

A QUANTITATIVE APPROACH FOR DEFINING RARITY

Lee Anne Jacobs Reilly

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology.

Chapel Hill
2010

Approved by:

Robert K. Peet (Chair)

Peter S. White

Alan S. Weakley

Dean L. Urban

Joel G. Kingsolver

John F. Bruno

©2010
Lee Anne Jacobs Reilly
ALL RIGHTS RESERVED

ABSTRACT

Lee Anne Jacobs Reilly:

A QUANTITATIVE APPROACH FOR DEFINING RARITY

(Under the direction of Robert K. Peet)

The concept of rarity is critical for understanding patterns of species abundance in the landscape, the ecological processes that drive those patterns, and for conservation of species that become threatened. Multiple definitions of rarity exist in the literature, as well as multiple approaches to the study of rarity. However, no universally accepted definition exists. Without a clear definition of rarity, approaches used to study rarity are inconsistent and communication of results becomes difficult.

My research proposes a comprehensive, quantitative definition of rarity. First, I review existing studies to create a comparative analysis of current definitions and approaches to the study of rarity. From this analysis, I develop a conceptual model for a comprehensive approach to define rarity, which forms the foundation of subsequent chapters. Second, using several large-scale vegetation datasets, I develop a quantitative definition of rarity for vascular plants of the Southeastern United States and characterize the patterns of rarity revealed within the region. Finally, I develop a simulation model to examine the effects of an impact such as habitat loss on plants of different forms of rarity.

My results indicate that measures of multiple criteria, such as local abundance, geographic range size, habitat preference, frequency, and occupancy, must be included to have a comprehensive understanding of the distributional patterns of species. For vascular plants of the Southeast, rarity is three-dimensional, consisting of abundance, range, and habitat volume. The forms of rarity defined in this analysis are largely congruent with existing lists of threatened and endangered plant species. However, the current analysis reveals species that exhibit distribution patterns similar to threatened species, but that do not yet have conservation status. Finally, simulation results indicate that species of different forms of rarity do respond differently to impacts, indicating that protection of threatened species may need to be approached differently depending on the form of rarity each species exhibits. Taken together, these results indicate that rarity is multidimensional and must be defined as such, and that the information provided by a comprehensive definition of rarity will allow more accurate and efficient development of conservation strategies.

To my parents for teaching me the value of education and hard work, and always
encouraging me to follow my dreams

To my sister, for helping me keep things in perspective and making me laugh

To Jason and Ethan - you are my heart and soul

As accumulation of knowledge continues, we eventually find facts that will not fit properly into any established pigeon hole. This should at once be the sign that possibly our original arrangement of pigeon holes was insufficient and should lead us to a careful examination of our accumulated data. Then we may conclude that we would better demolish our whole system of arrangement and classification and start anew with hope of better success.

- H.A. Gleason, 1926

ACKNOWLEDGEMENTS

This work would not have been possible without the contributions of many talented people. My graduate committee has provided advice and guidance every step of the way. My advisor, Bob Peet, has spent countless hours over the years to meet with me, puzzle over design of methods and interpretation of results, and read multiple drafts as the work began to take shape. Peter White's passionate and pragmatic approach to conservation has helped me interpret my findings more meaningfully. Alan Weakley's brilliance with botany and taxonomy has been an invaluable resource. Dean Urban's technical expertise provided me with the tools of analysis that made this work possible. Joel Kingsolver's wisdom, experience, and breadth of knowledge has helped me place this work in a context beyond plant ecology. John Bruno's enthusiasm for the study of rarity has encouraged me to pursue these and other questions relentlessly.

Jack Weiss has a level of genius for quantitative biology that few can imagine. While my work does not do justice to the training I have received from him, the tools and theory he has taught me and the many hours we have shared discussing the nuances of analysis have been invaluable.

The many volunteers who have contributed to the Carolina Vegetation Survey provided the data that made this work possible. Michael Lee and Forbes Boyle have done a

tremendous job maintaining the database, and are always willing to answer questions about data as well as brainstorm about possible analyses.

Gwen Davis and Mary Russo were willing to provide and explain NatureServe conservation status data.

Several sources of funding have provided support for my graduate work, including the Graduate School, the Biology Department, the Mrs. W.C. Coker Fellowship, the A.H. Beers Scholarship, the W.C. Coker Summer Fellowship, the Future Faculty Fellowship Program, Sigma Xi Grant in Aid of Research, the Center for the Study of the American South, and the H.V. Wilson Award.

The members of the Plant Ecology lab have been great comrades over the years. Dave Vandermast mentored me as a teacher and an ecologist. Pat Corry has been a great role model, and can always make me laugh. Todd Jobe has a clarity of analytic thought that has guided my thinking on this and other projects. Dane Kuppinger has opened my mind to new perspectives in ecology and in life. Kristin Taverna, Sarah Marcinko, Liz Matthews, Jeff Ott, Brooke Wheeler, and Amanda Senft have provided wonderful friendship and support. We have worked together, played together, taught together, and struggled through exams together. This experience would not have been the same without them.

Most important has been the support from family. My parents, Mike and Barbara Jacobs, have always encouraged me to pursue my dreams, have pushed me to do my best, and have provided so many opportunities that have prepared me to do so. My sister, who probably knows me as well as anyone, helps me keep things in perspective and isn't afraid to tell it like it is. She is always able to make me laugh. Debbie Reilly and Marty McClelland

have cooked for us and cared for us when we're sick to allow me to stay focused on work. Jason, my husband, has provided support in countless ways, from helping with household tasks to discussing approaches to analysis. His love, support, and encouragement has kept me going through the most difficult times. Ethan, our son, brings joy to every moment. Having him in our lives makes it all worth while.

TABLE OF CONTENTS

| | |
|---|------|
| LIST OF TABLES | xii |
| LIST OF FIGURES | xiii |
| CHAPTER | |
| 1. A COMPREHENSIVE FRAMEWORK TO DEFINE RARITY | 1 |
| Abstract | 1 |
| Introduction | 1 |
| The need for consistent communication | 2 |
| The challenge of defining rarity | 4 |
| Toward a comprehensive definition | 6 |
| The proposed framework | 10 |
| The comprehensive framework in practice: The condition of sparseness | 11 |
| A call for research priorities | 16 |
| Conclusions | 18 |
| References | 19 |
| 2. A QUANTITATIVE APPROACH TO DETERMINE THE DIMENSIONALITY OF RARITY | 22 |
| Abstract | 22 |
| Introduction | 23 |
| Methods | 25 |

| | |
|---|----|
| <i>Data preparation</i> | 25 |
| <i>Analysis</i> | 33 |
| Results and discussion | 34 |
| <i>Dimensionality</i> | 34 |
| <i>Analytic interpretation of dimensions</i> | 38 |
| <i>Biological interpretation of dimensions</i> | 40 |
| <i>Interpretation of groups</i> | 44 |
| <i>Regional results</i> | 59 |
| Conclusions | 61 |
| References | 63 |
| 3. A SIMULATION TO PREDICT THE EFFECTS OF HABITAT LOSS ON RARE PLANT SPECIES | 65 |
| Abstract | 65 |
| Introduction | 66 |
| Methods | 68 |
| <i>Data preparation</i> | 68 |
| <i>Simulation of habitat loss</i> | 68 |
| <i>Analysis</i> | 69 |
| Results | 76 |
| <i>Effects of habitat loss on across the forms of rarity</i> | 76 |
| <i>Biological interpretation</i> | 80 |
| Discussion | 83 |
| Conclusions | 84 |

| | |
|---|----|
| References | 85 |
| GENERAL CONCLUSIONS | 87 |
| Appendix 1: <i>Plot descriptions for 2424 Carolina Vegetation Survey plots</i> | 89 |
| Appendix 2: <i>Environmental data for 2424 Carolina Vegetation Survey plots</i> | 90 |
| Appendix 3: <i>Species list</i> | 91 |

LIST OF TABLES

| | | |
|-------------------|---|-----------|
| Table 1.1. | <i>Comparison of studies examining the competitive abilities of rare species.....</i> | 3 |
| Table 1.2. | <i>Proposed mechanisms that maintain sparse, “somewhere-abundant,” and habitat restricted forms of rarity and allow shifts between them in ecological time.....</i> | 15 |
| Table 2.1. | <i>Calculated species attributes and data sources.....</i> | 31 |
| Table 2.2. | <i>Correlations among all species. Attributes that were removed from the analysis are in grey, and the correlation values that caused those attributes to be removed are underlined.....</i> | 32 |
| Table 2.3. | <i>Definitions of NatureServe Global Conservation Ranks (G-ranks).....</i> | 33 |
| Table 2.4. | <i>Factor loadings for each attribute for three-factor solution. Factor 1 is interpreted as increasing with attributes representing habitat volume, factor 2 is interpreted as increasing with attributes representing geographic range, and factor 3 is interpreted as increasing with attributes representing local abundance. The code “ml” indicates variables that were log transformed prior to analysis.....</i> | 38 |
| Table 2.5. | <i>Interpretation of rarity for each of eight groups created using hierarchical clustering with flexible beta = -0.25 linkage method.....</i> | 51 |
| Table 2.6. | <i>Descriptive statistics for each of eight groups created using hierarchical clustering with flexible beta = -0.25 linkage method.....</i> | 51 |
| Table 2.7. | <i>Percentages of high G-rank species captured within the rarity groups.....</i> | 52 |
| Table 2.8. | <i>Total numbers of species of each G-rank within each rarity group.....</i> | 52 |
| Table 2.9. | <i>Factor loadings for each attribute for three-factor solution within each region.....</i> | 60 |
| Table 3.1. | <i>Framework defining the forms of rarity using the criteria of geographic range size, habitat specificity, and local population size, or abundance. (Rabinowitz, 1981).....</i> | 66 |
| Table 3.2. | <i>Calculated species attributes and data sources.....</i> | 68 |
| Table 3.3. | <i>Rarity groups as defined by division at the median of each of three attributes. High and low indicate above or below the median value. While the groups are not completely ordinal from most to least rare, Groups 1-4 can be considered more rare, with Group 1 the most rare (low on all attributes), while Groups 5-8 can be considered less rare, with Group 8 being common (high on all attributes).....</i> | 72 |

LIST OF FIGURES

| | | |
|--------------------|--|-----------|
| Figure 1.1. | <i>A comprehensive, multidimensional, mechanistic framework to define rarity. Points represent species in attribute space; arrows represent ecological mechanisms allowing species to shift among forms of rarity.....</i> | <i>11</i> |
| Figure 1.2. | <i>Representative mechanisms from Table 1.2 illustrate how species locations in each realm of rarity space might be maintained as well as how species might shift through the space.....</i> | <i>16</i> |
| Figure 2.1. | <i>Map of US counties included in analysis. The western boundary was determined by the ecoregion boundaries that created the most homogeneous representation of flora for the eastern US and the most consistent extent in all directions from the focal region (North Carolina and surrounding states) to avoid truncation of data by northern and southern political boundaries.....</i> | <i>28</i> |
| Figure 2.2. | <i>Scree plot of stress values from step-down nonmetric multidimensional scaling ordination, indicating an abrupt reduction in stress with the addition of a third axis, and very little reduction in stress with additional axes.....</i> | <i>35</i> |
| Figure 2.3. | <i>Barplot of eigen values from principal coordinates ordination, indicating that the first three components account for most variability in these data.....</i> | <i>36</i> |
| Figure 2.4. | <i>Shepard diagrams of NMDS and PCO 3-dimensional ordination results. For each analysis, the correlation between the ordination distances and the computed (Euclidean) distances is high ($r = 0.948$ for NMDS solution and $r = 0.943$ for PCO solution), indicating that either solution is a reliable representation of these data.....</i> | <i>36</i> |
| Figure 2.5. | <i>Proportional and cumulative variance with each additional factor. The three-factor solution explained 83.2% of variance; including the fourth factor added only 5.6% variance, indicating that three factors are sufficient for these data.....</i> | <i>37</i> |
| Figure 2.6. | <i>Plot of numerical solutions to scree test. All solutions indicate that three factors should be retained for these data (Raiche and Magis 2010; following methods described in Raiche et al. 2006).....</i> | <i>37</i> |

| | | |
|---------------------|--|----|
| Figure 2.7. | <i>Attributes plotted along factors 1, 2, and 3. (Plotted attribute codes are defined as follows: PlotCt.ml = log of plot count, CEGlct.ml = log of community count, Volume = Habitat volume, Area = range area, EWextent = east-west extent, NSextent = north-south extent, Avg.Cov.ml = log of mean abundance, Max.Cov = maximum abundance, Var.Cov.ml = variance in abundance). Factor 1 is interpreted as increasing with attributes representing habitat volume, factor 2 is interpreted as increasing with attributes representing geographic range, and factor 3 is interpreted as increasing with attributes representing local abundance.....</i> | 39 |
| Figure 2.8. | <i>Species plotted along factors 1 (habitat) and 2 (range). Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, QUERRUB = Quercus rubra, RHODCAT = Rhododendron catawbiense, RHODMAX = Rhododendron maximum, SMILROT = Smilax rotundifolia, TOXIRAD = Toxicodendron radicans, TRILDIS = Trillium discolor, VITIROT = Vitis rotundifolia. The linear pattern seen in the left side of the plot is due to the categorical nature of the variable of maximum abundance, measured as cover classes.....</i> | 41 |
| Figure 2.9. | <i>Species plotted along factors 1 (habitat) and 3 (abundance). Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, QUERRUB = Quercus rubra, RHODCAT = Rhododendron catawbiense, RHODMAX = Rhododendron maximum, SMILROT = Smilax rotundifolia, TOXIRAD = Toxicodendron radicans, TRILDIS = Trillium discolor, VITIROT = Vitis rotundifolia. The linear pattern seen in the plot is due to the categorical nature of the variable of maximum abundance, measured as cover classes.....</i> | 42 |
| Figure 2.10. | <i>Species plotted along factors 2 (range) and 3 (abundance). Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, QUERRUB = Quercus rubra, RHODCAT = Rhododendron catawbiense, RHODMAX = Rhododendron maximum, SMILROT = Smilax rotundifolia, TOXIRAD = Toxicodendron radicans, TRILDIS = Trillium discolor, VITIROT = Vitis rotundifolia. The linear pattern seen in the plot is due to the categorical nature of the variable of maximum abundance, measured as cover classes.....</i> | 43 |
| Figure 2.11. | <i>Group-contrast Mantel correlation values for each grouping level created using complete linkage method. The grouping level at which the maximum correlation value occurs (10 groups) is indicated by the vertical dotted line.....</i> | 45 |

| | | |
|---------------------|---|-----------|
| Figure 2.12. | <i>Group-contrast Mantel correlation values for each grouping level created using Ward's linkage method. The grouping level at which the maximum correlation value occurs (5 groups) is indicated by the vertical dotted line.....</i> | <i>45</i> |
| Figure 2.13. | <i>Group-contrast Mantel correlation values for each grouping level created using flexible beta linkage method. The grouping level at which the maximum correlation value occurs (8 groups) is indicated by the vertical dotted line.....</i> | <i>46</i> |
| Figure 2.14. | <i>Species plotted along factors 1 (habitat) and 2 (range). Groups were defined using hierarchical clustering with flexible beta = -0.25 linkage method. Group 1 is most rare, Group 8 is most common.....</i> | <i>48</i> |
| Figure 2.15. | <i>Species plotted along factors 1 (habitat) and 3 (abundance). Groups were defined using hierarchical clustering with flexible beta = -0.25 linkage method. Group 1 is most rare, Group 8 is most common.....</i> | <i>49</i> |
| Figure 2.16. | <i>Species plotted along factors 1 (habitat) and 3 (abundance). Groups were defined using hierarchical clustering with flexible beta = -0.25 linkage method. Group 1 is most rare, Group 8 is most common.....</i> | <i>50</i> |
| Figure 2.17. | <i>Factors 1 and 2, with species of each G-rank displayed. Species with G-ranks 1 and 2 are restricted on both range and habitat, and expand along both axes as G-rank increases.....</i> | <i>53</i> |
| Figure 2.18. | <i>Factors 1 and 3, with species of each G-rank displayed. Species with G-ranks 1 and 2 are restricted on habitat, but vary in local abundance; habitat increases as G-rank increases.....</i> | <i>54</i> |
| Figure 2.19. | <i>Factors 2 and 3, with species of each G-rank displayed. Species with G-ranks 1 and 2 vary in local abundance, but are restricted on range area, and increase in range as G-rank increases.....</i> | <i>55</i> |
| Figure 2.20. | <i>Rank abundance curve for 2233 species used in the current analyses. Abundance was averaged across all occurrences in the dataset for each species. The boundary of the tail of the curve is shown with the dotted line.....</i> | <i>57</i> |

- Figure 2.21.** *Species plotted along factors 1 (habitat) and 2 (range) and grouped into rare and common species defined by rank abundance. The dimensions of range and habitat were not use to define rarity in this case, so all information that would distinguish species on those attributes is lost, resulting in complete overlap of rare and common species. Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, QUERRUB = Quercus rubra, RHODCAT = Rhododendron catawbiense, RHODMAX = Rhododendron maximum, SMILROT = Smilax rotundifolia, TOXIRAD = Toxicodendron radicans, TRILDIS = Trillium discolor, VITIROT = Vitis rotundifolia.....*58
- Figure 2.22.** *Species plotted along factors 1 (habitat) and 3 (abundance) and grouped into rare and common species defined by rank abundance. Species with distribution patterns as different as Lindera melissifolia and Acer rubrum are both defined as common species. Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, QUERRUB = Quercus rubra, RHODCAT = Rhododendron catawbiense, RHODMAX = Rhododendron maximum, SMILROT = Smilax rotundifolia, TOXIRAD = Toxicodendron radicans, TRILDIS = Trillium discolor, VITIROT = Vitis rotundifolia.....*59
- Figure 3.1.** *Rank abundance curve for 2233 species used in the current analyses. Abundance was averaged across all occurrences in the dataset for each species. The boundary of the tail of the curve is shown with the dotted line.....*71
- Figure 3.2.** *Rarity groups as defined by Rank abundance. Species indicated by + symbols are those defined as rare (in the tails of the rank abundance curve, and also primarily at the lower end of each attribute axis). Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, QUERRUB = Quercus rubra, RHODCAT = Rhododendron catawbiense, RHODMAX = Rhododendron maximum, SMILROT = Smilax rotundifolia, TOXIRAD = Toxicodendron radicans, TRILDIS = Trillium discolor, VITIROT = Vitis rotundifolia.....*73
- Figure 3.3.** *Rarity groups as defined by division at the median of each of three attributes. Species indicated by + symbols are those defined as more rare. Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, QUERRUB = Quercus rubra, RHODCAT = Rhododendron catawbiense, RHODMAX = Rhododendron maximum, SMILROT = Smilax rotundifolia, TOXIRAD = Toxicodendron radicans, TRILDIS = Trillium discolor, VITIROT = Vitis rotundifolia.....*74

| | | |
|---------------------|--|----|
| Figure 3.4. | <i>Rarity groups as defined by clustering. Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, QUERRUB = Quercus rubra, RHODCAT = Rhododendron catawbiense, RHODMAX = Rhododendron maximum, SMILROT = Smilax rotundifolia, TOXIRAD = Toxicodendron radicans, TRILDIS = Trillium discolor, VITIROT = Vitis rotundifolia.....</i> | 75 |
| Figure 3.5. | <i>Average vector lengths for all species from 10% plot loss to 40% plot loss.....</i> | 76 |
| Figure 3.6. | <i>Average vector lengths for rarity groups defined by rank abundance and within protected community types of G-rank ≤ 2.....</i> | 77 |
| Figure 3.7. | <i>Average vector lengths for rarity groups defined by median and within protected community types of G-rank ≤ 2.....</i> | 78 |
| Figure 3.8. | <i>Average vector lengths for rarity groups defined by median and within protected community types of G-rank ≤ 3.....</i> | 78 |
| Figure 3.9. | <i>Average vector lengths for rarity groups defined by clustering and within protected community types of G-rank ≤ 2.....</i> | 79 |
| Figure 3.10. | <i>Average vector lengths for rarity groups defined by clustering and within protected community types of G-rank ≤ 3.....</i> | 80 |
| Figure 3.11. | <i>Representative species plotted along factors 1 and 2. Because the analysis is multidimensional, it is not possible to illustrate the direction of movement through the space. Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, RHODMAX = Rhododendron maximum, VITIROT = Vitis rotundifolia.....</i> | 81 |
| Figure 3.12. | <i>Representative species plotted along factors 2 and 3. Because the analysis is multidimensional, it is not possible to illustrate the direction of movement through the space. Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, RHODMAX = Rhododendron maximum, VITIROT = Vitis rotundifolia.....</i> | 82 |

CHAPTER 1

A comprehensive framework to define rarity

Abstract

Understanding patterns of species distribution and the mechanisms that shape those patterns has long been a fundamental objective in ecology and is critical for conservation planning, especially in the face of increasing rates of habitat fragmentation and global climate change, which may cause rapid and unpredictable changes in distribution patterns. However, despite many studies that have sought causes for rarity, few consistent trends have emerged. I argue that this lack of consistency is due, at least in part, to two factors. First, studies treat but often do not distinguish different forms of rarity (pattern), making comparisons among studies difficult. Second, many studies focus on inherent species traits as a basis of comparison, which neglects the effects of environmental and biotic context within which species interact (process). As a step toward a more cohesive approach to the study of rarity, I propose a novel framework to synthesize our current understanding of the pattern and process of rarity to facilitate communication of research results and guide future research priorities.

Introduction

The term rare has come to describe a range of conditions that are fundamentally different, produced and maintained by a range of different mechanisms (Fiedler and Ahouse 1992, Kunin and Gaston 1993, Gaston 1994, Binney and Bradfield 2000, Moora, Söber, and Zobel 2003). Numerous studies have examined causes of rarity, and many have provided valuable insights into mechanisms

causing distribution patterns of rare taxa. However, these results have yet to be synthesized in a way that provides a clear understanding of rarity. I propose an alternative approach to the study and communication of rarity, leading to a comprehensive framework to guide future research priorities.

The need for consistent communication

In a meta-analysis of studies comparing rare-common differences, Bevill and Louda (1999) stated that the data in the literature regarding related rare and common species is “scattered and unfocused” and has produced inconsistent results, yet they compared only studies of the same rarity type, defined as limited distribution and low abundance. If such inconsistency exists among studies investigating the same rarity type, the task of comparing studies when the term rare is not used consistently is daunting, if possible at all. I argue that there remains considerable inconsistency in the field due to variable use of the term rare as well as variable approaches to the study of rarity.

The case of rare plant competitive ability

One example of the difficulty of comparing research results comes from the study of competitive abilities of rare species. Early in the development of this line of work Griggs (1940) asserted that rare plants are rare because they are inferior competitors. Several recent studies have supported this hypothesis (Baskin and Baskin 1988; Walck, Baskin, and Baskin 1999; Binney and Bradfield 2000; Moora, Söber, and Zobel 2003), while several have contradicted it (Rabinowitz, Rapp, and Dixon 1984; Seabloom et al. 2003) or have given inconclusive results (Lloyd, Lee, and Wilson 2002). This collection of papers is often cited together as producing inconclusive results regarding the competitive abilities of rare plants. However, a close examination of these studies

reveals that they investigated different distribution patterns produced by different ecological processes (see **Table 1.1**), though all were simply called rare, resulting in the misleading interpretation that the results are inconsistent. Much of this inconsistency is due to variable use of the term rare and to variable methods (e.g., experimental comparison of rare-common congeners versus comparison of co-occurring species in context), rather than to actual contradiction of results. When the forms of rarity examined in each study are included (as in Table 1.1), consistency of conclusions is revealed. Therefore, we cannot directly compare the results of studies of rarity without first recognizing the distinctions among the forms of rarity examined. Reference to a comprehensive framework of definitions that includes and relates multiple definitions of rarity, combined with the mechanisms that produce them, will begin to resolve the contradiction and allow clear comparison and communication of results.

Table 1.1. *Comparison of studies examining the competitive abilities of rare species.*

| Study | Form of rarity considered | Evidence that rare species are inferior competitors? |
|------------------------------------|---------------------------|--|
| Baskin and Baskin (1988) | habitat restricted | yes |
| Walck, Baskin, and Baskin (1999) | habitat restricted | yes |
| Binney and Bradfield (2000) | habitat restricted | yes |
| Moora, Söber, and Zobel (2003) | habitat restricted | yes |
| Rabinowitz, Rapp, and Dixon (1984) | sparse | no |
| Seabloom et al. (2003) | sparse | no |
| Grubb (1986) | sparse | no |
| Lloyd, Lee, and Wilson (2002) | range restricted | Inconclusive |

The challenge of defining rarity

A multitude of definitions of rarity have been developed and are based on various criteria for delineating rarity. While existing definitions are too numerous to provide an exhaustive list here, I begin by identifying definitions that describe pattern (how things are rare) and those that address process (why things are rare).

Pattern - How things are rare

Geographic range restriction is perhaps the most common criterion used to define rarity, especially as it often includes habitat restriction and coincides with many species that are considered threatened or endangered. Due to fine-scale data limitations, range is often considered as total area occupied by a species. However, when data are available, some measure of occupancy, frequency, or density within the possibly-occupied range can be used to represent range more realistically (e.g., McGeoch and Gaston 2002). Likewise, habitat restriction can also be measured in an absolute sense or with some measure of occupancy, or proportion of occupied sites within all suitable sites (e.g., Rey Banayas et al. 1999). Abundance can be measured in many ways and varies by taxa and technique. For example, plant abundance may be measured as basal area or cover, and cover itself may be measured in a variety of ways. Rarity may then be defined as a maximum abundance value (by whichever method is used, e.g., a maximum cover value below which a taxon is considered rare) or rank abundance curves may be used, with varying thresholds applied to the tails of the curve to define rarity (Gaston 1994, Grime 1998, Murray et al. 1999).

While most studies focus on a single form of rarity, a few have combined forms into a relational framework to illustrate possible patterns. A classic example is the framework proposed by Rabinowitz (1981) which relates geographic range size, habitat restriction, and local abundance to produce seven forms of rarity. More recent frameworks have been proposed by Fielder and Ahouse

(1992), which includes two dimensions of rarity (geographic distribution and taxon age, both considered only for low-abundance species) and Rey Benayas et al. 1999, which adds occupancy as a fourth dimension to the three dimensions proposed by Rabinowitz.

Process - Why things are rare

Some of the earliest attempts to explain rarity focused on the criterion of range restriction and used the single mechanism of taxon age to explain restriction, such as rare taxa being new taxa that are geographically restricted because they simply had not had time to spread throughout the possible range (Willis 1922), or rare taxa being old taxa that are senescent and therefore decreasing in range as they approach extinction (Darwin 1859; Fernald 1922). Genetic diversity has also been used to explain rarity, if limited plasticity in the physiological requirements of a species limits the ability to expand into various locations or tolerate a range of conditions (Gleason 1924; Stebbins 1942). More recent studies have examined competitive ability, dispersal ability, reproductive traits, and species interactions such as seed predation and host-pathogen interactions (see Kunin and Gaston 1993; Rosenzweig and Lomolino 1997; Bevill and Louda 1999; Lloyd, Lee, and Wilson 2002; Moora, Söber, and Zobel 2003; Connell 1970, Janzen 1970, Hubbell 1979, Grubb 1986, Rabinowitz, Rapp, and Dixon 1984; Murray et al. 1999; and others).

Pattern and process together

Each pattern of rarity can be linked with possible mechanisms that produce it. Fiedler and Ahouse (1992) proposed one of the first frameworks to include and relate multiple criteria with multiple mechanisms to describe and explain patterns of rarity. The two criteria used are geographic distribution (narrow or wide) and persistence (short or long), producing a framework with four categories – species that have existed for a long time over wide or narrow ranges, and

those that have existed for only a short time over wide or narrow ranges. Mechanisms are then included to produce a hierarchy of possible causes of different forms of rarity. This application of mechanism is convincing in that it avoids the idiosyncratic approach of examining a single form of rarity in isolation and identifying single mechanisms that may produce it, and so provides a useful starting point for a comprehensive approach to investigating the causes of different distribution patterns.

Toward a comprehensive definition

I will describe several characteristics that a definition of rarity must possess to be useful for both research and practice, leading to the proposal of a comprehensive framework for defining rarity.

Multidimensional to capture pattern

The framework proposed by Fielder and Ahouse makes no distinction between geographic range restriction and habitat restriction. They describe the short/wide category (species that are rare over a wide range but have existed for only a short time) as being neoendemics, which implies habitat restriction, yet they also give examples in this category of species that are sparsely distributed, implying low abundance throughout the range, but not habitat restricted. Habitat restricted species may be widely distributed (predictable species such as *Rhizophora*) or narrowly distributed (such as narrow endemics). Likewise, widely distributed species may be habitat restricted (again, the predictable category) or not habitat restricted (sparse species). The example species *Rhizophora* is restricted by habitat (coastal fringe) but not by geographic range, and is not generally considered threatened, as compared to narrow endemics which are restricted by both criteria (according to Rabinowitz 1981), and are generally considered threatened and worthy of

investments of time and money to protect. Likewise, habitat restriction (including both predictable and endemic species) is created and maintained by a certain set of mechanisms (Baskin and Baskin 1988; Walck, Baskin, and Baskin 1999; Binney and Bradfield 2000; Moora, Söber, and Zobel 2003), whereas sparseness can be maintained by a quite different set of candidate causes (Rabinowitz, Rapp, and Dixon 1984; Seabloom et al. 2003). Removal of habitat restriction from the framework eliminates this level of resolution and therefore lumps patterns that are fundamentally different. The loss of resolution clouds our ability to discern specific mechanisms that would produce such different patterns.

Multiple dimensions of pattern also increase understanding of process. Stebbins (1980) considered several factors together to explain rarity. He asserted that rarity could be explained primarily by response to localized habitats, resulting in habitat restriction. Exceptions to this, such as related taxa that exhibit different degrees of habitat restriction, could then be explained by genetic variability within each taxon, allowing some species to tolerate a wider range of conditions and thus extend beyond localized habitat conditions. Exceptions to expected patterns based on one factor could be explained with another factor. The inclusion of multiple explanatory factors provides a more robust approach to understanding rarity.

Through the history of rarity research, authors have agreed that rarity is multidimensional (e.g., Fernald 1926, Stebbins 1942, Gleason 1924, Griggs 1940, Rabinowitz 1981, Fiedler and Ahouse 1992, Rey Benayas et al. 1999). Dimensions used to describe rarity have included abundance, geographic range size, habitat restriction, occupancy, and taxon age. As dimensions increase, so do the number of patterns and the resolution with which we can describe them; if dimensions are excluded, the patterns become less refined. Pattern motivates the investigation of mechanisms, so descriptions of pattern must be as clear and complete as possible before we can begin to understand what processes produce the pattern, as well as what processes are likely to change the pattern.

Mechanistic to capture process

Much of the purpose for describing patterns in nature is to understand the processes that produce them. For practical reasons most rarity frameworks are static and discrete (species are placed into categories and remain there); however, in reality species distributions are continuous and species do move within that continuous space - declining species become rare, invading species become common - in both evolutionary and ecological time (e.g., Fielder and Ahouse 1992). Here I refer to mechanism as the ecological process, or pathway, by which species either remain stable within a distribution pattern or shift within the continuous space.

Seabloom et al. (2003) examined the competitive abilities of native and exotic grasses in California. They hypothesized that the exotic grasses had been able to invade and persist due to superior competitive ability compared to native grasses. However, their results indicated that the native grasses were indeed strong competitors relative to the exotic species. The native grasses remained rare due to recruitment limitation, not direct competitive suppression. Experimental seed addition allowed the native grasses to expand and suppress the abundance of the exotic species (to shift from rare to more common). This example provides evidence of a specific ecological mechanism allowing a species to change distribution pattern, or to shift between categories of rarity. Additionally, and from a management perspective more importantly, this understanding of mechanism allows much more targeted and efficient interventions for restoration. The hypothesis of inferior competitive ability implicates eradication and containment of exotics as the necessary restoration strategy, a difficult and costly endeavor; seed addition is a much more feasible alternative (Seabloom et al. 2003; Lodge et al. 1998; Hobbs et al. 1995). Therefore, identifying the mechanism that produces the pattern gives us not only an understanding of the processes that allow species to shift between distribution patterns, but can greatly improve conservation biology.

Comprehensive to consider traits in context

Many studies of rarity have focused on identifying traits that produce the condition of rarity. Rabinowitz (1981) criticized this approach as being “rather monolithic,” pointing to past authors’ tendencies to attach consistent traits to certain forms of rarity, and thus making the condition of rarity an inherent state of being, or a species-specific attribute. Binney and Bradfield (2000) considered context experimentally in their study of rare-common differences. They did not compare a rare species and a common congener, rather a rare species and a species with which it co-occurs. The latter comparison tells us more about the ability of the rare species to persist within the context it exists, rather than attempting to elucidate the effects of inherent traits of rare and common species, the results of which are likely to be modified in context as the species encounters different neighbors in different locations.

Inherent traits certainly do affect species’ distribution patterns, as they determine fundamental growth and abundance responses, but ecological context determines the shape patterns take in the landscape. Most species do exhibit varying distribution patterns in the landscape. This variation is due to context - species with a given suite of traits will respond differently as they encounter different neighbors and different environmental conditions (see Murray et al. 1999). Traits determine the fundamental niche; context, in part, shapes the realized niche (Austin and Smith 1989).

We must recognize that rarity *per se* is not necessarily an inherent species trait, and that multiple mechanisms may be responsible for producing and maintaining varying distribution patterns, even within the same species (see Drury 1980; Mueller-Dombois and Ellenberg 1974). While the study of traits is useful in understanding rarity, traits alone cannot explain distribution patterns, and certainly should not be taken to have static effects. A context approach - comparing rare species responses to common species with which they co-occur in places where they co-occur,

rather than congeners, which they may not often encounter in nature and therefore have little influence on pattern - will provide a much more complete understanding (Rabinowitz et al. 1984; Binney and Bradfield 2000; Seabloom et al. 2003).

The proposed framework

Figure 1.1 presents a graphical synthesis of the approach I propose. Forms of rarity are shown relative to one another rather than in isolation, forms are linked with the mechanisms that produce and maintain them, and the continuous nature of species distributions is explicitly recognized, rather than assigning species to discrete *a priori* categories that may or may not exist. The three examples of forms of rarity shown (habitat restricted, sparse, and somewhere-abundant) are demarcated with dotted lines to indicate that those forms are not discrete, but rather represent realms within a continuous space. The framework is displayed in three-dimensional space, however, the dimensions shown (habitat, abundance, and range) as well as additional dimensions including frequency and occupancy, must be considered as possible dimensional that define rarity for a given dataset.

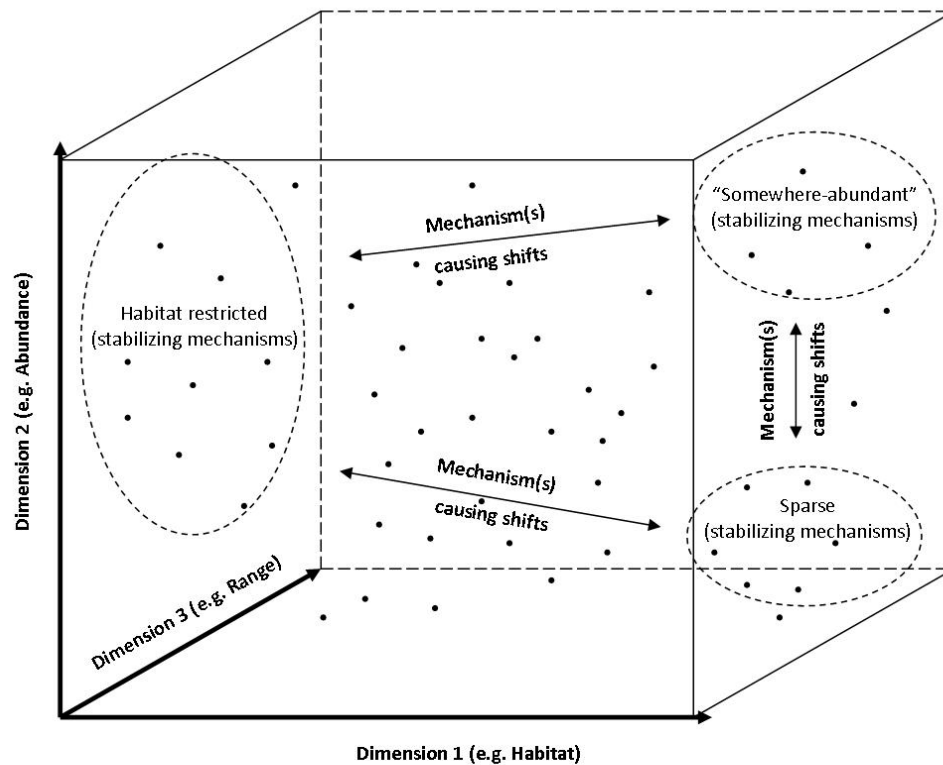


Figure 1.1. A comprehensive, multidimensional, mechanistic framework to define rarity. Points represent species in attribute space; arrows represent ecological mechanisms allowing species to shift among forms of rarity.

The comprehensive framework in practice: The condition of sparseness

To illustrate the application of a multidimensional, mechanistic, comprehensive framework, I consider the distribution pattern of sparseness, which I define as having wide geographic distribution across a range of habitat types, but never reaching high abundance at any given site. There has been a recurrence in the literature of paired descriptions of being sparse (having a wide range and broad habitat tolerance yet never reaching high abundance) and of being abundant somewhere (again having a wide range and broad habitat requirements, but reaching higher abundance at some point in the range). These patterns have been referred to or overlap with the concepts of sparse and pseudo-rare (Rabinowitz 1981), satellite and core species (Hanski 1982), suffusive and diffusive rarity (Schoener 1987) and “everywhere-sparse” and “somewhere-abundant”

(Murray et al. 1999), and have been explained by various mechanisms, such as mass effects (Shmida and Wilson 1985), source-sink population dynamics (Grubb, Kelly, and Mitchley 1982; Pulliam 1988), and range effects (e.g., Brown 1984; Brown, Mehlman, and Stevens 1995).

There are multiple pathways by which a species might shift among the patterns of sparseness, “somewhere-abundant” (sparse through most of the range but reaching higher abundance at some locations, though not restricted to a particular habitat type), and habitat restriction (occurring only within a certain habitat type, whether at high or low abundance) (**Table 1.2**). A comparison of these distribution patterns provides insights into what processes allow some species to increase in abundance in context while others are suppressed throughout a broad range.

Sparseness may be produced as one species is competitively suppressed but not excluded. Grubb (1986) demonstrated that interference competition holds some species sparse in species-rich chalk grasslands. The sparse species persist by being better able than the dominants to establish from seed in small microsites, as well as by the reduction of root and shoot competition from the dominants by grazing. This form of sparseness can transition to “somewhere-abundant” by the loss of the competitively dominant neighbor, or by conditions at a particular site that give the poorer competitor an advantage. Perhaps more important to consider is the reverse movement – if a “somewhere-abundant” species encounters a new and competitively dominant neighbor as the neighbor invades, the focal species may be driven to the sparse category through suppression. Conversely, Seabloom et al. (2003) demonstrated that rare native grasses in California were strong competitors relative to the more common invading exotic grasses, but were recruitment limited. The sparse distribution was maintained in this case by recruitment limitation, not competitive suppression, and movement to become “somewhere-abundant” occurred by seed addition. Therefore, it is possible for a single distribution pattern to be maintained by different mechanisms, and for shifts between patterns to occur by different pathways.

Sparseness may also be maintained by other mechanisms such as Janzen-Connell effects (self-repulsion of conspecifics through density-dependent effects of species-specific pathogens, herbivores, or seed predators) or apparency (Janzen 1970; Connell 1970; Fragoso, Silviu, and Correa 2003; Feeney 1976), and may shift to “somewhere-abundant” by the loss of or escape from the interacting taxa. In an unintentional experiment at a long-term study site in Peru, Silman, Terborgh and Kiltie (2003) discovered that tropical trees that had been sparsely distributed became significantly more dense at a location where a major seed predator, the white-lipped peccary, was lost from the area for a period of 12 years. Upon recovery of the peccary population, the tree distribution returned to its previous sparse spatial pattern. Escape from interacting taxa may also allow a change in this form of sparseness. Plant species that become naturalized outside their native range have been shown to carry lower fungal and viral pathogen loads, allowing greater spread of the invading plants in the non-native range (Mitchell and Power 2003). Similarly, Kilronomos (2002) demonstrated that plants tend to accumulate species-specific pathogens at different rates (rare plants tend to accumulate pathogens quickly whereas invasive species accumulate them more slowly; the author did not provide a specific definition for rarity), and that rare and invasive plants respond differently when grown in foreign versus home soil (rare plants experienced decreased growth in home soil where pathogens had accumulated, whereas invasive species did not). These examples further illustrate that inherent traits and ecological context must be considered together to fully understand factors affecting abundance patterns.

In a similar vein, habitat restricted species may be considered here. Several examples exist of species that are habitat restricted not because they require the specific habitat in which they occur, but because they tolerate it when other neighbors cannot. They are constrained to the less favorable habitat by an inability to compete with neighbors in more favorable conditions. Examples are provided by Goldberg (1982), who demonstrated that competition from deciduous oaks on

higher nutrient soils restricts evergreen oaks to nutrient-poor sites, and Baskin and Baskin (1988) who showed that species endemic to rock outcrops in the southeastern United States require high light levels and therefore can only persist on rock outcrops where light is available since competing species are excluded from the outcrops. Loss of the competing taxon may allow movement into the “somewhere-abundant” category if these species are able to establish in locations outside the restricted habitat.

Changing conditions that give an advantage to the restricted species may also allow an increase in abundance. Walck et al. (1999) found that outcrop-restricted endemics were strong root competitors but poor shoot competitors, and therefore persisted on drier sites where competing neighbors were excluded, whereas on wetter sites, the endemic was overtopped by neighbors. A series of extremely dry seasons may give the endemic an advantage if the dominant neighbors could not persist under drought conditions. Similarly, Binney and Bradfield (2000) demonstrated that a restricted grass species persisted in shallow soils because deep root systems stabilized the plants through freeze-thaw cycles whereas competitively dominant but shallow-rooted neighbors were uprooted from the shallow soils. An increase in severity of freeze-thaw disturbance that would affect deeper soils may give the restricted species an advantage as conditions change.

Table 1.2. *Proposed mechanisms that maintain sparse, “somewhere-abundant,” and habitat restricted forms of rarity and allow shifts between them in ecological time.*

| Sparse | Somewhere Abundant | Habitat Restricted |
|--|--|--|
| 1) Complete competitive suppression by same neighbor throughout | 1) Form of habitat restricted where loss of interacting taxon allows release | 1) Tolerance of marginal sites allows escape from competitively dominant neighbors (Goldberg 1982; Baskin and Baskin 1988; Walck et al. 1999; Binney and Bradfield 2000) |
| 2) Complete competitive suppression by different neighbors throughout (Grubb 1986) | 2) Form of sparse where loss of interacting taxon allows release throughout (in form 1) or at some locations (in form 2) | 2) Relictual |
| 3) Recruitment limitation (Seabloom et al. 2003) | 3) Form of sparse where management intervention allows release | |
| 4) Self-repulsion by direct competition (Rabinowitz, Rapp, and Dixon 1984) | | |
| 5) Self-repulsion by apparency or Janzen-Connell spacing effects (Janzen 1970; Connell 1970; Fragoso, Silvius, and Correa 2003; Feeney 1976) | 4) Loss of interacting taxa (Silman, Terborgh, and Kiltie 1999) | |
| | 5) Escape from interacting taxa (Kilronomos 2002; Mitchell and Power 2003) | |
| (any mechanism) | 6) Mass effects (e.g., Shmida and Wilson 1985) | (any mechanism) |

As shown in Table 1.2, a sparse distribution pattern may result from a variety of mechanisms.

Likewise, the same suite of traits is likely to produce different responses in different contexts (environmental conditions and presence /absence of interacting taxa), allowing shifts to other patterns by various pathways. Therefore, searching for single inherent explanatory factors by attempting to correlate rarity with specific traits is unlikely to capture the range of factors that drive distribution patterns. A few of the mechanisms proposed in Table 1.2 are superimposed on the three-dimensional framework shown in **Figure 1.2** to illustrate how these mechanisms might maintain and change species locations in the continuous rarity space.

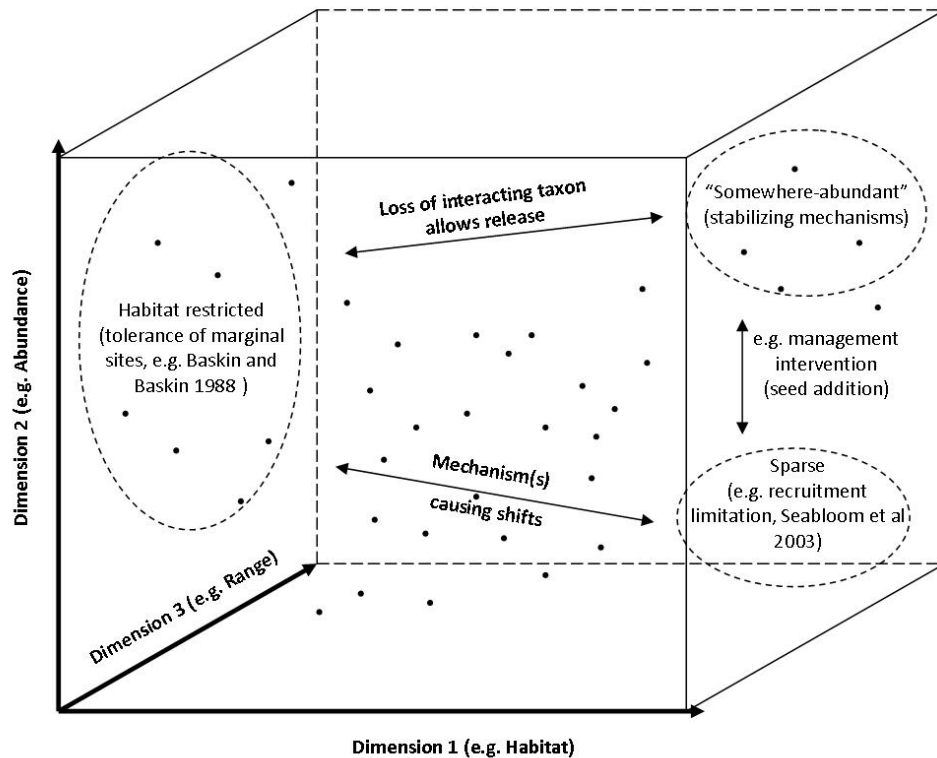


Figure 1.2. Representative mechanisms from Table 1.2 illustrate how species locations in each realm of rarity space might be maintained as well as how species might shift through the space.

A call for research priorities

Communicate consistently

First and foremost, authors must provide unambiguous definitions that are as complete as possible, include as many dimensions as possible, and provide clear descriptions of how each index was measured (e.g., abundance is meaningless unless it is clearly stated whether it was measured by cover or counts, and how the measurement was taken).

Gaston (1994) provides a thorough review of the many definitions and measurements for rarity. However, despite this wealth of terms for rarity, no single definition can be satisfying as a comprehensive description. Within an individual study of rarity, a single definition may suffice, and certainly the specific form examined should be described and defined, especially if a specific index is used, such as abundance, density, range, or frequency, each of which can be measured differently

and at different scales. The onus remains on the researcher to provide a clear definition of terms regardless of whether they use a singular definition or fit their work into an existing framework.

Determine dimensionality

Fielder and Ahouse (1992) propose two dimensions to describe rarity (geographic distribution and persistence). Rabinowitz (1981) proposed three (geographic range, habitat restriction, and local abundance). Rey Benayas, Scheiner, Sánchez-Colomer, and Levassor (1999) propose four, adding the criterion of habitat occupancy to range, abundance, and habitat. While there are compelling practical and empirical reasons behind each proposed framework, it remains unclear whether the proposed dimensions provide unique or redundant information. For example, abundance and occupancy may not be orthogonal, as they are derived from the same data matrix (He and Gaston 2003; Kunin 1998; Hartley, Kunin, Lennon, and Pocock 2004).

Dimensionality matters for several reasons. Dimensionality reduces ambiguity since it places distribution patterns clearly within context of each other and allows clear linkage with mechanisms so pattern and process may be synthesized. However, the goal should be for dimensionality to be complete but *efficient*, meaning determining the minimum number of dimensions required to sufficiently capture pattern. Efficiency matters from an empirical perspective to avoid data redundancy, but also from a practical standpoint for the same reason. We cannot afford to waste time or money measuring species attributes that do not contribute directly and clearly to capturing pattern and process (e.g., taxon age may be problematic for this reason, Bevill and Louda 1999). Identifying the core dimensions that comprise rarity will allow us to streamline research, monitoring, and restoration practices.

Identify mechanisms

I have proposed a comprehensive approach to study and describe rarity. However, the examples I provide address only three distribution patterns, and suggest only a sampling of mechanisms. Additional patterns and processes must be examined. Demographic data, for example, are largely missing from the literature (Bevill and Louda 1999). Other mechanisms, especially anthropogenic effects such as habitat loss and climate change, must be included to anticipate potentially rapid movement among forms. Only with a more complete and comprehensive understanding of rarity can we reverse or even predict and prevent rapid shifts in distribution as invading species become more common or rare species approach extinction.

Conclusions

I have argued that the study of rarity requires a multidimensional mechanistic approach in order to accurately represent and understand distribution patterns and to communicate results effectively. Figure 1.1 presents a graphical synthesis of the approach I propose. Forms of rarity are shown relative to one another rather than in isolation, forms are linked with the mechanisms that produce and maintain them, and the continuous nature of species distributions is explicitly recognized, rather than assigning species to discrete *a priori* categories that may or may not exist. While the framework is theoretical, the questions implied are clearly empirical and computationally intensive. As we enter the age of ecoinformatics, we now have the computing power, the analytic methods, and the large-scale publicly available datasets needed to extend Figure 1.1 with the appropriate number of dimensions and relevant mechanisms to describe and understand rarity for any given flora or fauna.

References

- Austin, M.P. and T.M. Smith. 1989. A new model for the continuum concept. *Vegetatio* 83:35 - 47.
- Baskin, J.M. and C.C. Baskin. 1988. Endemism in rock outcrop plant communities of unglaciated eastern United States: An evaluation of the roles of the edaphic, genetic and light factors. *Journal of Biogeography* 15:829 - 840.
- Bevill, R.L. and S.M. Louda. 1999. Comparisons of related rare and common species in the study of plant rarity. *Conservation Biology* 13:493 - 498.
- Binney, E.P. and G.E. Bradfield. 2000. An initial comparison of growth rates in the rare grass *Achnatherum hendersonii* and its common associate *Poa secunda*. *Ecological Research* 15:181 - 185.
- Brown, J.H., D.W. Mehlman, and G.C. Stevens. 1995. Spatial variation in abundance. *Ecology* 76:2028 - 2043.
- Brown, J.H. 1984. On the relationship between abundance and distribution of species. *American Naturalist* 124:255 - 279.
- Connell, J.H. 1970. On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. In *Dynamics of Populations: Proceedings of the Advanced Study Institute on Dynamics of Numbers in Populations* (eds. P.J. den Boer and G.R. Gradwell) Centre for Agricultural Publishing and Documentation, Netherlands. 298 - 310.
- Darwin, C. 1859. *The Origin of Species* New American Library, New York :.
- Drury, W.H. 1980. Rare species of plants. *Rhodora* 82:3 - 48.
- Feeny, P. 1976. Plant apparency and chemical defense. In *Biochemical Interaction between Plants and Insects* (eds. J.W. Wallace and R.L. Mansell) Plenum Press, NY. 1 - 40.
- Fernald, M.L. 1926. The antiquity and dispersal of vascular plants. *Quarterly Review of Biology* 1:212 - 245.
- Fernald, M.L. 1924. Isolation and endemism in northeastern America and their relation to the age-and-area hypothesis. *American Journal of Botany* 11:558 - 572.
- Fiedler, P.L. and J.J. Ahouse. 1992. Hierarchies of cause: Toward an understanding rarity in vascular plant species. In *Conservation Biology: The theory and practice of nature conservation, preservation, and management* (eds. P.L. Fiedler and S.K. Jain) Chapman and Hall, NY. 23 - 47.
- Fragoso, J.M.V., K.M. Silvius, and J.A. Correa. 2003. Long-distance seed dispersal by tapirs increases seed survival and aggregates tropical trees. *Ecology* 84:1998 - 2006.
- Gaston, K.J. 1994. *Rarity* Chapman and Hall, NY. 205.
- Gleason, H.A. 1924. Age and area from the viewpoint of phytogeography. *American Journal of Botany* 11:541 - 546.
- Goldberg, D.E. 1982. The distribution of evergreen and deciduous trees relative to soil type: An example from the Sierra Madre, Mexico, and a general model. *Ecology* 63:942 - 951.

- Griggs, R.F. 1940. The ecology of rare plants. *Bulletin of the Torrey Botanical Club* 67:575 - 594.
- Grime, J.P. 1998. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *Journal of Ecology* 86:902 - 910.
- Grubb, P.J. 1986. Problems posed by sparse and patchily distributed species in species-rich plant communities. In *Community Ecology* (eds. J. Diamond and T.J. Case) Harper and Row, NY. 207 - 225.
- Grubb, P.J., D. Kelly, and J. Mitchley. 1982. The control of relative abundance in communities of herbaceous plants. In *The Plant Community as a Working Mechanism* (ed. E.I. Newman) Blackwell, UK. 79 - 97.
- Hanski, I.A. 1982. Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos* 38:210 - 221.
- Hartley, S., W.E. Kunin, J.J. Lennon, and M.J.O. Pocock. 2004. Coherence and discontinuity in the scaling of species' distribution patterns. *Proceedings of the Royal Society of London Series B* 271:81 - 88.
- He, F. and K.J. Gaston. 2003. Occupancy, spatial variance, and the abundance of species. *American Naturalist* 162:366 - 375.
- Hobbs, R. J. and S.E. Humphries. 1995. An Integrated Approach to the Ecology and Management of Plant Invasions. *Conservation Biol.* 9:761-770.
- Hubbell, S.P. 1979. Tree dispersion, abundance, and diversity in a tropical dry forest. *Science* 203:1299 - 1309.
- Janzen, D.H. 1970. Herbivores and the number of tree species in tropical forests. *American Naturalist* 104:501 - 528.
- Kilronomos, J.N. 2002. Feedback with soil biota contributes to plant rarity and invasiveness in communities. *Nature* 417:67 - 70.
- Kunin, W.E. 1998. Extrapolating species abundance across spatial scales. *Science* 281:1513 - 1515.
- Kunin, W.E. and K.J. Gaston. 1993. The biology of rarity: patterns, causes and consequences. *Trends in Ecology and Evolution* 8:298 - 301.
- Lloyd, K.M., W.G. Lee, and J.B. Wilson. 2002. Competitive abilities of rare and common plants: Comparisons using *Acaena* (Rosaceae) and *Chionochloa* (Poaceae) from New Zealand. *Conservation Biology* 16:975 - 985.
- Lodge, D. M., R.A. Stein, K.M. Brown, A.P. Covich, C. Bronmark, J.E. Garvey, And S.P. Klosiewski. 1998. Predicting Impact of freshwater exotic species on native biodiversity: Challenges in spatial scaling. *Aust. J. Ecol.* 23:53-67.
- McGeoch, M. and K.J. Gaston. 2002. Occupancy frequency distributions: patterns, artefacts and mechanisms. *Biological Review* 77:311 - 331.
- Miller, R.S. 1967. Pattern and process in competition. *Advances in Ecological Research* 4:1 - 74.
- Mitchell, C.E. and A.G. Power. 2003. Release of invasive plants from fungal and viral pathogens. *Nature* 421:625 - 627.

- Moora, M., V. Söber, and M. Zobel. 2003. Responses of a rare (*Viola elatior*) and a common (*Viola mirabilis*) congeneric species to different management conditions in grassland - is different light competition ability responsible for different abundances?. *Acta Oecologia* 24:169 - 174.
- Mueller-Dombois, D. and H. Ellenberg. 1974. *Aims and methods of vegetation ecology* John Wiley & Sons, New York. 547.
- Murray, B.R., B.L. Rice, D.A. Keith, P.J. Myerscough, J. Howell, A.G. Floyd, K. Mills, and M. Westoby. 1999. Species in the tail of rank-abundance curves. *Ecology* 80:1806 - 1816.
- Pulliam, H.R. 1988. Sources, sinks, and population regulation. *American Naturalist* 132:652 - 661.
- Rabinowitz, D. 1978. Abundance and diaspore weight in rare and common prairie grasses. *Oecologia* 37:213 - 219.
- Rabinowitz, D. 1981. Seven forms of rarity. In *The Biological Aspects of Rare Plant Conservation* (ed. H. Synge) J. Wiley, Chichester, England. 205 - 217.
- Rabinowitz, D., J.K. Rapp, and P.M. Dixon. 1984. Competitive abilities of sparse grass species: Means of persistence or cause of abundance. *Ecology* 65:1144 - 1154.
- Rey Benayas, J.M., S.M. Scheiner, M.G. Sánchez-Colomer, and C. Levassor. 1999. Commonness and rarity: Theory and application of a new model to Mediterranean montane grasslands. *Conservation Ecology* 3:5.
- Rosenzweig, M. and M. Lomolino. 1997. Who gets the short bits of the broken stick?. In *The Biology of Rarity: Causes and Consequences of Rare-Common Differences* (eds. W. Kunin and K. Gaston) Chapman and Hall, London. 63 - 90.
- Schoener, T.W. 1987. The geographical distribution of rarity. *Oecologia* 74:161 - 173.
- Shmida, A. and M.V. Wilson. 1985. Biological determinants of species diversity. *Journal of Biogeography* 12:1 - 20.
- Silman, M.R., J.W. Terborgh, and R.A. Kiltie. 2003. Population regulation of a dominant rain forest tree by a major seed predator. *Ecology* 84:431 - 438.
- Stebbins, G.L. 1942. The genetic approach to problems of rare and endemic species. *Madrono* 6:241 - 258.
- Stebbins, G.L. 1980. Rarity of plant species: A synthetic viewpoint. *Rhodora* 82:77 - 86.
- Walck, J.L., J.M. Baskin, and C.C. Baskin. 1999. Relative competitive abilities and growth characteristics of a narrowly endemic and a geographically widespread *Solidago* species (Asteraceae). *American Journal of Botany* 86:820 - 828.
- Willis, J.C. 1922. *Age and Area* Cambridge, UK. 259.

CHAPTER 2

A quantitative approach to determine the dimensionality of rarity

Abstract

Understanding of the patterns and processes that drive rarity is crucial for the conservation of rare species as well as the control of invasive species (species that are rare in a landscape and quickly become common). Although several frameworks to categorize and describe rarity have been proposed, neither the categories nor the placement of species within them have been determined empirically for a regional set of organisms. An incomplete description of the patterns of rarity limits our ability to gain a more complete understanding of the causes and extent of different types of rarity. I used a multivariate ordination analysis and factor analysis for vascular plant species of the southeastern United States to identify the dimensions that define rarity for the region. I used nine species attributes, or dimensions, to define rarity, including various measures of local abundance, range area, range shape, habitat volume, and frequency. Starting from the nine original dimensions, the analyses consistently produced a three-dimensional solution representing abundance, range, and habitat volume. This result was robust for the full dataset (vascular flora of the southeastern US) as well as within each subregion (vascular flora of the mountains, piedmont, and coastal plain). The data show no inherent pattern of clustering into groups of rarity. While groups may be defined by various methods for further analysis and are useful for purposes of conceptual convenience and communication, they do not represent an inherent pattern in the data; species do not form consistent, distinct, or discrete groups in attribute space.

Introduction

It is generally accepted that rarity is defined by multiple axes, or dimensions, such as local abundance and geographic range (e.g., Gaston 1994; Rabinowitz 1981; Kunin and Gaston 1993; Reilly in preparation). Although several multidimensional frameworks of rarity have been proposed, the number and identity of dimensions remains uncertain. Fielder and Ahouse (1992) propose two dimensions to describe rarity (geographic distribution and persistence) with persistence defined as taxon age. Rabinowitz (1981) proposed a now classic framework of three dimensions (geographic range, habitat preference, and local abundance) to produce seven possible forms of rarity. A more recent framework proposed by Rey Benayas et al. (1999) adds occupancy to the three dimensions proposed by Rabinowitz to produce sixteen possible forms of rarity.

Several shortcomings of current rarity frameworks have been suggested. For example, taxon age can be difficult to determine with useful precision (Bevill and Louda 1999), and occupancy may provide redundant information in that occupancy and abundance are derived from the same data matrix and therefore are not independent (He and Gaston 2003; Kunin 1998; Hartley, Kunin, Lennon, and Pocock 2004). Additionally, studies to date have placed species into categories by simply dividing species lists according to high or low abundance, wide or narrow habitat preference, and wide or narrow geographic range (e.g., Kaye et al. 1997; Rey Benayas et al. 1999; Yu and Dobson 2000). Of necessity, such approaches to assigning species to categories make two assumptions – that the proposed categories actually exist in nature, and that species sort into discrete groups. The use of discrete categories, by definition, neglects the continuous nature of species distributions, and *a priori* definition assumes distribution patterns that may not exist in nature, or may not generalize to different systems. While such categorization is straightforward and heuristically useful, I argue that a multivariate approach will allow us to identify and represent patterns of rarity more realistically, and will lead to a more thorough understanding of the pattern and process of rarity.

Rather than assigning species to predefined categories, a multivariate approach will explicitly illustrate the continuous distribution of species relative to the dimensions that define rarity for a given dataset, revealing possible inherent patterns in the data. Once the inherent patterns in the data are identified, then categories or groups can be defined and interpreted based on questions of interest for a given dataset. This approach essentially reverses the order of operations - the data drive the identification of pattern, and groupings are simply used for convenience of interpretation, rather than allowing *a priori* groups to drive the definition of pattern.

Pattern and process are inextricably linked; to understand the processes that produce rarity, a necessary step is to describe patterns of rarity as clearly and completely as possible. For example, two species may be labeled rare, defined by their occurrence in the tails of the rank abundance curves for a given dataset data. However, one of those species may increase in abundance elsewhere in its range, meaning its rarity in the given data may be driven by mass effects or edge-of-range effects (e.g., Shmida and Wilson 1985, Brown 1984, Brown et al. 1995), whereas the other species is abundant nowhere in its range, indicating a truer form of rarity driven by different processes. Neglecting the dimension of range or examining data for only a restricted geographic area neglects aspects of pattern that distinguish the two species considered, and therefore gives an incomplete understanding of the processes that drive each distribution pattern.

Determining the dimensionality of rarity will produce more complete descriptions of pattern to facilitate clear communication as well as more comprehensive understanding of the processes that produce rarity. However, the goal should be for dimensionality to be complete but *efficient*, meaning determining the minimum number of dimensions to sufficiently capture pattern. Monitoring multiple species attributes can be expensive and time consuming, and the analysis of such data can be computationally intensive. Identifying the core dimensions that comprise rarity

will streamline research, monitoring, and restoration practices, and will result in a more heuristic mental framework.

Defining the dimensionality of rarity is ultimately an empirical question. In this study, I take advantage of publicly available vegetation data spanning the continental US, in combination with a large collection of vegetation plot data of the vascular flora of the southeastern US, to determine the dimensionality of rarity for the region. Rather than simply bifurcating species lists into predetermined categories (e.g., high or low abundance, wide or narrow range, etc), I take the approach of using multivariate analysis to reveal and interpret inherent patterns in the data.

Methods

Data Preparation

The dataset consists of 2233 vascular plant species sampled from 2424 plots in the southeastern US, following methods of the Carolina Vegetation Survey (CVS) (Peet, Wentworth, and White 1998). Plot size is equal to or greater than 100m². Plot size and shape vary to represent the scale of sampled habitat. For example, the 20x50m plot design may be reduced to 5x20m along river levees, or to 10x10m where habitat and vegetation is homogeneous but restricted in extent. Limiting data selection to uniform plot size or shape would systematically bias habitat and community type representation. Descriptions of the 2424 plots are included in Appendix 1 and Supplemental file Appendix1_Plots.csv.

Plots with incomplete data for soil nutrients, texture, elevation, and slope were not used in the analysis. Soil samples are taken from several locations within a plot, resulting in several values for each soil variable for a typical plot. For the current study, composite data for each plot were used such that each plot represents one sample unit. The mean value of soil texture variables was used. The median value for soil nutrient variables was used to avoid the disproportionate influence of

extreme values. A single value for slope and elevation are recorded for each plot. The dataset includes 25 soil nutrient variables, three soil texture variables, elevation, and slope for each of the 2424 plots. Environmental data for 2424 plots are described in Appendix 2 and included in Supplemental file Appendix2_EnvData.csv.

Taxonomy within the CVS data follows Weakley (2006). Taxonomy within the USDA data, from which range data were taken, mostly follows Kartesz (2002). In the current analysis, which combines CVS and USDA data, taxonomy follows Weakley (2006) with the exception of unresolved or ambiguous taxonomic relationships between Weakley and Kartesz. Varieties and subspecies were lumped to the species level, except in cases where only one variety or subspecies occurred within the dataset. The list of 2233 species as well as CVS and USDA species names and codes are described in Appendix 3 and included in Supplemental file Appendix3_SpeciesList.csv.

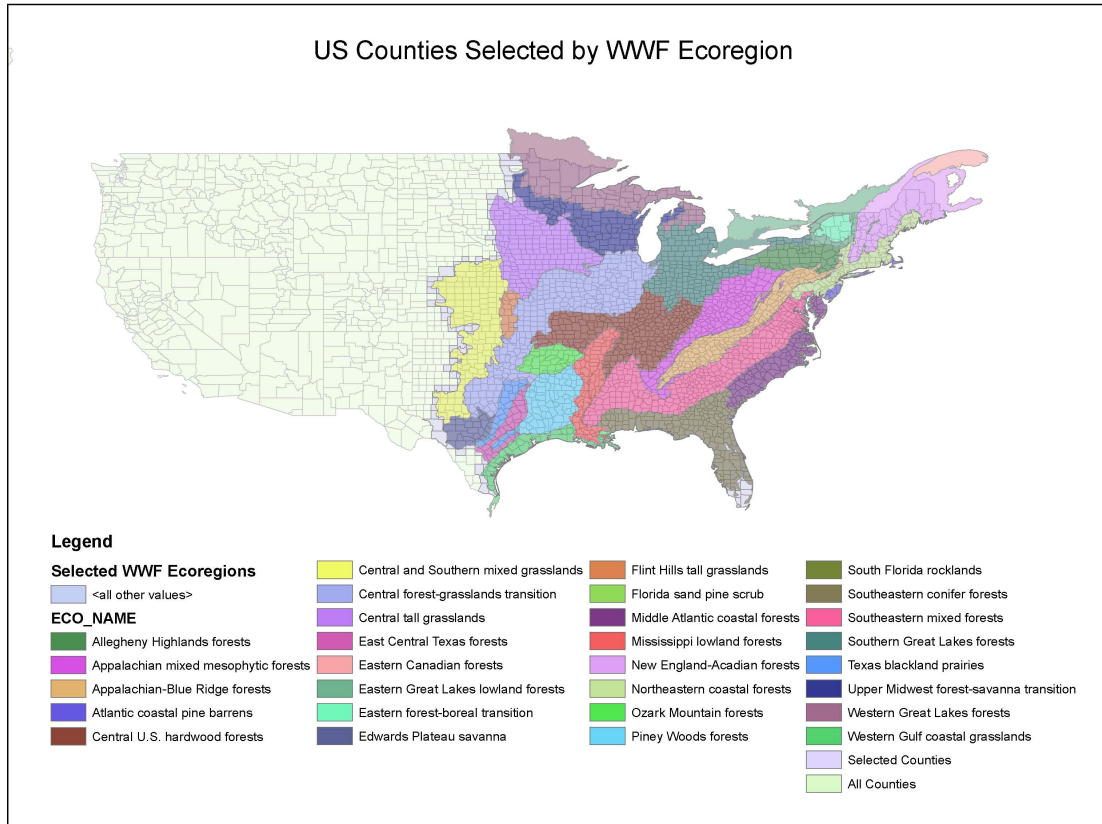
Using the above data sources, I created a species-by-attribute matrix with species as sample units and measures for range, local abundance, habitat preference, and frequency as species attributes. The calculation of each species attribute is described below.

Range

Four measures for range were calculated. Range data were derived from the USDA Plants Database, which includes the area in square miles of each US county where each species occurs. The database contains species occurrence data only for the United States, resulting in a truncation of data at northern and southern political boundaries. To account for this data limitation, and to provide equal consideration of species ranges with north-south and east-west spans, the USDA data were truncated at a western boundary based on World Wildlife Fund ecoregions. Only US counties contained within the boundary were considered for analysis (**Figure 2.1**). **Range area** for each species was calculated as the sum of areas of all US counties within the truncated range where each

species occurs. An additional limitation of the USDA data is a mismatch between state and federal records of occurrence for certain species, with several states being undersampled in the federal records. To account for this undersampling, I calculated **area ratio** as the sum of areas of US counties where the species occurs, divided by the total available area, calculated as the sum of areas of states missing subtracted from the sum of all US counties within the truncated range. **East-west** and **north-south extent** were included to account for the various processes that determine range size and shape (Gaston 2003, Ruggiero 2001, Ruggiero et al. 1999). Using the county centroid as the point location of each county (determined by the polygon centroid in GIS) (ESRI ArcInfo 9.2 1999-2006), extent was measured as the distance between the northern-most and southern-most county centroids, or eastern- and western-most county centroids, for each species.

Figure 2.1. Map of US counties included in analysis. The western boundary was determined by the ecoregion boundaries that created the most homogeneous representation of flora for the eastern US and the most consistent extent in all directions from the focal region (North Carolina and surrounding states) to avoid truncation of data by northern and southern political boundaries.



Local Abundance

Four measures of local abundance were calculated: mean abundance, maximum abundance, variance in abundance, and range in abundance. Local abundance was measured as cover class following CVS protocol. Each measure was then calculated by taking the mean, maximum, variance, or range in abundance for each species across all plots within which the species occurs. Preliminary analyses revealed no difference between using arithmetic mean, geometric mean, or mode for average abundance.

Habitat Preference

Two measures of habitat preference were calculated: a count of NatureServe Vegetation Associations (**unique CEGl codes**) where each species occurs within the CVS data, and **habitat volume** of each species. Habitat volume was defined following a procedure similar to that used by Rey Benayas et al. (1999). In each plot, 26 environmental variables were measured. Nonmetric multidimensional scaling was used to reduce the dimensionality of the data, resulting in a three dimensional solution consisting of plots as points in environmental space. Thus, each plot had an ordination score that defined its location along each of the three ordination axes. The ordination scores for the plots at the maximum and minimum location along each axis were used to calculate a three dimensional volume for each species. Species that occurred in only one plot, and therefore had a volume of zero, were assigned a volume smaller than the smallest calculated volume.

Frequency

Three measures of frequency were calculated: a count of all counties in the truncated range where each species occurs, a count of CVS plots in which a species occurs, and a weighted plot count in which each plot was counted in proportion to the number of plots in that region (mountains, piedmont, coastal plain/fringe) to account for disproportionate representation of plots in each region. For example, because there were 1162 mountain plots included in the 2424 total plots used, the weighted plot count tallies each mountain plot as 0.52 instead of 1.0, to account for the disproportionate number of mountain plots. Each piedmont plot was tallied as 0.89 (263 plots), and each coastal plain/fringe plot was tallied as 0.59 (999 plots). Measures of frequency were included to provide measure of occurrence independent of county area or plot size.

Preliminary data screening revealed that several calculated species attributes were mathematically redundant. For example, the count of US counties where each species occurs was

very highly correlated ($r=0.99$) with range area, indicating that county size was very consistent within the truncated range, such that adding each county added a nearly equivalent area. Four redundant attributes (range of abundance, area ratio, US county count, weighted CVS plot count) were removed from the 13 original calculated species attributes, leaving nine attributes used in the final analyses. The list of calculated species attributes is shown in **Table 2.1**, and correlations among all variables is shown in **Table 2.2**.

Table 2.1. *Calculated species attributes and data sources.*

| | |
|----------------------------------|---|
| Mean abundance | CVS – average cover class across all plots where each species occurs |
| Maximum abundance | CVS – Maximum cover for each species |
| Variance in abundance | CVS – Variance in cover for each species |
| Range in abundance (removed) | CVS – Range of cover for each species |
| Range area | USDA federal occurrence list – sum of the areas of the counties where each species occurs, within truncated range of counties |
| Area ratio (removed) | USDA state and federal species lists – ratio of the area occupied by a species to the area available to occupy |
| US county count (removed) | USDA – count of counties where each species occurs |
| North-South extent | USDA – distance between N-S extremes of range |
| East-West extent | USDA – distance between E-W extremes of range |
| Plot count | CVS – count of plots where each species occurs |
| Weighted plot count (removed) | CVS – count of plots where each species occurs, weighted by plot representation in each region (MT, PD, CP) |
| Community count | CVS – count of unique CEGl codes where each species occurs |
| Habitat volume | CVS – 3d volume defined by plots in environmental space |

Table 2.2. *Correlations among all species. Attributes that were removed from the analysis are in grey, and the correlation values that caused those attributes to be removed are underlined.*

| | Mean abundance | Maximum abundance | Variance in abundance | Range in abundance (removed) | US county count (removed) | Range area | Area ratio (removed) | East- West extent | North- South extent | Plot count | Weighted plot count (removed) | Community count | Habitat volume |
|----------------------------------|-------------------|----------------------|--------------------------|------------------------------------|---------------------------------|---------------|-------------------------|-------------------------|---------------------------|---------------|-------------------------------------|--------------------|-------------------|
| Mean abundance | 1 | 0.7093 | 0.6306 | 0.4901 | -0.0475 | -0.0402 | -0.0431 | -0.0400 | -0.0406 | 0.1608 | 0.1610 | 0.1255 | 0.1328 |
| Maximum abundance | 0.7093 | 1 | 0.7444 | <u>0.9469</u> | 0.1557 | 0.1477 | 0.1466 | 0.0755 | 0.0724 | 0.4610 | 0.4580 | 0.5099 | 0.5270 |
| Variance in abundance | 0.6306 | 0.7444 | 1 | 0.7356 | 0.0217 | 0.0225 | 0.0198 | 0.0208 | 0.0164 | 0.2079 | 0.2071 | 0.2031 | 0.2250 |
| Range in abundance (removed) | 0.4901 | <u>0.9469</u> | 0.7356 | 1 | 0.1992 | 0.1867 | 0.1850 | 0.0990 | 0.0969 | 0.4973 | 0.4939 | 0.5659 | 0.5817 |
| US county count (removed) | -0.0475 | 0.1557 | 0.0217 | 0.1992 | 1 | <u>0.9925</u> | <u>0.9849</u> | 0.7337 | 0.7507 | 0.3014 | 0.3097 | 0.3577 | 0.4013 |
| Range area | -0.0402 | 0.1477 | 0.0225 | 0.1867 | <u>0.9925</u> | 1 | <u>0.9914</u> | 0.7557 | 0.7706 | 0.2739 | 0.2816 | 0.3307 | 0.3710 |
| Area ratio (removed) | -0.0431 | 0.1466 | 0.0198 | 0.1850 | <u>0.9849</u> | <u>0.9914</u> | 1 | 0.7647 | 0.7789 | 0.2729 | 0.2805 | 0.3320 | 0.3737 |
| East-West extent | -0.0400 | 0.0755 | 0.0208 | 0.0990 | 0.7337 | 0.7557 | 0.7647 | 1 | 0.8824 | 0.1208 | 0.1265 | 0.1721 | 0.2097 |
| North-South extent | -0.0406 | 0.0724 | 0.0164 | 0.0969 | 0.7507 | 0.7706 | 0.7789 | 0.8824 | 1 | 0.1439 | 0.1493 | 0.1984 | 0.2364 |
| Plot count | 0.1608 | 0.4610 | 0.2079 | 0.4973 | 0.3014 | 0.2739 | 0.2729 | 0.1208 | 0.1439 | 1 | <u>0.9967</u> | 0.9088 | 0.8257 |
| Weighted plot count (removed) | 0.1610 | 0.4580 | 0.2071 | 0.4939 | 0.3097 | 0.2816 | 0.2805 | 0.1265 | 0.1493 | <u>0.9967</u> | 1 | 0.9171 | 0.8245 |
| Community count | 0.1255 | 0.5099 | 0.2031 | 0.5659 | 0.3577 | 0.3307 | 0.3320 | 0.1721 | 0.1984 | 0.9088 | 0.9171 | 1 | 0.8927 |
| Habitat volume | 0.1328 | 0.5270 | 0.2250 | 0.5817 | 0.4013 | 0.3710 | 0.3737 | 0.2097 | 0.2364 | 0.8257 | 0.8245 | 0.8927 | 1 |

In addition to the data sources described above, NatureServe global conservation ranks (G-ranks) for plant species were used to identify protected species in the final results. Rounded G-ranks were used, such that a species of rank G2G3 would have a rounded rank G2, and a species of rank G2G4 would have a rounded rank of G3. Because varieties and subspecies were lumped to species level, varieties and subspecies, which would usually receive a T-rank, were assigned the equivalent G-rank. G-rank definitions are shown in **Table 2.3**.

Table 2.3. *Definitions of NatureServe Global Conservation Ranks (G-ranks).*

| | |
|-----------|--|
| G1 | Critically imperiled - At very high risk of extinction due to extreme rarity (often 5 or fewer populations), very steep declines, or other factors. |
| G2 | Imperiled - At high risk of extinction or elimination due to very restricted range, very few populations, steep declines, or other factors. |
| G3 | Vulnerable - At moderate risk of extinction or elimination due to a restricted range, relatively few populations, recent and widespread declines, or other factors. |
| G4 | Apparently secure - Uncommon but not rare; some cause for long-term concern due to declines or other factors. |
| G5 | Secure - Common; widespread and abundant. |

Analysis

Nonmetric multidimensional scaling (NMDS) (R package Labdsv version 1.4, Roberts 2010; R package Ecodist, Goslee and Urban 2007) and factor analysis (R version 2.4.1) were used to ordinate species using measures of local abundance, habitat preference, range area, and frequency, resulting in an ordination of species as points in attribute space. Mean abundance, variance in abundance, CVS plot count, and CEGl count were log transformed to normalize the distributions of those variables (McCune and Grace 2002). As the raw values for some attributes, such as range area, were orders of magnitude larger than others, such as average cover class values, data were relativized to a common maximum for each attribute.

To determine the dimensionality of the data, a step-down run of NMDS was used starting from six axes. Distance matrices were created using Euclidean distance. The starting configuration of samples in the ordination space was determined with a random number generator. Subsequent

focal runs were set at three dimensions. Analyses were repeated using principal coordinates analysis (PCO) (R package labdsv version 1.4, Roberts 2010), also using Euclidean distance. Analyses were repeated a third time using factor analysis. Scores were calculated using regression and the final solution was rotated using varimax rotation (R version 2.4.1). To determine the robustness of the results, the data were divided by physiographic regions based on plot location (mountains, piedmont, and coastal plain/fringe) and the analyses were repeated for each region.

Polythetic hierarchical agglomerative clustering analysis (R version 2.4.1 and R package Cluster, version 1.13.1, Maechler et al. 2005) was used to detect groups of species based on attributes. The number of groups created ranged from 2 to 16 to encompass the 8 groups proposed in most classification schemes in the literature. A polythetic method was chosen to allow consideration of multiple species at each grouping division, and a hierarchical method was chosen as the inherent grouping structure of the data was not known, and a hierarchical method would allow a more intuitive result for interpreting an appropriate level of division among species with different distributional patterns based on the calculated attributes. Euclidean distance was used for clustering analysis. Linkage methods included complete, Ward's, and flexible beta = -0.25. A Mantel group-contrast test was used to assess grouping level (Legendre and Legendre 1998).

Results and Discussion

Dimensionality

The initial step-down NMDS ordination indicated a 3-dimensional solution with final stress of 3.4293, considered good for ecological data (McCune and Grace 2002). The range of stress values from six to two dimensions is shown in the scree plot in **Figure 2.2**. Stress decreases abruptly from two to three dimensions, followed by a gradual decrease from three to six dimensions. A plot of eigenvalues from the PCO ordination also indicates a 3-dimensional solution, shown in **Figure 2.3**.

The first three components account for most of the variability, with very little variability accounted for by the remaining components. Shepard diagrams for both analyses display the correlation between ordination distance and computed (Euclidean) distance for each data point (**Figure 2.4**). Correlations for both analyses are high, indicating that either solution is a reliable representation of these data.

Figure 2.2. *Scree plot of stress values from step-down nonmetric multidimensional scaling ordination, indicating an abrupt reduction in stress with the addition of a third axis, and very little reduction in stress with additional axes.*

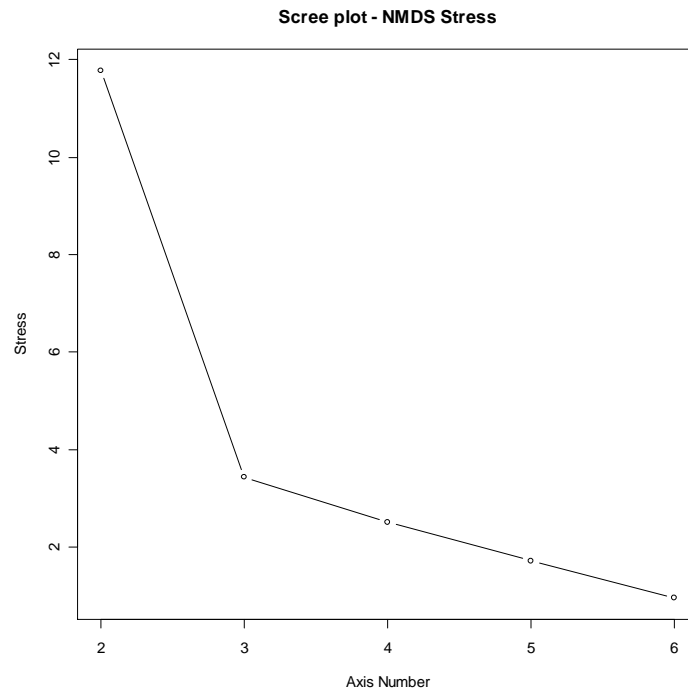


Figure 2.3. Barplot of eigenvalues from principal coordinates ordination, indicating that the first three components account for most variability in these data.

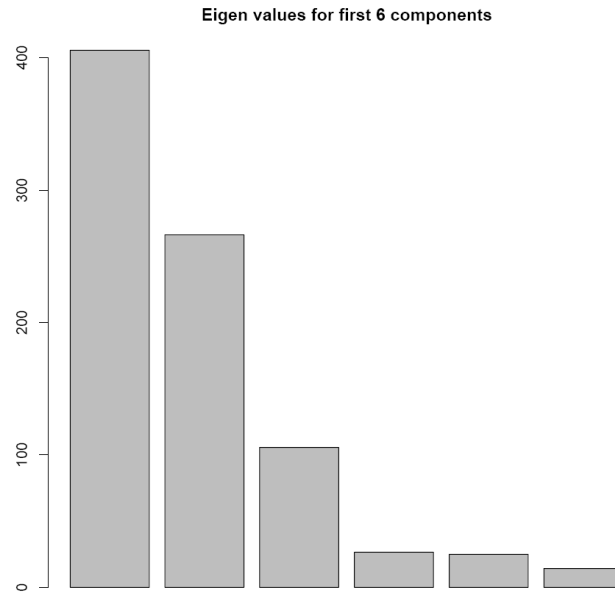
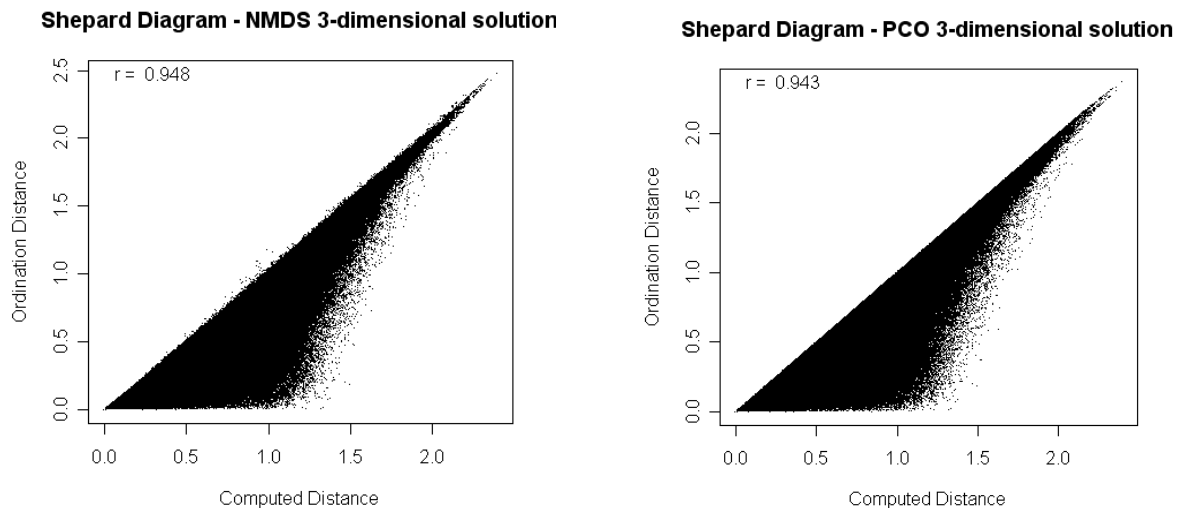


Figure 2.4. Shepard diagrams of NMDS and PCO 3-dimensional ordination results. For each analysis, the correlation between the ordination distances and the computed (Euclidean) distances is high ($r = 0.948$ for NMDS solution and $r = 0.943$ for PCO solution), indicating that either solution is a reliable representation of these data.



The cumulative variance of the first three factors explained 83.2% of variance; adding the fourth factor increased variance explained by only 5.6% (**Figure 2.5**). All solutions from the

numerical scree test indicate that three factors are sufficient for these data (**Figure 2.6**) (Raiche and Magis 2010; following methods described in Raiche et al. 2006).

Figure 2.5. *Proportional and cumulative variance with each additional factor. The three-factor solution explained 83.2% of variance; including the fourth factor added only 5.6% variance, indicating that three factors are sufficient for these data.*

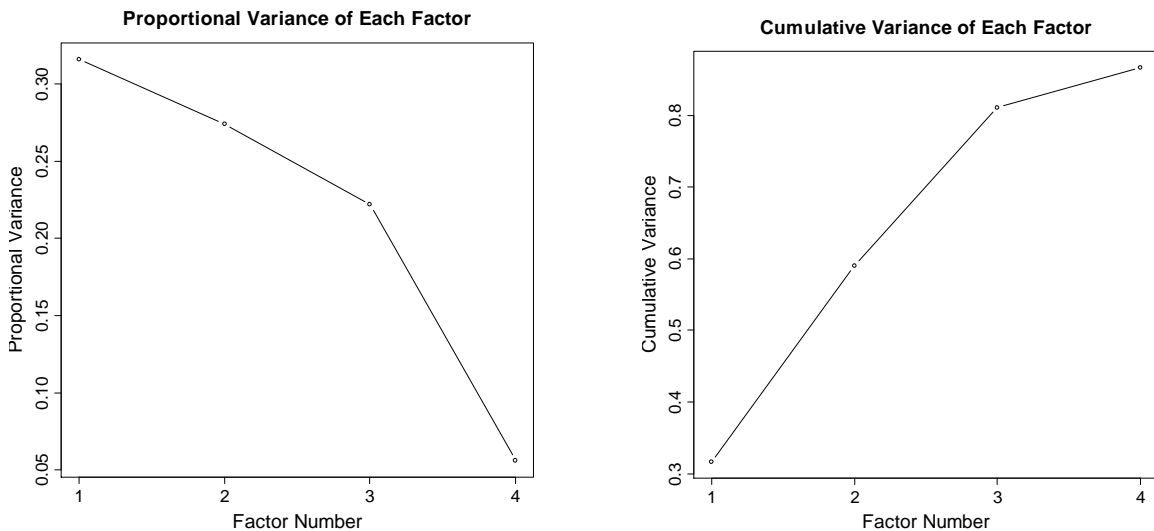
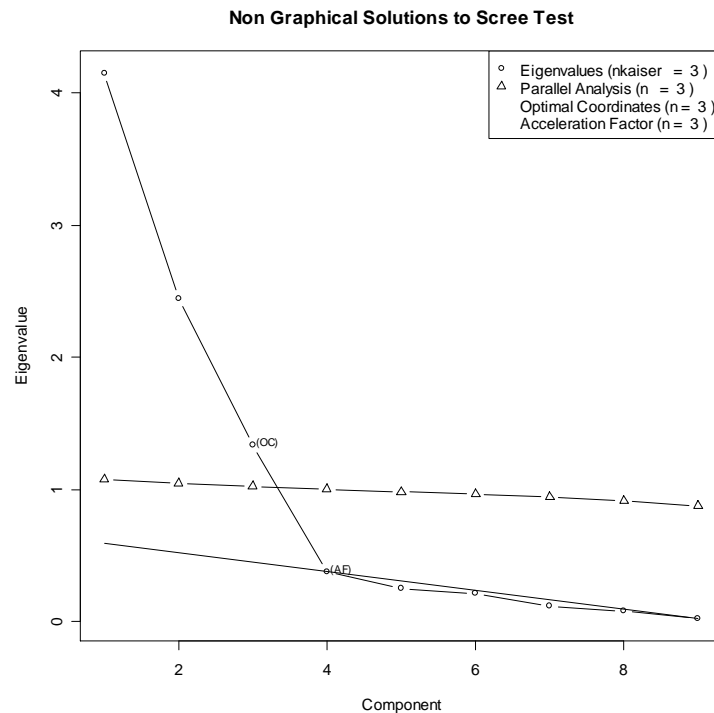


Figure 2.6. *Plot of numerical solutions to scree test. All solutions indicate that three factors should be retained for these data (Raiche and Magis 2010; following methods described in Raiche et al. 2006).*



As there is no widely accepted fixed indicator of the appropriate dimensionality of a given dataset (Kruskal and Wish 1978), and after consideration of the results and ease of interpretability of the preliminary analyses, I concluded that a three-dimensional solution was most appropriate to describe these data; therefore, the final analyses used 3 dimensions. For the remainder of results, including interpretation of dimensions and plotting of results, only results from factor analysis are reported.

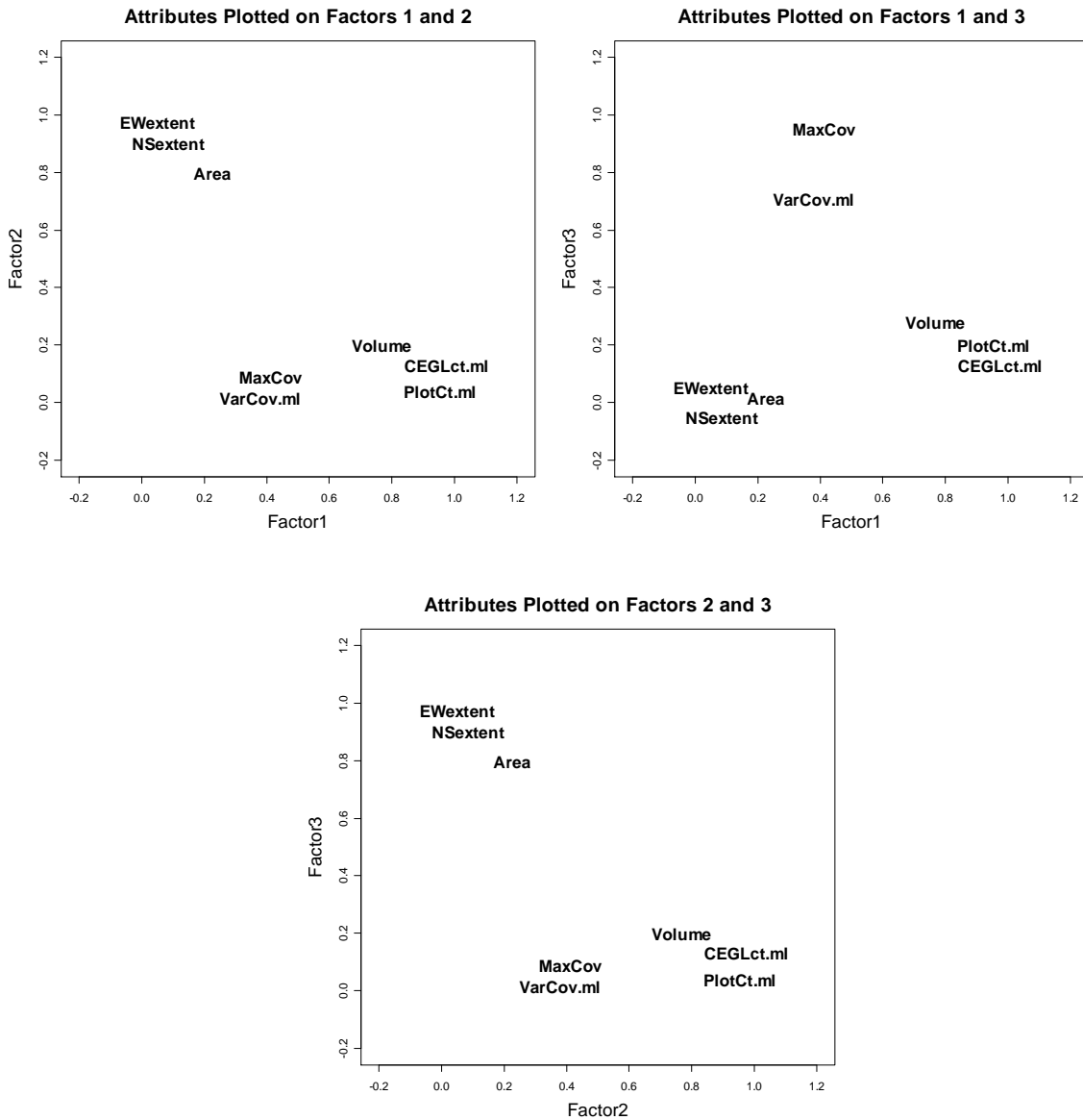
Analytic interpretation of dimensions

Table 2.4 shows the factor loadings of each attribute on each factor for the three-factor solution. All loadings are positive; all measures of habitat load strongly on factor 1, all measures of range load strongly on factor 2, and all measures of local abundance load strongly on factor 3. The attribute codes shown in Table 2.3 are displayed in the plots of attributes in factor space, **Figure 2.7**.

Table 2.4. *Factor loadings for each attribute for three-factor solution. Factor 1 is interpreted as increasing with attributes representing habitat volume, factor 2 is interpreted as increasing with attributes representing geographic range, and factor 3 is interpreted as increasing with attributes representing local abundance. The code "ml" indicates variables that were log transformed prior to analysis.*

| Attributes | Attribute codes | Factor 1 (Habitat) | Factor 2 (Range) | Factor 3 (Abundance) |
|-----------------------|-----------------|-----------------------|---------------------|-------------------------|
| Plot count | PlotCt.ml | 0.953 | | 0.203 |
| Community count | CEGLct.ml | 0.973 | 0.133 | 0.173 |
| Habitat volume | Volume | 0.766 | 0.201 | 0.239 |
| Range area | Area | 0.226 | 0.798 | |
| East-West extent | EWextent | | 0.932 | |
| North-South extent | NSextent | | 0.942 | |
| Mean abundance | Avg.Cov.ml | | | 0.834 |
| Maximum abundance | Max.Cov | 0.412 | | 0.907 |
| Variance in abundance | Var.Cov.ml | 0.378 | | 0.748 |

Figure 2.7. Attributes plotted along factors 1, 2, and 3. (Plotted attribute codes are defined as follows: *PlotCt.ml* = log of plot count, *CEGLct.ml* = log of community count, *Volume* = Habitat volume, *Area* = range area, *EWextent* = east-west extent, *NSextent* = north-south extent, *Avg.Cov.ml* = log of mean abundance, *Max.Cov* = maximum abundance, *Var.Cov.ml* = variance in abundance). Factor 1 is interpreted as increasing with attributes representing habitat volume, factor 2 is interpreted as increasing with attributes representing geographic range, and factor 3 is interpreted as increasing with attributes representing local abundance.



Biological interpretation of dimensions

Figures 2.8 - 2.10 show species plotted along each of the three factors, with several representative species displayed in the plots. One of the most restricted species in this dataset, *Hudsonia montana*, appears in the lower left corner of each plot, indicating that it is indeed restricted on habitat, range, and local abundance. *H. montana* occurs on rock outcrops in the Southern Appalachian mountains, and has low local abundance at those sites. *Trillium discolor* also has a very restricted geographic range, but occurs in slightly more habitats than *H. montana*, and occurs at greater local abundance. In contrast, *Lindera melissifolia* occurs at much higher local abundance, though it is also restricted on range and habitat.

The most common species in this dataset appear primarily at the upper right corner of each plot. *Acer rubrum*, *Quercus rubra*, *Toxicodendron radicans*, and *Smilax rotundifolia* all have high geographic ranges and occur in a variety of habitats. They differ primarily in local abundance, with the tree species (*A. rubrum* and *Q. rubra*) somewhat higher than the vines (*T. radicans* and *S. rotundifolia*). The other vine species displayed, *Vitis rotundifolia*, has a more restricted range as well as lower abundance than the tree species shown. The two rhododendron species, *R. catawbiense* and *R. maximum*, occur at relatively high abundance where they occur, but have more restricted geographic ranges than the more common species, with *R. catawbiense* having a more restricted range than *R. maximum*.

Figure 2.8. Species plotted along factors 1 (habitat) and 2 (range). Species codes are defined as follows: ACERRUB = *Acer rubrum*, HUDSMON = *Hudsonia montana*, LINDMEL = *Lindera melissifolia*, QUERRUB = *Quercus rubra*, RHODCAT = *Rhododendron catawbiense*, RHODMAX = *Rhododendron maximum*, SMILROT = *Smilax rotundifolia*, TOXIRAD = *Toxicodendron radicans*, TRILDIS = *Trillium discolor*, VITIROT = *Vitis rotundifolia*. The linear pattern seen in the left side of the plot is due to the categorical nature of the variable of maximum abundance, measured as cover classes.

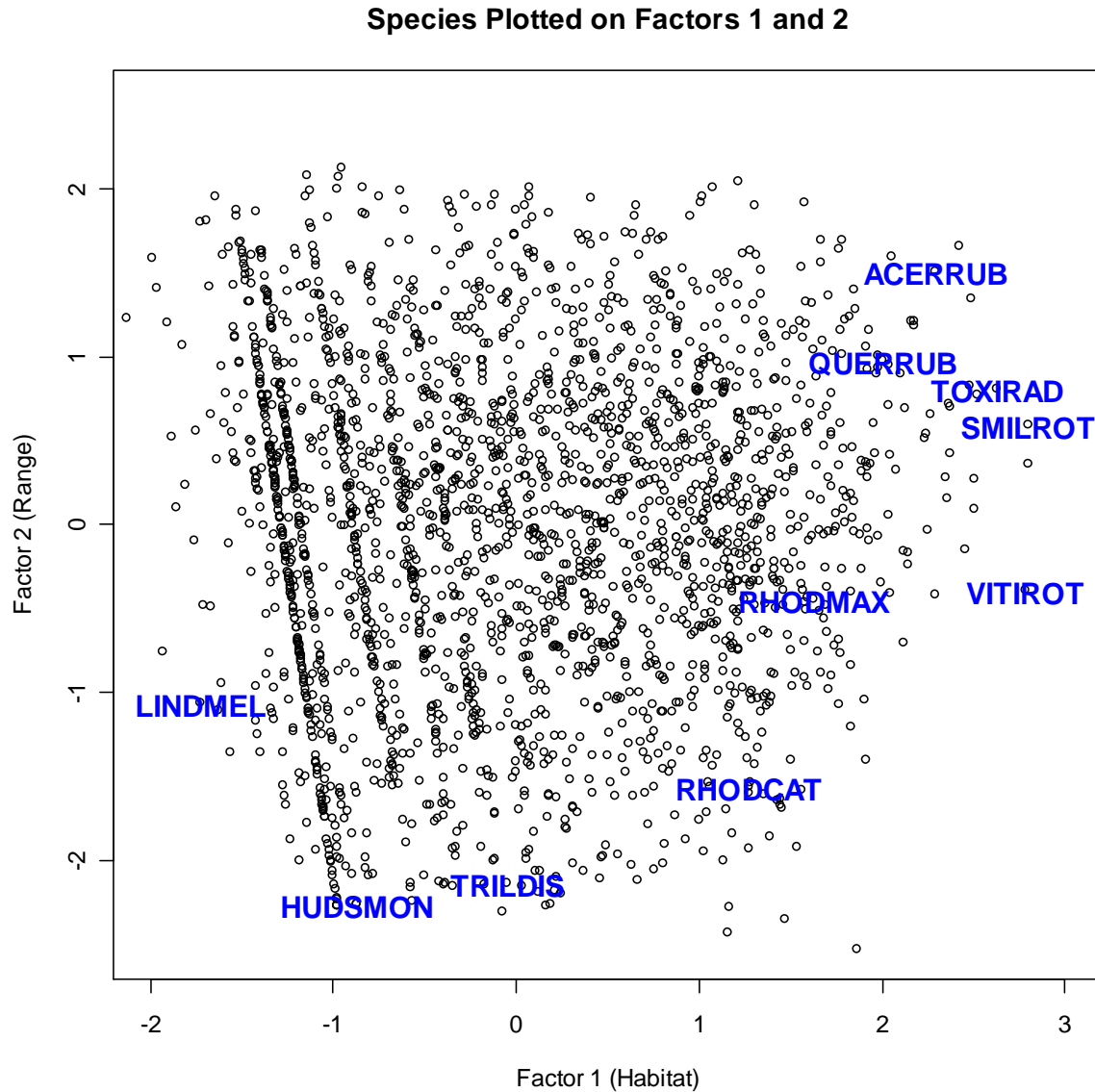


Figure 2.9. Species plotted along factors 1 (habitat) and 3 (abundance). Species codes are defined as follows: ACERRUB = *Acer rubrum*, HUDSMON = *Hudsonia montana*, LINDMEL = *Lindera melissifolia*, QUERRUB = *Quercus rubra*, RHODCAT = *Rhododendron catawbiense*, RHODMAX = *Rhododendron maximum*, SMILROT = *Smilax rotundifolia*, TOXIRAD = *Toxicodendron radicans*, TRILDIS = *Trillium discolor*, VITIROT = *Vitis rotundifolia*. The linear pattern seen in the plot is due to the categorical nature of the variable of maximum abundance, measured as cover classes.

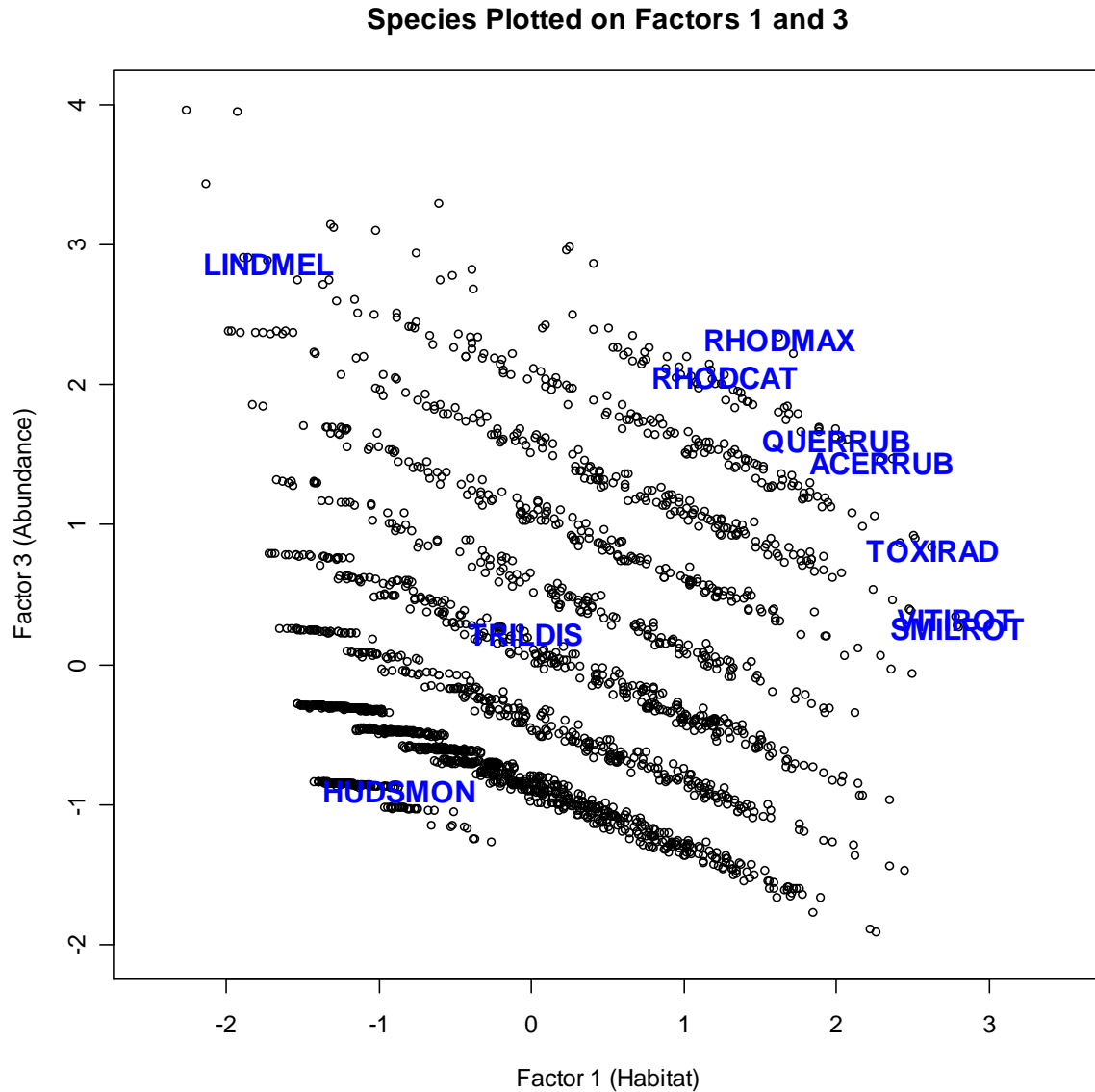
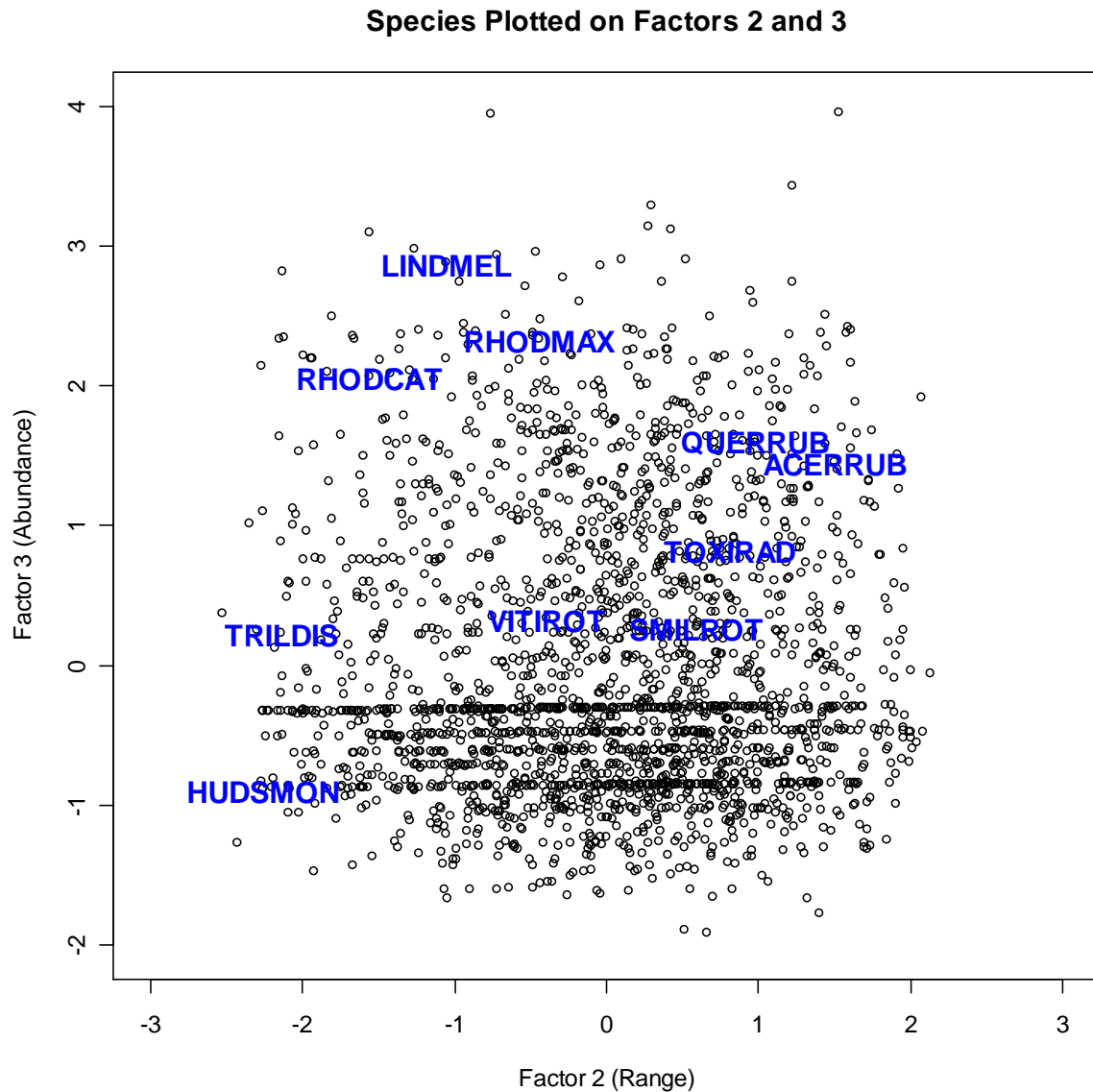


Figure 2.10. Species plotted along factors 2 (range) and 3 (abundance). Species codes are defined as follows: ACERRUB = *Acer rubrum*, HUDSMON = *Hudsonia montana*, LINDMEL = *Lindera melissifolia*, QUERRUB = *Quercus rubra*, RHODCAT = *Rhododendron catawbiense*, RHODMAX = *Rhododendron maximum*, SMILROT = *Smilax rotundifolia*, TOXIRAD = *Toxicodendron radicans*, TRILDIS = *Trillium discolor*, VITIROT = *Vitis rotundifolia*. The linear pattern seen in the plot is due to the categorical nature of the variable of maximum abundance, measured as cover classes.



Interpretation of groups

The results from the group-contrast Mantel tests are displayed in **Figures 2.11-2.13**. Each linkage method produced a different optimal number of groups - ten for complete linkage, five for Ward's linkage, and eight for flexible beta = -0.25. For complete linkage, the Mantel correlation values differ very little from five to ten groups, indicating that there is no clearly optimal group in that range. The optimal grouping level for Ward's and flexible beta are more clearly indicated by Mantel correlation values, however, the values maximize at different grouping levels (five and ten). Taken together, these results indicate that no grouping level is optimal to describe these data, and that grouping is arbitrary and useful primarily for purposes of interpretation. The remainder of the results will be displayed and discussed using eight groups produced by clustering. A grouping level of eight was chosen to coincide with a commonly used method to create groups of rarity (division of species lists at the mean or median values of each of (usually) three species attributes), which also results in eight groups (e.g., Kaye et al. 1997; Rey Benayas et al. 1999; Yu and Dobson 2000).

Figure 2.11. Group-contrast Mantel correlation values for each grouping level created using complete linkage method. The grouping level at which the maximum correlation value occurs (10 groups) is indicated by the vertical dotted line.

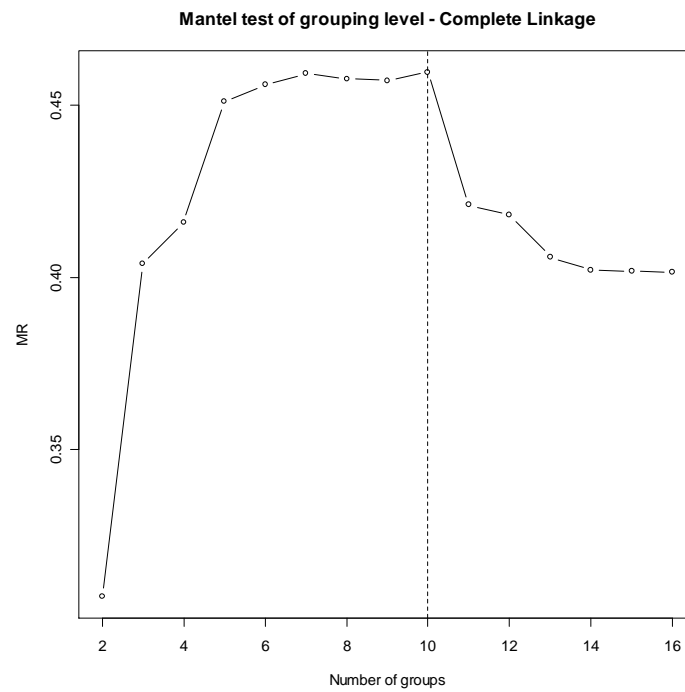


Figure 2.12. Group-contrast Mantel correlation values for each grouping level created using Ward's linkage method. The grouping level at which the maximum correlation value occurs (5 groups) is indicated by the vertical dotted line.

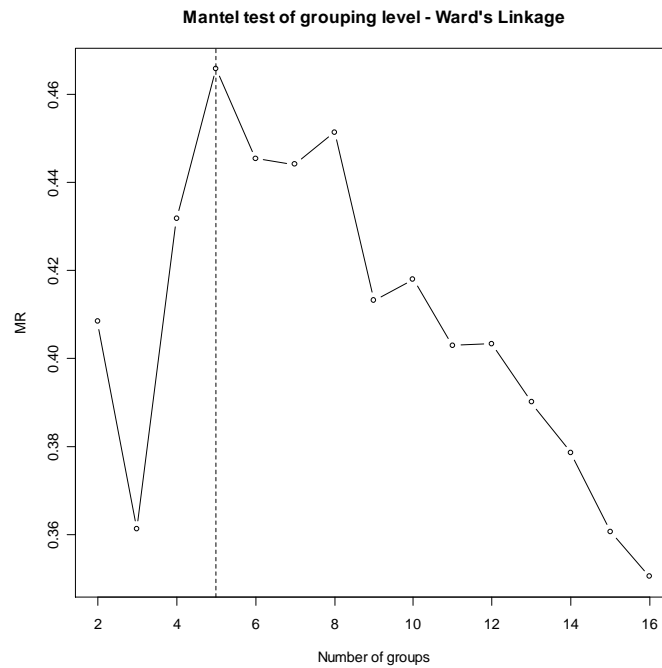
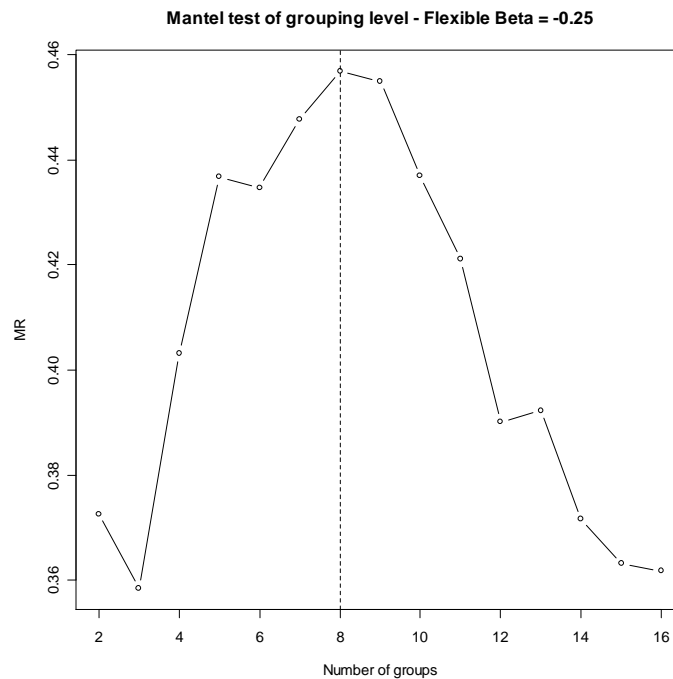


Figure 2.13. Group-contrast Mantel correlation values for each grouping level created using flexible beta linkage method. The grouping level at which the maximum correlation value occurs (8 groups) is indicated by the vertical dotted line.



Figures 2.14-2.16 display the factor analysis results for each pair of axes, with species as points in attribute space. The eight groups created using flexible beta = -0.25 linkage method are displayed in each plot. Based on each group's relative position in ordination space, the groups were interpreted as different forms and degrees of rarity. If a group was at the extreme low end of at least one species attribute, then that group was defined as being relatively rare. If a group was moderate to high on at least one attribute, then that group was define as being less rare. Because this method relies on three dimensions of rarity, the eight resulting categories are not completely ordinal from most to least rare. An interpretation of each group is offered in **Table 2.5**, and descriptive statistics of each group are offered in **Table 2.6**. Group membership is clearly distinguished along each axis, yet species are distributed continuously throughout the ordination space. The multivariate approach used in this study allows species to sort relative to one another

based on the attributes used in the analysis, revealing each species' relative location along each gradient. Groups were then superimposed on the resulting pattern. This approach not only indicates each species membership within a particular group, or form of rarity, but also the location of each species within the group, revealing whether the species is close to the boundary with another group.

Figure 2.14. Species plotted along factors 1 (habitat) and 2 (range). Groups were defined using hierarchical clustering with flexible beta = -0.25 linkage method. Group 1 is most rare, Group 8 is most common.

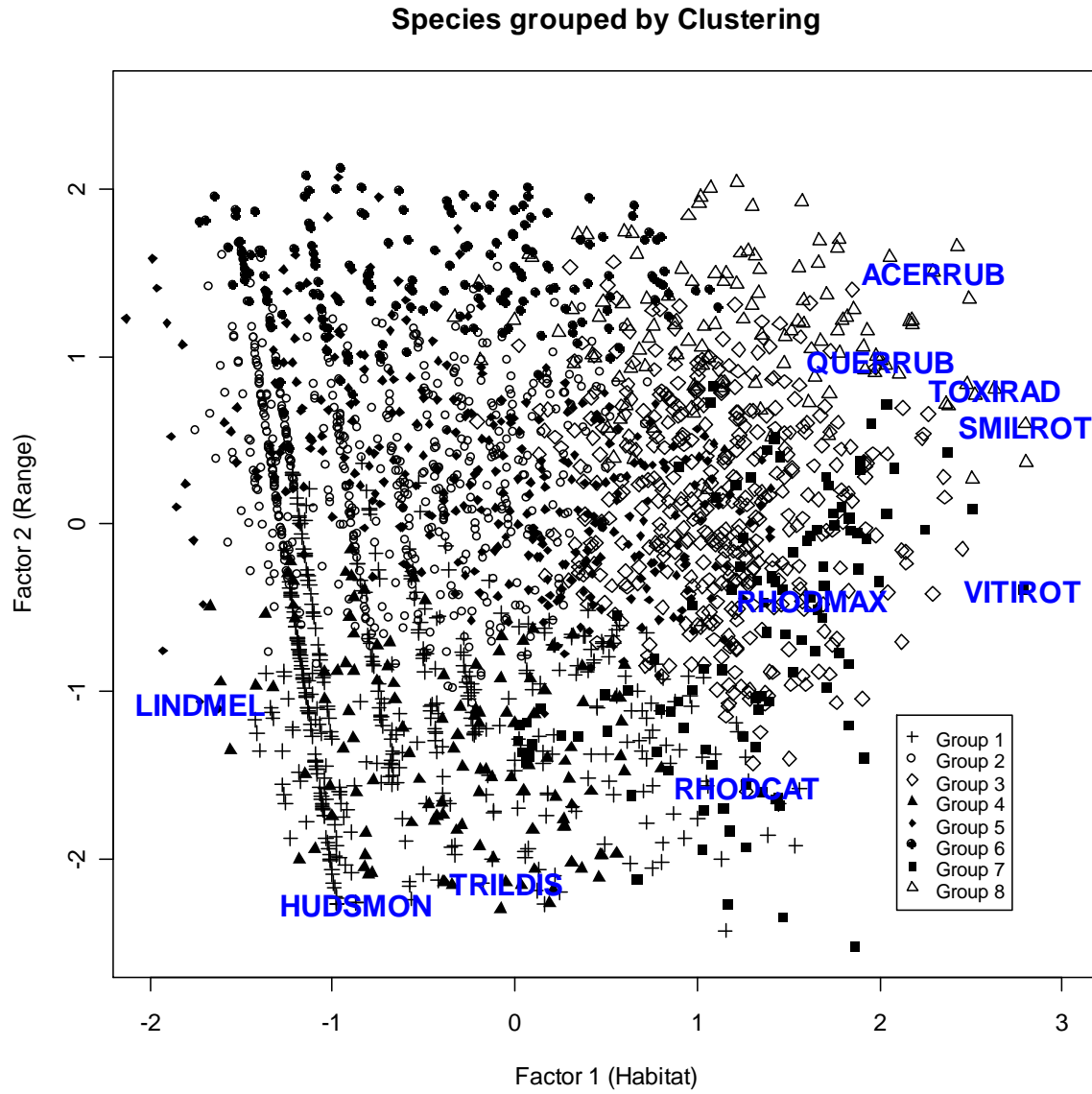


Figure 2.15. Species plotted along factors 1 (habitat) and 3 (abundance). Groups were defined using hierarchical clustering with flexible beta = -0.25 linkage method. Group 1 is most rare, Group 8 is most common.

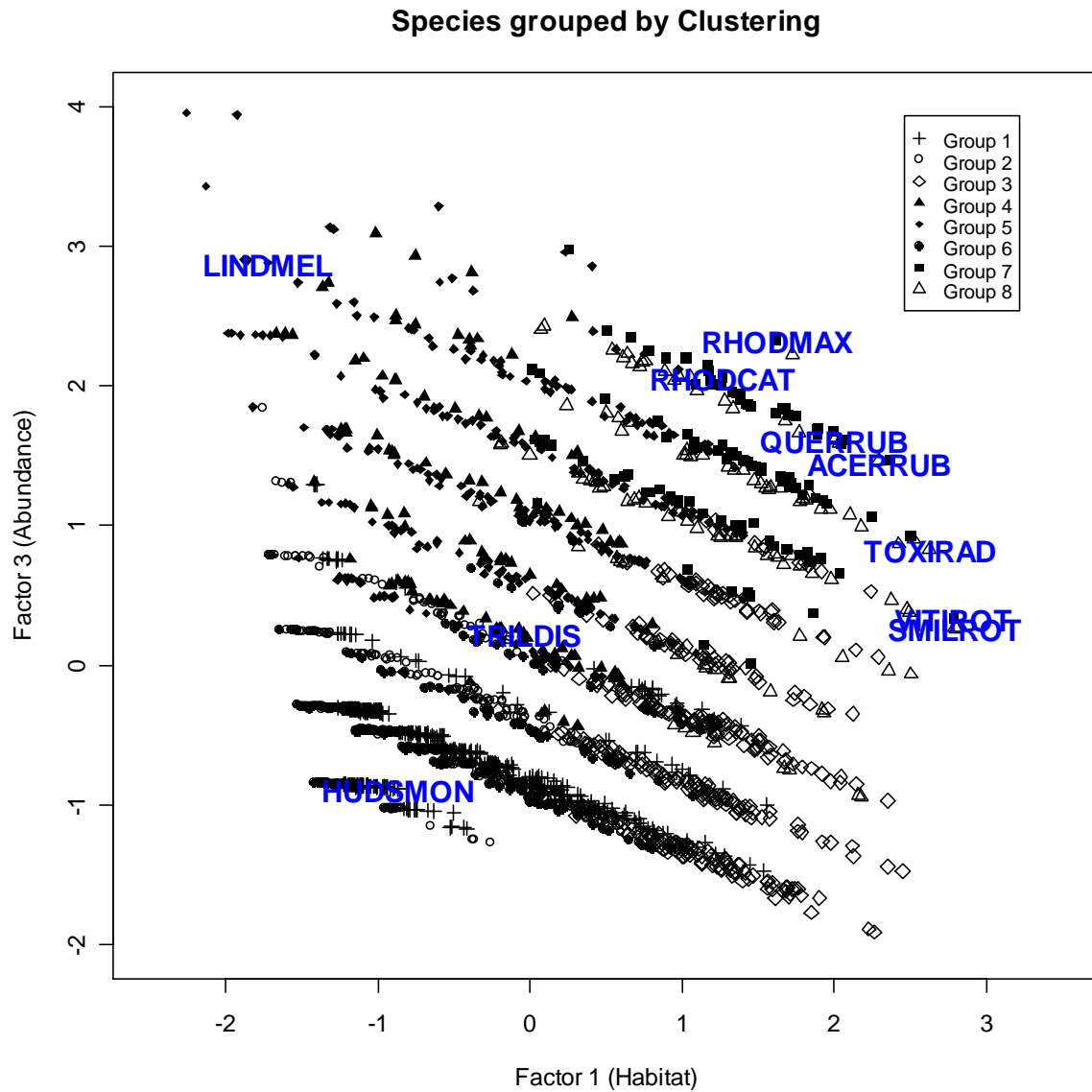


Figure 2.16. Species plotted along factors 2 (range) and 3 (abundance). Groups were defined using hierarchical clustering with flexible beta = -0.25 linkage method. Group 1 is most rare, Group 8 is most common.

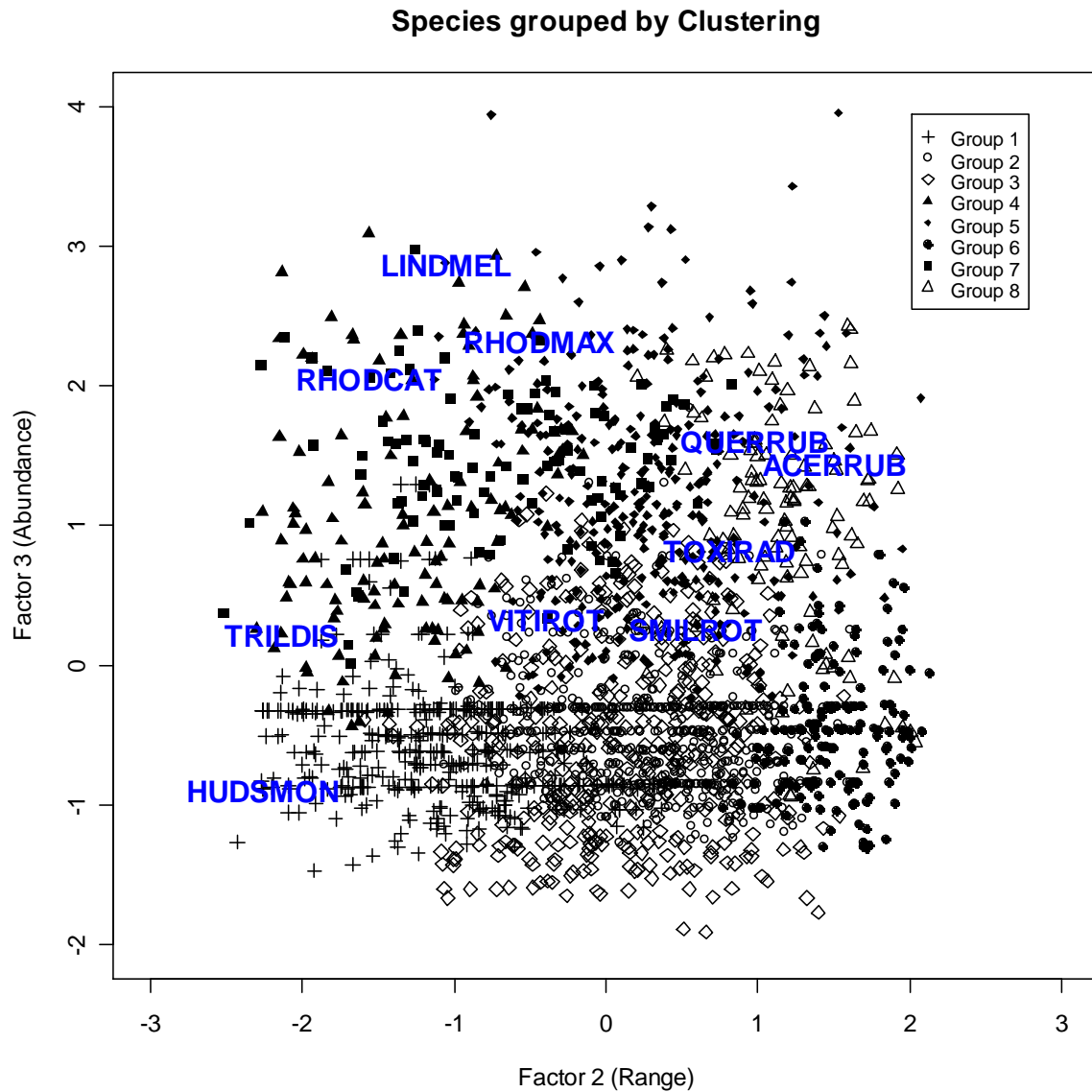


Table 2.5. Interpretation of rarity for each of eight groups created using hierarchical clustering with flexible beta = -0.25 linkage method.

| Group Number | Interpretation |
|--------------|---|
| 1 | Most rare; restricted on all attributes |
| 2 | Rare, but slightly less restricted on all attributes than group 1 |
| 3 | Moderate for most attributes, but some species can have high habitat volume |
| 4 | Restricted on habitat and range, but some species can have high abundance |
| 5 | Moderate for habitat and range, but some species can have high abundance |
| 6 | Restricted on habitat and abundance, but some species can have high range |
| 7 | High abundance and habitat, restricted on range |
| 8 | Least rare, little restricted on any attribute |

Table 2.6. Descriptive statistics for each of eight groups created using hierarchical clustering with flexible beta = -0.25 linkage method.

| Group Number | Average Cover | Maximum Cover | Range Area | Plot Count | CEGL Count | Habitat Volume |
|--------------|---------------|---------------|------------|------------|------------|----------------|
| 1 | 1.789796 | 2.052133 | 42669 | 9.338863 | 4.265403 | 0.1987146 |
| 2 | 1.839357 | 2.253188 | 208707 | 5.316940 | 3.697632 | 0.2241879 |
| 3 | 1.787973 | 3.652913 | 270694 | 90.184466 | 28.400485 | 2.3050511 |
| 4 | 2.964594 | 5.496000 | 35791 | 12.352000 | 5.888000 | 0.4453958 |
| 5 | 2.992047 | 6.168675 | 233797 | 32.885542 | 12.018072 | 1.0967697 |
| 6 | 1.724568 | 2.623529 | 660078 | 12.511765 | 7.211765 | 0.8019341 |
| 7 | 3.238617 | 7.869159 | 162512 | 222.803738 | 49.186916 | 3.8216631 |
| 8 | 2.664852 | 7.172414 | 667690 | 252.827586 | 59.163793 | 4.6121464 |

The groups defined using clustering effectively capture protected species as defined by NatureServe G-ranks. Groups 1 and 2, the most rare groups, captured high percentage of G1-G3 species (**Table 2.7**). The two G1 species captured in the less rare groups occurred in Group 4, which is one of the more rare groups (see Table 2.5 above). Total numbers of species of each G-rank within each group are shown in **Table 2.8**. As shown in **Figures 2.17-2.19**, species with low G-ranks (lower ranking equates to higher threatened status) tend to be restricted on range and habitat volume, but may occur at high, low, or moderate abundance.

Table 2.7. Percentages of high G-rank species captured within the rarity groups.

| | Groups 1-2 (Most rare, 971 species) | | | Groups 3-10 (Less rare to common, 1262 species) | | |
|------------------------|--|-----|-----|--|-----|-----|
| | G1 | G2 | G3 | G1 | G2 | G3 |
| Number | 1 | 22 | 79 | 2 | 7 | 32 |
| Total in G-rank | 3 | 29 | 112 | 3 | 29 | 112 |
| Percentage | 33% | 76% | 71% | 67% | 24% | 29% |

Table 2.8. Total numbers of species of each G-rank within each rarity group.

| Group | G-rank | | | | | | Row totals |
|----------------------|---------------|----------|----------|----------|----------|----------------|-------------------|
| | 1 | 2 | 3 | 4 | 5 | No rank | |
| 1 | 1 | 20 | 75 | 178 | 124 | 24 | 422 |
| 2 | 0 | 2 | 4 | 88 | 392 | 63 | 549 |
| 3 | 0 | 0 | 1 | 61 | 344 | 6 | 412 |
| 4 | 2 | 4 | 27 | 61 | 29 | 2 | 125 |
| 5 | 0 | 1 | 2 | 51 | 247 | 31 | 332 |
| 6 | 0 | 0 | 0 | 2 | 131 | 37 | 170 |
| 7 | 0 | 2 | 2 | 24 | 78 | 1 | 107 |
| 8 | 0 | 0 | 1 | 1 | 113 | 1 | 116 |
| Column totals | 3 | 29 | 112 | 466 | 1458 | 165 | |

Figure 2.17. Factors 1 and 2, with species of each G-rank displayed. Species with G-ranks 1 and 2 are restricted on both range and habitat, and expand along both axes as G-rank increases.

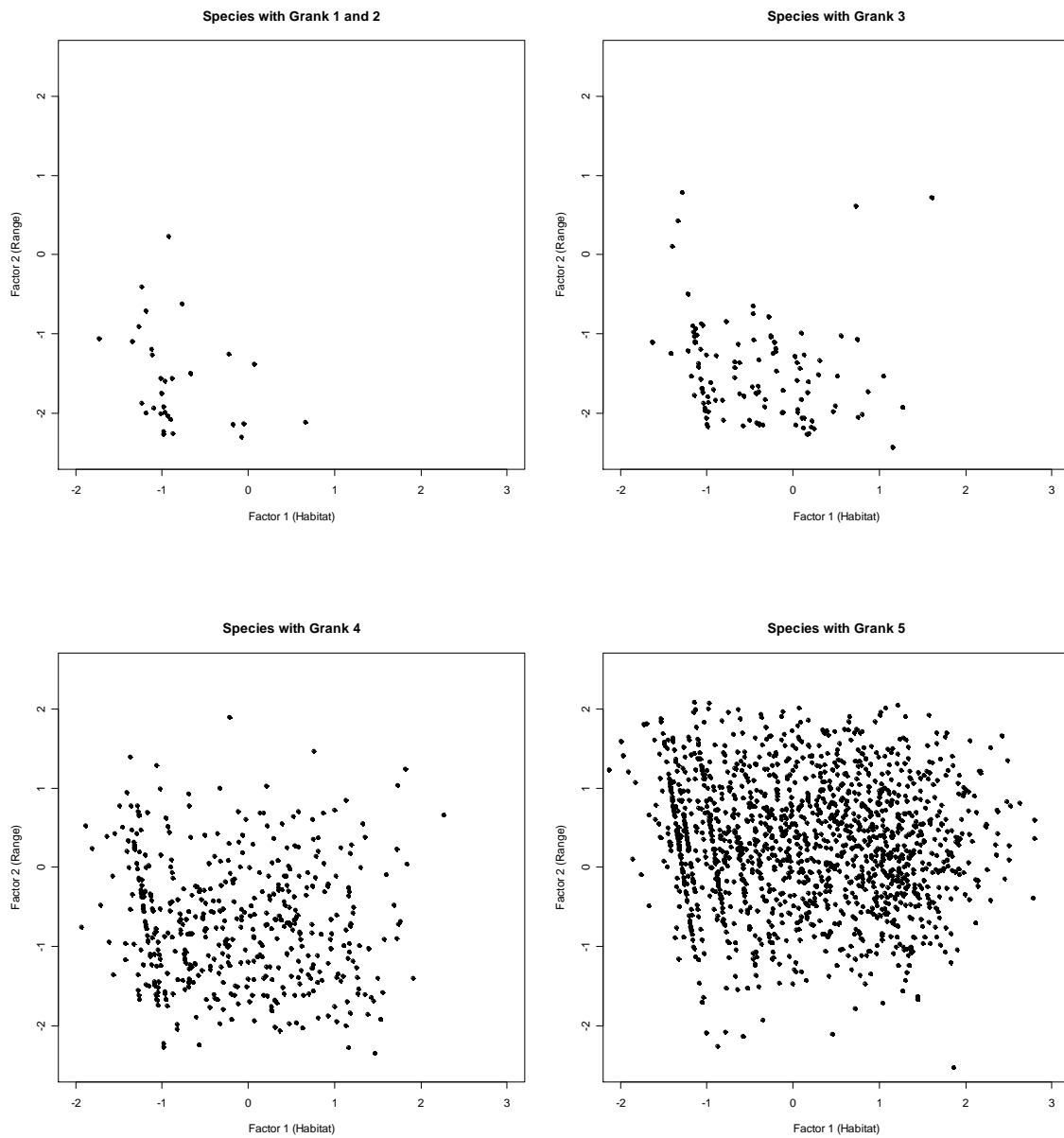


Figure 2.18. Factors 1 and 3, with species of each G-rank displayed. Species with G-ranks 1 and 2 are restricted on habitat, but vary in local abundance; habitat increases as G-rank increases.

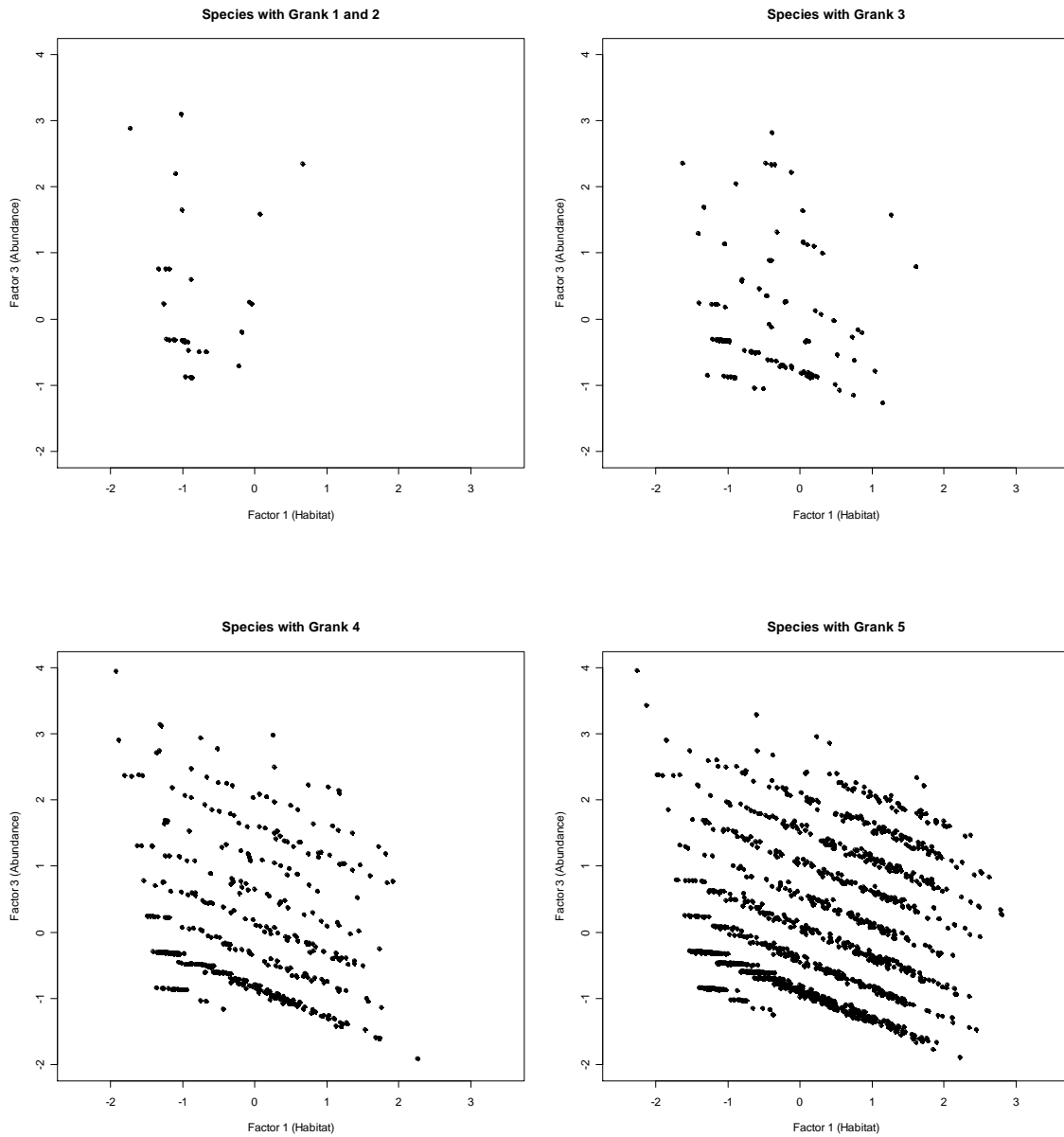
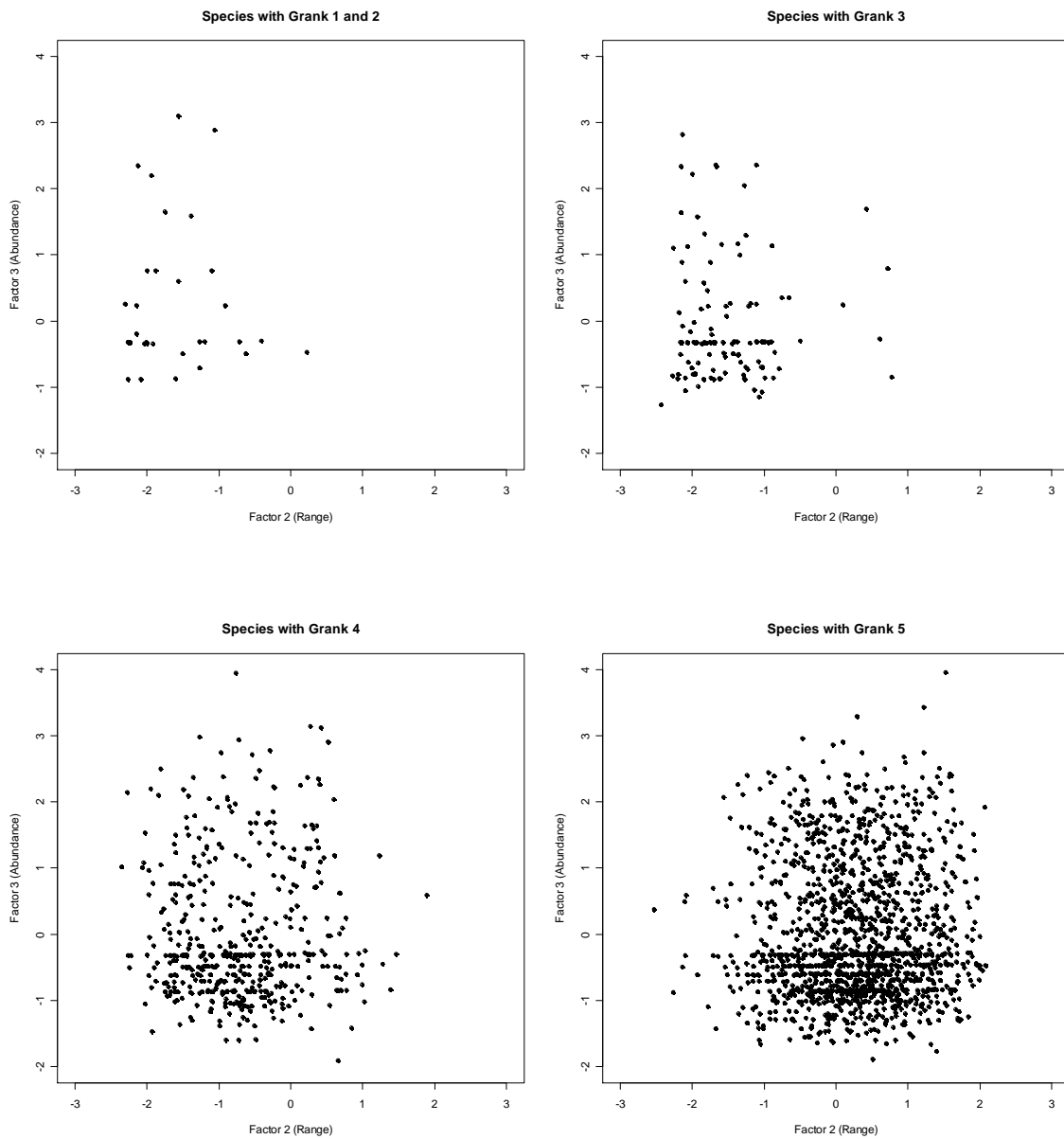


Figure 2.19. Factors 2 and 3, with species of each G-rank displayed. Species with G-ranks 1 and 2 vary in local abundance, but are restricted on range area, and increase in range as G-rank increases.



While Groups 1 and 2 do capture a high percentage of G1 - G3 species, Figures 2.17 - 2.19 show some overlap of species with different G-ranks in the ordination space - some G1-G3 species occur outside of the most restricted portion of the space (the lower left corner), while some G4-G5 species do occur there. The current analysis considers only pattern; mechanisms that contribute to rarity, such as aspects of population biology, and threats that contribute to vulnerability of a population or species, are not considered here. However, those aspects are considered when assigning G-ranks. Therefore, species with G-ranks 1-3 may exhibit a less restricted distribution pattern based on this analysis, yet could be highly vulnerable based on G-rank determinations. Likewise a species with G-rank 4-5 may exhibit a restricted distribution in this dataset but not be considered vulnerable based on other attributes considered when assigning ranks. A more likely explanation for the latter condition is the bias of the way the datasets of the current study were assembled. The extent of the CVS data, from which measures of local abundance and habitat volume were calculated, is the Southeastern US. The dataset includes species occurrences that are at the edges of the ranges for those species. Typically those occurrences are infrequent in the dataset, represent few habitat types, and are of low abundance. However, those species likely occur at higher abundance in more habitat types outside the extent of CVS sampling, so the attributes derived from CVS sampling are not necessarily representative of the actual distributions of those species. Therefore, it is not surprising to see G4 and G5 species distributed throughout the ordination space.

Another common method to define rarity uses the rank abundance curve for species in a dataset, defining the species in the tail of the curve to be rare. The rank abundance curve for species in this dataset is shown in **Figure 2.20**. By contrast to using clustering to produce groups, using only one attribute (abundance) to define rarity, much resolution of species distributions is lost. For example, as shown in **Figure 2.21**, rare and common species are completely intermingled along the axes of range and habitat. By definition, species do clearly sort along the axis of abundance

(Figure 2.22), however, species as different as *Acer rubrum* (a widespread abundant species) and *Lindera melissifolia* (a locally abundance species with very restricted range and habitat, protected with a G-rank of 2) would be group together as common species. While grouping is may be useful for interpretation, the method of grouping must be carefully chosen, and a multivariate approach must be used so a sensible grouping method can be applied based on the biological interpretation of a given dataset.

Figure 2.20. Rank abundance curve for 2233 species used in the current analyses. Abundance was averaged across all occurrences in the dataset for each species. The boundary of the tail of the curve is shown with the dotted line.

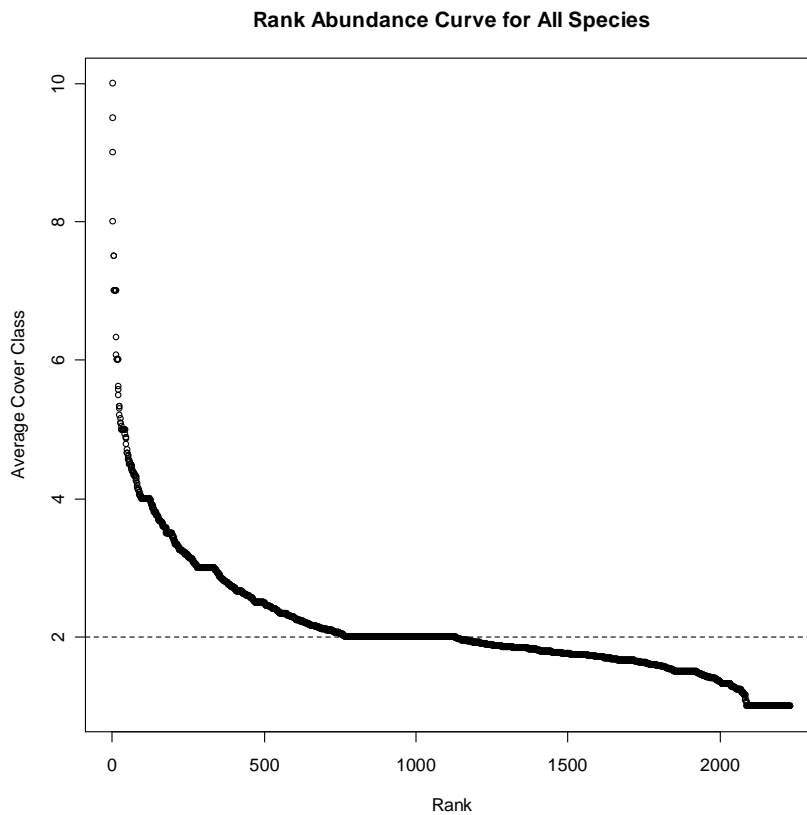


Figure 2.21. Species plotted along factors 1 (habitat) and 2 (range) and grouped into rare and common species defined by rank abundance. The dimensions of range and habitat were not used to define rarity in this case, so all information that would distinguish species on those attributes is lost, resulting in complete overlap of rare and common species. Species codes are defined as follows: ACERRUB = *Acer rubrum*, HUDSMON = *Hudsonia montana*, LINDMEL = *Lindera melissifolia*, QUERRUB = *Quercus rubra*, RHODCAT = *Rhododendron catawbiense*, RHODMAX = *Rhododendron maximum*, SMILROT = *Smilax rotundifolia*, TOXIRAD = *Toxicodendron radicans*, TRILDIS = *Trillium discolor*, VITIROT = *Vitis rotundifolia*.

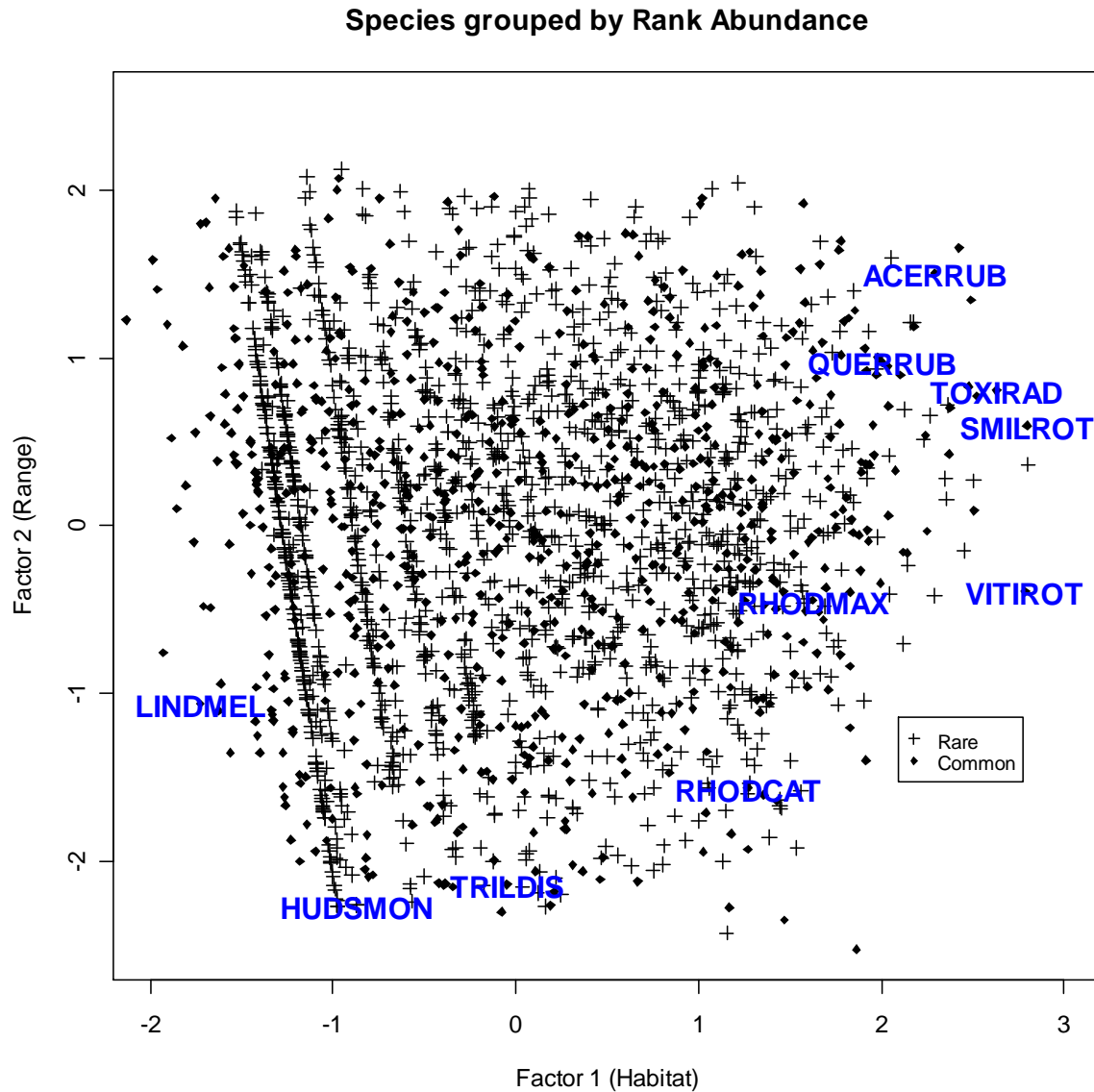
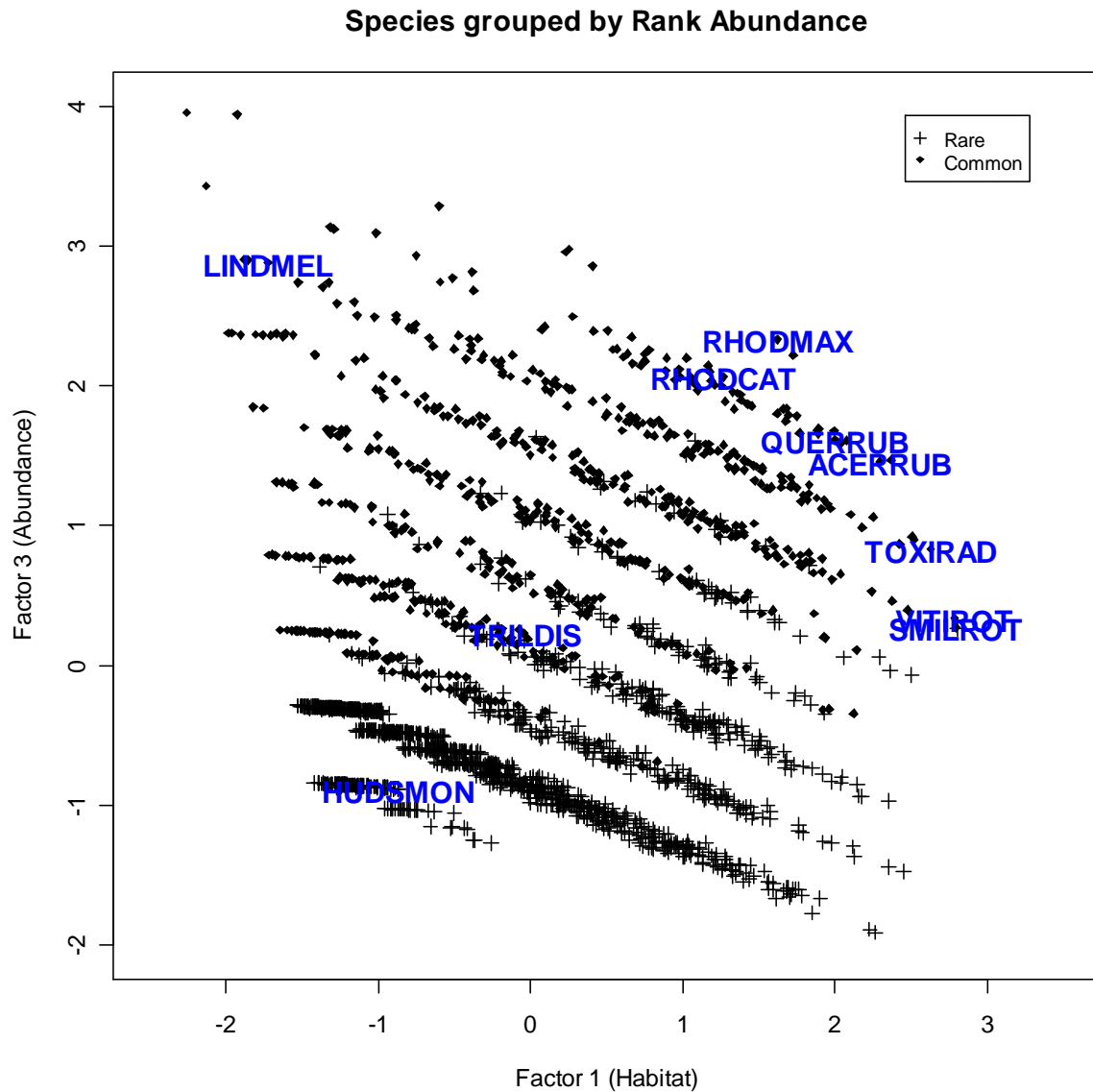


Figure 2.22. Species plotted along factors 1 (habitat) and 3 (abundance) and grouped into rare and common species defined by rank abundance. Species with distribution patterns as different as *Lindera melissifolia* and *Acer rubrum* are both defined as common species. Species codes are defined as follows: ACERRUB = *Acer rubrum*, HUDSMON = *Hudsonia montana*, LINDMEL = *Lindera melissifolia*, QUERRUB = *Quercus rubra*, RHODCAT = *Rhododendron catawbiense*, RHODMAX = *Rhododendron maximum*, SMILROT = *Smilax rotundifolia*, TOXIRAD = *Toxicodendron radicans*, TRILDIS = *Trillium discolor*, VITIROT = *Vitis rotundifolia*.



Regional results

The analyses were repeated for each of three physiographic regions represented within the dataset (mountains, piedmont, and coastal plain/fringe). For each region, the analysis converged on a

three-dimensional solution, and within each region, the axes had strongly similar correlation patterns (**Table 2.9**), indicating that the patterns detected for the full dataset are robust.

Table 2.9. Factor loadings for each attribute for three-factor solution within each region.

| Mountains | Factor 1 (Habitat) | Factor 2 (Range) | Factor 1 (Abundance) | Factor 4 (not used) |
|------------------------------|-------------------------------|-----------------------------|---------------------------------|--------------------------------|
| Plot count | 0.967 | | 0.169 | 0.122 |
| Community count | 0.962 | | 0.129 | 0.166 |
| Habitat volume | 0.862 | | 0.210 | |
| Range area | 0.169 | 0.833 | | |
| East-West extent | | 0.920 | | |
| North-South extent | | 0.958 | | |
| Mean abundance | | | 0.868 | |
| Maximum abundance | 0.498 | | 0.827 | 0.192 |
| Variance in abundance | 0.391 | | 0.631 | 0.665 |
| <i>Proportional variance</i> | <i>0.338</i> | <i>0.275</i> | <i>0.214</i> | <i>0.058</i> |
| <i>Cumulative variance</i> | <i>0.338</i> | <i>0.612</i> | <i>0.826</i> | <i>0.884</i> |

| Piedmont | Factor 1 (Habitat) | Factor 2 (Range) | Factor 1 (Abundance) | Factor 4 (not used) |
|------------------------------|-------------------------------|-----------------------------|---------------------------------|--------------------------------|
| Plot count | 0.925 | 0.151 | 0.240 | 0.161 |
| Community count | 0.959 | 0.141 | 0.168 | 0.167 |
| Habitat volume | 0.796 | 0.109 | 0.269 | |
| Range area | 0.207 | 0.843 | | |
| East-West extent | | 0.928 | | |
| North-South extent | | 0.942 | | |
| Mean abundance | 0.161 | | 0.835 | |
| Maximum abundance | 0.435 | | 0.863 | 0.247 |
| Variance in abundance | 0.401 | | 0.565 | 0.717 |
| <i>Proportional variance</i> | <i>0.315</i> | <i>0.280</i> | <i>0.214</i> | <i>0.071</i> |
| <i>Cumulative variance</i> | <i>0.315</i> | <i>0.595</i> | <i>0.808</i> | <i>0.879</i> |

| Coastal Plain/Fringe | Factor 1 (Habitat) | Factor 2 (Range) | Factor 1 (Abundance) | Factor 4 (not used) |
|------------------------------|-------------------------------|-----------------------------|---------------------------------|--------------------------------|
| Plot count | 0.958 | | 0.122 | |
| Community count | 0.980 | | 0.106 | |
| Habitat volume | 0.714 | 0.139 | 0.273 | |
| Range area | 0.109 | 0.804 | | |
| East-West extent | | 0.939 | | |
| North-South extent | | 0.938 | | |
| Mean abundance | | | 0.851 | |
| Maximum abundance | 0.405 | | 0.880 | 0.235 |
| Variance in abundance | 0.333 | | 0.590 | 0.730 |
| <i>Proportional variance</i> | <i>0.298</i> | <i>0.271</i> | <i>0.216</i> | <i>0.069</i> |
| <i>Cumulative variance</i> | <i>0.298</i> | <i>0.569</i> | <i>0.786</i> | <i>0.855</i> |

Conclusions

Repeated analyses of these data using various methods consistently produced stable three-dimensional solutions, indicating that rarity within the vascular flora of the southeastern US is three-dimensional. The three dimensions consistently loaded on factors representing habitat volume, geographic range, and local abundance, for the full dataset as well as each of three physiographic regions represented within the dataset. Measures of occupancy did not result in a fourth dimension for these data, as has been proposed by Rey Benayas et al. (1999) with their addition of occupancy to the three dimensions of range, abundance, and habitat proposed by Rabinowitz (1981). Likewise, one or two dimensions do not appear to be sufficient to describe distribution patterns for species in this dataset. Gaston (1994) cautions against seeking additional variables to define rarity to avoid generating “more, rather than less, confusion.” He emphasizes that rarity should be defined in terms of abundance or range size. However, the results of the current study indicate that three dimensions are needed to adequately describe species’ distributions for these data, and that use of a single dimension alone (abundance, defining rarity as the tail of the rank abundance curve) results in species with dramatically different distributions being inappropriately combined in the same rarity group.

Species were distributed continuously in ordination space along each dimension. Cluster analysis revealed that there is no inherent pattern of grouping for these data. While it is generally accepted that most variables that measure rarity are continuous, with the possible exception of counts of individuals (Gaston 1994), most definitions of rarity are dichotomous, even when the initial variables used in the definition are continuous (e.g., Rabinowitz 1981, Fiedler and Ahouse 1992, Gaston 1994, Kaye et al. 1997, Murray et al. 1999, Rey Benayas et al. 1999, Yu and Dobson 2000). One important reason this has been the case is the pragmatic need to categorize species as rare for legal or conservation purposes. I do not disagree with this need, however the results of the

current study suggest a more effective approach, which contributes two important improvements to a dichotomous approach. First, the multivariate approach used in this study reveals the relative locations of species in attribute space, allowing more complete interpretation of the patterns in the data, and only then superimposes groups that are biologically interpretable (and adjustable) based on the inherent patterns in the data. Defining categories *a priori* that may not capture the pattern for a given dataset may result in the inappropriate inclusion of species with very different distribution patterns in the same rarity group, as was the case for these data. Second, the use of multiple dimensions allows a more complete description of species' distribution patterns, allowing more biologically interpretable groups to be defined *a posteriori*, with greater discrimination between species with meaningfully different distributions. Rather than creating static points of division between a predetermined number of groups, the continuous nature of the data is explicitly recognized, allowing species to be parsed into any number of groups along each axis depending on the question of interest.

References

ESRI ArcInfo 9.2 1999-2006

Fiedler, P.L. and J.J. Ahouse. 1992. Hierarchies of cause: Toward an understanding of rarity in vascular plant species. In *Conservation Biology: The theory and practice of nature conservation, preservation, and management* (eds. P.L. Fiedler and S.K. Jain) Chapman and Hall, NY. 23 - 47.

Gaston, K.J. 1994. *Rarity* Chapman and Hall, NY. 205.

Gaston, K.J. 2003. *The Structure and Dynamics of Geographic Ranges*. Oxford, NY. 266.

Goslee, S.C. and Urban, D.L. 2007. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* 22(7):1-19.

Kartesz, J.T. and C.A. Meacham. 2002. *Syhtnesis of the Northa American Flora* CD-Rom Version 1. Chapel Hill, NC: North Carolina Cotanical Garden, University of North Carolina.

Kaye, T.N., R.J. Meinke, J. Kagan, S. Vrilakas, K.L. Chambers, P.F. Zika and J.K. Nelson. 1997. Patterns of rarity in the Oregon flora: Implications for conservation and management. In *Conservation and management of native plants and fungi* (eds. T.N. Kaye, A. Liston, R.M. Love, D.L. Luorna, R.J. Meinke and M.V. Wilson) Native Plant Society of Oregon, Corvallis, OR. 1-10.

Kruskal, J.B. and M. Wish. 1978. *Multidimensional Scaling*. Sage Publications, Beverly Hills, CA.

Legendre, P. and L. Legendre. 1998. *Numerical ecology*, 2nd English edition. Elsevier Science BV. Amsterdam, The Netherlands. 853.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. (2005). Cluster Analysis Basics and Extensions; R package version 1.13.1.

McCune, B., J.B. Grace, and D.L. Urban. 2002. *Analysis of ecological communities*. MjM Software Design, Gleneden Beach, Oregon, USA.

Peet, R.K., T.R. Wentworth, & P.S. White. 1998. A flexible, multipurpose method for recording vegetation composition and structure. *Castanea* 63:262-274.

R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Rabinowitz, D. 1981. Seven forms of rarity. In *The Biological Aspects of Rare Plant Conservation* (ed. H. Synge) J. Wiley, Chichester, England. 205 - 217.

Raiche, G. and Magis, D. 2010. nFactor: Parallel Analysis and Non Graphical Solutions to the Cattell Scree Test. R package version 2.3.2. Link: <http://cran.r-project.org/web/packages/nFactors/index.html>.

Raiche, G., Riopel, M. and Blais, J.-G. (2006). Non graphical solutions for the Cattell's scree test. Paper presented at the International Annual meeting of the Psychometric Society, Montreal. Link: http://www.er.uqam.ca/nobel/r17165/RECHERCHE/COMMUNICATIONS/2006/IMPS/IMPS_PRESENTATION_2006.pdf

- Rey Benayas, J.M., S.M. Scheiner, M.G. Sánchez-Colomer, and C. Levassor. 1999. Commonness and rarity: Theory and application of a new model to Mediterranean montane grasslands. *Conservation Ecology* 3:5.
- Roberts, D.W. 2010. Labdsv: Ordination and multivariate analysis for ecology. R package version 1.4-1. Link: <http://ecology.msu.montana.edu/labdsv/R>.
- Ruggiero, A., J.H. Lawton, and T.M. Blackburn. 1998. The geographic ranges of mammalian species in South America: Spatial pattern in environmental resistance and anisotropy. *Journal of Biogeography* 25:1093 - 1103.
- Ruggiero, A. 2001. Size and shape of the geographic ranges of Andean passerine birds. *Journal of Biogeography* 28:1281 - 1294.
- USDA Plants, <http://plants.usda.gov>
- Yu, J. and S. Dobson. 2000. Special paper: Seven forms of rarity in mammals. *Journal of Biogeography* 27:131-139.

CHAPTER 3

A simulation to predict the effects of habitat loss on rare plant species

Abstract

Effective conservation of rare species is dependent on a comprehensive understanding of species distributions in the landscape. To create a quantitative framework to define rarity, I used a multivariate analysis for plant species of the southeastern United States to determine how species distributions are defined by multiple measures of abundance, habitat, and range. I then extend the analysis to assess the impact of habitat loss on species distributions. To establish a null model of random habitat loss, I removed a random subset of plots from the dataset to simulate a 10% loss of habitat (repeated over 1000 iterations); this was repeated for 20%, 30%, and 40% plot loss. To test the effect on species of protecting plant community types designated as threatened by NatureServe G-rank definitions, I repeated the random plot removal over 1000 iterations, while protecting plots with threatened G-rank designations (G-rank ≤ 2 and G-rank ≤ 3). I then measured the distance of each species' movement through the multidimensional trait space from time 1 (the original dataset) to time 2 (after habitat loss). Using several approaches to divide species into rare and common groups, I compared the degree of impact of habitat loss on species with differing distribution patterns. When rarity was defined using only abundance (species in the tail of the rank abundance curve were defined as rare), rare species did move farther and had higher extinction rates than common species. When species were divided into eight forms of rarity, using either clustering or division at the median value of each attribute, in most cases the rare groups moved farther and had

higher extinction rates, however, not all rare groups experienced high impacts and some common groups experienced higher than expected impacts.

Introduction

The study of rarity often begins with the assignment of species to particular categories or forms of rarity. The rarity classification proposed by Rabinowitz (1981) (**Table 3.1**) is one of several frameworks that classify species distribution patterns based on measures of geographic range size, habitat specificity, and local abundance or population size (Gaston 1994, Fiedler & Ahouse 1992). Several studies have applied rarity classifications to regional datasets to determine how the species of a particular flora or fauna fall into the proposed categories of rarity (Kaye et al. 1997, Rey Benayas et al. 1999, Yu & Dobson 2000).

Table 3.1. Framework defining the forms of rarity using the criteria of geographic range size, habitat specificity, and local population size, or abundance. (Rabinowitz, 1981).

| Geographic Range | Large | | Small | |
|---|--|--|---|--|
| Habitat Specificity | Wide | Narrow | Wide | Narrow |
| Local population large, dominants | Common; locally abundant over a large range in several habitats | Predictable; locally abundant over a large range in a specific habitat | Unlikely; locally abundant in several habitats but restricted geographically | Endemic; Locally abundant in a specific habitat but restricted geographically |
| Local population small, non-dominant | Sparse; constantly sparse over a large range in several habitats | Constantly sparse in a specific habitat but over a large range | Non-existent? Constantly sparse and geographically restricted in several habitats | Endemic; Constantly sparse and geographically restricted in a specific habitat |

Studies to date have placed species into categories by dividing species lists according to high or low abundance, wide or narrow habitat preference, and wide or narrow geographic range. That approach assumes that the proposed categories actually exist in nature and forces species into specific categories. It also results in a discrete and static categorization, which creates two limitations on the usefulness of current classifications. First, current methods of classification do not reveal where within a rarity category each species occurs, and second, they cannot detect

movement within or shifts between categories. While it is widely accepted that species distribution patterns are in fact continuous and are not static (species can change from rare to common or common to rare), static and discrete rarity classifications have been a useful heuristic tool to understand and communicate about the phenomenon of rarity. The current study develops a method to reveal shifts through multivariate space within and between rarity categories in response to changing conditions such as habitat loss.

The categorization of species is in most cases only a means to an end; it is simply a foundation from which meaningful ecological questions can be extended. In the current study, I will test the effectiveness of a multivariate approach to detect *changes* in distribution patterns, in other words, to detect whether and how a species changes from one category of rarity to another. A method to detect a change in species distribution will have useful applications for conservation and management.

The current study addresses two specific questions regarding the impact of habitat loss on species distribution patterns. First, does a disturbance such as habitat loss cause species to shift between forms of rarity? By starting from a quantitative definition of rarity based on a specific flora, not only can species be divided into categories or forms of rarity, but the location of each species within each category can be detected, and therefore shifts within the multivariate rarity space can be quantified. Second, does habitat loss disproportionately affect rare species? By dividing the species into forms of rarity before habitat loss, the distance of shift can be compared between rare and less-rare species to determine which groups experienced greater shifts due to habitat loss.

Methods

Data preparation

The dataset consists of 2233 vascular plant species sampled from 2424 plots in the southeastern US, following methods of the Carolina Vegetation Survey (CVS) (Peet, Wentworth, and White 1998). Data preparation followed the methods described in Reilly (in preparation), with the exception that measures of range were excluded. Because habitat loss was simulated by eliminating plots from the dataset, the effect of loss on total geographic range could not be measured, so for these data, geographic range would not change, therefore measures of range were excluded from the analysis. The seven species attributes included in the final analyses are shown in **Table 3.2**.

Table 3.2. *Calculated species attributes and data sources.*

| | |
|------------------------------|--|
| Mean abundance | CVS – average cover class across all plots where each species occurs |
| Maximum abundance | CVS – Maximum cover for each species |
| Variance in abundance | CVS – Variance in cover for each species |
| Range in abundance (removed) | CVS – Range of cover for each species |
| Plot count | CVS – count of plots where each species occurs |
| Community count | CVS – count of unique CEGl codes where each species occurs |
| Habitat volume | CVS – 3d volume defined by plots in environmental space |

In addition to the data sources described above and in Reilly (in preparation) to test the null model of simulated habitat loss, NatureServe global conservation ranks (G-ranks) for plant associations were used to designate certain plots as protected in a conservation model of simulated habitat loss to measure the effect of varying levels of conservation status on the species that are protected within those communities.

Simulation of Habitat Loss

The full dataset consisting of 2233 vascular plant species was defined as “time 1”, or the dataset prior to habitat loss. To simulate a 10% loss of habitat, a random subset consisting of 90% of

the plots were selected and the species attributes were recalculated. These data were defined as “time 2”, or the dataset following habitat loss. Thus, a single iteration results in two sets of seven values (one values for each attribute) for each species, one set for time 1 and a set for time 2.

Mahalanobis distance was used to calculate the distance each species moved in this seven-dimensional space from time 1 to time 2. The formula for Mahalanobis distance is

$$D_{12}^2 = (x_1 - x_2)' V^{-1} (x_1 - x_2)$$

where x_1 is the matrix of seven attribute values for each species at time 1, x_2 is the matrix of seven attribute values for each species at time 2, and V^{-1} is the inverse of the covariance matrix of the time 1 species-by-attribute matrix (Manly 2005, Johnson and Wichern 1982).

The removal of plots and calculation of Mahalanobis distance was repeated 1000 times for each percentage of plots lost (10% through 40% loss), resulting in 1000 distances, or vector lengths, for each species for each percentage of loss. The mean vector length over 100 iterations was calculated for each species and was used in the final analyses to compare vector length between different rarity groups. In cases where species were lost from the dataset because all plot occurrences had been eliminated by random plot loss, the calculation of Mahalanobis distance resulted in missing values. In order to retain information about species lost from time 1 to time 2, the missing values were replaced with a value twice the maximum average vector length.

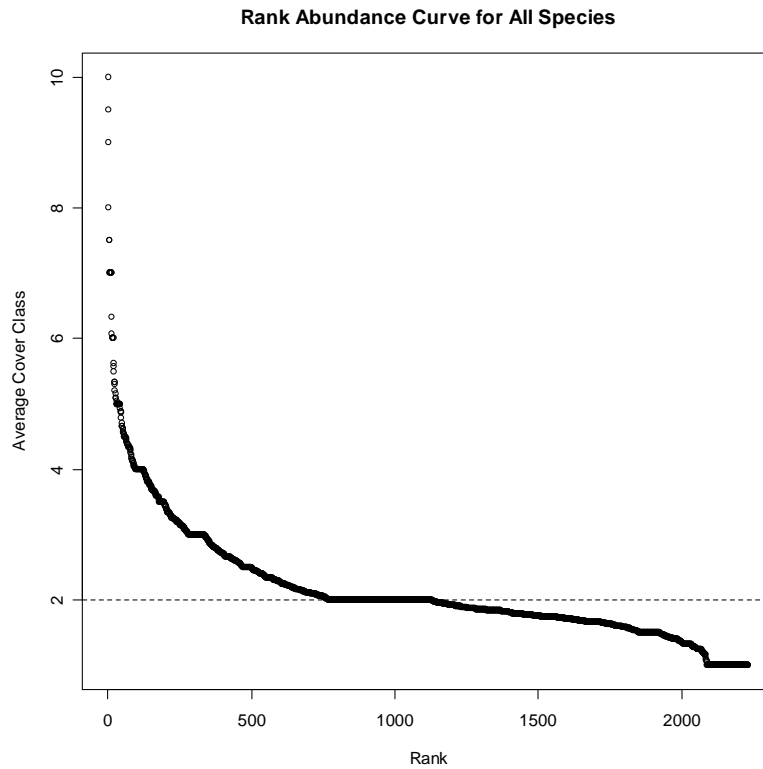
Analysis

Species were divided into rarity groups using three different methods - rank abundance, clustering, and division of species lists at the median value of each attribute. Previous analyses (Reilly in preparation) demonstrated that no number of groups captures the pattern of rarity in this dataset better than any other number of groups. For this reason grouping is arbitrary, and could be changed to any number of groups, or forms of rarity, that would be most appropriate for a given

dataset depending on the goals of the study and the investigator's knowledge of the flora or fauna of interest. Additionally, because the distributions of species along the attributes of range, abundance, and habitat preference are clearly continuous, as a convention no distinction will be made between "rare" and "common." Rather, the groups, or forms of rarity, will be referred to as more or less rare to reflect the continuous nature of species distributions. Two of the three methods chosen for the current analysis (rank abundance and median) represent common methods used to define rarity, and the third method used is hierarchical clustering (using Euclidean distance with flexible beta = -0.25 linkage method).

Rank abundance is one of the simplest methods to define rare species. In this case, species in the tail of the rank abundance curve, shown in **Figure 3.1**, were labeled as rare. This method uses only one dimension of rarity, so information about the habitat volume, geographic range, or frequency of species is ignored. Therefore, species at the high end of the rank abundance curve may have only a few occurrences in the dataset, or have a very restricted range, yet would be labeled as common because they have high abundance at the few location where they do occur.

Figure 3.1. Rank abundance curve for 2233 species used in the current analyses. Abundance was averaged across all occurrences in the dataset for each species. The boundary of the tail of the curve is shown with the dotted line.



Species were also labeled as rare by dividing the species list at the median value of each of three species attributes. The attributes most commonly used are measures of abundance, range, and habitat preference. In the current analysis, groups were defined using mean abundance, geographic range size (the sum of areas of US counties in the truncated range described in chapter 2), and count of unique NatureServe CEG codes were the attributed used. Dividing each of these three attributes into high and low values resulted in eight categories of rarity that are analogous to the categories defined by Rabinowitz (1981). Because this method relies on three dimensions of rarity, the eight resulting categories are not completely ordinal from most to least rare. However, they can be divided into four more rare (low on two or more attributed) and four less rare groups (high on two or more attributes). The groups are listed in **Table 3.3**.

Table 3.3. *Rarity groups as defined by division at the median of each of three attributes. High and low indicate above or below the median value. While the groups are not completely ordinal from most to least rare, Groups 1-4 can be considered more rare, with Group 1 the most rare (low on all attributes), while Groups 5-8 can be considered less rare, with Group 8 being common (high on all attributes).*

| Rarity Group | Geographic Range (sum of county areas) | Abundance (mean cover class) | Habitat (CEGL codes represented) |
|---------------------|--|--|--|
| 1 | Low | Low | Low |
| 2 | Low | Low | High |
| 3 | Low | High | Low |
| 4 | High | Low | Low |
| 5 | Low | High | High |
| 6 | High | Low | High |
| 7 | High | High | Low |
| 8 | High | High | High |

The final method used to define rarity was clustering. Polythetic hierarchical agglomerative clustering analysis (R version 2.4.1 and R package Cluster, version 1.13.1, Maechler et al. 2005) was used with the time 1 dataset to divide species into eight groups (using Euclidean distance with flexible beta = -0.25 linkage method). Factor analysis was used to ordinate species and plot species as points in attribute space. Scores were calculated using regression and the final solution was rotated using varimax rotation (R version 2.4.1). Based on each group's relative position in ordination space, the groups were interpreted as different forms of rarity. If a group was at the extreme low end of at least one species attribute, then that group was defined as being relatively rare. If a group was moderate to high on at least one attribute, then that group was define as being less rare. This resulted in Groups 1-4 being defined as more rare, and Groups 5-8 being defined as less rare. The species groups created by the three grouping methods are shown in **Figures 3.2-3.4**.

Figure 3.2. *Rarity groups as defined by Rank abundance. Species indicated by + symbols are those defined as rare (in the tails of the rank abundance curve, and also primarily at the lower end of each attribute axis). Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, QUERRUB = Quercus rubra, RHODCAT = Rhododendron catawbiense, RHODMAX = Rhododendron maximum, SMILROT = Smilax rotundifolia, TOXIRAD = Toxicodendron radicans, TRILDIS = Trillium discolor, VITIROT = Vitis rotundifolia.*

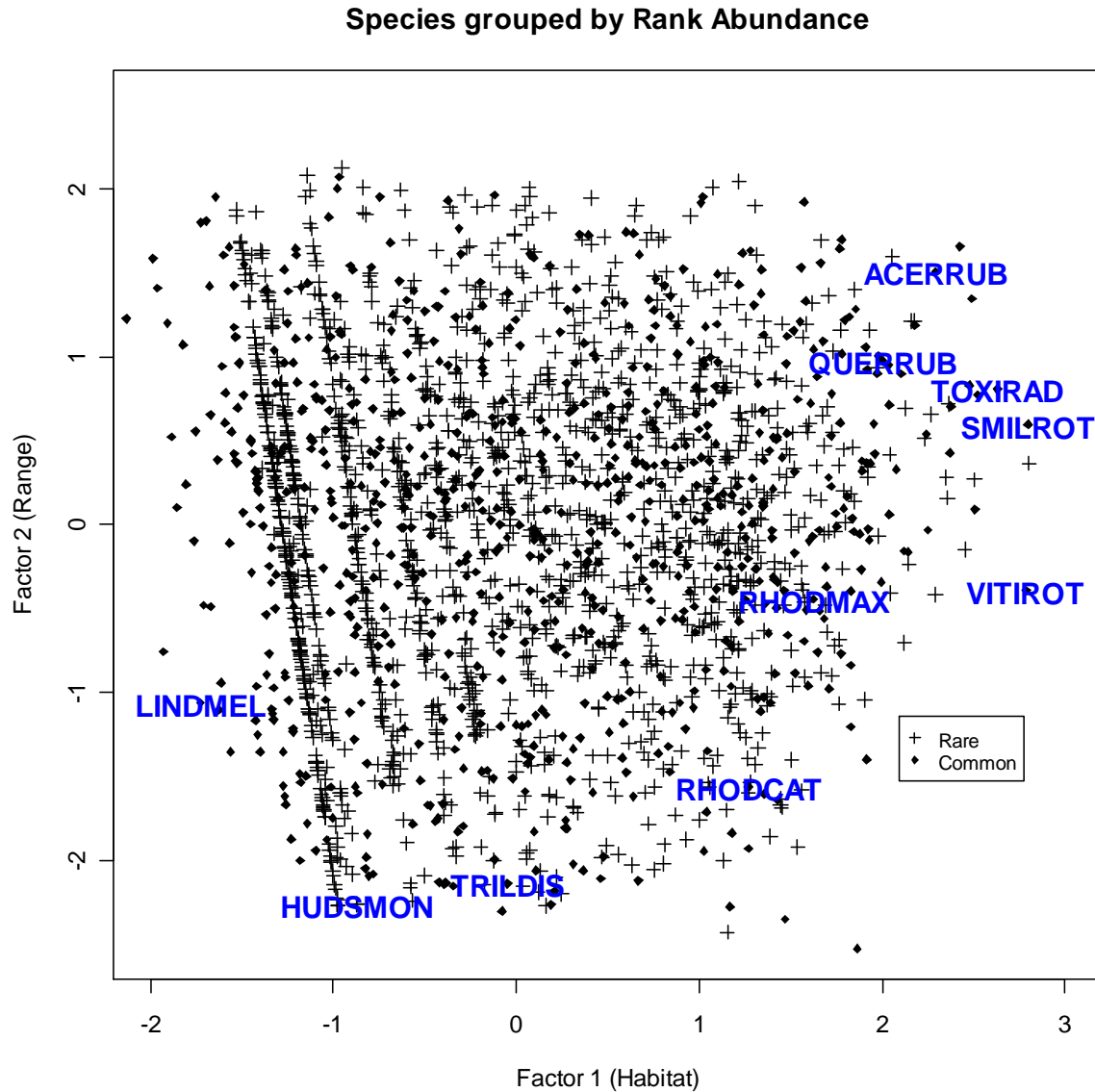


Figure 3.3. Rarity groups as defined by division at the median of each of three attributes. Species indicated by + symbols are those defined as more rare. Species codes are defined as follows: ACERRUB = *Acer rubrum*, HUDSMON = *Hudsonia montana*, LINDMEL = *Lindera melissifolia*, QUERRUB = *Quercus rubra*, RHODCAT = *Rhododendron catawbiense*, RHODMAX = *Rhododendron maximum*, SMILROT = *Smilax rotundifolia*, TOXIRAD = *Toxicodendron radicans*, TRILDIS = *Trillium discolor*, VITIROT = *Vitis rotundifolia*.

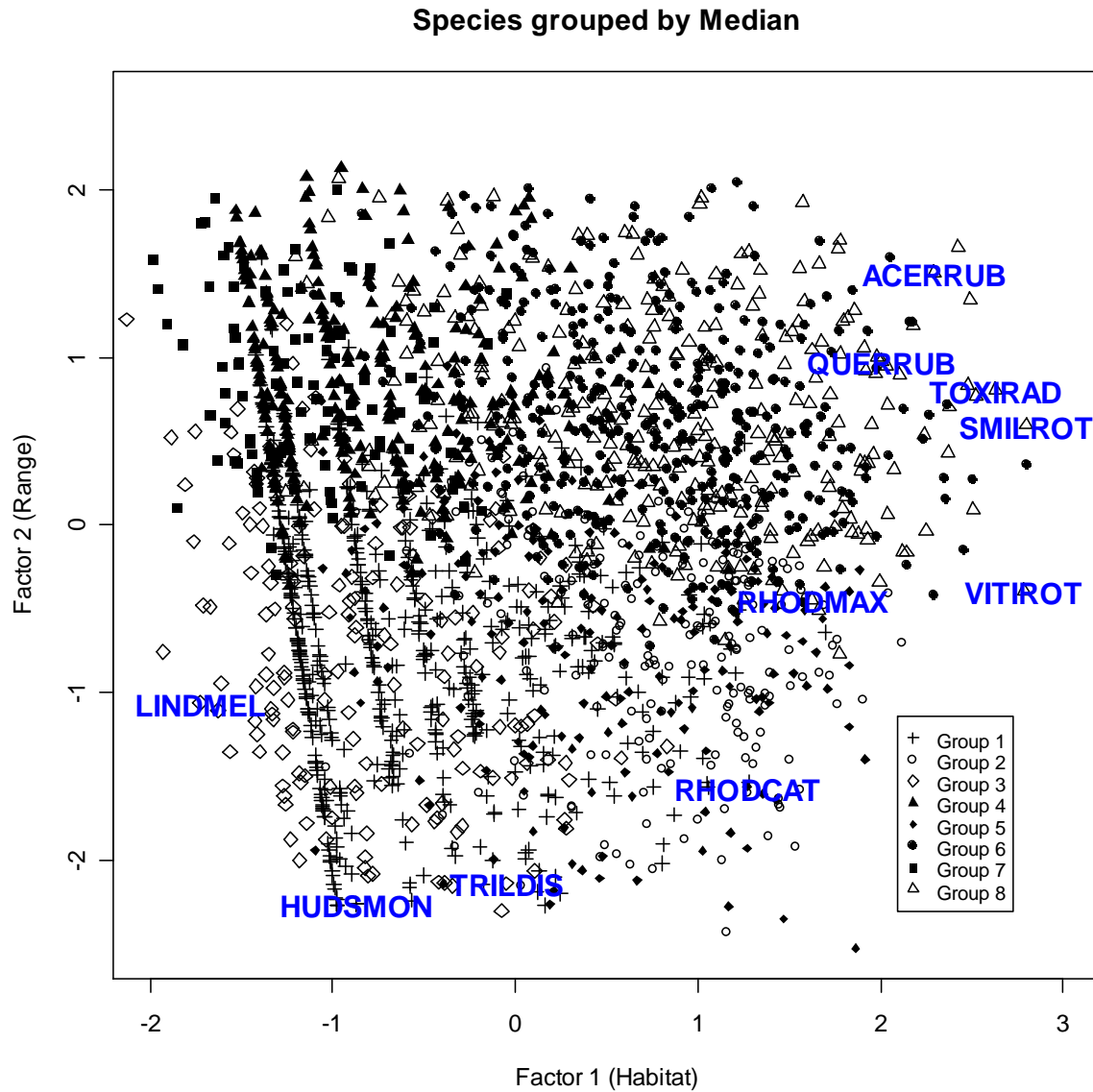
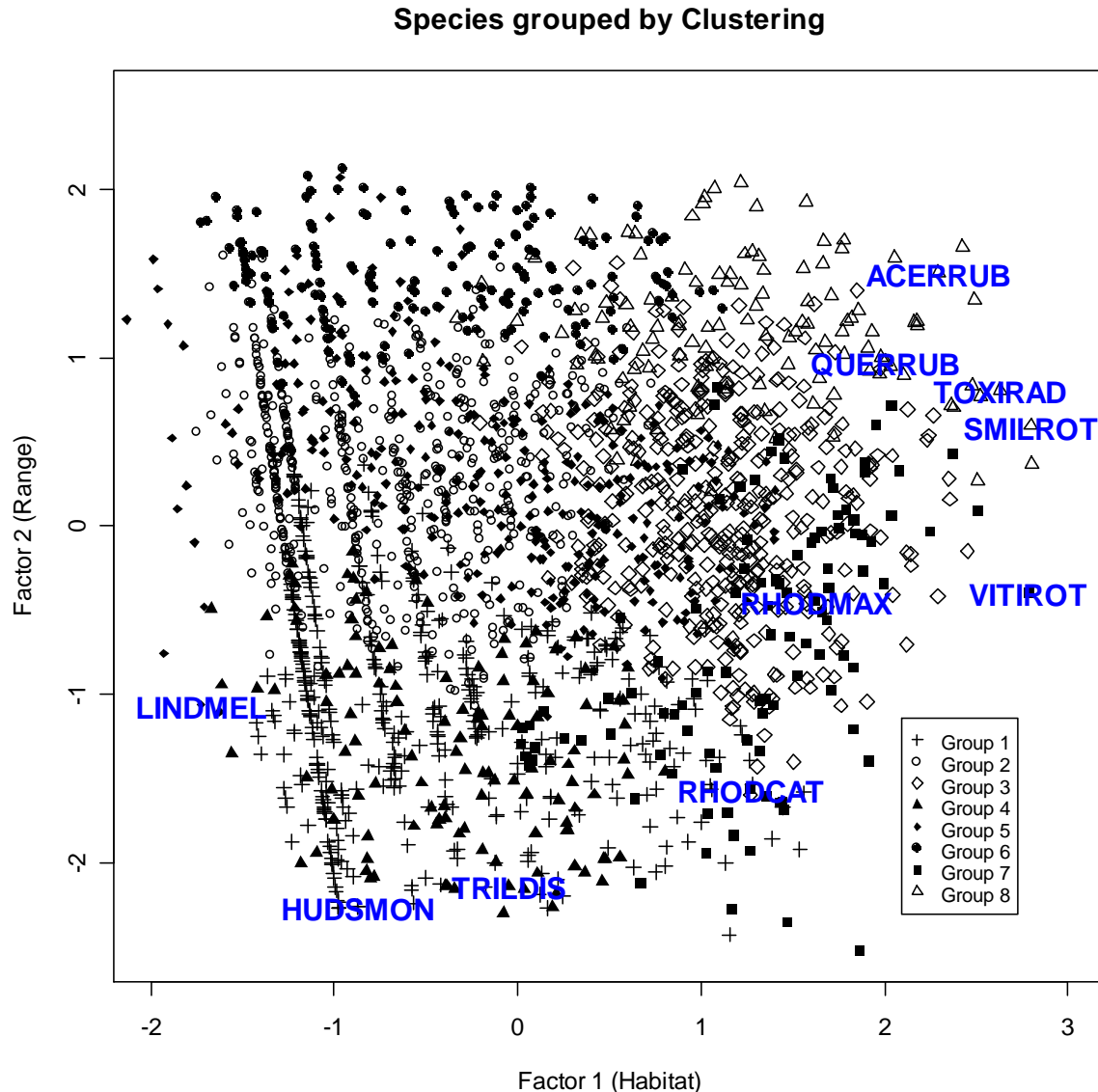


Figure 3.4. *Rarity groups as defined by clustering. Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, QUERRUB = Quercus rubra, RHODCAT = Rhododendron catawbiense, RHODMAX = Rhododendron maximum, SMILROT = Smilax rotundifolia, TOXIRAD = Toxicodendron radicans, TRILDIS = Trillium discolor, VITIROT = Vitis rotundifolia.*



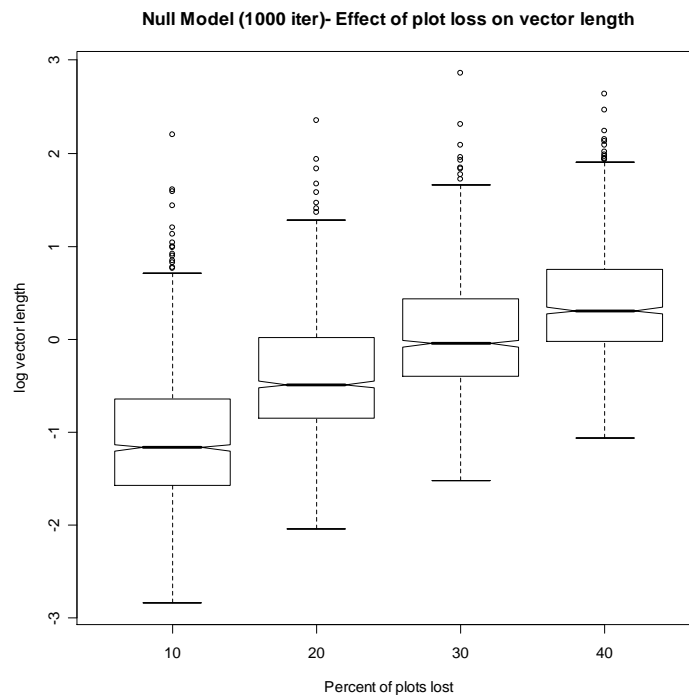
To compare the distance of movement through the ordination space due to simulated habitat loss, the mean vector lengths (averaged across 1000 iterations and log transformed) of all species were compared among each of the grouping methods. To test the effect of protecting plots that occur in community types designated as threatened by NatuerServe G-ranks, the mean vector

lengths were compared between the null model and the protected model for each grouping method.

Results

The null model revealed that increasing the percentage of plots lost increased mean vector length (species movement through the multidimensional attribute space measured by D^2 , Mahalanobis distance). The distribution of vector lengths is similar for each percentage of loss and the increase in movement is similar between each increase in loss (**Figure 3.5**), so the following results will present results for 10% plot loss.

Figure 3.5. Average vector lengths for all species from 10% plot loss to 40% plot loss.

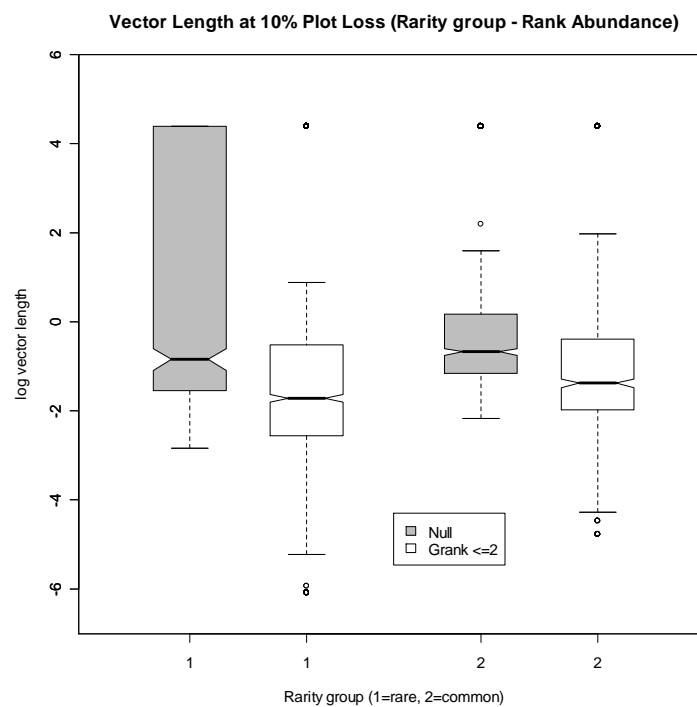


Effects of habitat loss across the forms of rarity

Rare groups defined by species in the tail of the rank abundance curve did have greater mean vector lengths than less rare species, and fewer extinctions occurred in the less rare group (**Figure**

3.6). Protecting G1 and G2 communities resulted in a reduction of vector length for both groups, and in the case of the rare group, resulted in many fewer extinctions. Due to the way extinction was measured (replacement of missing values with a value twice the maximum average vector length) groups with high extinction rates will have higher mean vector lengths. Therefore, the discussion of results will focus primarily on rates of extinction.

Figure 3.6. Average vector lengths for rarity groups defined by rank abundance and within protected community types of $G\text{-rank} \leq 2$.



Similarly, for the groups defined by median the more rare groups had greater vector lengths than the less rare groups (**Figure 3.7**). Three of the four rare groups experienced high extinction rates, while only one less rare groups had high extinction, even with protection of G1 and G2 communities. Increasing protection to G3 communities decreased extinction for all groups, but most notably for the two more rare groups, Groups 1 and 3 (**Figure 3.8**). It is notable that all four groups that experienced the highest extinction rates (Groups 1, 3, 4, 7) were the four groups that had low CEGL counts (Table 3.3).

Figure 3.7. Average vector lengths for rarity groups defined by median and within protected community types of $G\text{-rank} \leq 2$.

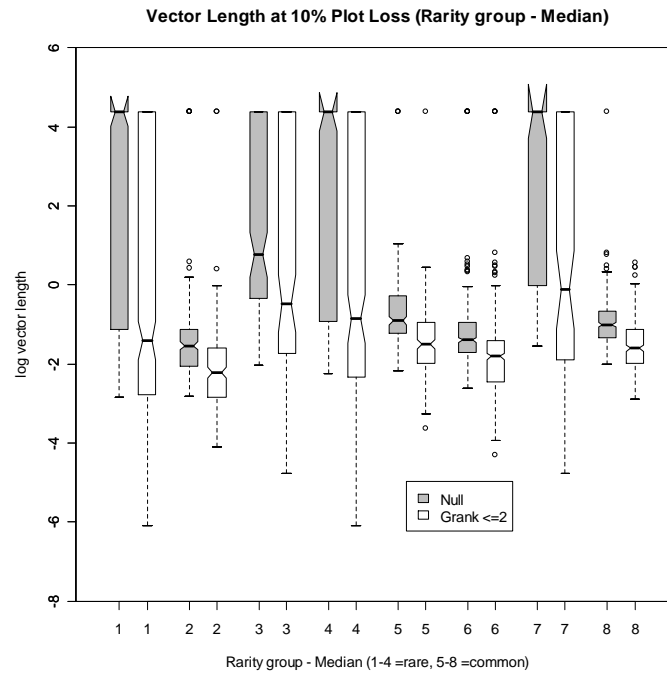
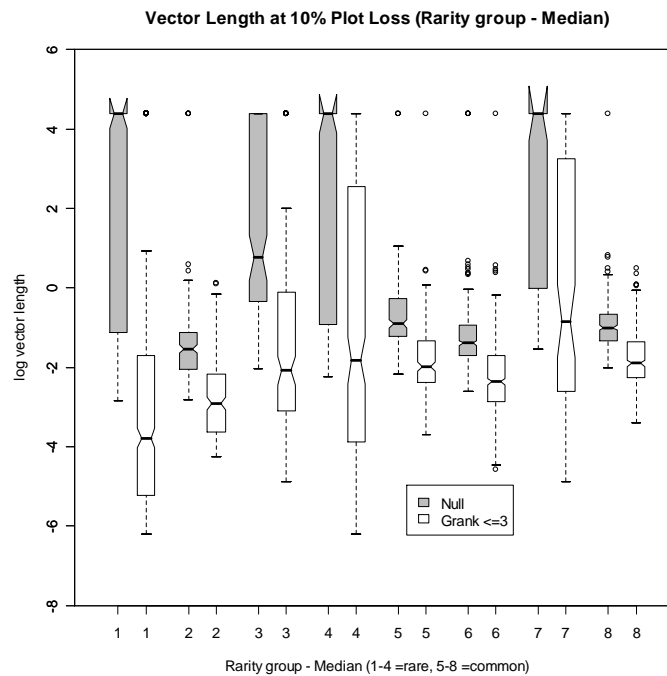


Figure 3.8. Average vector lengths for rarity groups defined by median and within protected community types of $G\text{-rank} \leq 3$.



Groups defined by clustering also showed disproportionate movement of rare species. The two most rare groups, Groups 1 and 2, both had high extinction rates, while one less rare groups had high extinction rates (**Figure 3.9**), even G1 and G2 communities. Increasing protection to G3 communities reduced vector lengths as well as extinction for all groups (**Figure 3.10**).

Figure 3.9. Average vector lengths for rarity groups defined by clustering and within protected community types of $G\text{-rank} \leq 2$.

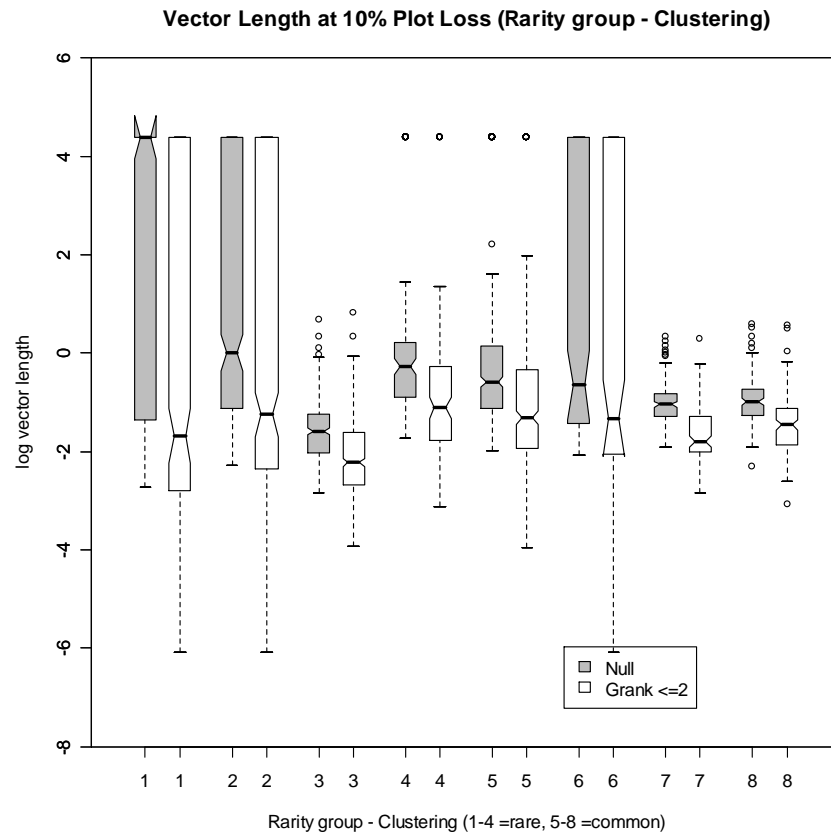
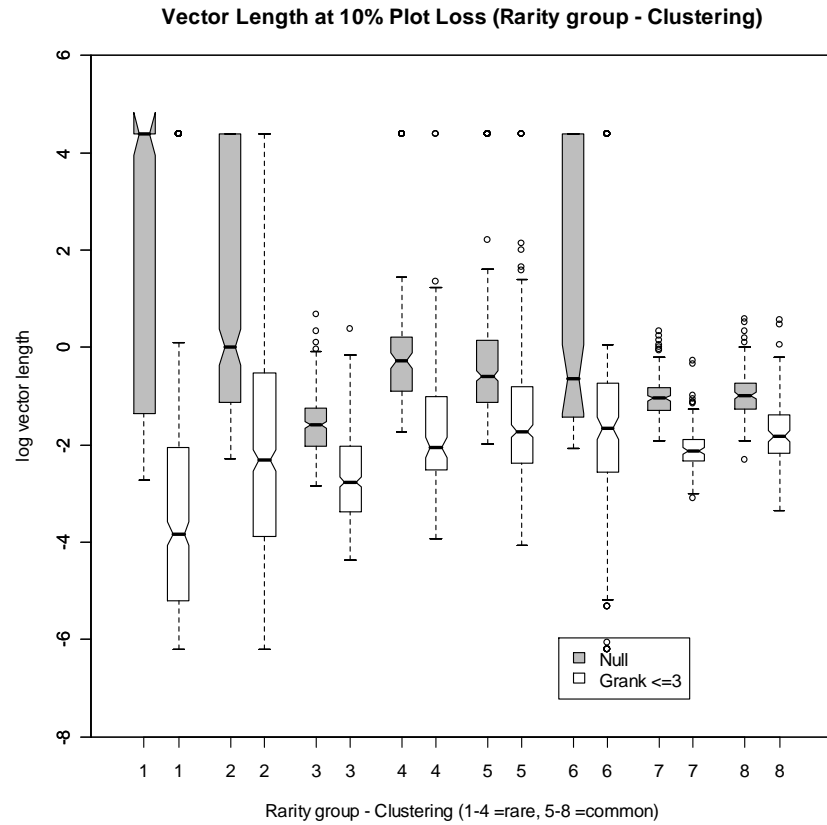


Figure 3.10. Average vector lengths for rarity groups defined by clustering and within protected community types of $G\text{-rank} \leq 3$.



Biological interpretation

The distance of movement through the multivariate space is plotted for five representative species in **Figures 3.11** and **3.12**. Each species occurs in a different realm of the space. Because the analysis is multidimensional, it is not possible to illustrate the direction of movement through the space. However, the degree of impact is indicated for each species.

Figure 3.11. Representative species plotted along factors 1 and 2. Because the analysis is multidimensional, it is not possible to illustrate the direction of movement through the space. Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, RHODMAX = Rhododendron maximum, VITIROT = Vitis rotundifolia.

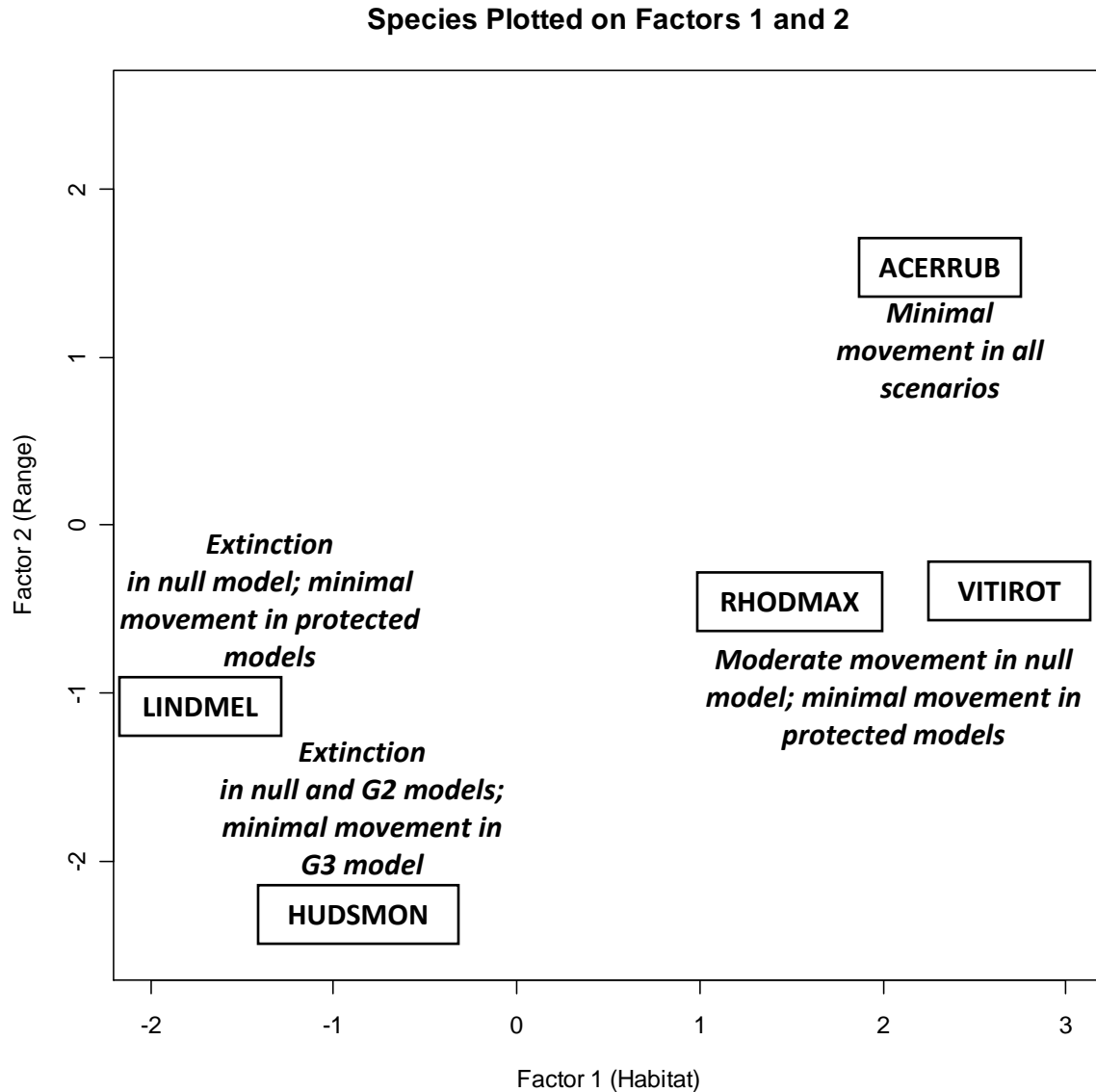
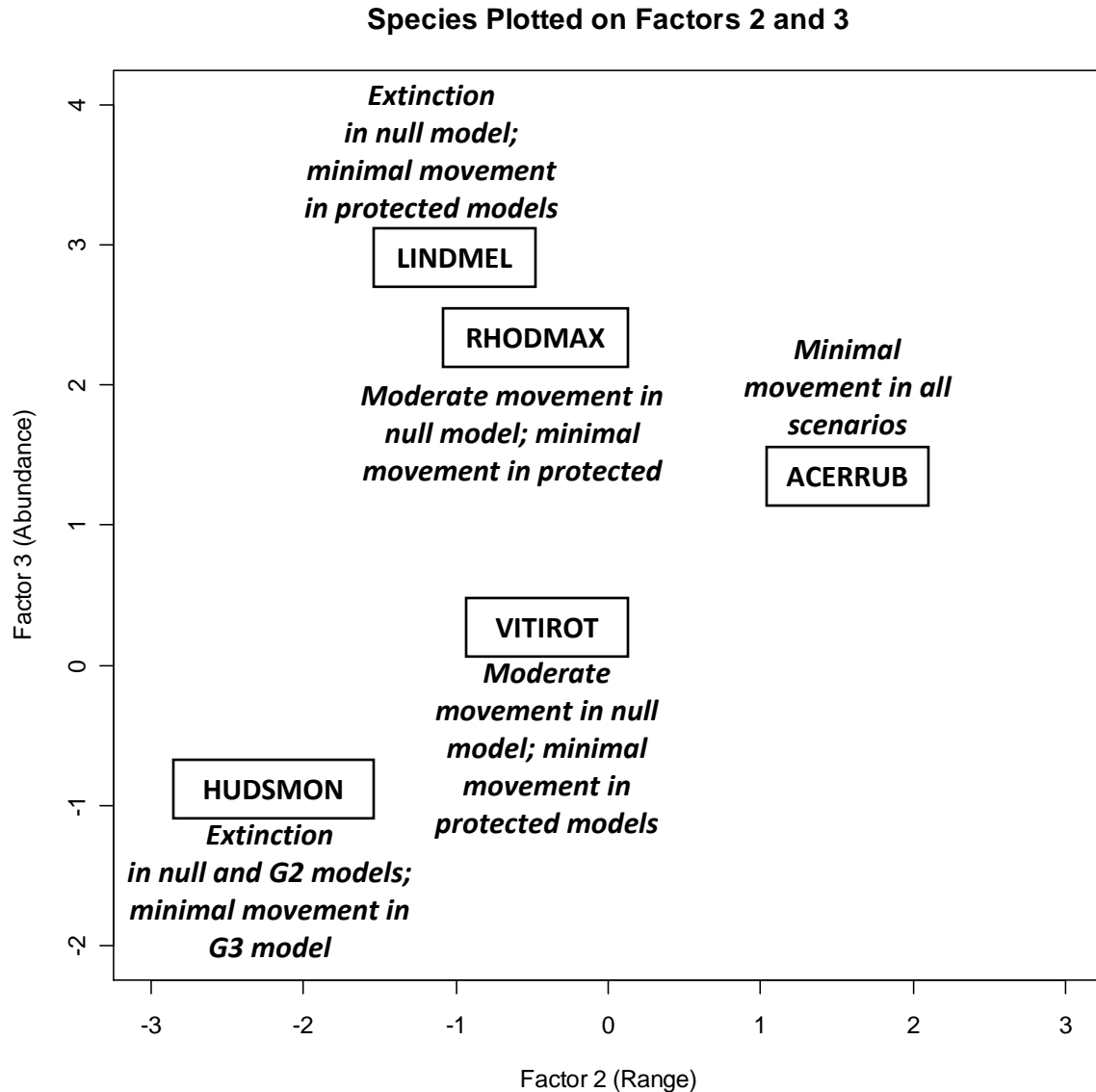


Figure 3.12. Representative species plotted along factors 2 and 3. Because the analysis is multidimensional, it is not possible to illustrate the direction of movement through the space. Species codes are defined as follows: ACERRUB = Acer rubrum, HUDSMON = Hudsonia montana, LINDMEL = Lindera melissifolia, RHODMAX = Rhododendron maximum, VITIROT = Vitis rotundifolia.



Discussion

Habitat loss does disproportionately affect rare species. While most species experienced some movement in ordination space, species in more rare groups were more likely to move and moved farther in ordination space, and experienced higher rates of extinction.

While it is generally accepted that species distributions are continuous (e.g., Gaston 1994), existing definitions of rarity are discontinuous and static, creating categories of rarity and filling the categories with lists of species that do not shift between categories (e.g., Rabinowitz (1981, Fiedler & Ahouse 1992, Kaye et al. 1997, Rey Benayas et al. 1999, Yu & Dobson 2000). It is important to note, however, that species are distributed continuously and that the boundaries between the forms of rarity, or the groups in ordination space, are arbitrary and only used to facilitate interpretation. The groups and the boundaries between them could be defined differently as would be appropriate for a given flora or fauna. Regardless of how groups are defined and whether they are labeled more rare, less rare, or common, it remains that species in certain regions of the ordination space tended to be less stable and experienced larger shifts than species in other regions. This finding provides a means of predicting which species within a given flora may tend to be more vulnerable to disturbances than other species. By indicating the distance of change through time, the degree or severity of impact can be detected and quantified.

This study demonstrates that habitat loss does cause species to shift between forms of rarity within a multivariate space defined by species attributes. While it is not surprising that habitat loss does affect species distributions, this analysis allows multiple species attributes to be considered simultaneously for an entire flora, and allows changes in attributes to be quantified simultaneously to reveal shifts in distribution pattern that may go unnoticed when species attributes are monitored and examined separately. The data required by this approach are not novel and would be currently

available for flora or fauna that are monitored; it is the use of the data in this approach that allows a more comprehensive view of species distributions as they change through time.

Conclusions

Studies of rare species have included developments of conceptual frameworks to classify species into forms of rarity, as well approaches to monitor species. This study brings together these pursuits by applying a quantitative method to classify species into forms of rarity as well as predict possible effects of habitat loss on those species

This approach provides three important contributions. First, it provides a more precise view of species locations within forms of rarity than has previously been possible and develops a method to detect movement within and between forms of rarity in response to disturbance. Second, it provides a method to measure the distance of movement within the multivariate attribute space. . Finally, it demonstrates that species with certain distribution patterns tend to be more vulnerable to disturbances such as habitat loss, and may therefore warrant higher levels of protection than species in less rare groups. Dividing species into only two groups based on only one dimension, such as defining rarity by rank abundance, neglects the information provided by a multivariate approach. Using multiple dimensions to divide species into multiple groups reveals more subtle patterns of impact experienced by species within each distribution pattern and allows a more precise targeting of conservation effort.

References

- Fiedler, P.L. and J.J. Ahouse. 1992. Hierarchies of cause: Toward an understanding of rarity in vascular plant species. In *Conservation Biology: The theory and practice of nature conservation, preservation, and management* (eds. P.L. Fiedler and S.K. Jain) Chapman and Hall, NY. 23 - 47.
- Gaston, K.J. 1994. *Rarity* Chapman and Hall, NY. 205.
- Johnson, R.A. and D.W. Wichern. 1982. *Applied multivariate statistical analysis*. Prentice Hall, NJ. 578.
- Kartesz, J.T. (1999) A synonymized checklist and atlas with biological attributes for the vascular flora of the United States, Canada, and Greenland. 1st ed. Synthesis of the North American Flora, Version 1.0 (ed. By J.T. Kartesz & C.A. Meacham). North Carolina Botanical Garden, Chapel Hill, NC, USA.
- Kaye, T.N., R.J. Meinke, J. Kagan, S. Vrilakas, K.L. Chambers, P.F. Zika and J.K. Nelson. 1997. Patterns of rarity in the Oregon flora: Implications for conservation and management. In *Conservation and management of native plants and fungi* (eds. T.N. Kaye, A. Liston, R.M. Love, D.L. Luorna, R.J. Meinke and M.V. Wilson) Native Plant Society of Oregon, Corvallis, OR. 1-10.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. (2005). Cluster Analysis Basics and Extensions; R package version 1.13.1.
- Manly, B.F.J. 2005. *Multivariate statistical methods: A primer*. 3rd ed. Chapman and Hall, NY. 207.
- McCune, B., J.B. Grace, and D.L. Urban. 2002. *Analysis of ecological communities*. MjM Software Design, Gleneden Beach, Oregon, USA.
- Peet, R.K., T.R. Wentworth, & P.S. White. 1998. A flexible, multipurpose method for recording vegetation composition and structure. *Castanea* 63:262-274.
- R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rabinowitz, D. 1981. Seven forms of rarity. In *The Biological Aspects of Rare Plant Conservation* (ed. H. Syngé) J. Wiley, Chichester, England. 205 - 217.
- Reilly, L.A.J. In preparation. A quantitative approach to determine the dimensionality of rarity.
- Rey Benayas, J.M., S.M. Scheiner, M.G. Sánchez-Colomer, and C. Levassor. 1999. Commonness and rarity: Theory and application of a new model to Mediterranean montane grasslands. *Conservation Ecology* 3:5.
- Roberts, D.W. 2010. Labdsv: Ordination and multivariate analysis for ecology. R package version 1.4-1. Link: <http://ecology.msu.montana.edu/labdsv/R>.
- USDA & NRCS (2004) The Plants Database, Version 3.5 (<http://plants.usda.gov>). National Plant Data Center, Baton Rouge, LA, USA.
- Weakley, A.S. (2005) Flora of the Carolinas, Virginia, and Georgia, Working Draft of January 2007. North Carolina Botanical Garden. <http://www.herbarium.unc.edu/flora.htm>

Yu, J. and S. Dobson. 2000. Special paper: Seven forms of rarity in mammals. *Journal of Biogeography* 27:131-139.

GENERAL CONCLUSIONS

I have observed that, for the flora of the Southeastern US, rarity is three-dimensional, those dimensions represent measures of local abundance, habitat volume, and geographic range, and species do not segregate into distinct groups or forms of rarity. Using fewer than three dimensions for these data, and dividing species into discrete groups defined *a priori*, results in species with different distribution patterns being inappropriately grouped together. While it is not uncommon to divide continuous variables into discrete groups, and while the practice can be useful for purposes of interpretation and communication, I assert that the definition of rarity is multidimensional and continuous and that groups must be defined *a posteriori* and interpreted biologically.

While the proposal of Rabinowitz (1981) that rarity is three dimensional has been supported by these findings and extended to explicitly recognize the continuous nature of species distributions, this work provides a foundation for further questions and applications for conservation. The three-dimensional solution is robust for the flora of the Southeastern US, however, dimensionality and inherent grouping patterns have yet to be tested for other systems and other taxa. The current availability of large-scale publicly available datasets and computing resources allow these aspects of rarity to be tested more rigorously than has been previously possible. Additionally, the multivariate approach used in the current study provides a precise location for each species within rarity space and reveals movement within that space due to impacts. Species in different areas of the ordination space did respond differently to simulated habitat loss. Other scenarios such as climate change and land use could be included in the model to predict how species with different distribution patterns

will respond to those impacts, allowing more strategic and biologically relevant conservation approaches to be designed.

APPENDIX 1
Plot descriptions for 2424 Carolina Vegetation Survey plots
Supplemental file: Appendix1_Plots.csv

The dataset includes 2424 plots selected from the Carolina Vegetation Survey (CVS). The descriptions contain the unique plot identification number assigned to each plot, the physiographic region in which the plot occurs, the NatureServe CEG code assigned to each code based on qualitative and quantitative assessment of community association, the scaled NatureServe Global Conservation Rank assigned to each plot, and the name of the community association to which each plot belongs.

APPENDIX 2
Environmental data for 2424 Carolina Vegetation Survey plots
Supplemental file: Appendix2_EnvData.csv

The dataset includes 25 soil nutrient variables, three soil texture variables, elevation, and slope for each of the 2424 plots selected from the Carolina Vegetation Survey (CVS). The suffix “ppm” indicates parts per million; the prefix “per” indicates percent. A typical CVS plot is 50m x 20m, consisting of 10m x 10m modules. Soil samples are taken from four modules within the plot, resulting in four values for each soil variable for a typical plot. For the current study, composite data for each plot were used such that each plot represents one sample unit. The mean value of soil texture variables was used. The median value for soil nutrient variables was used to reduce the effect of outlier values. A single value for slope and elevation are recorded for each plot. See Peet et al. (1998) for a detailed description of plot layout and data collection methods.

Peet, R.K., T.R. Wentworth, & P.S. White. 1998. A flexible, multipurpose method for recording vegetation composition and structure. *Castanea* 63:262-274.

APPENDIX 3
Species list
Supplemental file: Appendix3_SpecistList.csv

The dataset includes 2233 vascular plant species that occur in the 2424 plots selected from the Carolina Vegetation Survey (CVS). The list includes the CVS species codes and scientific names and the USDA equivalent and subequivalent codes and names, as well as group membership for each species as determined by rank abundance, median value, and cluster analysis. Varieties and subspecies were lumped to the species level, except in cases where only one variety or subspecies occurred within the dataset. Lumping is indicated by the assignment of the same CVS species code.

A typical CVS plot is 50m x 20m, consisting of 10m x 10m modules. Species cover values are taken within each module within the plot. For the current study, composite data for each plot were used such that each plot represented one sample unit with a single cover value for each species for each plot.