10})$	-0.821	0.349	0.019
Aspect (ASPEC1×10)	1.3/0	0.308	0.000
Fine Litter (FLITTER)	1.380	0.288	0.000
Coarse Litter (CLITTER)	0.960	0.335	0.004
Table 4.3 Estimated combined indirect effects on endogenous latent variables in the overstory-understory SR model, along with corresponding standard errors and p-values. Statistical tests were based on a *t*-distribution with 1181 degree of freedom as the effects of the nesting structure on standard errors have been adjusted.

Path Model	Indirect Effect	Standard Error	P-value
Late-Seral Herb Cover (LSHERB) =			
Overstory Cover (TREE)	-0.050	0.053	0.347
Mean DBH (QMD)	-0.037	0.031	0.240
Stand Density (log(TPH)×10)	-0.284	0.094	0.003
Stand Age (AGE)	0.010	0.005	0.057
Aspect (ASPECT×10)	0.301	0.129	0.020
Fine Litter (FLITTER)	0.241	0.085	0.005
Coarse Litter (CLITTER)	0.168	0.084	0.046
Understory Cover (UNDER) =			
Overstory Cover (TREE)	0.164	0.038	0.000
Mean DBH (QMD)	-0.098	0.064	0.128
Stand Density (log(TPH)×10)	-0.383	0.164	0.020
Stand Age (AGE)	0.026	0.011	0.013
Aspect (ASPECT×10)	0.159	0.091	0.079

Table 4.4 Estimated total effects on endogenous latent variables in the overstoryunderstory SR model, along with corresponding standard errors and p-values. Statistical tests were based on a *t*-distribution with 1181 degree of freedom as the effects of the nesting structure on standard errors have been adjusted.

Path Model	Total Effect	Standard Error	P-value
Late-Seral Herb Cover (LSHERB) =			
Overstory Cover (TREE)	-0.242	0.066	0.000
Mean DBH (QMD)	-0.037	0.031	0.240
Stand Density (log(TPH)×10)	-0.370	0.216	0.087
Stand Age (AGE)	0.165	0.043	0.000
Aspect (ASPECT×10)	0.725	0.187	0.000
Fine Litter (FLITTER)	0.518	0.202	0.011
Coarse Litter (CLITTER)	0.377	0.214	0.078
Understory Cover (UNDER)	0.175	0.047	0.000
Fine Litter (FLITTER) =			
Overstory Cover (TREE)	0.119	0.033	0.000
Mean DBH (QMD)	-0.071	0.046	0.127
Stand Density $(log(TPH) \times 10)$	-0.391	0.134	0.004
Aspect (ASPECT×10)	0.115	0.053	0.031
Coarse Litter (CLITTER) =			
Stand Density (log(TPH)×10)	0.164	0.053	0.002
Stand Age (AGE)	0.027	0.011	0.014
Understory Cover (UNDER) =			
Overstory Cover (TREE)	-0.474	0.169	0.005
Mean DBH (QMD)	-0.098	0.064	0.128
Stand Density (log(TPH)×10)	-1.204	0.339	0.000
Stand Age (AGE)	0.026	0.011	0.013
Aspect (ASPECT×10)	1.537	0.420	0.000
Fine Litter (FLITTER)	1.380	0.287	0.000
Coarse Litter (CLITTER)	0.960	0.335	0.004

Table 4.5 Estimated direct effects on the endogenous latent variables in the final overstory-understory equivalent SR model, along with corresponding standard errors and p-values (see Figure 4.9). Statistical tests were based on a *t*-distribution with 1181 degrees of freedom as the effects of the nesting structure on standard errors have been adjusted.

Path Model	Direct Effect	Standard Error	P-value
Late-Seral Herb Cover (LSHERB) =			
Overstory Cover (TREE)	-0.191	0.075	0.011
Tree Density $(\log(TPH) \times 10)$	-0.083	0.168	0.620
Stand Age (AGE)	0.156	0.044	0.000
Aspect (ASPECT×10)	0.423	0.162	0.009
Fine Litter (FLITTER)	0.273	0.181	0.132
Coarse Litter (CLITTER)	0.205	0.219	0.351
Understory Cover (UNDER)	0.176	0.047	0.000
Fine Litter (FLITTER) =			
Overstory Cover (TREE)	0.171	0.032	0.000
Tree Size (QMD)	-0.102	0.038	0.007
Tree Density $(log(TPH) \times 10)$	-0.400	0.128	0.002
Aspect (ASPECT×10)	0.000	0.040	0.999
Understory Cover (UNDER)	0.072	0.012	0.000
Coarse Litter (CLITTER) =			
Tree Density $(log(TPH) \times 10)$	0.169	0.055	0.002
Stand Age (AGE)	0.030	0.011	0.007
Understory Cover (UNDER) =			
Overstory Cover (TREE)	-0.508	0.181	0.005
Tree Density (log(TPH)×10)	-0.989	0.391	0.012
Aspect (ASPECT×10)	1.543	0.418	0.000
Coarse Litter (CLITTER)	-0.284	0.281	0.313

Table 4.6 Estimated combined indirect effects on the endogenous latent variables in the final overstory-understory equivalent SR model, along with corresponding standard errors and p-values. Statistical tests were based on a *t*-distribution with 1181 degree of freedom as the effects of the nesting structure on standard errors have been adjusted.

Path Model	Indirect Effect	Standard Error	P-value
Late-Seral Herb Cover (LSHERB) =			
Overstory Cover (TREE)	-0.053	0.057	0.353
Tree Size (QMD)	-0.028	0.023	0.235
Tree Density $(log(TPH) \times 10)$	-0.277	0.092	0.003
Stand Age (AGE)	0.005	0.006	0.478
Aspect (ASPECT×10)	0.302	0.128	0.018
Coarse Litter (CLITTER)	-0.056	0.053	0.295
Understory Cover (UNDER)	0.020	0.014	0.165
Fine Litter (FLITTER) =			
Overstory Cover (TREE)	-0.036	0.016	0.022
Tree Density $(log(TPH) \times 10)$	-0.074	0.028	0.007
Stand Age (AGE)	-0.001	0.001	0.327
Aspect (ASPECT×10)	0.111	0.037	0.003
Coarse Litter (CLITTER)	-0.020	0.020	0.296
Understory Cover (UNDER) =			
Tree Density (log(TPH)×10)	-0.048	0.047	0.308
Stand Age (AGE)	-0.009	0.009	0.338

Table 4.7 Estimated total effects on the endogenous latent variables in the final overstory-understory equivalent SR model, along with corresponding standard errors and p-values. Statistical tests were based on a *t*-distribution with 1181 degree of freedom as the effects of the nesting structure on standard errors have been adjusted.

Path Model	Total Effect	Standard Error	P-value
Late-Seral Herb Cover (LSHERB) =			
Overstory Cover (TREE)	-0.244	0.069	0.000
Tree Size (QMD)	-0.028	0.023	0.235
Tree Density $(log(TPH) \times 10)$	-0.361	0.216	0.095
Stand Age (AGE)	0.160	0.043	0.000
Aspect (ASPECT×10)	0.725	0.187	0.000
Fine Litter (FLITTER)	0.273	0.181	0.132
Coarse Litter (CLITTER)	0.149	0.216	0.490
Understory Cover (UNDER)	0.196	0.046	0.000
Fine Litter (FLITTER) =			
Overstory Cover (TREE)	0 135	0.031	0.000
Tree Size (OMD)	-0.102	0.031	0.000
Tree Density (log(TPH)×10)	-0 474	0.120	0.000
Stand Age (AGE)	-0.001	0.001	0.327
Aspect (ASPECT×10)	0.111	0.053	0.038
Coarse Litter (CLITTER)	-0.020	0.020	0.296
Understory Cover (UNDER)	0.072	0.012	0.000
Coarse Litter (CLITTER) =			
Tree Density (log(TPH)×10)	0.170	0.055	0.002
Stand Age (AGE)	0.030	0.011	0.007
Understory Cover (UNDER) =			
Overstory Cover (TREE)	-0.508	0.181	0.005
Tree Density (log(TPH)×10)	-1.037	0.400	0.010
Stand Age (AGE)	-0.009	0.009	0.337
Aspect (ASPECT×10)	1.543	0.418	0.000
Coarse Litter (CLITTER)	-0.284	0.281	0.313

Block	HARVEST	Yarding	Non-merchantable	Temporary	Tops Attached	Slash Piled and
		Method	trees felled	Road		Burnt
WF	0	Shovel loader	Yes	Yes	No	Yes
DP	1	Helicopter	No	No	Yes	No
BU	1	Helicopter	No	No	Yes	No
LWS	1	Helicopter	No	No	Yes	No
PH	0	Shovel loader	Yes	Yes	Yes	No

Table 4.8 Harvest method and other activities in each experimental block (adapted from Halpern and McKenzie 2001).

Table 4.9 The 11 observed variables used in modeling and their corresponding units, definitions, minimum, mean and
maximum values. The variables were summarized by 15% retention (i.e., 15%A and 15%D treatments combined) and 40%
retention (i.e., 40%A and 40%D treatments combined).

Observed	Unit	Definition	1:	15% Retention			40% Retention		
Variable			Min	Mean	Max	Min	Mean	Max	
dSHANNON	_	Difference in pre- and post-	-1.83	-0.33	0.87	-1.75	-0.12	0.96	
		harvest Shannon diversity							
		index							
RENKONEN	%	Renkonen similarity index of	0	66.06	100	0	80.53	98.82	
		the community between pre-							
		and post-harvest							
dUNDER	%	Difference in pre- and post-	-89.45	-16.82	12.24	-76.16	-9.70	24.82	
		harvest understory shrubs and							
		trees cover							
dCROWNVOL	m²/ha	Difference in pre- and post-	-560,900	-209,960	167,200	-443,100	-84,959	131,000	
		harvest crown volume							
dTPH	trees/ha	Difference in pre- and post-	-1475	-382	1925	-1450	-232	475	
		harvest tree density							
ASPECT	_	Cosine-transformed aspect	-1	0.02	1	-1	0.02	1	
SLASHCOV	%	Post-harvest slash cover	0	65.62	100	0	45.15	100	
SLASHDEP	cm	Post-harvest slash depth	0	9.19	45.62	0	5.50	37.4	
CWDVOL	m³/ha	Post-harvest coarse woody	0	213	1651	0	153	1446	
		debris volume							
DISTSOIL	%	Post-harvest disturbed	0	4.46	49	0	4.59	37.92	
		mineral soil cover							

Table 4.10 Estimated direct effects on the endogenous latent variables, difference in diversity (dSHANNON) and composition (RENKONEN), and on the composite variable, Logging Residuals, in the final SR model for 15% retention, along with corresponding standard errors and p-values (see Figure 4.15).

Path Model	Direct Effect	Standard Error	P-value
Diversity (dSHANNON) =			
Crown Volume (dCROWNVOL/10000)	1.206	0.478	0.012
Aspect (ASPECT×10)	1.628	0.430	0.000
Understory Cover (dUNDER)	0.049	0.321	0.878
Disturbed Soil (DISTSOIL×5)	-0.272	0.095	0.004
Logging Residuals	-0.130	0.168	0.439
Composition (RENKONEN) =			
Crown Volume (dCROWNVOL/10000)	0 744	0.145	0.000
$\Delta \text{spect} (\Delta \text{SPECT} \times 10)$	1 093	0.145	0.000
Understory Cover (dUNDER)	-0.168	0.130	0.000
Disturbed Soil (DISTSOIL×5)	-0.068	0.051	0.170
Logging Residuals	-0.088	0.135	0.515
Understory Cover (dUNDER) =			
Crown Volume (dCROWNVOL/10000)	-0.115	0.257	0.654
Aspect (ASPECT×10)	-1.391	0.230	0.000
Tree Density (dTPH/10)	0.091	0.086	0.290
Slash Cover (SLASHCOV) =			
Understory Cover (dUNDER)	-0 566	0.150	0.000
Crown Volume (dCROWNVOL/10000)	-1 454	0 198	0.000
HARVEST	10.79	6.531	0.098
Slash Depth (SLASHDEP) =			
Understory Cover (dUNDER)	-0.611	0.112	0.000
Crown Volume (dCROWNVOL/10000)	-0.758	0.209	0.000
HARVEST	25.22	3.601	0.000
Coarse Wood (CWDVOL/10) =			
Tree Density (dTPH/10)	-0.224	0.081	0.006
HARVEST	10.38	7.891	0.188

Table 4.10 (Continued).

Path Model	Direct Effect	Standard Error	P-value
Disturbed Soil (DISTSOIL $\times$ 5) =			
Tree Density (dTPH/10)	-0.190	0.039	0.000
HARVEST	-14.32	6.864	0.037
Logging Residuals =			
Slash Cover (SLASHCOV)	-1.169	1.066	0.273
Slash Depth (SLASHDEP)	1.000	_	_
Coarse Wood (CWDVOL/10)	0.249	1.117	0.824

Table 4.11 Estimated combined indirect effects on the endogenous latent variables difference in diversity (dSHANNON) and composition (RENKONEN) in the final SR model for 15% retention, along with corresponding standard errors and p-values.

Path Model	Indirect Effect	Standard Error	P-value
Diversity (dSHANNON) =			
Crown Volume (dCROWNVOL/10000)	-0.127	0.325	0.695
Aspect (ASPECT×10)	-0.060	0.371	0.873
Understory Cover (dUNDER)	-0.006	0.083	0.938
Tree Density (dTPH/10)	0.063	0.051	0.216
Slash Cover (SLASHCOV)	0.152	0.305	0.619
Slash Depth (SLASHDEP)	-0.130	0.168	0.439
Coarse Wood (CWDVOL/10)	-0.032	0.152	0.832
HARVEST	1.930	3.190	0.545
Composition (RENKONEN) =			
Crown Volume (dCROWNVOL/10000)	-0.063	0.152	0.679
Aspect (ASPECT×10)	0.239	0.150	0.110
Understory Cover (dUNDER)	-0.004	0.049	0.929
Tree Density (dTPH/10)	0.002	0.032	0.945
Slash Cover (SLASHCOV)	0.103	0.134	0.444
Slash Depth (SLASHDEP)	-0.088	0.135	0.515
Coarse Wood (CWDVOL/10)	-0.022	0.072	0.761
HARVEST	-0.356	1.836	0.846

Table 4.12 Estimated total effects on the endogenous latent variables difference in diversity (dSHANNON) and composition (RENKONEN) in the final SR model for 15% retention, along with corresponding standard errors and p-values.

Path Model	Total Effect	Standard Error	P-value
Diversity (dSHANNON) =			
Crown Volume (dCROWNVOL/10000)	1.079	0.201	0.000
Aspect (ASPECT×10)	1.569	0.409	0.000
Understory Cover (dUNDER)	0.043	0.269	0.873
Tree Density (dTPH/10)	0.063	0.051	0.216
Disturbed Soil (DISTSOIL×5)	-0.272	0.095	0.004
Slash Cover (SLASHCOV)	0.152	0.305	0.619
Slash Depth (SLASHDEP)	-0.130	0.168	0.439
Coarse Wood (CWDVOL/10)	-0.032	0.152	0.832
HARVEST	1.930	3.190	0.545
Composition (RENKONEN) -			
Composition (REINCONEN) –	0 (01	0 1 4 7	0.000
Crown volume (dCROWNVOL/10000)	0.681	0.14/	0.000
Aspect (ASPECT×10)	1.332	0.261	0.000
Understory Cover (dUNDER)	-0.172	0.095	0.070
Tree Density (dTPH/10)	0.002	0.032	0.945
Disturbed Soil (DISTSOIL×5)	-0.068	0.051	0.177
Slash Cover (SLASHCOV)	0.103	0.134	0.444
Slash Depth (SLASHDEP)	-0.088	0.135	0.515
Coarse Wood (CWDVOL/10)	-0.022	0.072	0.761
HARVEST	-0.356	1.836	0.846

Table 4.13 Estimated direct effects on the endogenous latent variables, difference in diversity (dSHANNON) and composition (RENKONEN), and on the composite variable, Logging Residuals, in the final SR model for 40% retention, along with corresponding standard errors and p-values (see Figure 4.16).

Path Model	Direct Effect	Standard Error	P-value
Diversity (dSHANNON) =			
Crown Volume (dCROWNVOL/10000)	0.777	0.355	0.029
Aspect (ASPECT×10)	0.422	0.294	0.152
Understory Cover (dUNDER)	-0.118	0.175	0.499
Disturbed Soil (DISTSOIL×5)	-0.078	0.046	0.092
Logging Residuals	0.214	0.186	0.252
Composition (RENIZONEN) -			
Composition (KENKONEN) –	0 505	0.164	0.002
Crown volume ( $dCROWNVOL/10000$ )	0.305	0.164	0.002
Aspect (ASPEC1×10)	0.225	0.156	0.149
Understory Cover (dUNDER)	-0.119	0.032	0.000
Disturbed Soil (DISTSOIL×5)	-0.048	0.027	0.073
Logging Residuals	-0.061	0.091	0.506
Understory Cover (dUNDER) =			
Crown Volume (dCROWNVOL/10000)	0.879	0.525	0.094
Aspect (ASPECT×10)	-0.583	0.192	0.002
Tree Density (dTPH/10)	-0.125	0.124	0.316
Slash Cover (SLASHCOV) =			
Understory Cover (dUNDER)	0.687	0.125	0.000
Crown Volume (dCPOWNVOL /1000)	-0.087	0.123	0.000
Tree Density (dTDH/10)	-1.074	0.042	0.000
Inter Defisity (ulf H/10)	-0.208	0.078	0.001
HARVESI	-3.112	9.439	0.742
Slash Depth (SLASHDEP) =			
Understory Cover (dUNDER)	-0.783	0.258	0.002
Crown Volume (dCROWNVOL/10000)	-1.401	0.305	0.000
HARVEST	7.982	4.780	0.095
Coarse Wood (CWDVOL /10) -			
$T_{ree} D_{ansity} (dTDU/10) =$	0 060	0.046	0.134
	1 002	5 205	0.134
ΠΑΚΥΕδΙ	1.905	5.295	0./19

Table 4.13 (Continued).

Path Model	Direct Effect	Standard Error	P-value	
Disturbed Soil (DISTSOIL $\times$ 5) =				
Tree Density (dTPH/10)	-0.173	0.101	0.086	
HARVEST	-13.05	11.54	0.258	
Logging Residuals =				
Slash Cover (SLASHCOV)	-1.079	0.471	0.022	
Slash Depth (SLASHDEP)	1.000	_	_	
Coarse Wood (CWDVOL/10)	-0.078	0.269	0.773	

Path Model	Indirect Effect	Standard Error	P-value
Diversity (dSHANNON) =			
Crown Volume	0.021	0.175	0.905
(dCROWNVOL/10000)			
Aspect (ASPECT×10)	0.074	0.070	0.290
Understory Cover (dUNDER)	-0.009	0.098	0.927
Tree Density (dTPH/10)	0.090	0.059	0.131
Slash Cover (SLASHCOV)	-0.230	0.184	0.210
Slash Depth (SLASHDEP)	0.214	0.186	0.252
Coarse Wood (CWDVOL/10)	-0.017	0.060	0.781
HARVEST	3.398	2.179	0.119
Composition (RENKONEN) =			
Crown Volume	-0.140	0.082	0.087
(dCROWNVOL/10000)			
Aspect (ASPECT×10)	0.068	0.043	0.114
Understory Cover (dUNDER)	0.003	0.028	0.928
Tree Density (dTPH/10)	0.006	0.035	0.871
Slash Cover (SLASHCOV)	0.065	0.095	0.494
Slash Depth (SLASHDEP)	-0.061	0.091	0.506
Coarse Wood (CWDVOL/10)	0.005	0.276	0.782
HARVEST	-0.059	1.057	0.956

Table 4.14 Estimated combined indirect effects on the endogenous latent variables difference in diversity (dSHANNON) and composition (RENKONEN) in the final SR model for 40% retention, along with corresponding standard errors and p-values.

Table 4.15	Estimated t	total effects of	n the end	logenous	latent var	iables di	fference	in
diversity (a	ISHANNON	I) and compose	sition (R	ENKONE	EN) in the	e final SR	R model t	for
40% retent	tion, along w	ith correspon	ding star	ndard erro	rs and p-	values.		

Path Model	Total Effect	Standard Error	P-value
Diversity (dSHANNON) =	10001211000	Standard 20101	1 / 010/0
Crown Volume (dCROWNVOL/10000)	0.798	0.254	0.002
Aspect (ASPECT×10)	0.496	0.311	0.112
Understory Cover (dUNDER)	-0.127	0.088	0.150
Tree Density (dTPH/10)	0.090	0.059	0.131
Disturbed Soil (DISTSOIL×5)	-0.078	0.046	0.092
Slash Cover (SLASHCOV)	-0.230	0.184	0.210
Slash Depth (SLASHDEP)	0.214	0.186	0.252
Coarse Wood (CWDVOL/10)	-0.017	0.060	0.781
HARVEST	3.398	2.179	0.119
Composition (RENKONEN) =			
Crown Volume (dCROWNVOL/10000)	0.365	0.157	0.020
Aspect (ASPECT×10)	0.293	0.177	0.097
Understory Cover (dUNDER)	-0.117	0.048	0.016
Tree Density (dTPH/10)	0.006	0.035	0.871
Disturbed Soil (DISTSOIL×5)	-0.048	0.027	0.073
Slash Cover (SLASHCOV)	0.065	0.095	0.494
Slash Depth (SLASHDEP)	-0.061	0.091	0.506
Coarse Wood (CWDVOL/10)	0.005	0.276	0.782
HARVEST	-0.059	1.057	0.956



Figure 4.1 A hypothetical path model for abundance of a small mammal species in mature forests. Single arrowheads depict direct effects and double arrowheads depict unanalyzed associations.



Figure 4.2 A hypothetical measurement model for species diversity. The circle depicts a latent variable and the rectangles depict observed variables. Arrows from the latent variable to observed variables represent direct effects and thicker arrows on the left of observed variables represent measurement errors.



Figure 4.3 A SR model for the overstory-understory relationship in mature Douglasfir forests. Measurement errors are fixed with *a priori* values as depicted by the numerical values pointing towards the observed variables. All factor loadings are fixed at a value of 1.0. The exogenous latent variables (and associated observed variables) are: (1) Aspect (ASPECT×10), (2) Overstory Cover (TREE), (3) Tree Size (QMD), (4) Stand Age (AGE), and (5) Stand Density (log(TPH)×10). The endogenous latent variables (and associated observed variables) are: (1) Fine Litter (FLITTER), (2) Coarse Litter (CLITTER), (3) Understory Cover (UNDER), and (4) Late-Seral Herb Cover (LSHERB).



Figure 4.4 An equivalent SR model for the overstory-understory relationship in mature Douglas-fir forests with a different path; i.e., Understory Cover  $\rightarrow$  Fine Litter.



Figure 4.5 Bivariate relationships and Pearson correlations between late-seral herb cover (LSHERB) and its six predictors: fine litter cover (FLITTER), coarse litter cover (CLITTER), understory cover (UNDER), overstory tree cover (TREE), stand density (log(TPH)), and aspect (ASPECT).



Figure 4.6 Bivariate relationships and Pearson correlations between understory cover (UNDER) and its five predictors: fine litter cover (FLITTER), coarse litter cover (CLITTER), overstory tree cover (TREE), stand density (log(TPH)), and aspect (ASPECT).



Figure 4.7 Bivariate relationships and Pearson correlations between fine litter cover (FLITTER) and coarse litter cover (CLITTER) and their corresponding predictors: overstory tree cover (TREE), stand density (log(TPH)), tree size (QMD), and stand age (AGE).



Figure 4.8 The final fitted SR model with unstandardized parameter estimates. The observed variables are omitted to simplify the diagram. Estimated unstandardized path coefficients (direct effects) are placed alongside each path. The squared multiple correlations ( $R^2$ ) are included for each endogenous latent variable. The thickness of a single arrowhead corresponds to the strength of its estimates (based on its standardized estimate). The dashed arrowheads depict non-significant paths. The double arrowheads depict the unstandardized error variance between the two latent variables.



Figure 4.9 The fitted equivalent SR model with unstandardized parameter estimates. The observed variables are omitted to simplify the diagram (refer to caption of Figure 4.8).



Figure 4.10 Fitted equivalent SR model with the lower limit (0.025 quantile) of percent measurement errors (refer to Table 4.1 and caption of Figure 4.8).



Figure 4.11 Fitted equivalent SR model with the upper limit (0.975 quantile) of percent measurement errors (refer to Table 4.1 and caption of Figure 4.8).



Figure 4.12 A SR model for immediate post-harvest responses of late-seral herbs to variable-retention harvesting. *A priori* measurement errors expressed as a percentage of the observed variance are depicted by the values pointing towards the observed variables. All factor loadings are fixed at a value of 1.0. The composite variable Logging Residuals is depicted with a hexagon, and Slash Depth determines its scale of measurement by having its path coefficient fixed at a value of 1.0.



Figure 4.13 Bivariate relationships and Pearson correlations (*r*) between difference in pre- and post-harvest species diversity (dSHANNON) and four direct predictors, and between community similarity (RENKONEN) and the same four predictors in the 15% retention treatments (i.e., 15%A and 15%D treatments combined). Predictors include difference in pre- and post-harvest crown volume (dCROWNVOL), difference in pre- and post-harvest understory shrubs and trees cover (dUNDER), aspect (ASPECT), and disturbed mineral soil (DISTSOIL).



Figure 4.14 Bivariate relationships and Pearson correlations (*r*) between difference in pre- and post-harvest species diversity (dSHANNON) and four direct predictors, and between community similarity (RENKONEN) and the same four predictors in the 40% retention treatments (i.e., 40%A and 40%D treatments combined). Predictors include difference in pre- and post-harvest crown volume (dCROWNVOL), difference in pre- and post-harvest understory shrubs and trees cover (dUNDER), aspect (ASPECT), and disturbed mineral soil (DISTSOIL).



Figure 4.15 The final 15% retention SR model with unstandardized parameter estimates. The observed variables, insignificant parameter estimates, and error covariances are omitted to simplify the diagram. Refer to Figure 4.8 for further description of the symbols.



Figure 4.16 The final 40% retention SR model with unstandardized parameter estimates. The observed variables, insignificant parameter estimates, and error covariances are omitted to simplify the diagram. Refer to Figure 4.8 for further description of the symbols.



Figure 4.17 The alternative 40% retention SR model after removal of the HARVEST variable and with unstandardized parameter estimates. The observed variables, insignificant parameter estimates, and error covariances are omitted to simplify the diagram. Refer to Figure 4.8 for further description of the symbols.



Figure 4.18 Bivariate relationships and Pearson correlations (*r*) between: (A) preharvest understory shrub and tree cover versus aspect for 15% retention, (B) postharvest versus pre-harvest understory shrub and tree cover for 15% retention, (C) preharvest understory shrub and tree cover versus aspect for 40% retention, and (D) postharvest versus pre-harvest understory shrub and tree cover for 40% retention. Pre- and post-harvest understory shrub and tree cover values were identified by three classes of ASPECT: ASPECT < -0.5 (southwestern aspect,  $\blacksquare$ ), ASPECT between -0.5 and 0.5 (×), and ASPECT > 0.5 (northeastern aspect,  $\blacktriangle$ ).



Figure 4.19 The reduced final 15% retention SR model after removal of the composite variable Logging Residuals and with unstandardized parameter estimates. The observed variables, insignificant parameter estimates, and error covariances are omitted to simplify the diagram. Refer to Figure 4.8 for further description of the symbols.



Figure 4.20 The reduced final 40% retention SR model after removal of the composite variable Logging Residuals and with unstandardized parameter estimates. The observed variables, insignificant parameter estimates, and error covariances are omitted to simplify the diagram. Refer to Figure 4.8 for further description of the symbols.



Figure 4.21 The reduced alternative 40% retention SR model after removal of composite variables Logging Residuals and HARVEST, and with unstandardized parameter estimates. The observed variables, insignificant parameter estimates, and error covariances are omitted to simplify the diagram. Refer to Figure 4.8 for further description of the symbols.
# **5 CHAPTER 5: CONCLUSIONS**

The overall goal of this dissertation was to synthesize information about responses of multiple forest taxa to variable-retention harvesting. As is typical of large experiments on forest biodiversity, statistical methods were needed for making inferences about distributions with excessive zero observations and about systems with complex interactions. Each chapter featured a distinctive methodology for addressing a specific challenge to rigorous statistical analysis. Other approaches may deal effectively with the same issues, and provide slightly different insights into their resolution. This dissertation formed the basis for systematically recognizing an analysis problem and assessing the suitability of alternative statistical models for accommodating it. By no means were the approaches explored exhaustive or comprehensive. Much ground remains to be explored in a future study.

In Chapter 2, known data generating processes created a variety of simulated data distributions that were then modeled with zero modified models. The relative performance of zero-inflated and hurdle models could then be evaluated by their ability to fit the simulated distributions and estimate parameters in the data generating process. Results from this simulation analysis suggested the need to exercise caution when fitting these models to data representing response of species with low frequencies to habitat variables; under these conditions, parameter estimates were unstable. Results from the simulation analysis were also discouraging with regard to the efficacy of goodness of fit statistics for inferring underlying processes that generated a set of data. A risk of misidentifying the dgp is misinterpretation of the

role that habitat predictors and hence silvicultural treatments have on a species. The choice of the most appropriate statistical model should be made to the fullest extent possible in the context of known ecological processes and the spatial and temporal scale of the study.

In Chapter 3, habitat associations were modeled for four forest floor small mammal species with seven models specified as Generalized Linear Models (GLMs) and/or Generalized Linear Mixed Models (GLMMs) in a Bayesian framework. The GLMs have advanced our understanding of treatment effects by conditioning mean captures within a treatment unit on habitat variables, under the conditions of both high and low capture rates. The GLMMs further refined this understanding by recognizing the relationship between captures and habitat variables at several levels, including individual treatment units and blocks. A main consequence of not accounting for nested subsampling and random block effects in the GLMs was bias in the variance of estimates and, hence, improper statistical inferences about influential treatments and habitat conditions. Interpreting results from multiple models provided an opportunity to observe the level of consistency in estimated effects, drawing attention to those predictors requiring further clarification.

In Chapter 4, a conceptual framework for applying Structural Equation Modeling (SEM) to forestry was presented. The methodology was then used to model overstory-understory relationships of late-seral herb species both under mature, undisturbed forest conditions and immediately after disturbance created by variableretention harvesting. A basic understanding of principles, limitations and assumptions of SEM was established, so that potential applications can be recognized and results from SEM correctly interpreted. Under mature forest conditions, overstory vegetation structure controls the late-seral herb community through light attenuation, belowground resource competition, and structural changes associated with increasing stand age, as well as through microclimatic effects associated with aspect. After variable-retention harvesting, SEMs indicated that microclimatic stresses likely played a more important role in affecting late-seral herb community than did logging disturbances.

A recurring theme throughout the dissertation is the advantage of multiple perspectives provided by a range of statistical models. Different analyses can complement one another and provide more comprehensive insights into the complex processes driving managed forest ecosystems.

# **5.1 FUTURE RESEARCH**

# 5.1.1 Simulation Study with Covariates

Management of biodiversity in managed forest ecosystems requires identifying the abundance of species conditional on habitat structure. The simulation analysis established the biased and highly unstable parameter estimates of zero-inflated and hurdle models when true mean abundance was low; however, the degree to which these results could be extended to conditional distributions that predict low mean abundance at a given level of one or more covariates was not addressed. If the observed bias and instability do occur, it is not clear how they would propagate through a series of conditional distributions with increasing mean abundance. Other issues including the number and choice of covariates and required sample sizes may also affect the relative performances of these models with covariates. Many empirical studies were based on conditional distributions, suggesting that a carefully crafted simulation study that includes covariates would be a valuable comparison. Such a study would also yield a better understanding of the robustness of zero-modified models.

## 5.1.2 Alternative Model Specifications for Infrequent Species

From the perspective of finite mixture models, no restrictions are necessary on the level of mixing for the zero-inflated and hurdle models. Therefore, one could declare the definitions of p and  $\pi$  according to the practical issue being addressed or the assumed properties of the statistical model. Furthermore, there is also no restriction on the number and type of distributions for the mixing. Undoubtedly, increased flexibility has the price of increased complexity in model specification and estimation. Future research could first focus on establishing a body of theory behind mechanisms driving the observed count of infrequent species, and explore the utility of N-mixture model suggested by Royle (2004) and Royle et al. (2005). If finite mixture models do not fit the proposed theories, other tools such as small area statistics (SAS) (Rao 2003), or extreme value models (Cole 2001) are also an option. However, application of these tools has been limited in natural sciences, so future studies may have to focus on understanding the properties and applicability of these models.

# **5.1.3 Zero-Modified Mixed Effects Models**

Mixed effects models establish inference validity by properly accounting for non-independence in the data. Although specification of a zero-modified mixed effects model is considerably more complex, the advent of Bayesian modeling has facilitated the research into this area, particularly in other disciplines such as medical research. These methods will likely be more fully developed, and future research could facilitate their transfer to the natural sciences.

### 5.1.4 Structural Equation Modeling

Structural Equation Modeling (SEM) in the dissertation was only cursory. Although SEM has as long a history as multiple regression and ANOVA, the methodology is still not that well known in the natural sciences. The methodology will likely continue to grow in this field given its intuitive appeal for modeling complex forest ecosystems. However, many statistical issues emerge from its application in natural sciences, particularly required sample size and its ability to accommodate the structure of designed experiments. A conventional SEM that is geared towards psychometric and econometric data analysis may be easily adapted to studying natural systems and complex designs from field experiments. However, a Bayesian approach to SEM (Lee 2007) may be more effective in accounting for complex study designs and improving inferences. Future studies could focus on further understanding of conventional SEM and its expansion under a Bayesian framework.

The current applications focused on a single guild of forest taxa; i.e., late-seral herb species. Future studies could consider building models for several interacting guilds, complex food webs, and effects of variable-retention harvesting on these interactions and webs. However, building such a model for datasets like those generated by DEMO will require reconciliation of the differing designs by which different taxa were sampled. Perhaps the most straightforward SEM model possible for the DEMO study would account for the effects of variable-retention harvesting on vegetation structures and, in turn, the effect of residual vegetation structure on forest floor small mammals. This could be explored by extending the GLMs and GLMMs in Chapter 3 to an SEM context. This modeling would help synthesize information on simultaneous responses of multiple taxa to variable-retention harvesting.

#### BIBLIOGRAPHY

- Affleck, D.L.R. 2006. Poisson mixture models for regression analysis of stand-level mortality. Canadian Journal of Forest Research 36, 2994–3006.
- Anderson, D.R. 2001. The need to get the basics right in wildlife field studies. Wildlife Society Bulletin 29, 1294–1297.
- Anderson, R.C., Loucks, O.L., and Swain, A.M. 1969. Herbaceous response to canopy cover, light intensity, and throughfall precipitation in coniferous forests. Ecology 50, 255–263.
- Asparouhov, T., and Muthén, B. 2006. Multilevel modeling of complex survey data. Proceedings of the Joint Statistical Meeting in Seattle, August 2006. ASA Section on Survey Research Methods, 2718-2726.
- Aubry, K.B., Amaranthus, M.P., Halpern, C.B., White, J.D., Woodward, B.L., Peterson, C.E., Lagoudakis, C.A., and Horton, A.J. 1999. Evaluating the effects of varying levels and patterns of green-tree retention: experimental design of the DEMO study. Northwest Science 73 (special issue), 12–26.
- Aubry, K.B., Halpern, C.B., and Peterson, C.E. 2009. Variable-retention harvests in the Pacific Northwest: a review of short-term findings from the DEMO study. Forest Ecology and Management 258, 398–408.
- Austin, M.P., and Meyers, A.J. 1996. Current approaches to modelling environmental niche of *Eucalypts*: implications for management of forest biodiversity. Forest Ecology and Management 85, 95–106.
- Bailey, J.D., Mayrsohn, C., Doescher, P.S., St. Pierre, E., and Tappeiner, J.C. 1998.Understory vegetation in old and young Douglas-fir forests of western Oregon. Forest Ecology and Management 112, 289–302.
- Bailey, J.D., and Tappeiner, J.C. 1998. Effects of thinning on structural development in 40- to 100-year-old Douglas-fir stands in western Oregon. Forest Ecology and Management 108, 99–113.
- Banerjee, S., Carlin, B.P., and Gelfand, A.E. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, Boca Raton FL, USA. 452 p.

- Barbier, S., Gosselin, F., and Balandier, P. 2008. Influence of tree species on understory vegetation diversity and mechanism involved – a critical review for temperate and boreal forests. Forest Ecology and Management 254, 1–15.
- Bartemucci, P., Messier, C., and Canham, C.D. 2006. Overstory influences on light attenuation patterns and understory plant community diversity and composition in southern boreal forests of Quebec. Canadian Journal of Forest Research 36, 2065–2079.
- Battles, J.J., Shlisky, A.J., Barrett, R.H., Heald, R.C., and Allen-Diaz, B.H. 2001. The effects of forest management on plant species diversity in a Sierran conifer forest. Forest Ecology and Management 146, 211–222.
- Baughman, A.L. 2007. Mixture model framework facilitates understanding of zeroinflated and hurdle models for count data. Journal of Biopharmaceutical Statistics 17, 943–946.
- Bauhus, J., Puettmann, K., and Messier, C. 2009. Silviculture for old-growth attributes. Forest Ecology and Management 258, 525–537.
- Beers, T.W., Dress, P.E., and Wensel, L.C. 1966. Aspect transformation in site productivity research. Journal of Forestry 64, 691–692.
- Berger, A.L., and Puettmann, K.J. 2000. Overstory composition and stand structure influence herbaceous plant diversity in the mixed aspen forest of northern Minnesota. American Midland Naturalist 143, 111–125.
- Bollen, K.A. 1989. *Structural Equations with Latent Variables*. John Wiley & Sons, New York, NY, USA. 514 p.
- Bowman, J., Forbes, G. J., and Dilworth, T. G. 2001. The spatial component of variation in small-mammal abundance measured at three scales. Canadian Journal of Zoology 79, 137–144.
- Brooks, S., and Gelman, A. 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7, 434–456.
- Browne, M.W., and Cudeck, R. 1993. Alternative ways of assessing model fit. In: *Testing Structural Equation Models*, Bollen, K.A., and J.S. Lang (eds.) pp. 294–316. Sage Publications, Newbury Park, CA, USA.

- Browne, W., and Draper, D. 2000. Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. Computational Statistics 15, 391–420.
- Burnham, K.P., and Anderson, D.R. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2<sup>nd</sup> ed. Springer, New York, NY, USA.
- Burnham, K.P., and Overton, W.S. 1979. Robust estimation of population size when capture probabilities vary among animals. Ecology 60, 927–936.
- Cameron, A.C., and Trivedi, P.K. 1998. *Regression Analysis of Count Data*. 1<sup>st</sup> ed. Cambridge University Press, Cambridge, UK.
- Cannell, M.G.R., and Grace, J. 1993. Competition for light: detection measurement, and quantification. Canadian Journal of Forest Research 23, 1969–1979.
- Carey, A.B., and Johnson, M.L. 1995. Small mammals in managed, naturally young, and old-growth forests. Ecological Applications 5, 336–352.
- Chen, J. 1991. Edge effects: microclimatic pattern and biological responses in oldgrowth Douglas-fir forests. Dissertation. University of Washington, Seattle, WA, USA.
- Chib, S., and Winkelmann, R. 2001 Markov Chain Monte Carlo analysis of correlated count data. Journal of Business and Economic Statistics 19, 428–435.
- Cody, M.L. 1986. Diversity, rarity, and conservation in Mediterranean climate regions. In: *Conservation Biology: The Science of Scarcity, and Diversity*, M. Soulé (ed.) pp. 122–152. Sinauer Associates, Sunderland, MA, USA.
- Cohen, A.C. 1963. Estimation in mixtures of discrete distributions. In: *Proceedings of the International Symposium on Discrete Distributions*, Montreal, Quebec, Canada.
- Cole, S. 2001. An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag, New York, NY, USA.
- Cole, E.C., McComb, W.C., Newton, M., Leeming, J.P., and Chambers, C.L. 1998. Response of small mammals to clearcutting, burning, and glyphosate application in the Oregon coast range. Journal of Wildlife Management 62, 1207–1216.

- Congdon, P. 2003. *Applied Bayesian Modelling*. 1<sup>st</sup> ed. John Wiley & Sons, Chichester, UK. 478 p.
- Congdon, P. 2005. *Bayesian Models for Categorical Data*. 2<sup>nd</sup> ed. John Wiley & Sons, Chichester, UK. 425 p.
- Congdon, P. 2006. *Bayesian Statistical Modelling*. John Wiley & Sons, Chichester, UK. 573 p.
- Corn, P.S., and Bury, R.B. 1990. Sampling methods for terrestrial amphibians and reptiles. USDA Forest Service General Technical Report PNW-GTR-256, Portland, OR, USA. 34 p.
- Cragg, J.G. 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica 39, 829–844.
- Craig, V.J., Klenner, W., Feller, M.C., and Sullivan, T.P. 2006. Relationships between deer mice and downed wood in managed forests of southern British Columbia. Canadian Journal of Forest Research 36, 2189–2203.
- Cunningham, R.B., and Lindenmayer, D.B. 2005. Modeling count data of rare species: some statistical issues. Ecology 86, 1135–1142.
- De Vries, P.G. 1986. Sampling Theory for Forest Inventory: a Teach-yourself Course. Springer-Verlag, Berlin, Germany. 399 p.
- DeGranpré, L., Gagnon, D., and Bergeron, Y. 1993. Changes in the understory of Canadian southern boreal forest after fire. Journal of Vegetation Science 4, 803–810.
- Dias, P. C. 1996. Sources and sinks in population biology. Trends in Ecology & Evolution 11, 326–330.
- Dubrasich, M.E., Hann, D.W., Tappeiner, J.C. 1997. Methods for evaluating crown area profiles of forest stands. Canadian Journal of Forest Research 27, 385–392.
- Dupouey, J.L., Dambrine, E., Laffite, J.D., and Moares, C. 2002. Irreversible impact of past land use on forest soils and biodiversity. Ecology 83, 2978–2984.
- Dzwonko, Z., and Gawroński, S. 2002. Influence of litter and weather on seedling recruitment in a mixed oak-pine woodland. Annals of Botany 90, 245-251.

- Elith, J., and Burgman, M.A. 2002. Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. In: *Predicting Species Occurrences: Issues of Scale and Accuracy*, Scott, J.M., Heglund, P.J., Morrison, M., Raphael, M., Haufler, J., and B. Wall (eds.) pp. 303–314. Island Press, Covello, CA, USA.
- Ellsworth, J., Harrington, R., and Fownes, J. 2004. Seedling emergence, growth, and allocation of oriental bittersweet: effects of seed input, seed bank, and forest floor litter. Forest Ecology and Management 190, 255–264.
- Fekedulegn, D., Hicks Jr., R.R., and Colbert, J.J. 2003. Influence of topographic aspect, precipitation and drought on radial growth of four major tree species in an Appalachian watershed. Forest Ecology and Management 177, 409–425.
- Fisher, R.A. 1926. *The Design of Experiments*. 1<sup>st</sup> ed. Oliver and Boyd. Edinburgh, UK. 250 p.
- Fortin, M., and DeBlois, J. 2007. Modeling tree recruitment with zero-inflated models: the example of hardwood stands in southern Québec, Canada. Forest Science 53, 529–539.
- Franklin, J.F., Berg, D.R., Thornburgh, D.A., and Tappeiner, J.C. 1997. Alternative silviculture approaches to timber harvesting variable retention systems. In: *Creating a Forestry for the 21<sup>st</sup> Century*, Kohm, K.A., and J.F. Franklin (eds.) pp. 111–140, Island Press, Washington, DC, USA.
- Gao, K., and Khoshgoftaar, T.M. 2007. A comprehensive empirical study of count models for software fault prediction. IEEE Transactions on Reliability 56, 223–236.
- Garber, S.M., Monserud, R.A., and Maguire, D.A. 2008. Crown recession patterns in three conifer species of the northern Rocky Mountains. Forest Science 54, 633–646.
- Gazol, A., and Ibáñez, R. 2009. Different response to environmental factors and spatial variables of two attributes (cover and diversity) of the understory layers. Forest Ecology and Management 258, 1267–1274.
- Gelman, A., and Hill, J. 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Cambridge, UK. 625 p.

- George, L.O., and Bazzaz, F.A. 1999. The fern understory as an ecological filter: Growth and survival of canopy-tree seedlings. Ecology 80, 846–856.
- Gitzen, R.A., West, S.D., Maguire, C.C., Manning, T., and Halpern, C.B. 2007. Response of terrestrial small mammals to varying amounts and patterns of green-tree retention in Pacific Northwest forests. Forest Ecology and Management 251, 142–155.
- Grace, J.B. 2006. *Structural Equation Modeling and Natural Systems*. Cambridge University Press, Cambridge, UK. 365 p.
- Grace, J.B. 2008. Structural equation modeling for observational studies. Journal of Wildlife Management 72, 14–22.
- Grace, J.B., and Bollen, K.A. 2006. The interface between theory and data in structural equation models. U.S. Geological Survey Open-File Report 2006-1363, Reston, VA, USA. 33 p.
- Grace, J.B., and Bollen, K.A. 2008. Representing general theoretical concepts in structural equation models: the role of composite variables. Environmental and Ecological Statistics 15, 191–213.
- Grace, J.B., Youngblood, A., and Scheiner, S.M. 2009. Structural equation modeling and ecological experiments. In: *Real World Ecology: Large-Scale and Long-Term Case Studies and Methods*, Miao, S.L., Carstenn, S., and M. Nungesser (eds.) pp. 19–46. Springer Science+Business Media, New York, NY, USA.
- Gracia, M., Montané, F., Piqué, J., and Retana. 2007. Overstory structure and topographic gradients determining diversity and abundance of understory shrub species in temperate forests in central Pyrenees (NE Spain). Forest Ecology and Management 242, 391–397.
- Gray, B.R. 2005. Selecting a distributional assumption for modelling relative densities of benthic macroinvertebrates. Ecological Modelling 185, 1–12.
- Greene, W.H. 1994. Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Working Paper EC-94-10, Department of Economics, Stern School of Business, New York University, New York NY, USA.
- Greenwood, D.L. and Weisberg, P.J. 2008. Density-dependent tree mortality in pinyon-juniper woodlands. Forest Ecology and Management 255, 2129–2137.

- Hall, D.B. 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics 56, 1030–1039.
- Hall, D.B., and Zhang, Z. 2004. Marginal models for zero inflated clustered data. Statistical Modelling 4, 161–180.
- Halpern, C.B., and McKenzie, D. 2001. Disturbance and post-harvest ground conditions in a structural retention experiment. Forest Ecology and Management 154, 215–225.
- Halpern, C.B., McKenzie, D., Evans, S., and Maguire, D.A. 2005. Early responses of forest understories to timber harvest with varying levels and patterns of greentree retention. Ecological Applications 15, 175–195.
- Halpern, C.B., and Raphael, M.G. (eds.) 1999. Retention harvests in northwestern forest ecosystems: the Demonstration of Ecosystem Management Options (DEMO) study. Northwest Science 73 (special issue), 1–125.
- Halpern, C.B., and Spies, T.A. 1995. Plant Species Diversity in Natural and Managed Forests of the Pacific Northwest. Ecological Applications 5, 913–934.
- Harrington, C.A., and Reukema, D.L. 1983. Initial shock and long-term stand development following thinning in a Douglas-fir plantation. Forest Science 29, 33–46.
- Hayduk, L.A. 1987. *Structural Equation Modeling with LISREL: Essentials and Advances*. The John Hopkins University Press, Baltimore, MD, USA. 405 p.
- Heithecker, T.D., and Halpern, C.B. 2006. Variation in microclimate associated with disperse-retention harvests in coniferous forests of western Washington. Forest Ecology and Management 226, 60–71.
- Hutchinson, G.E. 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology 22, 415–427.
- Jennings, S.B., Brown, N.D., and Sheil, D. 1999. Assessing forest canopies and understory illumination: canopy closure, canopy cover and other measures. Forestry 72, 59–73.
- Johnson, N.L., and Kotz, S. 1969. *Discrete Distributions*. Houghton Mifflin, Boston, MA, USA.

- Johnson, R.A., and Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, Upper Saddle River, NJ, USA. 773 p.
- Johnstone, J., Boby, L., Tissier, E., Mack, M., Verbyla, D., and Walker, X. 2009. Postfire seed rain of black spruce, a semiserotinous conifer, in forests of interior Alaska. Canadian Journal of Forest Research 39, 1575–1588.
- Jöreskog, K.G. 1973. A general method for estimating a linear structural equation system. In: *Structural Equation Models in the Social Sciences*, Goldberger, A.S., and O.D. Duncan (eds.) pp. 85–112. Seminar Press, New York, NY, USA.
- Jöreskog, K.G. 1993. Testing structural equation models. In: *Testing Structural Equation Models*, Bollen, K.A., and J.S. Lang (eds.) pp. 294–316. Sage Publications, Newbury Park, CA, USA.
- Jöreskog, K.G., and Sörbom, D. 2006. *LISREL 8.8 for Windows* [Computer software]. Scientific Software International Inc., Lincolnwood, IL, USA.
- Jöreskog, K.G., Sörbom, D., du Toit, S., and du Toit, M. 2000. *LISREL 8: New Statistical Features*. Scientific Software International Inc., Lincolnwood, IL, USA. 248 p.
- Kelloway, E.K. 1998. Using LISREL for Structural Equation Modeling: a Researcher's Guide. Sage Publications, Thousand Oaks, CA, USA. 145 p.
- Kenny, D.A. 1979. *Correlation and Causality*. John Wiley & Sons, New York, NY, USA. 277 p.
- Kenny, D.A., and Judd, C.M. 1984. Estimating the nonlinear and interactive effects of latent variables. Psychological Bulletin 96, 201–210.
- King, G. 1989. Event count models for international relations: generalizations and applications. International Studies Quarterly 33, 123–147.
- Klenner, W., and Sullivan T.P. 2003. Partial and clear-cut harvesting of high-elevation spruce-fir forests: implications for small mammal communities. Canadian Journal of Forest Research 33, 2283–2296.
- Kline, R.B. 2005. *Principles and Practice of Structural Equation Modeling*. 2<sup>nd</sup> ed. The Guilford Press, New York, NY, USA. 366 p.

- Kneeshaw, D.D., H. Williams, E. Nikinmaa, and Messier, C. 2002. Patterns of aboveand below-ground response of understory conifer release 6 years after partial cutting. Canadian Journal of Forest Research 32, 255–265.
- Krebs, C.J. 1999. Ecological Methodology. Harper & Row, New York, NY, USA.
- Kuehl, R.O. 2000. Design of Experiments: Statistical Principles of Research Design and Analysis. 2<sup>nd</sup> ed. Duxbury Press, Pacific Grove, CA, USA. 666 p.
- Lam, T.Y., Huso, M., and Maguire, D.A. 2010. Modeling abundance of infrequent species: a simulation study for clarifying theoretical and ecological interpretation of zero-modified models. In: *Exploration of Statistical Methods for Synthesizing the Effects of Variable-Retention Harvesting on Multiple Taxa*. Dissertation, Oregon State University, Corvallis, OR, USA.
- Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34, 1–14.
- Laughlin, D.C., and Abella, S.R. 2007. Abiotic and biotic factors explain independent gradients of plant community composition in ponderosa pine forests. Ecological Modelling 205, 231–240.
- Laughlin, D.C., Abella, S.R., Wallace, C.W., and Grace, J.B. 2007. Species richness and soil properties in Pinus ponderosa forests: a structural equation modeling analysis. Journal of Vegetation Science 18, 231–242.
- Lee, S.Y. 2007. *Structural Equation Modeling: a Bayesian Approach*. John Wiley and Sons, New York, NY, USA. 432 p.
- Lieffers, V.J., Messier, C., Stadt, K.J., Gendron, F., and Comeau, P.G. 1999. Predicting and managing light in the understory of boreal forests. Canadian Journal of Forest Research 29, 796–811.
- Lindenmayer, D.B., and Franklin, J.F. 2002. *Conserving Forest Biodiversity: A Comprehensive Multiscaled Approach*. Island Press, Washington, DC, USA. 351 p.
- Lindgren, P.M.F., and Sullivan, T.P. 2001. Influence of alternative vegetation management treatments on conifer plantation attributes: abundance, species diversity, and structural diversity. Forest Ecology and Management 142, 163– 182.

- Lindh, B.C., Gray, A.N., and Spies, T.A. 2003. Responses of herbs and shrubs to reduced root competition under canopies and in gaps: a trenching experiment in old-growth Douglas-fir forests. Canadian Journal of Forest Research 33, 2052–2057.
- Lindh, B.C., and Muir, P.S. 2004. Understory vegetation in young Douglas-fir forests: does thinning help restore old-growth composition? Forest Ecology and Management 192, 285–296.
- Lindsey, J.K., and Jones, B. 1998. Choosing among generalized linear models applied to medical data. Statistics in Medicine 17, 59–68.
- LISREL Documentation. 2005. Analysis of Structural Equation Models for Continuous Random Variables in the Case of Complex Survey Data. http://www.ssicentral.com/lisrel/techdocs/compsem.pdf [Last Accessed 4 February 2010].
- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., and Schabenberger, O. 2006. *SAS for Mixed Models*. 2<sup>nd</sup> ed. SAS Publishing, Cary, NC, USA. 840 p.
- Lord, D., Washington, S.P., and Ivan, J.N. 2005. Poisson, Poisson-gamma and zeroinflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention 37, 35–46.
- Lord, D., Washington, S.P., and Ivan, J.N. 2007. Further notes on the application of zero-inflated models in highway safety. Accident Analysis and Prevention 39, 53–57.
- Luoma, D.L., Eberhart, J.L., Molina, R., and Amaranthus, M.P. 2004. Response of ectomycorrhizal fungus sporocarp production to varying levels and patterns of green-tree retention. Forest Ecology and Management 202, 337–354.
- Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. 2000. WinBUGS a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10, 325–337.
- MacArthur, R. H., and Wilson, E. O. 1967. *The Theory of Island Biogeography*. Princeton University Press, Princeton, NJ, USA.
- MacCracken, J.G. 2005. Effects of uneven-aged timber harvest on forest floor vertebrates in the Cascade Mountains of southern Washington. Forest Ecology and Management 208, 123–135.

- MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A., and Langtimm, C.A. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83, 2248–2255.
- Maguire, D.A., and Forman, R.T.T. 1983. Herb cover effects on tree seedling patterns in a mature hemlock-hardwood forest. Ecology 64: 1367–1380.
- Maguire, D.A., Halpern, C.B., and Phillips, D.L. 2007. Changes in forest structure following variable-retention harvests in Douglas-fir dominated forests. Forest Ecology and Management 242, 708–726.
- Maguire, D.A., Mainwaring, D., and Halpern, C.B. 2006. Stand dynamics after variable retention harvesting in mature Douglas-fir forests of western North America. Allgemeine Forst und Jagdzeitung 177, 120–131.
- Manning, J.A., and Edge, W.D. 2004. Small mammal survival and downed wood at multiple scales in management forests. Journal of Mammalogy 85, 87–96.
- Martin, K.J., and McComb, W.C. 2002. Small mammal habitat associations at patch and landscape scales in Oregon. Forest Science 48, 255–264.
- Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., and Possingham, H.P. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecology Letters 8, 1235–1246.
- Matthews, J.D. 1989. *Silvicultural Systems*. Oxford University Press, Oxford, UK. 284 p.
- Mayer, A.L., Cameron, G.N. 2003. Consideration of grain and extent in landscape studies of terrestrial vertebrate ecology. Landscape and Urban Planning 65, 201–217.
- McCullagh, P., and Nelder, J.A. 1989. *Generalized Linear Models*. 2<sup>nd</sup> ed. Chapman & Hall/CRC, London, UK.
- McCune, B., and Grace. J.B. 2002. *Analysis of Ecological Communities*. MjM Software Design, Gleneden Beach, OR, USA. 300 p.
- McKenzie, D., and Halpern, C.B. 1999. Modeling the distributions of shrub species in Pacific Northwest forests. Forest Ecology and Management 114, 293–307.

- McKenzie, D., Halpern, C.B., and Nelson, C.R. 2000. Overstory influences on herb and shrub communities in mature forests of western Washington, USA. Canadian Journal of Forest Research 30, 1655–1666.
- McLachlan, G., and Peel, D. 2000. *Finite Mixture Models*. 1<sup>st</sup> ed. Wiley-Interscience, New York NY, USA.
- Medin, D.E. 1986. Small mammal responses to diameter-cut logging in an Idaho Douglas-fir forest. USDA Forest Service Research Note INT-362, Ogden UT, USA. 6 p.
- Meffe, G.K., and Carroll, C.R. 1997. *Principles of conservation biology*. 2<sup>nd</sup> ed. Sinauer Associates, Sunderland, MA, USA.
- Miao, S.L., Carstenn, S., and Nungesser, M. 2009. Introduction unprecedented challenges in ecological research: past and present. In: *Real World Ecology: Large-Scale and Long-Term Case Studies and Methods*, Miao, S.L., Carstenn, S., and M. Nungesser (eds.) pp. 1–18. Springer Science+Business Media, New York, NY, USA.
- Miller, D.H., and Getz, L.L. 1973. Factors influencing the local distribution of the redback vole, *Clethrionomys gapperi* in New England. II. Uni. Conn. Occas. Pap. Biol. Sci. Ser. 2. pp. 159–180.
- Min, Y., and Agresti, A. 2005. Random effect models for repeated measures of zeroinflated count data. Statistical Modelling 5, 1–19.
- Minami, M., Lennert-Cody, C.E., Gao, W., and Román-Verdesoto, M. 2007. Modeling shark bycatch: the zero-inflated negative binomial regression model with smoothing. Fisheries Research 84, 210–221.
- Mitchell, J.E., and Bartling, P.N.S. 1991. Comparison of linear and nonlinear overstory-understory models for ponderosa pine. Forest Ecology and Management 42, 195–204.
- Moses, R.A., and Boutin, S. 2001. The influence of clear-cut logging and residual leave material on small mammal populations in aspen-dominated boreal mixedwoods. Canadian Journal of Forest Research 31, 283–495.
- Mullahy, J. 1986. Specification and testing of some modified count data models. Journal of Econometrics 33, 341–365.

- Muthén, B. 1983. Latent variable structural equation modeling with categorical data. Journal of Econometrics 22, 43–65.
- Muthén, B. 1984. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika 49, 115–132.
- Muthén, L.K., and Muthén, B.O. 2007. Mplus User's Guide. 5th ed. Los Angeles, CA USA.
- Muzika, R.M., Grushecky, S.T., Liebhold, A.M., and Smith, R.L. 2004. Using thinning as a management tool for gypsy moth: the influence on small mammal abundance. Forest Ecology and Management 192, 349–359.
- Nemati, N., and Goetz, H. 1995. Relationships of overstory to understory cover variables in a Ponderosa pine/Gambel oak ecosystem. Vegetatio 119, 15–21.
- New, T.R. 2000. *Conservation Biology: An Introduction for Southern Australia*. Oxford University Press, Melbourne, Australia.
- Ntzoufras, I. 2009. *Bayesian Modeling Using WinBUGS*. John Wiley & Sons, Hoboken, NJ, USA. 492 p.
- Oliver, C.D., and Larson, B.C. 1990. *Forest Stand Dynamics*. John Wiley & Sons, New York, NY, USA. 467 p.
- Ping, R.A. 1996. Interaction and quadratic effect estimation: a two-step technique using structural equation analysis. Psychological Bulletin 119, 166–175.
- Potts, J.M., and Elith, J. 2006. Comparing species abundance models. Ecological Modelling 199, 153–163.
- Powell, G.W., and Bork, E.W. 2006. Aspen canopy removal and root trenching effects on understory vegetation. Forest Ecology and Management 230, 79–90.
- Prescott, C.E. 2002. The influence of the forest canopy on nutrient cycling. Tree Physiology 22, 1193–1200.
- Pugesek, B.H., Tomer, A., and von Eye, A. 2003. Structural Equation Modeling: Applications in Ecological and Evolutionary Biology. Cambridge University Press, Cambridge, UK. 424 p.

- Quinby, P.A. 2000. First-year impacts of shelterwood logging on understory vegetation in an old-growth pine stand in central Ontario, Canada. Environmental Conservation 27, 229–241.
- R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <u>http://www.R-project.org</u>.
- Rabinowitz, D., Cairns, S., and Dillon, T. 1986. Seven forms of rarity, and their frequency in the flora of British Isles. In: *Conservation Biology: The Science of Scarcity, and Diversity*, M. Soulé (ed.) pp. 184–204. Sinauer Associates, Sunderland, MA, USA.
- Rao, J.N.K. 2003. Small Area Estimation. Wiley, Hoboken, NJ, USA.
- Raphael, M.G. 1984. Wildlife populations in relation to stand age and area in Douglasfir forests of northwestern California. In: *Fish and Wildlife Relationships in Old-growth Forests: Proceedings of a Symposium*, Meehan, K.R., Merrell Jr., T.R., and T. Hanley (eds.) pp. 259–274. 12–15 April 1982, Juneau, Alaska, USA.
- Raykov, T., and Marcoulides, G.A. 2000. *A First Course in Structural Equation Modeling*. Lawrence Erlbaum Associates, Mahwah, NJ, USA. 216 p.
- Rider, P.R. 1961. Estimating the parameters of mixed Poisson, binomial and Weibull distributions by methods of moments. Bulletin de l'Institut International de Statistiques 38, Part 2.
- Riegel, G.M., Miller, R.F., and Krueger, W.C. 1992. Competition for resources between understory vegetation and overstory Pinus ponderosa in northeastern Oregon. Ecological Applications 2, 71–85.
- Rose, C.E., Martin, S.W., Wannemuehler, K.A., and Plikaytis, B.D. 2006. On the use of zero-inflated and hurdle models for modelling vaccine adverse event count data. Journal of Biopharmaceutical Statistics 16, 463–481.
- Rosenvald, R., and Lõhmus, A. 2008. For what, where and when is green-tree retention better than clearcutting? A review of the biodiversity aspects. Forest Ecology and Management 255, 1–15.
- Royle, J.A. 2004. *N*-mixture models for estimating population size from spatially replicated counts. Biometrics 60, 108–115.

- Royle, J.A., Nichols, J.D., and Kery, M. 2005. Modelling occurrence and abundance of species when detection is imperfect. Oikos 110, 353–359.
- Runciman, J.B., and Sullivan, T.P. 1996. Influences of alternative conifer release treatments on habitat structure and small mammal populations in south central British Columbia. Canadian Journal of Forest Research 26, 2023–2034.
- Saetre, P., Saetre, L.S., Brandtberg, P.O., Lundkvist, H., and Bengtsson, J. 1997. Ground vegetation composition and heterogeneity in pure Norway spruce and mixed Norway spruce–birch stands. Canadian Journal of Forest Research 27, 2034–2042.
- Sariyildiz, T., Anderson, J.M., Kucuk, M. 2005. Effects of tree species and topography on soil chemistry, litter quality, and decomposition in Northeast Turkey. Soil Biology & Biochemistry 37, 1695–1706.
- Schabenberger, O., and Gotway, C.A. 2005. *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL, USA. 488 p.
- Schabenberger, O., and Pierce, F.J. 2001. *Contemporary Statistical Models for the Plant and Soil Sciences*. Taylor & Francis, Boca Raton, FL, USA. 738 p.
- Schlaepfer, M. A., Runge, M. C., and Sherman, P. W. 2002. Ecological and evolutionary traps. Trends in Ecology & Evolution 17, 474–480.
- Schumacker, R.E., and Lomax, R.G. 2004. *A Beginner's Guide to Structural Equation Modeling*. 2<sup>nd</sup> ed. Taylor & Francis Group, New York, NY, USA. 498 p.
- Sharpe, F., Shaw, D.C., Rose, C.L., Sillett, S.C., and Carey, A.B. 1996. The biologically significant attributes of forest canopies to small birds. Northwest Science 70, 86–93.
- Shipley, B. 2000. Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference. Cambridge University Press, Cambridge, UK. 336 p.
- Shipley, B. 2009. Confirmatory path analysis in a generalized multilevel context. Ecology 90, 363–368.
- Sileshi, G. 2008. The excess-zero problem in soil animal count data and choice of appropriate models for statistical inference. Pedobiologia 52, 1–17.

- Smith, A., and Gelfand, A. 1992. Bayesian statistics without tears: a samplingresampling perspective. The American Statistician 46, 84–88.
- Spearman, C. 1904. General intelligence, objectively determined and measured. American Journal of Psychology 15, 201–293.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B 64, 583-640.
- Spies, T.A., and Franklin, J.F. 1991. The structure of natural young, mature, and oldgrowth Douglas-fir forests in Oregon and Washington. In: *Wildlife and Vegetation of Unmanaged Douglas-fir Forests*, Ruggiero, L.F., Aubry, K.B., Carey, A.B., and M.H. Huff (eds.) pp. 91–109. USDA Forest Service General Technical Report PNW-GTR-285.
- Stage, A.R., and Salas, C. 2007. Interactions of elevation, aspect, and slope in models of forest species composition and productivity. Forest Science 53, 486–492.
- Staudhammer, C.L., and LeMay, V.M. 2001. Introduction and evaluation of possible indices of stand structural diversity. Canadian Journal of Forest Research 31, 1105–1115.
- Steiger, J.H., and Fouladi, R.T. 1997. Noncentrality interval estimation and the evaluation of statistical models. In: *What if there were no significance test?* Harlow, L.L., Mulaik, S.A., and J.H. Steiger (eds.) pp. 221-257. Lawrence Erlbaum Associates, Mahwah, NJ, USA. 472 p.
- Steventon, J.D., MacKenzie, K.L., and Mahon, T.E. 1998. Response of small mammals and birds to partial cutting and clearcutting in northwest British Columbia. The Forestry Chronicle 74, 703–713.
- Sullivan, T.P., and Sullivan, D.S. 2001. Influence of variable retention harvests on forest ecosystems. II. Diversity and population dynamics of small mammals. Journal of Applied Ecology 38, 1234–1252.
- Sullivan, T.P., Sullivan, D.S., and Lindgren, P.M.F. 2008. Influence of variable retention harvests on forest ecosystems: plant and mammal responses up to 8 years post-harvest. Forest Ecology and Management 254, 239–254.
- Sullivan, T.P., Sullivan, D.S., Lindgren, P.M.F., and Ransome, D.B. 2005. Long-term responses of ecosystem components to stand thinning in young lodgepole pine

forest II. Diversity and population dynamics of forest floor small mammals. Forest Ecology and Management 205, 1–14.

- Suzuki, N., and Hayes, J.P. 2003. Effects of thinning on small mammals in Oregon coastal forests. Journal of Wildlife Management 67, 352–371.
- Tappeiner, J.C., Maguire, D.A., and Harrington, T.B. 2007. Silviculture and Ecology of Western U.S. Forests. Oregon State University Press, Corvallis, OR, USA. 440 p.
- Thompson, B. (ed.) 2003. *Score Reliability: Contemporary thinking on reliability issues*. Sage Publications, Thousand Oaks, CA, USA. 296 p.
- Tuchmann, E.T., Connaughton, K.P., Freedman, L.E., and Moriwaki, C.B. 1996. The Northwest Forest Plan: a report to the President and Congress. USDA Forest Service, Pacific Northwest Research Station, Portland, OR, USA.
- Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K., and Possingham, H.P. 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. Ecological Applications 13, 1790–1801.
- Ure, D.C., and Maser, C. 1982. Mycophagy of red-backed voles in Oregon and Washington. Canadian Journal of Zoology 60, 3307–3315.
- Uresk, D.W., and Severson, K.E. 1989. Understory-overstory relationships in ponderosa pine forests, Black Hills, South Dakota. Journal of Range Management 42, 203–208.
- USDA and USDI. 1994. Record of decision for amendments to Forest Service and Bureau of Land Management planning documents with the range of the Northern Spotted Owl. USDA Forest Service, Pacific Northwest Research Station, Portland, OR, USA.
- Venables, W.N., and Ripley, B.D. 2002. *Modern Applied Statistics with S.* 4<sup>th</sup> ed. Springer-S+Business Media Inc., New York, NY, USA. 495 p.
- Von Trebra, C., Lavender, D.P., and Sullivan, T.P. 1998. Relationships of small mammal populations to even-aged shelterwood systems in sub-boreal spruce forest. Journal of Wildlife Management 62, 630–642.
- Waldien, D.L., Hayes, J.P., and Huso, M.M.P. 2006. Use of downed wood by Townsend's chipmunks (Tamias townsendii) in western Oregon. Journal of Mammalogy 87, 454–460.

- Warton, D.I. 2005. Many zeros does not mean zero inflation: comparing the goodnessof-fit of parametric models to multivariate abundance data. Environmetrics 16, 275–289.
- Weisberg, S. 2005. *Applied Linear Regression*. John Wiley & Sons, Inc., Hoboken, NJ, USA. 310 p.
- Welsh, A.H., Cunningham, R.B., Donnelly, C.F., and Lindenmayer, D.B. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecological Modelling 88, 297–308.
- Wenger, S.J., and Freeman, M.C. 2008. Estimating species occurrence, abundance, and detection probability using zero-inflated distributions. Ecology 89, 2953– 2959.
- West, S.D. 1991. Small mammal communities in the southern Washington Cascade Range. In: *Wildlife and Vegetation of Unmanaged Douglas-fir Forests*, Ruggiero, L.F., Aubry, K.B., Carey, A.B., and M.H. Huff (eds.) pp. 269–285. USDA Forest Service General Technical Report PNW-GTR-285. Pacific Northwest Research Station, Portland, OR, USA.
- Whitney, G.G., and Foster, D.R. 1988. Overstory composition and age as determinants of the understory flora of woods of central New England. Journal of Ecology 76, 867–876.
- Wiens, J.A. 1989. Spatial scaling in ecology. Functional Ecology 3, 385–397.
- Wilson, S.M., and Carey, A.B. 2000. Legacy retention versus thinning: influences on small mammals. Northwest Science 74, 131–145.
- Winkelmann, R. 2008. *Econometric Analysis of Count Data*. 5<sup>th</sup> ed. Springer-Verlag, Berlin, Germany.
- Wright, S. 1921. Correlation and causation. Journal of Agricultural Research 20, 557–585.
- Wright, S. 1934. The method of path coefficients. Annals of Mathematical Statistics 5, 161–215.
- Wright, S. 1960. Path coefficients and path regressions: alternative or complementary concepts? Biometrics 16, 189–202.

- Yau, K.K.W., and Lee, A.H. 2001. Zero-inflated Poisson regression with random effects to evaluate an occupations injury prevention programme. Statistics in Medicine 20, 2907–2920.
- Youngblood, A., Grace, J.B., and McIver, J.D. 2009. Delayed conifer mortality after fuel reduction treatments: interactive effects of fuel, fire intensity, and bark beetles. Ecological Applications 19, 321–337.
- Yuan, K.H., and Bentler, P.M. 2000. Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In: *Sociological Methodology*, Sobel, M.E., and M.P. Becker (eds.) pp. 165–200. Washington, DC, USA.
- Zeileis, A., Kleiber, C., and Jackman, S. 2008. Regression models for count data in R. Journal of Statistical Software 27, 1–24.
- Zorn, C.J.W. 1998. An analytic and empirical examination of zero-inflated and hurdle Poisson specifications. Sociological Methods & Research 26, 368–400.
- Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., and Smith, G.M. 2009. *Mixed Effects Models and Extensions in Ecology with R.* Springer Sciences+Business Media, New York, NY, USA. 574 p.