University of Tennessee, Knoxville

Trace: Tennessee Research and Creative Exchange

Doctoral Dissertations   Graduate School

12-2006

The Tangled Web of Community Ecology: Making Sense of Complex Data

Monica Lynn Beals
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Monica Lynn Beals entitled "The Tangled Web of Community Ecology: Making Sense of Complex Data." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Ecology and Evolutionary Biology.

Susan Riechert, Major Professor

We have read this dissertation and recommend its acceptance:

Daniel Simberloff, James Drake, James Fordyce, Hamparsum Bozdogan

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Monica Lynn Beals entitled "The tangled web of community ecology: making sense of complex data." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Ecology and Evolutionary Biology.

Susan Riechert

Major Professor

We have read this dissertation
and recommend its acceptance:

Daniel Simberloff

James Drake

James Fordyce

Hamparsum Bozdogan

Accepted for the Council:

Linda Painter

Interim Dean of Graduate Studies

(Original signatures are on file with official student records)

THE TANGLED WEB OF COMMUNITY ECOLOGY: MAKING SENSE OF COMPLEX DATA

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Monica Lynn Beals
December 2006

ACKNOWLEDGEMENTS

ABSTRACT

Ecological communities are governed by complicated processes that give rise to observable patterns. Making sense of these patterns, much less inferring the underlying processes, has proved challenging for several reasons. Manipulative experiments in natural communities may not be feasible due to large numbers of variables, lack of adequate replication, or the risk of undesirable consequences (e.g., introducing an invasive species). The multivariate nature of ecological datasets presents analytical problems as well; many statistical techniques familiar to ecologists have difficulty handling large numbers of potentially collinear variables. I present results from three studies of spider communities in which I employ a combination of familiar and less familiar statistical approaches to elucidate the factors influencing community structure in spiders. These approaches include null model analyses, nonmetric multidimensional scaling (NMS) for variable reduction of predictor and response data matrices, multiple regression, and observed variable structural equation modeling (SEM). While NMS has been employed as a multivariate descriptive analysis, examples of its use in further analyses are rare. SEM is a technique widely applied in other fields, but has only recently been used in ecological studies. General results from analyses of these three studies suggest that: 1) significant patterns of spider species co-occurrence based on null model analyses are consistent with a hypothesis of shared habitat preferences rather than one of species interactions, 2) in multiple regressions using NMS axes as predictor and response variables to compare the roles of plant species composition and habitat architecture in influencing spider species composition, the plants explained as much or

more variation as the architecture, and 3) based on SEM analyses using NMS axes for spider species, plant species, arthropod orders and habitat architecture as variables, plant species composition acts both indirectly (through its effect on arthropods and architecture) and directly. The combination in these analyses of a traditionally descriptive multivariate approach (NMS) with null models, a classic regression approach, and SEM permits the analysis of otherwise statistically intractable datasets (the original data matrices). This suite of approaches provides new insights into spider community structure, and can be applied by ecologists working in other systems as well.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Problem statement

Ecological communities have been likened to a "black box" in which observable patterns, such as abundance of individuals, local distributions, or species diversity, are shaped by complicated processes hidden from the observer (Bender et al. 1984, Shipley 2000). Collectively these patterns can be termed community structure (Swihart and Slade 1990), and the mechanisms by which they arise (i.e., intra- and interspecific interactions, environmental influences) are community organization (Landres and MacMahon 1983). Although these patterns may be fairly easy to measure, the underlying mechanisms and processes are more difficult to discern. This is because the specific attributes of a community are determined by multiple, potentially collinear, environmental and ecological factors, and direct and indirect interactions among these factors present both conceptual and methodological challenges (Wootton 1994a, McCune and Grace 2002, Graham 2003, Wootton and Emmerson 2005).

The basis of community organization is generally viewed as a combination of habitat preferences and species interactions such as competition, predation and mutualism (Wellborn et al. 1996). An organism's biotic environment includes both conspecific and heterospecific individuals, some of which may serve as resources (e.g., food, shelter, mates) or potential competitors for resources. For example, plants are a component of an animal's environment that can provide nutritive or spatial/structural resources and may ameliorate abiotic conditions by creating microclimates. Other animals may be prey, mates, competitors, or predators. At the community level, individuals interact both

1

directly with members of other species within and across trophic levels (direct effects), and indirectly when interactions are mediated through one or more additional species (indirect effects) (Wootton 1994a).

Questions about the roles of these different processes can be framed in the context of habitat selection. In a broad sense, habitat selection can viewed as a mechanism enabling species coexistence through the differential recognition and use of patches within a habitat, which then determines the number of species that can occupy that habitat (e.g., Rosenzweig 1981, Brown 1990, Morris 2003). For example, differences in the grain at which habitat selection occurs (i.e., the number of sub-habitats recognized by populations) can influence species diversity (MacArthur et al. 1966). This view considers habitat selection on an evolutionary scale by considering how selective forces have acted to permit species coexistence (Brown 1990).

Habitat selection can also be thought of at a finer, ecological scale as an outcome of biotic interactions (i.e., actual choice of patches is influenced by these interactions). Within suitable areas, in the absence of competition or predation risks, individuals' choice of habitat will be governed by resource availability (prey, space, etc.) and preferences for particular habitat features. More realistically, these choices are modified by the need to avoid risks from heterospecifics or conspecifics (e.g., Spence 1981, Gorman 1988, Dick 1996), with these interactions sometimes leading to occupation of otherwise suboptimal habitats (Rosenzweig 1991). At an ecological scale, these interacting processes that shape habitat selection give rise to the patterns of local spatial distributions of species within a community.

In this dissertation, I use spider communities to investigate community patterns and the potential processes that result in these patterns. I also explore statistical approaches designed for multivariate data and hypotheses with the goal of presenting and/or refining methods that yield meaningful and interpretable results from complicated data. Spiders are a well-studied and diverse group of arthropods that are abundant in terrestrial ecosystems (reviewed in Wise 1993), making them suited for studies of complicated community dynamics. Further, they play important roles as terrestrial arthropod predators and have been shown to regulate other arthropod populations (Finke and Denno 2002, Rosenheim et al. 2004, Denno et al. 2004) which can result in cascading effects on plant diversity and biomass (reviewed in Halaj and Wise 2001; see also Schmitz 2003, Schmitz et al. 2004), influence community succession (Hodkinson et al. 2001, Schmitz et al. 2006), and affect ecosystem properties such as nutrient cycling (Schmitz 2006).

Mechanisms of animal community organization

The role of plant communities

Habitat selection by animals is often influenced by the physical structure of the environment, which may be created by the physical structure of the vegetation. Habitat architecture has repeatedly been shown to have a strong influence on community structure in a variety of taxa (McCoy and Bell 1991, Tews et al. 2004; see Langellotto and Denno 2004 for meta-analysis). Habitat architecture can be described by single measures such as vegetation height diversity (MacArthur and MacArthur 1961, Rosenzweig and Winakur 1969, Murdoch et al. 1972, M'Closkey and Lajoie 1975,

Greenstone 1984), vegetation density (Kelaher 2003), plant surface area (Parker et al. 2001), biomass (Heck and Wetstone 1977, Rigby and Lawton 1981), fractal geometry (Williamson and Lawton 1991, Gunnarsson 1992), or the presence/abundance of specific features (for example, perches [Rand 1964, Moermond 1979], leaf litter [Bultman and Uetz 1982, Friend and Cellier 1990], nest sites [Tscharntke et al. 1998]). Other authors have used multiple measurements to describe the heterogeneity or complexity of habitat architecture, often employing variable reduction techniques such as principal components analysis, discriminant analysis, or cluster analysis to represent the data in a statistically tractable manner while incorporating more information than a single metric can provide (M'Closkey 1976, Rotenberry and Wiens 1980, August 1983, Koen and Crowe 1987, Brose 2003).

Habitat architecture, often explains more variation than other components of the environment. Numerous studies have compared the influence of habitat architecture versus that of prey availability (Greenstone 1984), temperature (Abele 1974, Rypstra 1986), species interactions (Dueser and Porter 1986), flooding regime (Brose 2003) or the plant community (MacArthur and MacArthur 1961, M'Closkey and Lajoie 1975, Heck and Wetstone 1977, Rotenberry 1985, Parker et al. 2001, Brose 2003), and have generally found that architecture is a better predictor of animal community structure.

Comparing attributes of the floristic community to habitat architecture, or using these attributes as sole predictors, to investigate their role in habitat selection and consequent influence on animal community structure commonly involves representing the plant community by a diversity measure (e.g., MacArthur and MacArthur 1961, Murdoch et al. 1972, Heck and Wetstone 1977, Panzer and Schwartz 1998, Siemann

1998, Siemann et al. 1998, Edwards and Otis 1999, Brose 2003, Hawkins and Pausas 2004), which some authors refer to as plant species composition (e.g., M'Closkey and Lajoie 1975). The relationship between plant diversity and animal community structure, however, has been inconsistent. In some studies plant diversity measures have been important influences on animal community structure (Murdoch et al. 1972, Panzer and Schwartz 1998, Tscharntke et al. 1998, Edwards and Otis 1999), while in others the relationships have been weak (Siemann 1998, Siemann et al. 1998, Parker et al. 2001, Hawkins and Pausas 2004) or non-significant (MacArthur and MacArthur 1961, Heck and Wetstone 1977, Koen and Crowe 1987, Brose 2003).

These conflicting results may reflect the use of a single variable, plant diversity, to represent the plant community. Plant diversity metrics may not adequately represent the taxonomic features of the plant community to which animals respond. Two sites could have comparable diversity values without necessarily sharing species in common. I use plant species composition here to refer to measures that incorporate specific plant species or functional groups of plant species. Plant species composition in this sense has been an important predictor of animal community structure as well (Ewert 1982, Rotenberry 1985, Martínez-Vilalta et al. 2002, Beals 2006). Rotenberry (1985) used the dry weight of major taxonomic groups in his analyses and Ewert (1982) compared frequencies of different plant taxa among his study areas. In the studies cited above, the plant community was represented by a single variable, plant diversity.

It is also reasonable to expect that plant species and habitat architecture will be correlated (Rotenberry 1985). For example, in studies where the dependent variables are plant diversity and foliage height diversity (MacArthur and MacArthur1961, Murdoch et

al. 1972, etc.), this correlation is expected since more diverse assemblages of plants are likely to exhibit a greater diversity of growth forms. MacArthur and MacArthur (1961) do not report the correlation between these variables, but Murdoch et al. (1972) found that plant diversity and foliage height diversity were highly correlated with one another and were equally correlated with insect diversity. Problems arise because few studies satisfactorily account for the relationship between plant species composition (or diversity) and habitat architecture (but see Rotenberry 1985). If plant species and plant architecture are strongly correlated, then statistically they will have similar effects on community structure and cannot readily be distinguished. For example, MacArthur and MacArthur (1961) concluded that plant diversity does not explain *additional* variation beyond that explained by habitat architecture, but never tested whether the reverse may be true. Their conclusions may result from the order in which they tested sequential sums of squares, rather than from the greater importance of architecture to bird communities, yet this study is cited frequently in papers as a classic example of the importance of habitat architecture.

Species interactions

In addition to preferences for architectural or floristic characteristics, habitat selection can also be influenced by species interactions. Observational and experimental evidence have indicated that both predation and competition influence community organization, although opinions about the relative importance of these interactions have shifted from an emphasis on competition to an emphasis on predation (Sih et al. 1985). Through the 1970s the primacy of interspecific competition in shaping communities was

a dominant paradigm in ecological research (Sih et al. 1985). Several reviews have addressed the evidence for interspecific competition from field experiments (e.g., Schoener 1983, Connell 1983) and concluded that competition was demonstrated in a majority of studies. Connell (1983) cautions, however, that this may in part be an artifact of ecologists investigating competition only in systems where they expect it to occur, and not publishing negative results. While some studies have provided evidence for competition as a contemporary force (e.g., Conley 1976, Stiling and Strong 1984), evidence from others has not (e.g., Rotenberry and Wiens 1980, Strong 1982, Dueser and Porter 1986). Patterns that appear to be indicative of competition may in fact be the "ghost of competition past," rather than evidence of contemporary competition (Connell 1980). Several authors have also pointed out that observed patterns could arise as a result of habitat variation (Simberloff 1983), variation in food availability (Schluter 1982), natural enemies (Strong 1982) or by random processes (Connor and Simberloff 1979, Ulrich 2004, Bell 2005), without a need to invoke current competition as the mechanism. Other studies have found that when "secondary associations" (Schluter 1984) are taken into account (e.g., associations with food resources or habitat characteristics), putative competition is less important in explaining distributions of populations (Schluter 1982, Dueser and Porter 1986, Schoener and Adler 1991, Sfenthourakis et al. 2005).

While the existence of interspecific competition is not generally disputed, its significance relative to other factors (e.g., disturbance, intraspecific competition, predation) has been questioned (Paine 1981, Peterson 1982, Stapp 1997). Stochastic disturbance, for example, may reduce the importance of competition by keeping population levels below thresholds at which competition occurs (Connell 1978, Huston

1979, Paine 1981), as may parasitism (Faeth and Simberloff 1981). Habitat

heterogeneity may reduce the chances of interspecific encounters and subsequent

potential for competition (Huston 1994). Intraspecific competition may function

similarly by limiting population densities and thereby reducing the chances of encounter

with heterospecifics (Anon. 1944 [cited in May 1984], Riechert 1981, 1982).

Predation, and its avoidance, has also been shown to have a strong influence on

habitat selection. Numerous studies have found that predation is as important or more

important than competition in structuring a broad variety of animal communities (e.g.,

Sinclair 1985 [ungulates], Hairston 1986 [salamanders], Stapp 1997 [mice]). Dayton and

Fitzgerald (2001) suggested that among anuran species in ephemeral breeding sites, the

importance of competition versus predation depended on relative duration of the

existence of the site. The strongest competitor was the most susceptible to predation, and

so may have been excluded from less ephemeral sites, which have higher probabilities of

colonization by predators (Dayton and Fitzgerald 2001). Predation and competition may

interact, with one mediating the effect of the other (Gurevitch et al. 2000, Chase et al.

2002). For example, predation may reduce the importance of competition among prey

species (Sih et al. 1985) by reducing one or more of their population sizes. In the

absence of predators, competition may have a greater impact than when predators are

present (Gurevitch et al. 2000). In some systems, the presence of competitors is

predicted to increase predation pressure by limiting refuge availability (Garvey et al

1994). Intensity of competition has also been found to vary with daily cycles of

predation risk (Hill and Lodge 1994). Perhaps the best example of the interaction of

predation and competition is that of intraguild predation, in which an individual both

acquires a food resource and reduces the potential for competition with another individual (Polis et al. 1989).

In addition to interactions between predation and competition, interactions and indirect effects among species interactions and environmental factors have been demonstrated. For example, structural heterogeneity of the habitat may influence predator-prey interactions by affecting the number of refuges for prey (e.g., Gilinsky 1984, Wywialowski 1987) or predator success (Clark and Messina 1998, Rypstra et al. 1999), or may influence competitive interactions by affecting the number of available niches (Sih et al. 1985). Predator-predator interactions may be mediated by habitat architecture if structural characteristics reduce or enhance the impacts of intraguild predation (Finke and Denno 2002). Taken together, the interactions of these potential mechanisms of community organization determine habitat selection among community members.

Studies of community organization in spiders

Potential mechanisms of habitat selection have been studied extensively in spiders. Empirical studies of habitat associations that have attempted to distinguish among the many factors potentially influencing spider communities have generally concluded that vegetation architecture is the best predictor of spider community structure. This is due in part to the fact that spiders exhibit an array of foraging strategies, requiring various substrate configurations for different web types or hunting modes (reviewed in Uetz 1991). Habitat architecture has been shown to have fitness consequences as well; grazing-mediated changes in vegetation structure affect the success of overwintering

juvenile wolf spiders (Bonte et al. 2000). Architectural parameters such as vegetation structure (Hatley and MacMahon 1980, Robinson 1981, Gunnarsson 1988, 1990, Halaj et al. 1998, McNett and Rypstra 2000), vegetation heterogeneity (Greenstone 1984), density (Balfour and Rypstra 1998), and litter depth (Bultman and Uetz 1982) have all been shown to influence both species diversity and abundance.

Both observational and manipulative studies have demonstrated this relationship between habitat architecture and spider community structure. Hatley and MacMahon (1980) found that species and guild diversities were higher in shrubs that were experimentally manipulated to increase density and architectural complexity than in clipped or control shrubs. Increases in spider abundance and diversity are also associated with increasing complexity in the physical structure of tree branches (needle density and orientation, twig biomass, etc.) (Gunnarsson 1988, 1990, Halaj et al. 1998, Halaj et al. 2000). Balfour and Rypstra (1998) found that higher weed densities were correlated with higher web spider densities, indicating that more web attachment sites were available as the vegetation became more structurally complex. Successional changes in vegetation lead to changes in architecture and associated changes in spider communities; older habitats with more complex vegetation structure have greater diversities of spiders (Lowrie 1948, Barnes 1953, Luczak 1959, 1963, 1966 [cited in Uetz 1991], Hurd and Fagan 1992). Studies that have discriminated between the effects of habitat architecture and prey availability have generally concluded that while prey availability plays a role (e.g., Ward and Lubin 1993), habitat architecture is a better overall predictor of spider diversity and density (Bultman and Uetz 1982, Greenstone 1984, Rypstra 1986, Bradley 1993, Halaj et al. 2000, Fournier et al. 2003).)

Habitat selection in response to plant species composition has not been studied extensively in spiders. This is because spiders are strictly predaceous, and most species are generalist predators and therefore unlikely to be associated with specific plant species based on specialization on a specialist herbivore. Although some researchers have observed associations between spiders and particular plant species or plant assemblages, these have been attributed to the differing architectural qualities of the flora (Barnes 1953, Riechert and Reeder 1970, Stratton et al. 1979; see also succession literature cited above). Neither Barnes nor Stratton et al. explicitly quantified vegetation architecture. Barnes speculated in his discussion that increases in vegetation stratification created more niches, while Stratton et al. investigated spider communities associated with three coniferous tree species considered to differ in physical structure. Riechert and Reeder, on the other hand, further investigated associations with plants by analyzing spider associations with plant species grouped together based on measured architectural features. Four of the nine species studied were significantly associated with physiognomically similar plant species.

The role of competition in habitat selection has been difficult to assess in spider communities. Many experiments to detect interspecific competition in spiders (reviewed in Wise 1984, 1993) have found little evidence to support interspecific competition as a mechanism organizing spider communities (e.g., Horton and Wise 1983, Riechert and Cady 1983), although in a study of interspecific competition between two web building species Spiller (1984) found that one species located its webs higher in the vegetation in the absence of the other species, suggesting that selection of web locations is influenced by the presence of a heterospecific. One potential problem with interspecific competition

experiments in spiders is that removal of competitors may also serve to remove a prey resource, since intraguild predation is common among spiders (Riechert and Cady 1983, Wise 1993 and citations therein). More recent experimental studies have found both evidence supporting the occurrence of interspecific competition (e.g., Balfour et al. 2003) and evidence for a lack of interspecific competition (e.g., Buddle 2002). Population densities of spiders are often below levels at which competition might occur, possibly as a result of abiotic factors and natural enemies (e.g., predators, parasites) (Wise 1993). Riechert (1981, 1982, and pers. comm.) has suggested that territoriality may limit population sizes and thereby reduce both interspecific and intraspecific exploitative competition (see also Riechert and Gillespie 1986). Demonstrating intraspecific competition has also been difficult (reviewed in Wise 1993). For example, although Wise (1983) found evidence of intraspecific competition in agonistic encounters between web residents and invaders, the effect on survival and egg production was not significant. Many studies of spider communities have not investigated competition per se, but have inferred it as a possible mechanism to explain observed patterns (Post & Reichert 1977; Hatley & MacMahon 1980; Robinson 1981; Rypstra 1986). However, as discussed above, numerous other mechanisms could explain these patterns.

Predation can also influence habitat selection in spiders. Spiders are generalist predators that exhibit intraguild predation, often preying on confamilial (Balfour et al. 2003) and heterofamilial spiders (Turner and Polis 1979, Wise and Chen 1999, Denno et al. 2004) and other predatory arthropods (Finke and Denno 2002, Rosenheim et al. 2004. Spiders also exhibit cannibalism (Hallander 1970, Turner and Polis 1979, Wagner and Wise 1996, Denno et al. 2004). Densities of conspecifics have been shown to influence

selection of microhabitats; increasing densities of a web building species resulted in more even spacing of spiders on leaves of individual plants (Miyashita et al. 1998). Both intraguild and intraspecific predation are thought to be important in structuring spider communities (reviewed in Wise 1993). Further, spiders have many non-spider natural enemies, and different architectural configurations can influence spiders' susceptibility to predators (e.g., birds [Gunnarsson 1990, 1996] or lizards [Spiller and Schoener 1988]).

As in other animal communities, habitat selection in spiders is determined to varying degrees by multiple interacting factors, including habitat architecture, microclimate, prey availability, species interactions and predation risks (reviewed in Riechert and Gillespie 1986 and in Wise 1993). Riechert and her colleagues have demonstrated habitat selection in response to these factors in multiple spider taxa (e.g., Riechert and Tracy 1975, Kronk and Riechert 1979, Riechert 1981, 1982). Given that spider communities are both influenced by the same types of processes that affect other animal communities and subject to interactions among these processes, they are an ideal taxon in which to investigate the complicated nature of community structure and organization.

Dissertation overview

Considerable research has been undertaken in an effort to understand the processes governing community structure, but the complicated nature of factors determining the organization of ecological communities has led many researchers to adopt a reductionist approach (Billick and Case 1994, Inchausti 1994, Wootton 1994a). This is reflected in studies in which only one or a few environmental variables are

measured, but is also true in the case of species interactions, where pairwise interactions are the most commonly investigated (Kareiva 1994, Stanton 2003, Strauss and Irwin 2004). This tendency to reduce the scope of community analyses by: 1) *a priori* data simplifications (selection of abundant species for analyses, use of diversity measures, etc.), 2) a reliance on what are essentially univariate hypotheses, and 3) the lack of incorporation of indirect interactions has slowed progress in understanding communities (Drake et al. 1996, Malaeb et al. 2000, Grace in prep). Many processes of interest to ecologists cannot be measured meaningfully by a single variable (Arhonditsis et al. 2006), and the interdependence among organisms in communities necessitates consideration of multivariable hypotheses in which species interact both directly and indirectly with other species (Wootton 1994a, 1994b, Stanton 2003) and their environment (Grace and Pugesek 1997, Pugesek 2003b).

Several practical considerations have contributed to this problem. Experimental manipulation of entire communities is rarely feasible due to the sheer numbers of variables involved and difficulties with replication (Wootton 1994b, Hilborn and Mangel 1987, Shipley 2000). Observational studies present difficulties both in terms of inferring underlying processes and distinguishing between random and nonrandom patterns (Gotelli and Graves 1996). Ecologists have used null model approaches to investigate large-scale patterns of species co-occurrence, body-size distributions, niche overlap, etc., and to determine the probability that an observed pattern arose by chance (reviewed in Gotelli and Graves 1996). While the mechanisms underlying nonrandom patterns cannot be conclusively determined from such analyses, the results can suggest avenues for future

research (Haukisalmi and Henttonen 1998, Peres-Neto et al. 2001, Collins and Simberloff in prep).

An additional practical matter in the analysis of community data is that many complex statistical analyses require large amounts of computing power, which until relatively recently limited their use (McCune and Grace 2002). The nature of ecological datasets has also posed analytical problems. Ecological data matrices tend to have high β-diversity among sample units as a result of habitat heterogeneity and/or large numbers of rare species, resulting in sparse data matrices (i.e., matrices with many zeros). Most traditional statistical techniques, including some multivariate approaches (e.g. Principal Components Analysis, cluster analysis, and factor analysis), have difficulty handling such matrices (Legendre and Legendre 1998). The use of multivariate methods such as nonmetric multidimensional scaling (NMS), which is amenable to use with sparse data matrices, has been advocated and has seen increasing use by ecologists in recent years as computing capabilities have expanded (McCune and Grace 2002).

Additionally, there is often high collinearity among ecological variables, which causes problems for traditional techniques such as multiple regression (Graham 2003). Indirect effects, in which relationships between two species are mediated by a third species or an environmental variable, are difficult to explicitly incorporate in familiar statistical analyses as well (Wootton 1994b, Arhonditsis et al. 2006). Path analysis, which has the ability to test complicated interdependencies among correlated variables, was introduced in the ecological and evolutionary literature in the early half of the last century. The methods have since been refined and extended (e.g., d-separation, structural equation modeling), but despite widespread application in other fields (psychology,

15

sociology, economics [Malaeb et al. 2000, Tomer 2003]), the use of these powerful techniques in ecology has been limited (Grace and Pugesek 1998, McCune and Grace 2002, Arhonditsis et al. 2006, Grace in prep).

In Chapter 1, I investigate patterns of spider species co-occurrence in three communities. Using a traditional approach to co-occurrence analysis, I show that, across the entire community species are generally distributed randomly with respect to one another, but that within subsets of the community (guilds and families) some groups show aggregated, or congruent, distributions. I suggest that these results are consistent with a hypothesis of shared habitat affinities in the three communities studied, influenced by the patchy distribution of habitat features previously shown to be preferred by spiders. In this chapter I also develop a null model to examine patterns of species avoidance or coexistence in species pairs within guilds that, based on NMS analyses, contribute to the variation in species composition within sample quadrats. Spider species pairs showing significant nonrandom patterns (i.e., avoidance) are generally consistent with a differential habitat preference hypothesis rather than a hypothesis of interspecific competition.

In Chapter 2, I employ NMS for variable reduction of both predictor and response variables. The resulting NMS axes are used in multiple regression to investigate the influences of vegetation architecture and plant species on community composition in spiders. Variable reduction via an ordination technique such as NMS, rather than dropping raw variables, allows retention of the information in the original dataset while producing statistically tractable variables for use in further analyses. NMS reduced the number of variables for predictor (habitat architecture and plant species) and response

(spider species) data matrices, and I then use these new variables in multiple regression analyses. Because the NMS axes can be interpreted based on their correlations with the original variables, this approach allows the recovery of biologically meaningful information from regressions. The advantage of this approach is that the important variables are determined by the data themselves, rather than by *a priori* assumptions of the researcher. Contrary to expectations based on previous work in spiders and other animals, plant species composition explains more variation in spider communities than does habitat architecture, and is also a stronger predictor of other community structure variables (overall abundance, species richness, and species diversity).

Chapter 3 expands upon the work in Chapter 2, attempting to elucidate the nature of the relationship between plant and spider species composition in one of the old field communities from Chapter 2. In Chapter 2, I suggest that this relationship is an indirect one in which plants influence the composition of non-spider arthropods, which in turn influence the spiders through their roles as either prey or potential competitors or predators. Plant species may also be confounded with fine-scale architectural features of the habitat, having an indirect effect on spiders owing to their structure or additionally owing to the influence of plant species composition on biomass. Using data collected in control and treatment quadrats, in which I manipulated the architecture of a focal plant species, I construct and test path models to explore the direct and indirect effects of plant species, habitat architecture, biomass, temperature and non-spider arthropods (direct effects only) on spider species composition. As in Chapter 2, I use NMS for variable reduction of the plants, habitat variables, arthropods, and spiders, allowing meaningful interpretations of the results. The use of path analytic models enables me to entertain

multiple hypotheses about direct and indirect effects simultaneously, as opposed to the traditional statistical approach of investigating a null hypothesis against a single alternative hypothesis. The results in the control quadrats are as expected; the influence of plant species on spider species composition acts indirectly through their effects on vegetation architecture and non-spider arthropod composition. In the treatment quadrats these indirect effects are still present, but the plant species also have direct effects on the spiders, and these path models have a greater number of significant pathways than in the control quadrats.

In Chapter 3 I also present regression models in which the non-spider arthropods are grouped into non-predatory and predatory groups. These results indicate that herbivore and scavenger arthropods (potential prey) and predator and parasite arthropods (potential competitors/predators) influence spider communities, and in some cases this influence varies between treatment and control quadrats.

CHAPTER 1

Species co-occurrence and habitat partitioning in heterogeneous environments

Introduction

Both biotic interactions and environmental constraints act to structure animal communities. Abiotic and biotic environmental factors (e.g., microclimate, vegetation characteristics) delimit the range of suitable or preferred habitats, while biotic interactions such as competition and predation are thought to shape the observed or realized distributions of the constituent species of a community (Wellborn et al. 1996). Differential use of space by animals within a community is thus assumed to reflect partitioning of the habitat along one or more resource gradients (Conley1976) occurring within the context of potential biotic interactions. In a deterministic framework, this spatial structuring of local distributions of species may result from active habitat selection in heterogeneous environments based on resource requirements, interspecific interactions, or some combination of these mechanisms. In this study, I examine spatial distributions (i.e., co-occurrence patterns) among species in three spider communities to ask the question: do species occur independently of one another? If not, are the patterns of species co-occurrence consistent with a hypothesis of negative, interspecific interactions or a hypothesis of shared habitat preferences?

Analyses of species co-occurrence have been used to infer processes influencing community structure or assembly, based on the assumptions that distributions of the species themselves implicitly incorporate all relevant environmental or resource information (Schoener 1974, Conley 1976, Bell 2005), as well as incorporating

19

information about species interactions (Levins 1968, Stone and Roberts 1992). The response to either environmental factors or other species varies in different contexts. Thus, communities that are structured by species interactions should exhibit lower average co-occurrence among species pairs than expected if interspecific competition is important (Gotelli 2000) while communities in which positive species interactions (e.g., facilitation or mutualism) are important will tend to show higher than average co-occurrence (Schluter 1984). On the other hand, negative or positive patterns of association among species may be largely determined by species' affinities for dissimilar or similar habitats, different resource requirements, or differing species characteristics, resulting from dispersal abilities or traits governed by phylogenetic constraints (Schoener and Adler 1991, Gotelli and McCabe 2002, Peres-Neto 2004).

Null models have been used extensively in ecology to determine whether observed co-occurrence patterns are non-random or are simply the product of chance (Gotelli 2000, 2001). However, given that different mechanisms can give rise to similar patterns, null models do not necessarily provide insight into which mechanisms may be operating. Inferring the processes underlying non-random patterns has been the subject of controversy (see Gotelli and Graves 1996 for review). One approach to this problem is to limit analyses to groups in which we would expect a particular ecological mechanism to be operating, permitting a narrower scope of inference.

Some authors have suggested restricting analyses to ecologically similar assemblages (i.e., guilds [Root 1967]) or taxonomically related groups within a community because this is the level at which ecologically significant interactions are expected to occur (Stone and Roberts 1992, Graves and Gotelli 1993, Collins and

Simberloff, in prep). Closely related species may have similar resource requirements that arise from recent shared ancestry (i.e., phylogenetic constraints), and by definition the same is true for guilds. Therefore, we might reasonably expect species in these groups to be more likely to exhibit interspecific competition than those with differing resource requirements (Stone and Roberts 1992, Graves and Gotelli 1993). Conversely, due to similar resource requirements, species in these groups may have similar habitat preferences (Schluter 1984, Peres-Neto et al. 2001). A significant pattern of negative associations among species within a guild or taxon is consistent with a hypothesis of interspecific competition, while a pattern of positive associations is consistent with a hypothesis of shared habitat preferences.

In addition to the difficulties of elucidating the mechanisms underlying species co-occurrence patterns, traditional analyses use presence/absence matrices and do not take into account differences in species abundances (Graves and Gotelli 1993). Two species may appear to co-occur randomly based on presence/absence, but may show nonrandom patterns in abundance (e.g., positive or negative covariance in abundances). Incorporating the actual numerical composition of species within sample units, thus, may reveal finer-scale patterns that are consistent with either species interactions or responses to habitat features. A common technique in community ecology that can include either presence/absence or abundance data is ordination. Ordination techniques such as nonmetric multidimensional scaling use compositional dissimilarities among sample units to represent the original species data matrix in a reduced-dimensional space. The correlations between species abundances in the original data matrix and the scores for the sample units in the reduced dimensions (i.e., ordination axis scores) reveal both co-

occurrence patterns and the simultaneous responses of each species to underlying (unmeasured) environmental gradients (McCune and Grace 1992, Beals 2006). Species responding in opposite directions to one another may be partitioning the habitat to avoid competition or predation, or may simply have different resource requirements or habitat preferences (reviewed in Gotelli and Graves 1996). Species responding similarly may be partitioning the habitat with respect to prey size, diel activity, or spatial or other resource requirements, etc. (Turner and Polis 1979).

Analyses of species co-occurrence patterns have a long history in community ecology (Gotelli 2000). Despite their limitations in inference, null model analyses of species co-occurrence serve as a good starting point for investigating community structure, as any patterns that appear can be further investigated experimentally (Haukisalmi and Henttonen 1998, Peres-Neto et al. 2001). Here I employ two null-model approaches to investigate patterns of species co-occurrence and habitat partitioning in three spider communities. The first approach involves analyzing presence/absence matrices using the C- and T-scores, which measure exclusiveness and congruence, respectively (Stone and Roberts 1990, 1992) for each community as a whole, within guilds, and within families. The second approach uses ordination (nonmetric multidimensional scaling [NMS]) of the species abundance data in each community to examine patterns in species distributions among the sample units, which could potentially result from species interactions or responses to underlying environmental gradients. The results of the NMS analyses are then used to construct null models for species pairs within broadly-defined guilds based on similar foraging strategies (e.g., pairs of active hunters or web builders). This latter null model analysis was motivated by an observed

pattern in the results from previous ordination analyses of these spider communities
(unpublished data). Pairs of spider species within guilds that contributed to the variation
in composition among the sample units along ordination axes tended to be correlated with
the ordination axes in opposite directions. The null distributions of differences in
correlation coefficients for all species pairs were used to test for nonrandom patterns.

Methods

Study areas

The study area consisted of three neighboring sites at the University of Tennessee
Woodlot in Knox County, Tennessee, USA, described in Beals 2006. Two were old-field
sites: the "sloped field" (SF), maintained by periodic mowing, and the "tussock grass
field" (TGF) maintained by yearly late-spring flooding. The third site, the "deciduous
woodland" (DW), was located in a deciduous forest dominated by tulip poplar
(*Liriodendron tulipifera* L.) and northern red oak (*Quercus rubra* L.). I sampled spiders
in these sites during two periods (July and late August/early September) in 1997 and
1998.

Circular quadrats (0.1 m^2) were located using a random walk technique (Catana
1955). Initial sample sizes were set at 20 per sampling period in all three sites in the first
year. In the second year sample sizes were set based on taking samples until the standard
error of the mean number of individuals per quadrat no longer decreased as samples were
added (Post and Riechert 1977). This led to increased sample sizes in the SF and TGF
sites in the second year. A total of 108 quadrats were sampled in the SF, 100 in the TGF,

and 80 in the DW. Spiders were sampled with a hand-held aspirator following the

protocol of Beals (2006) and Post and Riechert (1977).

Previous analyses did not detect differences in the spider communities across

sampling periods (Beals 2006), so the data were grouped by site for analyses. The SF

and TGF each had about 60 species (numbers are approximate because of juveniles that

could be identified only to genus), while the DW had about 50 species. As is typical in

natural communities, most species were rare; 60-70% of the species occurred in fewer

than five percent of the quadrats within a site, and about 30% of species in each site were

represented by a single individual. Spider species were grouped into three broadly-

defined guilds: web-builders, ambushers/stalkers, and active hunters. These guilds are

loosely based on the broad guilds of Uetz et al. (1999), but I have separated the

ambusher/stalker guild from the active hunters as they exhibit an intermediate strategy

between web-builders (sit-and-wait) and active hunting spiders (foraging occurs without

a web). Analyses were also conducted at the family level, as spider families have often

been used as proxies for guilds (Uetz et al. 1999).

Co-occurrence

Typical null model analyses of species co-occurrence patterns involve a Monte

Carlo randomization algorithm in which the original species presence/absence matrix is

permuted to generate a specified number of null matrices (Gotelli and Graves 1996). The

permutations are often constrained to maintain either fixed row (species) or column

(sample units) sums, or both (Gotelli 2000, Miklós and Podani 2004). In general, fixing

both marginal totals is recommended for both ecological and statistical reasons (Connor

and Simberloff 1979, Gotelli 2000). Several "swap" algorithms are available for generating the randomized matrices; here I employed Miklós and Podani's (2004) trial-swap method (see below for details). This method has been shown to sample null matrices equiprobably from the universe of all possible matrices. (Earlier swap algorithms [e.g., the sequential swap and independent swap] resulted in biased distributions of the null matrices [Gotelli and Entsminger 2001, Zaman and Simberloff 2004, Miklós and Podani 2004].) Before randomization, a co-occurrence metric is calculated for the observed matrix. This value is then compared to the null distribution of the metric calculated from each of the randomized matrices to test for non-random patterns in the data.

Two common measures of co-occurrence are the C-score and T- (or Togetherness) score. The C-score measures the average number of "checkerboard units" within a presence/absence matrix, in which one species is present in the absence of the other (Stone and Roberts 1990). A checkerboard unit is any 2×2 submatrix of the form

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

in which rows are species and columns are sample units. The T-score (Stone and Roberts 1992) measures the number of 2×2 submatrices of the form

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

These two scores ostensibly indicate patterns of negative or positive associations among species; a significantly high C-score is interpreted as indicating a large number of species pairs with exclusive distributions, while a significantly high T-score is interpreted as

25

indicating a large number of species pairs with congruent distributions. However, the C- and T-scores have a linear relationship if calculated for the entire presence/absence matrix when marginal totals (i.e., row and column sums) are fixed (Stone and Roberts 1992, Peres-Neto 2004, Collins and Simberloff, in prep); thus a community with a significantly high C-score, indicating exclusive distributions, will at the same time have a significantly high T-score, indicating congruent distributions. With fixed marginal totals, then, significant scores simply indicate a non-random pattern, but not whether associations are negative or positive (Collins and Simberloff, in prep). This is also illustrated by Haukisalmi and Henttonen (1998) who found that in an artificial, structured data matrix with similar numbers of positive (55) and negative (50) associations, the C-score was higher than expected by chance. This result is typically interpreted as indicating overall exclusive distributions, but it is clear in this example that this conclusion would be incorrect.

The relationship between these two measures can be decoupled by randomizing (through the trial-swap) the entire matrix and then calculating the scores for sub-groups (such as guilds or families) within this matrix (Stone and Roberts 1992, Collins and Simberloff, in prep.). Limiting co-occurrence analyses to ecologically similar or taxonomically related species is generally agreed to be a more appropriate approach than analyzing all species within a community together (Graves and Gotelli 1993, Collins and Simberloff, in prep). Species within guilds and families are theoretically more likely to compete because of ecological and taxonomic similarities. Thus, analyzing these subsets of the community addresses more relevant and interesting biological questions (Collins and Simberloff, in prep). This approach also avoids the problem where, if only some

subsets of species within the community are actually interacting or exhibiting nonrandom patterns, the signal of this interaction will be swamped by all the species pairs that do not exhibit exclusive patterns (the "dilution effect;" Diamond and Gilpin 1982).

C and T-scores for broadly-defined guilds (hunters, ambusher/stalkers, and web-builders) were calculated within each site, as well as those for each spider family represented by two or more species. Analyses were done using C++ code written by M. Collins, which performs the trial-swap algorithm of Miklós and Podani (2004). 10,000 randomized matrices were generated, with 500,000 attempted swaps (trial swaps) between sampled matrices. Each swap involves randomly selecting a 2×2 submatrix (by randomly selecting two rows and two columns) and, if the submatrix is a checkerboard unit (see above), swapping the rows or columns (Gotelli and Entsminger 2003). This method eliminates the "neighbor" bias created by traditional swap algorithms (Zaman and Simberloff 2004) by counting the number of attempted swaps rather than the number of actual swaps (Miklós and Podani 2004). Both C- and T-scores calculated for the null matrices were treated as one-tailed tests when comparing them to the scores for the observed matrix, so only scores that were significantly high ($\alpha=0.05$) were considered (e.g., a significantly *low* C-score did not count as evidence for congruent distributions). In addition to these scores, the number of species pairs exhibiting perfect checkerboards (CHECKER) (i.e., species pairs for which every 2×2 submatrix is a checkerboard unit) (Diamond 1975, Gotelli 2000) was calculated for the entire matrix. This measure is extremely stringent (and prone to Type II error), because a species pair that is exclusive in all but a single sample unit cannot be counted as a checkerboard (Gotelli 2000).

Habitat partitioning

In Beals (2006) I used nonmetric multidimensional scaling (NMS) implemented in PC-ORD (PC-ORD 4, 1999, MjM Software Design, Gleneden Beach, Oregon) (see Beals 2006 for details). This method reduced the number of dimensions of the data matrices and produced scores (the NMS axes) for each quadrat within a site based on spider species composition. Correlations between species and the NMS axes can be interpreted as representing the direction and strength (sign and magnitude) of a given species' response to the underlying environmental gradient(s) (biotic or abiotic). Because individuals were not equitably distributed among the different species, each species was relativized by its maximum occurrence in the quadrats within a given site (Faith et al. 1987, Legendre and Legendre 1998). In this way, responses of less abundant species were not swamped by responses of super-abundant ones. Very rare species, occurring in fewer than 5% of the quadrats in a site, were omitted from the analyses to reduce noise in the data (McCune and Grace 2002).

I found that species pairs within broadly defined guilds that were at least moderately associated with NMS axes (Kendall's $\tau \geq |0.30|$) tended to have opposite signs of their correlation coefficients (Beals 2006) (Table 1.1; all Tables and Figures are in Appendix A). I assume here that differing responses among spider species within broadly defined guilds indicate that they may be spatially partitioning the habitat, based on either dissimilar habitat preferences or species interactions. Three spider axes (S3 in both the SF and TGF and S1 in the DW) either had only a single species associated (TGF and DW), or species associated with the axis were not in the same guilds; these axes

could not be included in the analyses. A total of 14 species pairs identified from the previous analyses discussed above were used here.

The difference in spider species' responses can be quantified by the magnitude of difference between their correlations with a given NMS axis. In other words, a species pair in which one species is strongly positively correlated with the NMS axis and one is strongly negatively correlated will have a greater difference between their correlation coefficients than a pair with weak positive and negative correlations, or where the correlations have the same sign. I compared pairwise differences between correlation coefficients of 14 species pairs to null distributions generated by Monte Carlo simulations to ascertain whether apparent patterns of habitat partitioning were significantly non-random. MATLAB (MATLAB 6.5.1, 1984-2000, The Mathworks, Inc., Natick, Massachusetts) was used to randomly generate 10,000 within-guild species pairs by resampling the correlation coefficients from all species within the focal guild in the original data matrix. The correlation coefficient differences were calculated for each null pair to generate distributions. Observed values within guilds were compared to these null distributions to obtain *p*-values.

Results

Hunting spiders were numerically dominant in terms of number of individuals in both the SF (56%) and TGF (50%), a pattern primarily driven by the abundance of wolf spiders (family Lycosidae), which comprised 47% and 40% of individuals in the SF and TGF, respectively. The DW was dominated by web-building spiders (69%); these were primarily individuals in the families Linyphiidae (25%) and Theridiidae (28%). The

29

remaining hunting and web-building families (13 in the SF and TGF and 11 in the DW)

each made up less than 15% of the individuals in the old field sites and less than 10% of

the individuals in the woodland site, while ambush spider families (Salticidae and

Thomisidae) each made up less than 16% of individuals in all three sites.

Co-occurrence

Community-wide, both the C- and T-scores indicated that species occur randomly

with respect to one another in all three sites (Tables 1.2-1.4). The CHECKER score,

however, indicated a significantly higher number of perfect checkerboards than expected

by chance in the TGF (p=0.007), and weaker evidence for this in the DW (p=0.09)

(Tables 1.2-1.4). Within guilds and families the majority of analyses of co-occurrence

patterns were non-significant, but significant nonrandom patterns that were found all

indicated congruent distributions. In total, 21 out of 35 sub-matrices analyzed had

observed T-scores that were higher than the mean scores from the randomized matrices,

with 7 of these sub-matrices having significantly (α=0.05) higher scores (Tables 1.2-1.4).

Nine sub-matrices had higher but non-significant C-scores, and 5 had non-significant

Togetherness and C-scores that were either both higher or both lower. Of the guilds and

families with significant non-random patterns, all showed congruent distributions; species

in these groups tended to co-occur more often than expected by chance (Tables 1.2-1.4).

In the sloped field, both the ambush guild as a whole and the family Salticidae (in the

ambush guild) had significantly high T-scores (*p*=0.05 and 0.03, respectively) (Table

1.2). The Lycosidae (hunting spiders) also showed a trend towards congruent

distributions, though this was not significant (*p*=0.10). All of the guilds in the tussock

grass field showed random patterns, but within the family Thomisidae (ambush guild) there was evidence of positive associations (*p*=0.05) (Table 1.3). In the deciduous woodland the web-building guild and two web-building families (Araneidae and Theridiidae) also showed congruent patterns (*p*=0.03, 0.05 and 0.03, respectively) (Table 1.4). Although I performed multiple tests I did not apply a Bonferroni correction because this is an observational study, and therefore it is undesirable to inflate Type II error (see Moran 2003). Further, the Bernoulli probability of obtaining exactly 7 significant results (at α=0.05) out of 35 by chance is 0.001.

Habitat partitioning

Of the 14 species pairs tested in this analysis, seven pairs were hunting spiders and seven pairs were web-builders (Table 1.1). Seven species pairs (five web-building pairs and two hunting pairs) showed significantly (α=0.05) large differences in their correlation coefficients compared to the null distributions (Table 1.5). Additionally, two species of hunting spiders had small but non-significant p-values for the differences in their correlation coefficients (p=0.09). An example graph of a null distribution of differences in correlation coefficients plotted with the observed value is given in Figure 1.1 for the species pair *Bathyphantes pallida* and *Gea heptagon* associated with the first NMS axis in the SF. As above, I have not used a Bonferroni correction; here the probability of 7 spurious results out of 14 at α=0.05 is 0.000002.

Discussion

The results of both null model analyses show interesting nonrandom patterns in some groups (guilds, families, and species pairs) within these spider communities, despite a prevalence of random patterns in the majority of groups. As discussed above, the primary utility of null model analyses is the detection of nonrandom patterns; the mechanisms that may give rise to these patterns cannot be conclusively inferred (Connor and Simberloff 1979, Gotelli and Graves 1996). However, constraining analyses to groups of species among which we might reasonably expect to see particular mechanisms operating (based on guild membership, etc., as here for example, or habitat affinities [Peres-Neto et al. 2001, Peres-Neto 2004]) may allow us to entertain a smaller range of mechanistic explanations. If patterns are in fact nonrandom, these potential explanations can subsequently be tested experimentally (Haukisalmi and Henttonen 1998, Peres-Neto 2004).

The first null model approach used here evaluates patterns of co-occurrence in species presence/absence matrices. At the community level, spider species in all three communities appear to be distributed independently of one another based on the C- and T-scores. This may be due in part to the "dilution effect" (Diamond and Gilpin 1982), in which the inclusion of many species pairs that would not be expected to exhibit nonrandom patterns drowns out the signal of pairs that might. Within guilds and families in this study, the six submatrices that did show significantly nonrandom patterns all indicated more co-occurrence (higher T-scores) than expected by chance. Of those submatrices with non-significant patterns, 15 had higher T-scores than expected, 9 had

higher C-scores, and in 5 the C- and T-scores could not be decoupled. These results are generally consistent with a hypothesis of shared habitat preferences within these groups.

In both the SF and TGF, the ambush/stalker spiders were the only ones to show significant positive associations (the ambush/stalker guild and family Salticidae in the SF, and the family Thomisidae in the TGF). Salticids have been shown to prefer more open vegetation architecture (Robinson 1981) and vegetation structures with horizontal orientation (Heikkinen and MacMahon 2004). Thomisidae have been shown to prefer dense (Hatley and MacMahon1980) and herbaceous (Abraham 1983) vegetation. Previous analyses of the SF and TGF indicated that these types of architectural characteristics were distributed patchily throughout the sites (Beals 2006), so the significant congruent distributions of members of these two families may reflect their shared affinities for vegetation architectural features that are spatially heterogeneous.

In the DW, the web-building guild and two web-building families (Araneidae and Theridiidae) showed significant positive associations. In this site, the vegetation was sparse and patchily distributed; many sample units had no vegetation at all. Web-builders require attachment points for their webs, and in this study the majority of species in this guild build aerial webs (Uetz et al. 1999). Consequently it stands to reason that these species would aggregate in sample units that provide vegetation structure necessary to construct their webs.

The overall prevalence of random patterns in the presence/absence matrices from these three communities is generally consistent with results from studies of arthropods and other terrestrial invertebrates (e.g., Gotelli 2000, Gotelli and Ellison 2002, Ribas and Schoereder 2002, Escobar et al. 2005, Sfenthourakis et al. 2005), though some of these

33

studies also found significant nonrandom patterns in some matrices. Gotelli and McCabe's meta-analysis (2002) of presence/absence matrices found no evidence for nonrandom patterns among matrices of invertebrates (other than ants), and analyses of congeneric species by Sfenthourakis et al. (2005) found that all but one matrix of invertebrates (both arthropods and others) showed random patterns. There are few examples in the literature of analyses of spider species presence/absence matrices, but Peres-Neto et al. (2001) found that within an assemblage of hunting spiders, associations were positive overall. When they accounted for environmental heterogeneity, all pairwise negative interactions disappeared as did many of the positive ones, resulting in random patterns of association among most species. (Peres-Neto et al. did not constrain the column totals in their analysis, so the C-and T-scores were not coupled.) These results are qualitatively similar to those presented here, and provide strong support for a shared habitat preference hypothesis.

The second null-model approach used here incorporated relative abundance data in NMS ordinations, and then examined patterns between species within guilds whose abundances described moderate to large amounts of variation in species composition among the sample units. Analyses of the differences in correlation coefficients indicated that observed values for half of the within-guild species pairs (7 out of 14) were non-significant, a result that is consistent with results from the co-occurrence analyses. Some of these species pairs may co-exist through their use of different parts of the habitat within sample quadrats, or through temporal partitioning. For example, *C. abbotti* and *R. rabida* in the SF are both nocturnal, but *C. abbotti* is a foliage runner while *R. rabida* is a

ground runner (Uetz et al. 1999). In the TGF, *C. abbotti* and *P. undulata* may coexist through temporal partitioning (*P. undulata* is diurnal).

The remaining seven species pairs, however, did show significant negative patterns of association. Five of the seven significant comparisons involved web builders (one pair in the SF and four pairs in the DW). The six species in these comparisons belong to four different families, each of which builds a different type of web. Four of the five species pairs with significant p-values for their correlation differences (one in the SF and three in the DW) involve species from different web-building families, and it may be that different architectural requirements for each web type drive the observed pattern of negative co-occurrence. As in the SF and TGF, architectural features of the vegetation (e.g., maximum height and heterogeneity of open space within the vegetation) described considerable variation among the sample units (Beals 2006). However, members of one pair in the DW (*B. pallida* and *L. nebulosa*, S2) belong to the same family (Linyphiidae). The pattern for these two similar species may indicate a role for interspecific competition, as both species build their sheet webs low to the ground (Kaston 1981) suggesting that they have similar requirements for web attachments. Linyphiids are also known to forage off their webs (Uetz et al. 1999), and thus might be expected to have higher encounter rates with one another than other web-builders. The other two significant comparisons involved the same two hunting species (*P. milvina* and *C. abbotti*) in different dimensions (S1 and S2) in the TGF. In both dimensions, the species showed significant opposite responses. These two species exhibit differences in diel activity (*P. milvina* hunts diurnally while *C. abbotti* is nocturnal) and thus would not be expected to interact

strongly, supporting a hypothesis that their co-occurrence patterns are driven by dissimilar habitat preferences.

Previous work (Beals 2006) provides some insight into spider responses to particular habitat characteristics. The two web-building species in the SF (*B. pallida* and *G. heptagon*), for example, have been shown to have weak but significant responses to increasing cover of herbaceous or grass species, respectively. The significant opposite-responding pair of hunting spiders in the TGF showed weak but significant responses to different aspects of habitat architecture. For example, *P. milvina* responded positively to increasing heterogeneity in the heights of plants within a sample unit, while *C. abbotti* responded negatively. Spider species in the DW were predicted by plant species rather than vegetation architecture (Beals 2006). In the SF (above), spiders responded differently to plant species with different growth forms (grasses versus herbs), but in the DW the differences among the plant species to which spiders responded were less clear. Differences in finer-scale architecture or palatability to herbivores may explain in part the spatial patterns of the spiders examined here.

Taken together, the results presented here suggest that habitat preferences play a greater role in structuring spider communities than species interactions. This is not to say that mechanisms like competition and predation do not occur or are never important, at least for specific pairs or groups of species (Peres-Neto 2004). In general, however, habitat preferences seem likely to be stronger drivers of community-wide patterns in these spider communities. Interspecific competition among spiders has been difficult to demonstrate and is generally thought to be of limited importance in structuring spider communities (reviewed in Wise 1993). Riechert (1981, 1982) has

36

suggested that intraspecific competition acts to regulate spider populations at a level below that at which interspecific competition would occur. Wise (1993), on the other hand, has argued that intraspecific competition is negligible, at least among web-building spiders, because prey limitation acts as a density-independent limiting factor. This also results, however, in spider population densities being below competitive levels, so interspecific competition is therefore even less important than intraspecific competition (Wise 1993). Spider population densities may be limited by a combination of abiotic mortality and natural enemies (Wise 1993), and so we would not expect to see strong signals of species × species interactions in this group, but rather signals of species × habitat interactions (e.g., habitat preferences) (Gotelli and Graves 1996). Finally, it is important to note that the analyses presented here indicate that many species occur independently of one another. Although I have focused on deterministic explanations for the nonrandom patterns, stochastic processes also appear to be important in shaping these communities.

Null model analyses such as those presented here are useful tools for describing patterns of species co-occurrence (Gotelli 2001). The use and interpretation of null models has been criticized because numerous mechanisms may be invoked as potential explanations for nonrandom patterns. By constraining analyses to groups such as guilds, families, or other taxonomic units (e.g., Stone and Roberts 1992, Peres-Neto et al. 2001, Sanderson 2004, Sfenthourakis et al. 2005, Collins and Simberloff, in prep) and accounting for the scale of the study (e.g., islands versus sample units in an old field), some mechanistic explanations can be ruled out, permitting more robust inference. In a similar fashion, many authors have constrained their analyses based on environmental

factors, by either accounting for environmental heterogeneity (Peres-Neto et al. 2001,

Sfenthourakis et al. 2005) or using incidence functions to constrain reshuffled species in a

null matrix to occur within the range of sites in which they were actually observed

(Escobar et al. 2005, Collins and Simberloff in prep). While null models cannot be used

to demonstrate mechanisms conclusively, by limiting the number of potential mechanistic

explanations constraints such as these may suggest more fruitful directions in future

research.

CHAPTER 2
Understanding community structure: a data-driven multivariate approach

This chapter has been revised slightly from a paper by the same name published in the journal Oecologia in 2006 by Monica L. Beals:

Introduction

The inherent complicatedness of natural communities challenges our

understanding of how habitat influences the abundance of individuals, local species

distributions, and species diversity. Ecological datasets frequently contain large numbers

of variables that are highly collinear, leading to difficulties with both analysis and

interpretation (Graham 2003). Many community-level studies simply ignore this and

focus on the effects of a single component of the environment, or limit measurements to a

small number of variables. Given the potential for interactions and covariation among

ecological variables, a reductionist approach may yield statistically significant results

without elucidating the intricate nature of the relationships between communities and the

environment. Here I combine two statistical approaches to assess whole-community

responses to multiple environmental variables. I provide an example of this approach by

focusing on how communities of an important terrestrial arthropod predator (spiders)

respond to biotic components of the environment.

In animal communities, environmental variables may include simple measures of

habitat architecture or heterogeneity (such as vegetation density, foliage height diversity,

etc.) or plant species composition (e.g., species richness, diversity indices). While

heterogeneity of habitat architecture has been shown to influence community structure in diverse taxa (surveyed in Tews et al. 2004), many studies have represented habitat architecture with only one or two variables (e.g., MacArthur and MacArthur 1961; Rosenzweig and Winakur 1969; Estades 1997), and in some cases have found no significant relationships between vegetation structure and animal communities (e.g., Johnsingh and Joshua 1994). Studies that have investigated several variables have found that multiple aspects of habitat architecture (or other environmental components) contribute to species abundance and diversity (e.g., Rotenberry and Wiens 1980; August 1983; Brose 2003). Studies examining the role of plant species in structuring animal communities have been similarly equivocal. These typically use species richness or diversity to summarize the plant community, and often find weak (e.g., Siemann et al. 1998; Parker et al. 2001; Hawkins and Pausas 2004) or contradictory (Siemann 1998) relationships, or none at all (Heck and Wetstone 1977; Koen and Crowe 1987; Brose 2003). Studies using multivariate techniques to describe plant communities have generally met with greater success in detecting relationships between plant and animal communities (Rotenberry 1985; Sanderson et al. 1995; Martínez-Vilalta et al. 2002).

The problems discussed above relate to the independent, or predictor, variables measured in ecological studies. Multivariate approaches have been recommended for variable reduction of the predictors (Graham 2003), because information from the original data is retained and a smaller, statistically tractable number of variables can be used in further analyses such as regression (see Somershoe and Chandler 2004, for example). The same problems arise, however, when considering dependent, or response, variables. Many community-level studies either restrict the variables (i.e., species) to a

small subset of the larger assemblage, chosen *a priori*, analyze individual species'

responses separately, or use diversity measures to represent the community (Drake et al.

1996). While these are important components of community structure, they do not fully

address the complicated nature of community dynamics.

The importance of habitat architecture for spiders is well documented, both for

spider communities and for the distribution of individual species (reviewed in Uetz 1991

and Wise 1993; see Langellotto and Denno 2004 for meta-analysis). Quantitative studies

investigating spider associations with floristic composition, on the other hand, are

uncommon (but see Riechert and Reeder 1970). Because of their abundance and the ease

with which they can be studied, spiders are an ideal taxon for investigating the ways in

which habitat architecture and plant species composition structure communities. Spiders

are also important predators in terrestrial ecosystems, with impacts on processes

including nutrient cycling, trophic cascades, and insect population regulation (reviewed

in Wise 1993). In this study, I demonstrate a method by which multiple species'

responses to multiple architectural and floristic variables can be investigated

simultaneously, combining nonmetric multidimensional scaling and multiple regression

to analyze three spider communities.

Methods

Study area

The study areas were in Knox County, Tennessee (USA) in three adjacent sites

owned by the University of Tennessee. Each site ("sloped field", "tussock grass field",

and "deciduous woodland") represents a different habitat type. The sloped field (SF) is

an old field maintained by periodic mowing, with a slope ranging from 5-15%.

Dominant plant species tend to be distributed fairly evenly throughout the site. The

tussock grass field (TGF) is a level area separated from the sloped field by a band of

shrubs (approximately 7m wide). This area is not mowed but is generally flooded at least

once each spring and drains slowly, usually drying out by mid-June. Plant species tend to

be clumped, creating relatively homogeneous patches within this site that are dominated

by one or two species. The deciduous woodland (DW) is adjacent to the sloped field but

disjunct from the tussock grass field. This site consists of two facing slopes separated by

a narrow strip of wet-mesic bottomland (varying from approximately three to ten meters

in width). The herbaceous layer is sparse, with a small number of shrubs dispersed

throughout the site (spiders were not collected in the forest canopy).

I sampled the three sites twice annually in consecutive years: 3 July to 18 July

1997, 23 August to 8 September 1997, 10 July to 29 July 1998, and 24 August to 20

September 1998, resulting in a total of 12 samples. Within each site in a given sampling

period, I located circular quadrats (0.1 m^2) by a random walk. In the initial year of the

study, there were 20 quadrats in each site for both sampling periods. Sample sizes were

increased the following year. For both sampling periods, I sampled 34 quadrats in the

sloped field, 30 quadrats in the tussock grass field, and 20 quadrats in the deciduous

woodland (total number of quadrats sampled = 288). These larger sample sizes were

determined by adding quadrats until the standard error of the number of individual

spiders per quadrat no longer decreased with additional quadrats (Post and Riechert

1977).

Data collection

Measurement of habitat architecture variables and assessment of plant species composition occurred at least one day in advance of spider sampling to minimize disturbance to the spiders. In each quadrat I recorded the following data: percent cover for each plant species, an approximation of vegetation density, vertical structural heterogeneity, vegetation height, and horizontal structural heterogeneity. To approximate foliage density, I placed a wood dowel (0.3 cm diameter) marked in one-centimeter increments vertically in the center of each quadrat, and recorded the lowest height at which at least one full centimeter was visible through the vegetation. Vertical heterogeneity is a measure of habitat architecture modeled after the point-intercept method of estimating plant species cover. At 10-centimeter increments above the ground, I suspended small dowels (divided into seven five-centimeter segments) horizontally from poles on either side of the quadrat, and counted the number of segments in contact with vegetation at each height. I reduced these data to two vertical structural heterogeneity variables using nonmetric multidimensional scaling (an ordination method; see below), which represented the amount and variability of open space within the vegetation. Vegetation height included the height of the tallest plant in each quadrat, as well as the coefficient of variation of the heights of the tallest plants in each of four quarters within the quadrat (a measure of horizontal structural heterogeneity; quadrats with relatively uniform heights of the tallest plants will have a lower coefficient of variation than quadrats in which heights are high in one quarter and low in another, for example). In the deciduous woodland, I also measured litter depth (measured in the center of each quadrat and in each of the four quarters of the quadrat). I derived

additional habitat architecture variables in each quadrat by using the plant percent cover data to quantify the total surface area of plants grouped by physiognomic characteristics (hirsute or glabrous leaves, and growth form [i.e., vines, herbs, grasses, shrubs]) based on Gleason and Cronquist (1991). Although these variables are species-based, they are not species-specific and therefore are appropriately included as habitat architecture variables.

Spider collections were completed between 10am and 3pm on clear days in which no measurable precipitation occurred during the six hours prior to sampling. I sampled spiders based on the protocol of Post and Riechert (1977). Populations of almost all species were aggregated (s^2 > mean) at the sampling scale of 0.1 m^2 quadrats. Post and Riechert (1977) suggested that in cannibalistic taxa such as spiders, aggregated dispersion patterns indicate adequate sampling scale (i.e., the quadrat size is greater than spider's spacing of themselves). In each quadrat location I placed a metal cylinder (0.38 m high, 0.1 m^2) on the ground and pushed the base of it into the soil to prevent spiders from escaping. I removed spiders in the cylinder with a handheld aspirator and placed them in vials containing 70% ethyl alcohol. I also removed the vegetation (and/or litter) from each quadrat, which I then sorted in the lab to retrieve any remaining spiders. I identified individuals to species when possible; early instars that could not be identified to species were identified to genus. When a genus was represented by a single species in all three study areas during all four sampling periods, I assumed juveniles of that genus belonged to that same species. In total, 2,337 individuals were identified (1,119 in the sloped field, 825 in the tussock grass field, and 393 in the deciduous woodland), representing approximately 90 species, 75 genera, and 20 families.

Data analysis: variable reduction and regression

Many commonly used statistical techniques (e.g., regression) are unable to deal with sparse data matrices (i.e., matrices with many zeroes, as are often encountered with species abundance data) (Legendre and Legendre 1998), or become difficult to interpret in the presence of multicollinearity (Graham 2003), as is also often observed with ecological data (such as habitat data). Ordination techniques have been recommended for variable reduction of predictor variables as a solution to the problem of multicollinearity in multiple regression analyses (Graham 2003), which also addresses the problems of sparse data matrices. Very rarely, ecologists have used ordination to reduce the number of *response* variables for regression or other general linear models (e.g., Boyer and Fong 2005; Willis et al. 2005), which has several advantages over more traditional approaches. Because one can obtain correlations between the raw variables (e.g., species in this case) and the new reduced variables, information from the raw data can be recovered. Summary measures such as diversity lose information about individual species. Additionally, multiple tests of individual species are unnecessary; species responses can be analyzed simultaneously. Using variables from ordination analyses in multiple regression is conceptually similar to canonical correspondence analysis (CCA), but unlike CCA does not constrain the representation of community structure to only the measured environmental variables (McCune and Grace 2002).

Nonmetric multidimensional scaling (NMS) is an iterative ordination method that places sample units in k-dimensional space using the ranked distances between them. As in other ordination methods (e.g., principal components analysis, detrended correspondence analysis), sample units that are more similar to one another (based on

species composition, for example) have scores that are closer together than sample units that are less similar (Fasham 1977). Ordered positions of the sample units in the final configuration (i.e., reduced dimension) are optimized to maintain a monotonic relationship between the distances of the ordinated points and the distance matrix generated from the raw data (Legendre and Legendre 1998; McCune and Grace 2002). Departures from monotonicity are indicated by "stress" (higher values of stress suggest poorer fit between the raw data and the NMS configuration). In contrast to other ordination methods, the numbering of the axes is arbitrary; the first axis of an NMS ordination does not necessarily explain more variation among the sample units than the second, and so on (McCune and Grace 2002). Correlations between the k axes of the ordination and the original variables can be used to interpret the axes. Because NMS does not assume linearity or monotonicity of the underlying data structure, it is particularly appropriate with the kinds of ecological data in this study (Fasham 1977; Minchin 1987; McCune and Grace 2002).

To reduce the dimensionality of the plant, architecture and spider data matrices, I used NMS with the Sørenson distance measure (PC-ORD 4, 1999, MjM Software Design, Gleneden Beach, Oregon). PC-ORD implements global NMS (see Discussion). I used PC-ORD's autopilot mode, which performs 40 runs with the raw data and 50 runs with randomized data using a random starting configuration each time; the program calculates one-dimensional through six-dimensional solutions for each run and reports the recommended k-dimensional solution. I used raw percent cover data to perform ordinations of the plant data matrices. For the spider species ordinations, I adjusted species counts by the maximum number of individuals of each species occurring within a

site (Faith et al. 1987; Legendre and Legendre 1998). Singletons and species occurring in fewer than five percent of the quadrats were omitted from the ordinations (McCune and Grace 2002). Architecture ordinations included the following variables: vegetation density, maximum height, horizontal heterogeneity, vertical heterogeneity (two variables), and total surface area of hirsute and glabrous plants, grasses, vines, herbs, and shrubs. In the deciduous woodland the architecture data also included two litter variables: maximum litter depth and coefficient of variation of litter depth.

I then used the k axes from each NMS analysis as variables in multiple linear regressions using PROC REG in SAS (SAS 9.1, 2001, Sas Institute Inc., Cary, North Carolina), employing all possible subsets regression for variable selection. I used the plant and architecture axes in multiple linear regressions as predictors of three general components of spider community structure (number of individuals, species richness, and species diversity [Simpson's D']) and as independent variables in regressions with the spider axes as dependent variables. I selected initial candidate regression models (out of all possible subsets) based on Akaike's Information Criterion (the AIC option in SAS); the final models chosen were those in which all parameter estimates were significant at $\alpha=0.05$ and which had minimum AIC scores. Responses of individual spider species to raw plant and architecture variables were interpreted based on the signs of the correlation coefficients of both predictor and response variables with the NMS axes, and the signs of the parameter estimates from the final regression models. As this is an observational rather than experimental study, I did not use a sequential Bonferroni correction, due to the increased risk of Type II errors associated with it. Arguments against using this correction are detailed in Moran (2003). In particular, he argues that the probability of

finding multiple significant tests due to chance is very low. Further, the correction discourages ecologists from analyzing complicated communities in detail, because in simpler analyses (fewer tests) the correction is less stringent (Moran 2003).

Results

Variable reduction

Because initial NMS analyses with all 288 quadrats indicated separation of the sites based on all three data matrices (plant species, spider species, and habitat architecture), but no separation based on sampling month or year, I analyzed the data within sites (see Appendix B for alternate method of clustering quadrats by site). The spider and plant data were reduced to three dimensions (axes) and the architecture data to two dimensions (axes) in all three sites. I used Kendall's τ to interpret the NMS axes; original variables that were correlated with a given axis at $|\tau| \geq 0.30$ are included in the tables below. In total, 23 spider species were included in the NMS analyses for the sloped field, and 18 spider species were included for those in the other two sites. Analyses of the plant species data included totals of 38, 33, and 32 species in the sloped field, tussock grass field, and deciduous woodland, respectively. Eleven architecture variables were included for the sloped field and tussock grass field; the deciduous woodland included two additional litter variables. Correlations between the spider (S1, S2 and S3), plant (P1, P2 and P3) and architecture (A1 and A2) axes themselves did not exceed $|\tau| = 0.30$. All final configurations met the criterion of instability < 0.001 (instability measures changes in stress) with the exception of the plant ordination in the Tussock Grass Field (instability < 0.004). The spider ordination axes explained the least

48

variation among sample units (cumulative variation explained ranged from 39% to 49%), while the plant axes explained between 68% to 84% and the two habitat axes in all three sites explained 95% of the variation.   Following a similar pattern, stress (Kruskal's stress Formula 1 x 100) was greatest in the spider ordinations (24.33 [SF], 22.74 [TGF] and 21.36 [DW]) followed by the plant ordinations (16.57 [SF], 15.69 [TGF] and 19.22 [DW]) and the habitat ordinations (10.97 [SF], 10.49 [TGF] and 8.76 [DW]). While some of these stress values are considered high (i.e., those approaching or exceeding 20 [McCune and Grace 2002]), stress values are known to be higher with larger sample sizes or higher species counts (Clarke 1993; McCune and Grace 2002).

Regression analyses

Components of general community structure, such as number of individuals, species richness or species diversity, were significantly predicted (α=0.05) by either habitat architecture or plant species composition in all three sites (Table 2.1). All three measures were significantly predicted in the sloped field; two were significantly predicted in the tussock grass field (richness and diversity) and in the deciduous woodland (number of individuals and richness). In five of the seven significant regressions, whole model *p*-values were less than 0.01.

Responses to the original plant and architecture variables are presented in Table 2.2 as positive or negative, based on the signs of the correlation coefficients and regression parameter estimates. All three variables in the sloped field responded in the same way (i.e., all positively or all negatively) to the five architecture variables, but responded differently to the plant variables, if at all. Richness and diversity in the

49

tussock grass field, but not number of individuals, responded in the same way to individual plant species (Table 2.2). Similarly, in the deciduous woodland, both number of individuals and richness showed the same responses to each plant species. Number of individuals per quadrat in the sloped field and deciduous woodland showed the strongest associations with the predictor variables; these models explained 19% and 22% of the variation in number of individuals, respectively (SF R^2=0.19, DW R^2=0.22). The other models of general community structure in these two sites explained less than 10% of the variation ($R^2 \leq 0.10$).

Spider species composition was also significantly predicted by either habitat architecture or plant species composition in all three sites (Table 2.3). Regressions for two of the spider axes in all three sites were significant at α=0.05; in the deciduous woodland the regression for the third axis was also significant. The pattern in the sloped field and tussock grass field was for one of the spider axes to be predicted by an architecture axis with low R^2 values (architecture explained 4-7% of the variation along these axes), and the other spider axis was predicted by one or more plant axes with higher R^2 values of 0.15 and 0.18. In the deciduous woodland, one spider axis (S1) was predicted by an architecture axis, and the other two were each predicted by a plant axis; however, R^2 values were comparable for all three regression models in this site (R^2 values ranged from 0.09-0.11).

Individual spider species' responses to the original plant and habitat architecture variables are presented in Tables 2.4-2.6, following the format of Table 2.2. In the sloped and tussock grass fields, most species responded positively to most of the architecture variables (e.g., the number of individuals of a given species tended to increase with

increasing vertical heterogeneity, vegetation density, and herb cover) (Tables 2.4 and

2.5), but in the deciduous woodland the single spider species associated with S1

responded negatively to increasing horizontal and vertical heterogeneity (Table 2.6).

Clubiona abboti responded in the same direction to shared architecture variables in both

the sloped field and tussock grass field (Tables 2.4 and 2.5). Unfortunately, this was the

only species that was correlated in more than one site with spider axes that were

predicted by the same architecture variables (plant species composition varied

considerably from site to site, so there was little overlap of those axes across sites).

Spider species responses to plant composition were more variable; positive and negative

responses to individual plant species were evenly split among the 14 spider species

associated with NMS axes across all three sites. Some spider species were correlated

with more than one axis in each site, and so appeared to be responding to both

architecture and plant species (*C. abboti* in both the sloped and tussock grass fields,

Pardosa milvina in the tussock grass field, and *Theridion frondeum* in the deciduous

woodland), while others were only correlated with axes predicted by plant species

composition (*Bathyphantes pallida* in all three sites, and *Lepthyphantes nebulosa* in the

deciduous woodland) (Tables 2.4-2.6).

Discussion

The regression models involving habitat architecture generally support results

from previous studies in spiders (e.g., Hatley and MacMahon 1980; Rypstra 1986;

Gunnarsson 1990; Halaj et al. 1998). Plant species composition, however, was a

significant predictor in all but one of the regression models, and was the sole predictor in

almost two-thirds of them. The stronger response of both spider communities as a whole

and individual spider species to plant species composition relative to habitat architecture

was contrary to expectations based on previous studies in both spiders and other animal

groups (e.g., MacArthur and MacArthur 1961; Riechert and Reeder 1970; M'Closkey and

Lajoie 1975; Brose 2003). In general plant species composition explained more variation

in spider community composition than did habitat architecture.

Plant species may reflect habitat architecture at a finer scale than the architectural

variables considered here; for example, leaf pubescence has been shown to influence

arthropod community composition (Gruner et al. 2005). Further, many attributes of plant

communities are not species-specific; physiognomic characteristics of plants may be

typical of a particular species but not unique to it. These features may be correlated with

species composition, making it difficult to distinguish between an animal's association

with specific plant species (or species assemblages) and the physical structure created by

the vegetation. Some support for this possibility is provided by the sloped field data,

where individuals of each species responded in the same direction to the three upright

herbaceous species (*Desmodium* sp., *Diodia virginiana*, and *Verbesina* sp.), but in

opposite directions to the two graminoid species (*Dactylis glomerata* and *Microstegium*

vimineum) (Table 2.4).

On the other hand, plant species may have an indirect effect on spider community

structure via their influence on prey abundance or diversity. Prey availability has been

shown to explain variation in spider community structure in some cases (e.g., Rypstra

1986, Halaj et al. 1998) but not others (e.g., Greenstone 1984). In general, however,

these studies have demonstrated that habitat architecture explains more variation in

distributions of spiders than prey availability does. de Souza and Martins (2004) found that spiders were more abundant on both natural and artificial inflorescences than on vegetative branches, suggesting that the architectural features of the flowering branches were at least as important as their potential to attract prey.

The significance of both the habitat architecture and plant composition axes in these analyses underscores the complicated ways in which communities are organized, and the difficulties of teasing apart the relative contributions of different environmental factors. Previous studies have shown differing effects of plant community attributes on animal community structure, perhaps in part because the variables included (such as plant diversity) did not capture the complicated nature of community dynamics. In earlier analyses of the data presented here, plant diversity (Simpson's D) was never a significant predictor of any aspect of the spider communities. Unless one is interested in diversity *per se*, richness and diversity measures may not adequately represent species composition and the information contained therein (Rotenberry 1985). Wimp et al. (2005) found, for example, that while arthropod richness and abundance were very similar across genotype treatments, the actual species composition of the arthropod community was significantly different. Similarly, Cramer and Willig (2005) found that a measure of rodent β diversity based on the relative abundance of each species distinguished between habitat types, while one based on species presence/absence did not. In both cases, the inclusion of detailed information about the species provided a more complete picture of these communities.

The use of NMS axes in regression models allows incorporation of this detailed information about species composition. Ecologists increasingly use NMS as a tool for

descriptive multivariate data analysis, and the principles and mechanics have been well documented (e.g., McCune and Grace 2002). However, several points bear emphasizing here, in particular issues of sparse data matrices, choice of distance measures, the stress metric, and the nature of NMS axes, as well as variant forms of the procedure.

Sparse data matrices. NMS is well suited to community data, particularly when β diversity is high (i.e., the data matrix contains many zeroes) (Faith et al. 1987) and provides robust analysis of many data types. In analyses of simulated data with known gradients, NMS has shown superior ability to recover underlying data structure compared to principal components analysis, principal coordinates analysis, reciprocal averaging, and detrended correspondence analysis (Fasham 1977; Minchin 1987).

Choice of distance measure. Though the use of an appropriate distance measure is critical for any ordination analysis, in general NMS should not be sensitive to the choice of distance measure, since it uses ranked rather than proportional distances (as long as the distance measure used is appropriate for the type of data). Both metric and nonmetric measures can be used (Fasham 1977). Because NMS assumes a monotonic relationship between the distance measure and the ecological distances, however, any measure in which this assumption does not hold would be inappropriate (the chi-square measure, for example, violates this assumption [Faith et al. 1987, Legendre and Gallagher 2001]).

Stress. This assumption of monotonicity is evaluated in NMS by the stress metric (a loss function). As mentioned above, stress values exceeding 20 may indicate poor recovery of the underlying gradients (McCune and Grace 2002), but Clarke (1993) suggested that rules-of-thumb regarding stress values are overly simplistic. For example,

stress increases with increasing sample sizes; with a data matrix of 100 or more sample units, representing their inter-relationships in low-dimensional space becomes challenging (Clarke 1993). A potential solution to this problem would be to perform NMS on grouped subsets of the data. Stress values close to zero, on the other hand, should be viewed with suspicion, as this can indicate a degenerate matrix resulting in a configuration where sample units are tightly clustered in a few groups (Davison 1983, Cox and Cox 2001).

NMS axes. Axes in NMS behave differently than in other ordination techniques. Because NMS attempts to minimize stress across dimensions when determining the final configuration, the first axis of a solution in two dimensions will not be the same as the first axis in a three-dimensional solution (McCune and Grace 2002), so the choice of number of dimensions affects the interpretation of the scores for each axis.

Global versus local NMS. In global NMS all distances between points are evaluated simultaneously, and the final configuration attempts to maximize the rank order agreement between all points (Minchin 1987). In local NMS, this criterion is more relaxed; the rank order agreement is assessed for sample points in pairs, allowing for differences in the rate of change in compositional dissimilarity with environmental distance throughout the ordination space (Prentice 1977, Minchin 1987). Because of this flexibility, this form of NMS might yield lower stress values. Although both approaches have been evaluated with respect to other ordination techniques (Fasham 1977, global; Minchin 1987, local), the two approaches have not been systematically evaluated with respect to each other.

A number of ordination techniques are available in which environmental variables are incorporated along with species data (i.e., dependent variables), among them canonical correspondence analysis (CCA), biplots, canonical correlation, redundancy analysis, and co-inertia analysis. The latter three, however, are sensitive to nonlinearity in the underlying data structure (McCune and Grace 2002), and therefore may be inappropriate for community (i.e., species) data. CCA is a constrained ordination method that relates environmental variables to species data, but it ignores any variation in the species data that is not explained by the measured variables (McCune and Grace 2002); in other words, the results do not indicate what amount of variation in community structure is unexplained. This may result in misleading conclusions about the importance of the chosen variables (McCune 1997). McCune (1997) found that CCA can be very sensitive to noisy data, and that inclusion of variables that are biologically unimportant to the community can distort the representation of community gradients. Both McCune (1997) and McCune and Grace (2002) recommend relating environmental variables to species ordinations subsequent to the application of unconstrained ordination techniques such as NMS.

Unlike CCA, the species ordination in biplots represents unconstrained community structure. Biplots are a descriptive method that can be used to visualize multivariate data and detect patterns (Gower and Hand 1996) by superimposing trajectories of environmental variables on the species ordination. Classical biplots, however, assume a linear model for the data structure (Cox and Cox 2001). While nonlinear biplot methods have been proposed (e.g., Gower and Ngouenet 2005), there still exists a practical problem in the visual interpretation of biplots in more than two

dimensions (Jongman et al. 1995) or with large numbers of environmental variables overlaid on an ordination graph. Finally, both CCA and biplots incorporate a single matrix of environmental variables. While one could conceivably combine multiple data matrices, this would make it impossible to assess community responses to different types of environmental variables separately (for example, plant species composition versus habitat architecture). Ordinating data matrices of like variables individually allows interpretation of the separate influences of each type of variable.

In the approach I present here, the NMS axes representing spider species composition are secondarily related to NMS axes for environmental variables through multiple regression analyses. This approach has several advantages for community analyses over the methods discussed above. It provides the opportunity for statistical hypothesis testing of complicated data sets by incorporating as much information from the raw data as possible while producing statistically tractable datasets for regression analyses, rather than simply being descriptive. Because the assumptions in NMS about the underlying structure of the data (e.g., response shapes) are less restrictive, it is more likely to produce ecologically meaningful variables for further analyses. The regression models for the data in this study explain relatively low amounts of variation. This is likely because other factors that influence spider communities, such as abiotic factors or predation by vertebrates or other arthropods, were not incorporated. The low R^2 values indicate a problem with the data, not with the approach. The high stress values and comparatively lower variation explained in the spider ordinations suggest that the final configuration was not capturing the underlying data structure as effectively for this matrix, resulting in noise in the variables. In spite of the noisiness of the data, this

approach allowed an interpretable analysis of complicated data sets in which the data themselves determined which of the original variables were important in these communities. Being able to explain at least some of the variation in these spider communities allows a broader understanding of the roles of architecture and plant species composition. A final advantage is that by ordinating sample units based on actual species composition, this approach implicitly incorporates species associations in the regressions, thereby allowing individual species responses to be interpreted within the context of community interactions.

Animal communities are structured by a complicated set of interacting factors. Using an unconstrained, data-driven approach, this study has shown how multiple components of both habitat architecture and plant species are related to spider community structure, and elucidated the different responses of individual spider species. Understanding the simultaneous responses of multiple species to environmental variables constitutes an important step in community-level research. This study contributes not only to our knowledge of the community dynamics in an important group of terrestrial predators, but also to our ability to incorporate complicated community data in ecological research.

CHAPTER 3

Indirect and direct effects on habitat selection in spider communities: the roles of
arthropods, vegetation architecture and plant species

Introduction

Interactions among heterospecific individuals in a community occur within and

across trophic levels, and are often indirect; that is, they are mediated through

interactions with other species (Wootton 1994a). This combination of direct and indirect

effects leads to complicated dynamics among community members. Community

structure and organization are determined, in part, by both environmental features such as

microclimate or habitat architecture, as well as by ecological interactions between a focal

group and more distantly related species (e.g., predation, competition and mutualism)

(Landres and MacMahon 1983, Wellborn et al. 1996), but these factors do not act in

isolation from one another. One difficulty facing community ecologists has been the

dearth of analytical tools to develop and test multivariate hypotheses that both

incorporate direct and indirect effects and evaluate the simultaneous influences and

responses of community variables (Grace and Pugesek 1998, Malaeb et al. 2000,

Arhonditsis et al. 2006, Grace in prep). Even if multiple factors are included in

traditional analyses, treating them as independent predictors precludes an understanding

of how these factors may interact, and the roles of direct and indirect effects. Recently in

the ecological literature there has been a call for the use of multivariate models or

hypotheses as opposed to the more traditional application of multiple univariate models

to community ecology problems (Shipley 2000, Graham 2003, Arhonditsis et al. 2006,

Grace in prep). In this study I employ path analysis in a structural equation modeling

framework to investigate the structure of a spider community as it is influenced by the complicated and interacting roles of the non-spider arthropod community, the plant community and its architectural attributes, vegetation biomass and temperature.

At the community level, individuals interact both directly with members of other species within and across trophic levels (direct effects), and indirectly when interactions are mediated through one or more additional species (indirect effects) (Wootton 1994a, 1994b). Although the significance of indirect effects in structuring ecological communities is not well understood (Halaj and Wise 2001), indirect effects have been shown to have important influences on community attributes and processes in a variety of taxa. For example, indirect effects influence tenebrionid beetle abundance (Sánchez-Piñero and Polis 2000), genetic variation in predatory ladybird beetles (Astles et al. 2005), coexistence of aphids (van Veen et al. 2005), succession (Benedetti-Cecchi 2000) and species interactions (Wootton 1994) in intertidal assemblages, reproduction success in ducks (Blums et al. 2002) and plants (Vazquez and Simberloff 2004), and species diversity in butterflies (Hawkins and Porter 2003), bats (Stevens 2004), spiders (Halaj et al. 2000), and plants (e.g., Schmitz 2006, Schmitz et al. 2006). Given their ubiquity in ecological systems, ignoring indirect effects may result in inaccurate characterizations of communities and misleading conclusions about the nature of species interactions within them (Sih et al. 1985, Wootton 1994a, Grace in prep).

Local distributions of species within a community can arise from the direct effects of the physical structure of the environment, resource availability, and species interactions as well as the indirect effects of any of these factors. For example, in animal communities, vegetation architecture may simultaneously have a direct influence on a

focal taxon (McCoy and Bell 1991) and an indirect influence through its effects on prey availability by providing refuges or otherwise altering predator foraging efficiency (e.g., Gilinsky 1984, Schmitz et al. 2006). In some predatory groups, structurally complex vegetation may reduce the effects of intraguild predation by providing refuges for predators, thus having an indirect negative effect on prey species (Finke and Denno 2002).   In some cases, the directions of direct and indirect effects may differ; Malaeb et al. (2000) found a negative direct effect between two variables, but the indirect effect was positive and stronger than the direct effect, resulting in a net effect that was positive. This example underscores the need to incorporate the complicated dynamics of a system or pathway in community analyses.

As in other animal communities, local distributions of spider species are determined by a number of factors, including habitat architecture, microclimate, prey availability, and predation risks (reviewed in Riechert and Gillespie 1986 and in Wise 1993). Spiders' foraging modes are tightly linked to habitat architecture, in part because their various foraging strategies require different substrate configurations and different spiders may show preferences for more open or more dense vegetation depending on web type and/or hunting mode (reviewed in Uetz 1991). Temperature affects spider activity levels (e.g., Schmitz et al. 1997), and can vary at the microhabitat scale (i.e., from sampling unit to sampling unit). Spiders tend to be food-limited (Wise 1993), and prey availability has been shown to influence spider distributions (Olive 1982, Caraco and Gillespie 1986, Real and Caraco 1986, Gillespie and Caraco 1987). Finally, spider distributions may be influenced by avoidance of non-spider natural enemies (Wise 1993).

These factors likely interact, and affect species composition both directly and indirectly. Spider densities can be influenced by a combination of habitat complexity and prey availability (Rypstra 1983), and prey availability itself can vary with habitat structure. Susceptibility of spiders to bird predation has been shown to depend on architectural configuration of the vegetation as well (Gunnarsson 1990, 1996). Harwood et al. (2003) have suggested that web-building spiders locate their webs based on a compromise between habitat structure, prey availability, microclimate and predator avoidance and this compromise likely applies to active hunting spiders and ambushers as well. The interactions of these factors may explain why, for example, many studies find prey availability to be less important than other variables; spiders may be required to choose habitats that minimize other risks at the expense of high prey availability. Additionally, spider associations with habitat features such as plant species composition or habitat architecture may be confounded by similar associations between these factors and prey availability (Beals 2006). Associations between spiders and habitat features such as plant species composition may be indirect, mediated by direct associations with prey availability. Most studies that have examined indirect effects on spider communities have looked at how spiders influence the plant community through top-down indirect effects (i.e., trophic cascades) (e.g., the work of Schmitz and colleagues, Rosenheim et al. 2004, see also meta-analysis by Halaj and Wise 2001, but cf Halaj et al. 2000).

The current study expands on my previous work (Beals 2006) in which I found that plant species composition predicted spider community structure as well as or better than habitat architecture. I hypothesized that plant species might indirectly affect spiders through their influence on the presence of non-spider arthropods, through shared

association with microclimate factors, or that they might be confounded with architecture. To investigate the interacting roles of non-spider arthropod community composition, plant community composition, habitat architecture, vegetation biomass, and microclimate (temperature) with respect to spider community composition, I develop directed graphical models (sensu Shipley 2000) to test causal hypotheses about the structure of an old field spider community. Models specifying asymmetrical influences (i.e., causal relationships) between variables cannot prove cause, of course, but can be used to falsify causal hypotheses (Grace and Pugesek 1998) or determine whether causal inferences are consistent with the structure of the data (Bollen 1989, Shipley 2000).

Methods

Study area

The study was conducted in an old-field site in Knox County, Tennessee, USA. This site was described as the "sloped field" site in Beals (2006). The composition of the dominant plants was similar to that described in Beals (2006) (the grasses *Dactylis glomerata* and *Microstegium vimineum*, and the perennial herbs *Desmodium* sp., *Calystegia sepium*, *Verbesina* sp., *Vernonia gigantea*, and *Lysimachia nummularia*); additionally, *Glechoma hederacea*, *Solanum* sp., and *Stellaria media* were abundant during this study. While most of these species were distributed fairly evenly throughout the site, *C. sepium* was primarily found in the upper half of the site, while *G. hederacea* was limited to the lower half. Sampling was carried out during October 2004. Scientific names of plants are taken from Wofford (1989).

Data collection

Of the plant species in this site, *Desmodium* sp. was previously shown to have the strongest influence on spider community composition (Beals 2006). Quadrats (0.1m^2) were selected throughout the site based on the criterion of having ≥75% cover of *Desmodium* sp. To separate the effects of plant architecture from other plant qualities (e.g., palatability to herbivores), I manipulated *Desmodium* plants in 50 of the quadrats by removing every other leaf at its node. The remaining 50 quadrats were unmanipulated and served as controls. *Desmodium* senesced prematurely in some quadrats during the course of sampling, and it was necessary to exclud these quadrats from the analyses. As a result, a total of 54 quadrats (32 control and 22 treatment) were included in the final analyses

At least one day prior to sampling within a given quadrat, plant species composition and vegetation architecture data were collected. I recorded plant species composition in 5% increments for all species with cover ≥5%; species with less than 5% cover were counted for richness information. I measured vegetation architecture following the protocol described in Beals (2006). Variables measured included an index of vegetation density, the height of the tallest plant in each compass quarter of the quadrat, and a measure of open space within the vegetation of a quadrat based on the point-intercept method. I calculated the coefficient of variation of the plant heights to use as a measure of horizontal heterogeneity independent of mean height; quadrats with relatively uniform heights of the tallest plants have a lower coefficient of variation than those in which plant heights differed greatly. The open space data were reduced to two variables using nonmetric multidimensional scaling, which represented a measure of

vertical heterogeneity. Quadrats with similar configurations of open space had similar values for these variables.

Sampling was completed between 1000 and 1500 hrs EDST on days in which no measurable precipitation had occurred during the previous six hours. On the day of sampling I measured temperature and relative humidity in the center of each quadrat (near the ground) 10 minutes prior to collecting the arthropods. Following sampling, all vegetation was removed from the quadrat and taken back to the lab, where it was dried at 60º C for 48 hours and then weighed to measure aboveground dry weight biomass. Four missing biomass data points were replaced with the mean value of biomass (after Malaeb et al. 2000, Arhonditsis et al 2006).

Arthropod sampling

To determine the influence of potential prey, competitors, and predators/parasites on the spider community, I sampled spider and non-spider arthropods initially using a vacuum sampler to ensure capture of flying insects. Following the vacuum sampling, a $0.1m^2$ cylinder (open at both ends) was quickly placed around the quadrat and pushed into the ground to prevent escape of the remaining arthropods. I sampled these with a handheld aspirator. Arthropods were stored in Whirl-Pak bags in a cooler until return to the lab. I then placed them in a –20 º C freezer and later transferred them into vials containing 70% ethanol. The large numbers of Collembola collected with the vacuum sampler required that this group be sub-sampled. To remove soil particles and chaff from the samples, I floated the Collembola in a 1:1 (v:v) sucrose solution, removed them from the surface, rinsed them, and placed them back in ethanol. I then sub-sampled from a

65

gridded petri dish by randomly selecting 10 squares out of 68 (approximately 9 mm^2 each) and counting the number of Collembola in each square. Although I did not assess the possible bias introduced by this sub-sampling method, the bias should be comparable for any given sample and therefore should not influence subsequent analyses.

All non-spider arthropods were identified to order, with the exception of the Chilopoda (centipedes) and Diplopoda (millipedes), which were identified to class. The total numbers of individuals in each order (or class) were used in the primary analyses. For analyses of the separate influences of potential prey arthropods (herbivores and scavengers) and potential competitor or predator arthropods (predators and parasites), orders in which both trophic levels were represented were identified to family, and classified based on the predominant feeding mode of that family. Total numbers of herbivore/scavenger and predator/parasite individuals were tallied for each order (e.g., Coleopteran herbivores versus Coleopteran predators). Spiders were identified to species where possible; juveniles that could not be identified to species were identified to genus.

Variable reduction

The spider, arthropod, plant, and habitat architecture matrices contained large numbers of often collinear variables, and the species data matrices contained many zeros. To reduce the number of variables in each matrix I used nonmetric multidimensional scaling (NMS) (PC-ORD 4, 1999, MjM Software Design, Gleneden Beach, Oregon), which represents the data in reduced dimensions based on the underlying variables that describe the variation among sample units. In other words, the variable reduction is based on the information contained in the raw data rather than potentially subjective

66

decisions of the researcher (see Beals 2006). Prior to the NMS analyses, spider species

occurring in fewer than 5% of the quadrats were deleted, along with arthropod orders and

plant species occurring in only a single quadrat. Two different criteria were used because

there were many more spider species than either plant species or arthropod orders, and

higher numbers of entities/species affects the ability of NMS to accurately represent the

data in lower dimensions. It was therefore necessary to use a more stringent standard to

achieve reasonable ordinations of the spiders. For the spider, arthropod, and habitat

architecture matrices individual species or variables were relativized with respect to their

maximum value to account for the fact that species abundances varied widely, and the

architecture variables were not measured on the same scale. Plant percent cover was

arcsine square root transformed (McCune and Grace 2002). The Sørenson distance

measure was used for all NMS ordinations, and final configurations were selected using

PC-ORD's autopilot mode set on "medium", which performs 15 runs with the raw data

and 30 runs with randomized data using a random starting configuration each time to

determine the best configuration.

Data analysis

The resulting spider, arthropod, plant and architecture axes, along with

temperature and biomass data, were *z*-transformed so each variable had a mean of zero

and variance of one (recommended by Grace in prep). A nominal variable (C/T) was

included in analyses to represent the control and treatment quadrats. To investigate both

the direct and indirect influences of the arthropod and plant communities, vegetation

architecture, biomass, and temperature on spider species composition I used the structural

equation modeling (SEM) software program AMOS (Arbuckle 1999). SEM analyses are

an extension of path analysis (Wootton 2002, McCune and Grace 2002, Tomer 2003) that

typically include latent, or unmeasured variables, which are indicated by the measured

observed variables (Shipley 2000, Grace 2003, Tomer and Pugesek 2003). In its broadest

definition, however, it tests hypotheses about the direct and indirect relationships

between independent and dependent variables that can be either latent or observed

(Shipley 2000, Hershberger et al. 2003). Latent variables are theoretical constructs that

cannot be measured directly but can be represented by one or more indicator variables

(Hershberger et al. 2003, Pugesek 2003a). Body size, for example, is a latent variable

that can be represented by measuring an organism's length, weight, and circumference.

These measured variables are indicators of the abstract concept of "body size."

Because my variables were primarily NMS axes, I employed observed variable

SEM or "modern path analysis", *sensu* Grace (in prep), which was necessitated by two

issues, one conceptual and one technical. First, the NMS axes themselves are abstract

constructs derived from the observed data matrices; for example, the spider axes

represent the concept of the spider community, but are not observed variables. Second,

and more importantly, the indicators of a latent variable are required to covary in some

way or they cannot be considered as joint indicators of a single latent variable. While

NMS analyses do not necessarily produce strictly orthogonal axes (McCune and Grace

2002), for a given data matrix here they are insufficiently correlated with one another to

be used as indicators of a latent variable.

Grace's use of the terminology "modern path analysis" and "observed variable

SEM" serves to distinguish between this approach and path analysis as traditionally

implemented by ecologists. Traditional path analysis either uses the correlation matrix among variables to test models with direct and indirect effects (e.g., Wrightian path analysis) (Pugesek 2003a, Grace in prep), or uses a series of regression equations to estimate direct path coefficients between variables, using the partial regression coefficients to estimate indirect effects (McCune and Grace 2002; see, for example Wootton 1994b). In the SEM framework, the path coefficients are estimated simultaneously, with the goal of optimizing the fit between the observed covariance matrix and the covariance matrix implied by the path model being tested (Pugesek 2003, Grace in prep). In other words, in addition to simultaneously testing the significance of individual path coefficients, the power of SEM methods stems in part from the inherent capability to test whether the model, in its entirety, provides a reasonable fit to the data (Pugesek 2003a, Grace in prep).

Three general approaches to SEM have been discussed: confirmatory, alternate model testing, and exploratory or model generating (McCune and Grace 2002, Hershberger et al. 2003, Arhonditsis et al. 2006). In a confirmatory approach, a single theoretical model, specified *a priori* on the basis of existing theory or prior knowledge, is tested against the data and either rejected or accepted based on absolute measures of fit. Alternate model testing extends this approach to the testing of two or more models (again, specified *a priori* on the basis of theory or previous results). In this approach, competing models may be rejected or accepted either on the basis of absolute measures of fit or relative measures of fit. Relative fit measures select the best model in terms of general fit and parsimony regardless of whether the absolute measures indicate adequate fit. Thus, even if none of the alternate models fit well, one can still select the "best"

69

model from this group. Finally, in the exploratory or model generating approach a researcher develops one or more potential models based on the patterns in the data themselves. This can be done, for example, by specifying a model based on correlations among the variables. More commonly, models are generated following the rejection of an *a priori* model by modifying the path model based on theory, modification indices, and/or significance of individual path coefficients. Because a generated or modified model is tested against the data used to develop it, the final model is generally considered a new hypothesis requiring further testing (McCune and Grace 2002).

I used two of these SEM approaches to assess the relationships between biomass, temperature, and the arthropod, plant and spider communities. I began by testing two alternative models of the data specified *a priori* (illustrated in Figure 3.1). These competing models were developed based on my previous work (Beals 2006), which indicated that plant species composition predicted spider species composition and community structure as well as or better than did vegetation architecture. The first hypothesized model (Figure 3.1A) posits an indirect influence of plants through their effects on the arthropod community, vegetation architecture, and vegetation biomass. The second model (Figure 3.1B) additionally includes a direct effect of plants on spiders. I used a suite of commonly reported fit measures to assess both the absolute and relative fits of these models (global model chi-square, CMIN/DF, CFI, and RMSEA for absolute fit, AIC and CAIC for relative fit) (Tomer and Pugesek 2003). The chi-square test for the overall model will be rejected when the fit of the model is poor, so *p*-values greater than 0.05 indicate adequate fit between the model and the data. CMIN/DF is the chi-square statistic scaled by degrees of freedom, with values less than 2 indicating good fit. CFI is

70

the comparative fit index. Unlike the other absolute fit measures described here, CFI compares the fit of the specified model to the independence model, in which there are no relationships (paths) between the variables, rather than the saturated model in which all variables are related through direct paths. CFI values greater than 0.90 are generally considered an indication of adequate fit. RMSEA is the root-mean-square error of approximation; values less than or equal to 0.08 indicate adequate fit, with values less than 0.05 suggesting very good fit. Each of the absolute fit measures has strengths and weaknesses. Tomer and Pugesek (2003) recommend always reporting the global model chi-square, and CMIN/DF accounts for lack of parsimony based on the ratio of this chi-square value to degrees of freedom. The global chi-square, however, is conservative with small sample sizes, making a rejection of the hypothesis that the model fits less likely. CFI behaves well with small sample sizes, but because it compares the specified model to the perhaps unrealistic independence model it tends to overestimate the goodness-of-fit of the specified model. RMSEA performs less well with small sample sizes, but because it has a known sampling distribution one can test whether a value greater than the standard cut-off is significantly large.

The relative fit measures were AIC (Akaike Information Criterion) and CAIC (Consistent Akaike Information Criterion). These measures incorporate both the fit of the model to the data and a penalty term that increases as model complexity increases, so that models with lower values for these measures are considered better. AIC is familiar to ecologists, but the criterion CAIC, developed by Bozdogan (citation in Bozdogan 1986), is rarely used. CAIC extends AIC in such a way as to provide a more consistent penalty for over-fitting a model that takes sample size into account, always choosing the

simplest of the "true" models (Bozdogan 1986, 1987). For all models, path coefficients were estimated using maximum likelihood.

Following testing of the *a priori* models, I used an exploratory approach to refine and simplify them. I was specifically interested in what factors either directly or indirectly influenced the spiders, and whether there was an effect of the manipulation of *Desmodium* in treatment quadrats. Paths were deleted from the models one at a time based on the highest p-value (Hershberger et al. 2003); after each deletion I re-ran the model and checked the fit indices and modification indices (MIs; AMOS calculates these as paths that are absent from the model that would significantly improve the fit if they were included). Models were pruned and un-pruned in this way until the measures of absolute fit indicated poor fit after the removal of a path, or until all paths were significant at $\alpha=0.05$. I then evaluated the models having good absolute fit (those with and without non-significant paths) using the relative fit measures to select the final models.

Finally, to assess the influence of non-spider arthropods on spider communities, I used NMS analyses to reduce the variables for a data matrix of herbivores and scavengers (potential spider prey) and a data matrix of predators/parasites (potential spider predators and/or competitors). The spider NMS axes were regressed on the NMS axes for both of these data matrices using SAS PROC REG (SAS 9.1, 2001, Sas Institute Inc., Cary, North Carolina). A nominal variable (C/T) that identified control and treatment quadrats, and the interaction terms between this variable and each NMS axis, were also included in the regressions. Predictor variables for the final models were selected using all possible subsets regression in SAS.

Results

Variable reduction

NMS ordination of these data resulted in three axes (dimensions) for each of the original data matrices (spider, arthropod, plant, habitat structure, herbivore/scavenger and predator/parasite). Details of the NMS analyses (stress, instability, and proportion variance explained by each axis) are given in Table 3.1. Although the stress values (a "badness of fit" measure) for the spider, arthropod and herbivore/scavenger data are somewhat high (>15), an over-reliance on stress has been cautioned against by Clarke (1993) and discussed in McCune and Grace (2002), since stress increases as number of sample units and/or number of entities (e.g., species) increases. Correlations between the NMS axes and the original variables are in Table 3.2. These can be used to interpret response directions among the original variables following analyses using the NMS axes (see Beals 2006). Axes are referred to below using S for spiders, Arth for arthropods, P for plants, and Struct for habitat structure. This notation is followed by a number indicating the axis being considered: For the species (or order) data, each axis represents a measure of community composition.

A priori path models

Both of the *a priori* models tested exhibited poor absolute fit. The relative measure (CAIC), however, indicated that the model without direct pathways between plants and spiders (Model A) fit better than the model that included direct plant-spider pathways (Model B) (Table 3.3). For all spider axes, CAIC selected the independence

73

model (with no paths) over the specified model, likely as a result of the poor absolute fit of the specified model. The lack of parsimony due to the inclusion of many hypothetical paths also contributed to the selection of the independence model over the specified model; the saturated model, which includes all possible paths, was always rejected by CAIC.

For comparison I also include AIC values in Table 3.3. Like CAIC, AIC is minimized for Model A for S1 and S2. For S3, AIC indicates equivalence between the two models, although it is minimally lower for Model B (i.e., the model with the additional plant-spider paths). In contrast to CAIC, when comparing the independence model with the specified and saturated models, AIC always rejects the independence model. These results illustrate the tendency of AIC to over-fit and suggest that CAIC is to be preferred, particularly in the case of complex models.

Reduced path models

All final path models for each spider axis exhibited adequate fit based on standard fit measures (Table 3.4). All global p-values were greater than 0.05, indicating that the covariance structure implied by the final models was not significantly different from the covariance structure of the data. All models had CMIN/DF values less than 2 and CFI values of 1.00 (CFI should be ≥ 0.90; 1.00 is the maximum), indicating good fit. The RMSEA values also indicated very good fit in all three models. Graphical representations of the path models indicate the significance of each path but not the sign of the path coefficient, since the direction of the effects depends on the sign of the correlation coefficients of the original variables (Figures 3.2-3.4). Covariances between

the errors for Struct1 and Struct2, Struct1 and Struct3, and Struct2 and Struct3 were included in the models to improve fit (based on the modification indices provided by AMOS), but are omitted from the figures for clarity.

Coefficients of multiple determination (R^2) indicate varying amounts of variation explained in the models (Table 3.5). The model for S3 explained the greatest variation in spider species composition (46%), followed by the model for S2 (31%). These two spider axes reflected two to three times the variation in species composition than did S1 (Table 3.1). In the general model (i.e., excluding the spider variables) the greatest variation explained was for Struct3 (83%), followed by Struct2 (42%) and Arth2 (38%). Struct3 and Arth2 accounted for the greatest variation in the original data matrices in their respective NMS ordinations (Table 3.1). Biomass and Arth3 had the lowest R^2 values. It should be noted that these R^2 values are unadjusted, and therefore may be slightly inflated for variables with more predictors (R^2 naturally increases with increasing number of predictor variables even if these variables are not significant [Neter et al. 1996]).

Interpretation of these path models is facilitated by investigation of how the direct and indirect effects interact with one another. The total effect between two variables is calculated by adding the coefficient for the direct effect to the product of the two coefficients along an indirect or compound pathway (Grace in prep). The total effects for variables that only influence a variable of interest in one way or the other are the same as these component parts, but the total effects of variables that exert influence both directly and indirectly may be amplified or reduced compared to the direct and indirect effects. The total effects of each predictor variable in the general model are presented in Table

3.6. Control versus treatment (C/T) affected all three arthropod axes and Struct3, although the only direct effects were on Arth2 and Struct3 (Figures 3.2-3.4). Biomass, P1 and Struct3 each had effects on two to four variables, while P2 and P3 each affected six variables directly. P3 had indirect effects only on two additional variables (Arth1 and Arth3 through Struct3) (Figures 3.2-3.4). The strongest total effects in the general model were the effects of C/T on Arth2, P2 on Arth2 and Struct2, and P3 on Struct3 (Table 3.6).

All three spider axes were directly predicted by two to three variables (Figures 3.2-3.4). S1 responded directly to Arth1, Struct2 and temperature (Figure 3.2) and was indirectly influenced by P2 and Struct3 (through Arth1), by P2 and P3 (through Struct 2), and by P3 and biomass (through temperature). The effect on S1 of C/T was indirect through the compound pathway C/T → Struct3 → Arth1 → S1. S2 was directly predicted by two arthropod axes (Arth1 and Arth2), with indirect effects of P2 and Struct3 (through Arth1) and P2, P3 and C/T (through Arth2) (Figure 3.3). S3 was directly predicted by Arth1, P2 and C/T, with indirect effects of P2 and Struct3 through Arth1 (Figure 3.4).

The total, direct and indirect effects of predictor variables on each spider NMS axis are presented in Table 3.7. All three spider axes were influenced (either directly, indirectly, or in the case of S3 both) by C/T, Arth1, biomass, P1, P2, P3 and Struct3. S1 was additionally influenced by Struct2 and temp, and S2 was influenced by Arth2. The strongest effects on S1 were the direct effects of Arth1, Struct2 and temperature. The strongest effects on S2 were either direct (Arth1, Arth2) or indirect (C/T, P2). S3 was strongly influenced by C/T, Arth1 and P2, with C/T and P2 having both direct and

76

indirect effects with opposite signs. The indirect effects were small enough, however, that they did not reduce the direct effects by much.

Responses of individual spider species

Responses of individual spider species associated with S1 (i.e., those with correlations where $\tau \geq |0.30|$) to the original variables associated with NMS axes, as well as biomass and temperature, are shown in Table 3.8 as positive or negative. The direction of response is based on the signs of the correlations of a given spider species and original predictor variable translated through the sign of the path coefficient (see Beals 2006). For example, if the correlation of the response variable (spider species) is positive and that of the predictor variable negative, a positive path coefficient indicates that the spider responds negatively to the predictor, while a negative path coefficient indicates a positive response. The effect of the variable C/T was indirect and cannot be interpreted for individual spider species; therefore it is omitted from this table.

For S1, the hunting spider species (*Clubiona abbotti*) responded oppositely to the three web building species (*Eperigone trilobata*, *Florinda coccinea* and *Theridion frondeum*) (Table 3.8). In terms of the arthropods, *C. abbotti* responded positively to Orthoptera (Family Gryllidae, Subfamily Nemobiinae [ground crickets] with a single exception), while the web building species responded positively to Diplura and Opiliones. *Clubiona abbotti* responded negatively to biomass and temperature, and positively to horizontal heterogeneity, vine cover, density, grass cover and glabrous cover. Web builders associated with this axis responded in the opposite direction to these variables.

Responses to plant species were more difficult to interpret, as these sometimes conflicted in different plant dimensions (Table 3.8). For example, *C. abbotti* responded positively to *Desmodium* with P2, but negatively with P3; this conflict was found in responses to *Toxicodendron radicans* and *Calystegia sepium* (P1 and P2) and *Dactylis glomerata* (P1, P2 and P3) as well. (Similar conflicts occurred for all spider axes and are discussed below.) The total effects of plant species NMS axes, however, were fairly weak for this spider axis (Table 3.7).

The two spiders associated with S2 (the hunting spider *Pisaurina undulata* and the web builder *Theridion cheimatos*) also responded in opposite directions to one another (Table 3.9). *Pisaurina undulata*, like *C. abbotti* on S1, responded positively to Orthoptera, as well as Isopoda, and negatively to Diplura, Opiliones, Diptera, Hymenoptera, Hemiptera and Acari. Unlike *C. abbotti*, this hunting spider responded positively to biomass and showed no response to temperature. *Theridion cheimatos* responded in the opposite direction to these variables. For S2, the total effect of P2 was strong relative to the other plant axes (Table 3.7). *Pisaurina undulata* responded negatively to *Desmodium* sp. and *C. sepium*, and positively to the remaining plant species associated with P2, while *T. cheimatos* responded positively to *Desmodium* sp. and *C. sepium* (Table 3.9). As in Table 3.8, the C/T variable is omitted here.

Spiders associated with S3 were directly affected by the *Desmodium* treatment (C/T); the web builder *Bathyphantes pallida* was relatively less abundant in treatment quadrats, while the web builder *Eridantes erigonoides* and the hunting spider *Oxyopes* sp. were relatively more abundant (Table 3.10). Here, the two web builders responded in opposite directions to one another, and one (*E. erigonoides*) responded in the same

78

direction as the hunting spider. Like the hunting spiders associated with S1 and S2 (Table 3.8 and 3.9), *Oxyopes* sp. responded positively to Orthoptera and negatively to Diplura and Opiliones. *Oxyopes* sp. responded positively to biomass (like *P. undulata*) and to density, grass cover and glabrous cover (like both *C. abbotti* and *P. undulata*). The two web builders associated with this axis belong to the same family (Linyphiidae) and responded oppositely to one another; *B. pallida* tended to respond in the same direction as the two Linyphiids associated with S1, with the exception of its response to biomass.

Regression models

Regression models of the influence of herbivore/scavenger and predator/parasite arthropods on spiders were non-significant for S1 but highly significant for S2 and S3 ($p<0.0001$ for both) (Table 3.15). S2 was strongly predicted by Herb2 ($p<0.0001$) and C/T (control/treatment; $p=0.009$) but there was no interaction between the two. The effect (slope) of Herb2 was the same in both control and treatment plots, but C/T was significant because the intercept in the control quadrats was lower (Figure 3.8). The model for S3 showed significant main effects for Herb2 and Herb3, but there were also interactions between C/T and Herb2 and C/T and Pred3. In both cases, these NMS axes were only significant in treatment quadrats, and did not affect S3 in control quadrats. Herb3 had a significant effect in both groups of quadrats. Spider species responses to the original variables associated with Herb2, Herb3 and Pred3 are shown in Table 3.16.

A priori path models

My previous work (Beals 2006) indicated a strong role for plant species

composition, and in particular *Desmodium*, in influencing spider communities. I

hypothesized that this relationship might be due to indirect effects of the plants through

their direct effects on either arthropods or vegetation structure. Testing of the *a priori*

models, which either omitted or included direct paths between plant species and spider

species, indicated that in general the model without direct plant pathways fit the data

relatively better (or in one case equivalently), providing support for my initial hypothesis.

Reduced path models

The subsequent generation of reduced models with improved fit provided a more

detailed understanding of spider community structure. In the models for S1 and S2

(Figures 3.2 and 3.3) the direct relationship between plant species and spider species

disappeared when other variables were taken into account. Both the plant NMS axes and

Desmodium manipulation (represented by C/T) acted indirectly on S1 and S2 through

their effects on both arthropod composition and habitat architecture. Habitat architecture

(and plant species composition) additionally had an indirect effect through its effects on

the arthropods. These results provide further support for my initial hypothesis.

The reduced path model for S3, like those for S1 and S2, included a direct effect

of the arthropods and indirect effects of plants and architecture through Arth1 (Figure

3.4). Additionally in this model, *Desmodium* manipulation had a strong direct effect on

spider species composition. That changes in *Desmodium* architecture (and consequent

changes in biomass and cover) directly influenced spider species composition also provides support for my hypothesis that the finer-scale architecture of plants may help explain the influence of plant species composition on spiders found in Beals 2006. Overall plant species composition (P2), however, also had a direct effect on this spider axis. This effect may potentially be explained by finer-scale architecture of the plant species as well; P2 was associated with *Desmodium* (i.e., the architecturally manipulated species) and even more strongly with *Calystegia sepium* (bindweed), whose twining growth form changes the architecture of other plant species (pers. obs.).

The direct effect of habitat architecture on S1 is congruent with previous research. Vegetation growth form (e.g., grasses or herbs) has been shown to influence the types of prey captured by webs built on one type of growth form or the other (McReynolds 2000). In this study, grass cover, vine cover and cover of plants with glabrous leaves influenced the spider axes (Tables 3.8-3.10). Horizontal heterogeneity predicted one spider axis (S1), and density influenced all three axes. More complex and/or dense vegetation has been observed to support greater numbers of spiders (Rypstra 1986; Gunnarsson 1990; Halaj et al. 1998) or spider species (Hatley and MacMahon 1980) as well as other arthropod groups (reviewed in Langellotto and Denno 2004). For S1, the effect of habitat architecture was somewhat stronger than that of the arthropods (Table 3.7). Previous work in spiders has generally shown that vegetation architecture is a more important determinant of spider communities than prey availability (e.g., Greenstone 1984, de Souza and Martins 2004). Even in studies that have found a significant influence of prey availability, vegetation architecture has explained more of the variation in spider communities (e.g., Rypstra 1986, Halaj et al. 1998).

S1, however, explained the least variation in spider species composition of the three spider NMS axes (Table 3.1). Neither S2 nor S3 were directly predicted by vegetation architecture axes; S2 was predicted solely by arthropod axes (Figure 3.3) and S3, in addition to P2 and C/T, was predicted by Arth1 (Figure 3.4). The apparent discrepancy between these results and those of previous studies may reflect different approaches to representing the spider and prey communities. Greenstone (1984) and Halaj et al. (1998), for example, used diversity or abundance of spiders and prey, while here I used taxonomic composition. For the data here, there were only weak correlations between number of arthropod individuals and spider individuals, and arthropod diversity and spider diversity.

The effects of *Desmodium* manipulation, then, appear to affect overall spider species composition (represented by the three NMS axes) in several ways, both indirect and direct. It may be that the manipulation of *Desmodium* and accompanying changes in overall vegetation structure changed the dynamics of spider interactions with their prey and their potential predators or competitors. Differences in the structural complexity of vegetation have been shown to affect the susceptibility of prey to their predators (Gilinsky 1984, Schmitz et al. 2006). Further, interactions between habitat architectural complexity and prey availability affect not only the relationship between predators and prey, but the also the relationship among predators themselves (e.g., intraguild predation) (Hughes and Grabowski 2006, Finke and Denno 2006). The path model for S3, however, suggests that the effects on arthropods and vegetation architecture of manipulated *Desmodium* were not sufficient to account for the influence of plant species composition on the spider community. As discussed above and

previously (Beals 2006), the direct importance of C/T and P2 may reflect the influence of the finer-scale architecture of individual plant species.

One issue here involves the necessarily *a priori* decisions about the direction of the arrows in the initial models that were used to develop the reduced models. There is an extensive body of literature on the top-down effects of spiders in food webs (see, for example, Schmitz 1998, Schmitz and Suttle 2001, etc.), and it could be argued that some of the arrows in these models might just as easily be specified in the opposite direction (e.g., spider → arthropod → biomass). To ascertain whether a bottom-up or top-down model fit the data here better, I attempted to compare a model with effects of spiders on arthropods (and arthropods on both plant composition and biomass), with the two specified *a priori* models (Model A and Model B [Figure 3.1]).

Reversing the direction of only these arrows, however, resulted in a non-recursive model. A non-recursive model is one in which a variable can have an indirect influence on itself; for example, the path spider → arthropod → plant → architecture → spider would be non-recursive. Here, non-recursiveness resulted in under-identified models, which require imposing additional constraints (e.g., freeing up degrees of freedom by deleting paths) in order to achieve stable parameter estimates (Hershberger et al. 2003). Because all paths in the initial model were theoretically justified, there was no justification for the removal of one over another. Further, removing any paths would have resulted in a model that was not strictly comparable to the initial bottom-up model. Although I was unable to compare the relative fits of bottom-up versus top-down models, Schmitz (pers. comm.) has suggested that in systems that are not dominated by leaf-chewing orthopterans or coleopterans, such as this one, the bottom-up direction of the

arrows in the path models here is appropriate. That distinct changes in spider responses were evidenced in the plots in which a plant species was manipulated provides additional support to the idea of bottom-up influences in this community (see also Halaj et al. 2000, Moran and Scheidler 2002).

Responses of individual spider species

Individual spider responses to original variables were more difficult to interpret than the overall path models, in part because the direction of influence of variables associated with predictor axes was not always the same within a model for a given spider axis. For example, *Desmodium* was positively correlated with P2 and negatively correlated with P3 (Table 3.2). Because both axes were included in each model, and the total effects of these two axes had the same sign, the interpretation of the response direction of the spider species was different, depending on the plant axis (Tables 3.8-3.10). This apparent contradiction is initially perplexing, but it is important to understand that each NMS axis represents a different dimension of the underlying (e.g., plant) data matrix. As with the direct and indirect effects, which can have opposite signs, it may be that spider responses to a given variable differ in different dimensions. These differing responses may be analogous to interaction effects, in that the suite of species (or other variables) defining variation in one dimension of a predictor axis may result in a different response to one of those species than when that species is defining variation with a different suite of species in another dimension. In other words, the response of the spider community to a variable on a given axis may depend on the other variables associated with that axis. The total effects of P2 and P3 were weak for S1, while for S2 and S3 the

magnitude of the total effect of P2 was greater than that of P3 (Table 3.7); these effects should perhaps be given more weight.

Patterns of responses of individual spider species within guilds or families varied across path models. The three hunting spider species (*C. abbotti*, *P. undulata* and *Oxyopes* sp.) did not always respond in the same direction. The web building spiders in the genus *Theridion* associated with S1 and S2 responded similarly to some predictors (Arth1, P1 and Struct3) but not others (biomass, P2 and P3) (Tables 3.8 and 3.9). Two of the web building species associated with S1 belonged to the same family (Linyphiidae) and responded in the same direction as one another, but did not always respond in the same direction as the two Linyphiids associated with S3, which always responded in opposite directions to one another (Tables 3.8 and 3.10). The differing responses within guilds and families support the idea that organisms within the same trophic level do not necessarily exhibit the same influences or responses in the context of community dynamics (Schmitz and Suttle 2001, Rosenheim and Corbett 2003), implying that the incorporation of species composition in analyses, as I have done here and in Beals (2006), may be necessary to truly understand community structure.

Regression models

In addition to the path model analyses, I also regressed the spider NMS axes on axes for herbivore/scavenger and predator/parasite individuals. Both significant models (for S2 and S3) included effects of the variables associated with potential prey (i.e., the herbivores and scavengers). S2 responded similarly to the one predictor axis (Herb2) in both control and treatment quadrats, although the intercepts varied. S3 responded to

85

Herb3 in the same way in both groups of quadrats, but only responded to Herb2 and Pred3 in treatment quadrats. As mentioned above, changes in architecture can affect interactions between predators and prey, and among predators. Changes in *Desmodium* biomass may also have had a direct effect on the herbivore species. Spider species associated with S3 responded in the same direction to the non-predatory and predatory dipterans, but in opposite directions to two groups of hemipterans. There was a single family of Diptera classified as a predator (Ephydridae), but some members of this family are scavengers, so perhaps should have been classified with the non-predatory dipterans. The families of Hemiptera classified as predators (Anthocoridae and Reduviidae), on the other hand, are dominated by predaceous species. Arthropods associated with Pred3 may interact with spiders in the treatment quadrats in several ways; they may compete with spiders for other arthropod prey, they may prey on spiders, or they may serve as potential prey for the spiders. Intraguild predation, as in the latter two examples, has been well documented in spiders, involving both predation among spider taxa (e.g., Balfour et al. 2003, Denno et al. 2004) and with other predatory arthropod taxa (e.g., Finke and Denno 2002, Rosenheim et al. 2004).

Conclusions

Clearly the manipulation of *Desmodium* caused changes in the way various aspects of this community influenced spider species, but the exact mechanisms by which this occurred are not entirely clear. My goal in this study was to determine whether the previously observed effect of plant species on spiders (Beals 2006) was due to indirect effects through arthropods and architecture or the direct effects of the architecture of a

particular plant species (*Desmodium*). The effects of overall plant species composition and the manipulation of *Desmodium* architecture were indirect for two dimensions of the spider community (S1 and S2). The direct effects exhibited in the path model for S3, however, suggest that the finer-scale architecture Desmodium and other plant species also have a direct influence on the spider community.

I used path analysis in an SEM framework to a) test a priori hypotheses about the different factors influencing spider community structure, and b) generate new models to provide insights into the relative strengths of these factors. Path modeling approaches, particularly in an SEM framework, are powerful methods for testing and developing complicated multivariate hypotheses (Grace and Pugesek 1998, Hershberger et al. 2003, Grace in prep). A further advantage of the approach here, in which NMS axes incorporating information about species composition are used in path analysis, is that it allows a more complete understanding of real-world community interactions than that obtained by analyses of species pairs. Theoretical and empirical work that considers only two-species interactions fails to account for the variety of more complicated dynamics that can arise when multiple species and trophic levels are considered (Sih et al. 1985, Wootton 1994a). The use of modern path analysis with species composition data permits an evaluation of the different ways in which species respond both directly and indirectly to other species and environmental factors.

CONCLUSIONS

The three chapters of this dissertation combine to further our understanding of spider communities by investigating the evidence for species interactions among spiders, and spider responses to other arthropod and plant species within and across trophic levels, as well as to the physical structure of the habitat. The research presented here not only addresses community structure in this important group of terrestrial arthropod predators, but also contributes a suite of statistical approaches that can be used by community ecologists in other systems. A novel contribution in all three chapters is the use of nonmetric multidimensional scaling (NMS) axes to represent the spider community as a whole (as well as other variables), and the use of these axes in subsequent analyses. Compared to other ordination techniques, NMS has shown superior performance in recovering the underlying gradients of ecological data matrices. Representing spider species composition in this way, and using NMS axes for both predictor and response variables, creates a more complete picture of the spider community and the influences of other arthropods, plant species and physical structure on it. Additionally, in Chapter 3 I employ path analysis in a structural equation modeling (SEM) framework, which has seen limited use in community ecology.

Chapter 1 examines patterns of spider species co-occurrence using two approaches. The first approach, analyzing species presence/absence matrices, has been used extensively in ecology, although comparatively few studies have investigated terrestrial invertebrate communities and even fewer have involved spiders. Further, the standard metric used in these analyses (the C-score) has been misinterpreted in much of

the literature; a significant C-score calculated for an entire matrix when row and column sums are fixed in the randomization runs (generally recommended and the default in the software EcoSim [Gotelli and Entsminger 2001]) can only indicate nonrandom patterns, not whether species distributions are exclusive or congruent (Stone and Roberts 1992, Collins and Simberloff in prep). Interestingly, despite citing Stone and Robert's 1992 paper, most authors fail to acknowledge this and conclude that a significantly high C-score is evidence for exclusive distributions. Calculating the C-score (and T- or Togetherness score) for submatrices within the entire randomized matrix allows more meaningful interpretations of significant nonrandom patterns.

I used C++ code provided by M. Collins to analyze patterns of species co-occurrence in three spider communities. At the community-wide level (i.e., the entire matrix) spider species overall appeared to be distributed randomly with respect to one another, although somewhat paradoxically a significant number of species pairs exhibited perfectly exclusive distributions in one community (and evidenced a weak tendency in this direction in a second). At the guild and family levels (i.e., submatrices within the entire matrix) most species within these groups showed random patterns, but a few groups showed patterns of positive association, which is consistent with a hypothesis of shared habitat preferences. The significant exclusive distributions at the community-wide level then may reflect differences in the habitat preferences of some guilds or families, rather than interspecific interactions. The conclusions that these communities are structured randomly with respect to species interactions among spiders, while habitat affinities may explain patterns within some groups is consistent with previous work. Competitive interactions among spiders are generally thought to be fairly weak (reviewed

in Wise 1993), while the importance of habitat features (specifically habitat architecture) has been well-documented (reviewed in Uetz 1991). It should be noted that other possible explanations of patterns of positive and negative co-occurrence have been proposed, such as differences or similarities in dispersal routes or abilities, colonization histories and speciation, particularly for species on islands. These processes, however, are unlikely to be operating at the spatial scale of this study. The second null model approach in Chapter 1 investigates patterns of species responses to underlying environmental variables based on individual species' correlations with NMS axes. Although species pairs within guilds often showed opposite responses to the underlying variables, reflecting differing spatial distributions within the habitat, these results too were generally consistent with a habitat preference hypothesis (in this case, a hypothesis of differing habitat preferences).

Chapters 2 and 3 address the question of what aspects of the biotic and physical environment influence spider community structure. Chapter 2 represents a first attempt at addressing this question by introducing the use of NMS axes as predictor and response variables in a multiple regression analysis to distinguish between the influences of plant species composition and vegetation architecture. Although habitat architecture predicted some of the variation in spider species composition, consistent with a large body of literature on spider communities, the important role of plant species composition was unexpected. In particular, the legume *Desmodium* sp. appeared to be strongly associated with spider composition. The role of plant species composition per se has never been directly investigated in spider communities. I suggest in Chapter 2 that this relationship may be a result of either fine-grained architectural features that are confounded with plant

species composition or an indirect effect of plants through the association of non-spider arthropods with them. Multiple regression analyses are not amenable to the examination of these indirect effects. Multiple regression accounts for the partial effects of one predictor on a response variable given the other predictors in the model, but cannot explicitly incorporate indirect effects.

Chapter 3 expands on the work of Chapter 2 in several ways. First, it includes additional potential influences on the spider community such as non-spider arthropods, total plant biomass and temperature. Second, the influence of the plant *Desmodium* sp. was examined further through manipulation of its architecture. Finally, a statistical analysis was employed that tested multivariate hypotheses, thus allowing an assessment of the roles of indirect effects in structuring a spider community. This analytical approach yielded insights into the complicated relationships across trophic levels in this system. Overall, models that included only indirect paths between plants and spiders fit better than or as well as models that included direct effects. This result supported my hypothesized explanation for the results from the regression analyses presented in Chapter 2. In two of the simplified models the effect of plant species composition on spider composition was indirect, operating through the plants' effects on both habitat structure and non-spider arthropod composition. The strong responses of spider species to *Desmodium* in Chapter 2 may, in part, have reflected its influence on these other variables.

In the third model, both plant species composition and *Desmodium* manipulation had direct effects on spider species composition. This result suggested that the finer scale architecture of *Desmodium* and other plant species themselves influenced spider

community structure. It seems reasonable, then, to conclude that the manipulation of

Desmodium architecture affected the role of plant species composition on spider

community structure in three ways: indirectly though the effects on both arthropods and

coarse-grained architecture, and directly through the changes in the fine-grained

architecture of *Desmodium* itself. Finally in Chapter 3, I show that spider composition,

while predicted by the arthropod community in general, specifically responds to

components of the non-predatory arthropod community (i.e., potential prey) and in one

case also responds to the composition of predatory arthropods.

The use of NMS axes in all of these analyses represents a data-driven approach

that facilitates an understanding of community structure by allowing the information in

the data to determine which of the original variables (spider species or architectural

measurements, for example) contribute to the overall variation in species composition (or

architectural configuration). This approach also permits the analysis of complicated,

otherwise statistically intractable, data matrices. The use of path analytic methods to

both test hypothesized models specified a priori and develop new reduced models using

NMS axes as variables further refines our understanding of community structure. The

development and testing of complicated multivariate hypotheses represents an important

advance in community ecology (Grace in prep). This method permits a confrontation of

models with data (Hilborn and Mangel 1997), asking whether a proposed model is

consistent with the data rather than asking whether the data are consistent with a model.

This shift from a more traditional univariate frequentist approach to hypothesis testing to

one in which multiple models incorporating multivariate hypotheses can be entertained

and evaluated (based on maximum likelihood and/or information criteria) is made

possible with the use of SEM-related procedures.

LITERATURE CITED

Abraham, B.J. 1983. Spatial and temporal patterns in a sagebrush steppe spider community (Arachnida, Araneae). Journal Of Arachnology 11:31-50.

Arhonditsis, G.B., C.A. Stow, L.J. Steinberg, M.A. Kenney, R.C. Lathrop, S.J. McBride and K.H. Reckhow. 2006. Exploring ecological patterns with structural equation modeling and Bayesian analysis. Ecological Modelling 192:385-409.

Astles, P.A., A.J. Moore and R.F. Preziosi. 2005. Genetic variation in response to an indirect ecological effect. Proceedings of the Royal Society B 272:2577-2581.

August, P.V. 1983. The role of habitat complexity and heterogeneity in structuring tropical mammal communities. Ecology 64:1495-1507.

Balfour, R.A. and A.L. Rypstra. 1998. The influence of habitat structure on spider density in a no-till soybean agroecosystem. Journal of Arachnology 26:221-226.

Balfour, R.A., C.M. Buddle, A.L. Rypstra, S.E. Walker and S.D. Marshall. 2003. Ontogenetic shifts in competitive interactions and intra-guild predation between two wolf spider species. Ecological Entomology 28:25-30.

Barnes, R.D. 1953. The ecological distribution of spiders in non-forest maritime communities at Beaufort, North Carolina. Ecological Monographs 23:315-337.

Beals, M..L. 2006. Understanding community structure: a data-driven multivariate approach. Oecologia DOI 10.1007/s00442-006-0551-8.

Bell, G. 2005. The co-distribution of species in relation to the neutral theory of community ecology. Ecology 86:1757-1770.

Bender, E.A., T.J. Case, and M.E. Gilpin. 1984. Perturbation experiments in community ecology: theory and practice. Ecology 65:1-13.

Benedetti-Cecchi, L. 2000. Predicting direct and indirect interactions during succession in a mid-littoral rocky shore assemblage. Ecological Monographs 70:45-72.

Bollen, K.A. 1989. Structural equation models with latent variables. Wiley, New York, USA.

Bonte, D., J.P. Maelfait and M. Hoffmann. 2000. The impact of grazing on spider communities in a mesophytic calcareous dune grassland. Journal of Coastal Conservation 6:135-144.

Boyer, K.E. and P. Fong. 2005. Co-occurrence of habitat-modifying invertebrates: effects on structural and functional properties of a created salt marsh. Oecologia 143:619-628.

Bozdogan, H. 1986. Multi-sample cluster analysis as an alternative to multiple comparison procedures. Bulletin of Informatics and Cybernetics 22:95-130.

Bozdogan, H. 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika 52:345-370.

Brose, U. 2003. Bottom-up control of carabid beetle communities in early successional wetlands: mediated by vegetation structure or plant diversity? Oecologia 135:407-413.

Brown, J.S. 1990. Habitat selection as an evolutionary game. Evolution 44:732-746.

Buddle, C.M. 2002. Interactions among young stages of the wolf spiders *Pardosa moesta* and *P. mackenziana* (Araneae: Lycosidae). Oikos 96:130-136.

Bultman, T.L. and G.W. Uetz. 1982. Abundance and community structure of forest floor spiders following litter manipulation. Oecologia 55:34-41.

Caraco, T. and R.G. Gillespie. 1986. Risk-sensitivity - foraging mode in an ambush predator. Ecology 67:1180-1185.

Chase, J.M., P.A. Abrams, J.P. Grover, S. Diehl, P. Chesson, R.D. Holt, S.A. Richards, R.M. Nisbet and T.J. Case. 2002. The interaction between predation and competition: a review and synthesis. Ecology Letters 5:302-315.

Clark, T.L. and F.J. Messina. 1998. Foraging behavior of lacewing larvae (Neuroptera: Chrysopidae) on plants with divergent architectures. Journal Of Insect Behavior 11:303-317.

Clarke, K.R. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18:117-143.

Collins, M.D. and D. Simberloff. *in prep*. An examination of community-wide measures of co-occurrence.

Conley, W. 1976. Competition between *Microtus*: A behavioral hypothesis. Ecology 57:224-237.

Connell, J.H. 1978. Diversity in tropical rainforests and coral reefs. Science 199:1302-1310.

Connell, J.H. 1980. Diversity and the coevolution of competitors, or the ghost of competition past. Oikos 35:131–138.

Connell, J.H. 1983. On the prevalence and relative importance of interspecific competition: evidence from field experiments. American Naturalist 122:661-696.

Connor, E.F. and D. Simberloff. 1979. The assembly of species communities: chance or competition? Ecology 60:1132-1140.

Cox, T.F. and M.A.A. Cox. 2001. Multidimensional scaling. Chapman and Hall/CRC, Boca Raton, Florida, USA.

Cramer, M.J. and M.R. Willig. 2005. Habitat heterogeneity, species diversity and null models. Oikos 108:209-218.

Davison, M.L. 1983. Multidimensional scaling. John Wiley and Sons, New York, USA.

Dayton, G.H. and L.A. Fitzgerald. 2001. Competition, predation, and the distribution of four desert anurans. Oecologia 129:430-435.

de Souza, A.L.T. and R.P. Martins. 2004. Distribution of plant-dwelling spiders: inflorescences versus vegetative branches. Austral Ecology 29:342-349.

Denno, R.F., M.S. Mitter, G.A. Langellotto, C. Gratton and D.L. Finke. 2004. Interactions between a hunting spider and a web-builder: consequences of intraguild predation and cannibalism for prey suppression. Ecological Entomology 28:566-577.

Diamond, J.M. 1975. Distributional ecology and habits of some Bougainville birds (Solomon-Islands). Condor 77:14-23.

Diamond, J.M. and M.E. Gilpin. 1982. Examination of the null model of Connor and Simberloff for species co-occurrences on islands. Oecologia 52:64-74.

Dick, J.T.A. 1996. Post-invasion amphipod communities of Lough Neagh, Northern Ireland: influences of habitat selection and mutual predation. Journal of Animal Ecology 65:756-767.

Drake, J.A., G.R. Huxel and C. Hewlett. 1996. Diversity and higher levels of organization. Pages 149-166 *in* K. Gaston, editor. Biodiversity: A Biology of Numbers and Differences. Blackwell, London, UK.

Dueser, R.D. and J.H. Porter. 1986. Habitat use by insular small mammals: relative effects of competition and habitat structure. Ecology 67:195-201.

Edwards, N.T. and D.L. Otis. 1999. Avian communities and habitat relationships in South Carolina Piedmont beaver ponds. American Midland Naturalist 141:158-171.

Escobar, F., J.M. Lobo, and G. Halffter. 2005. Altitudinal variation of dung beetle (Scarabaeidae: Scarabaeinae) assemblages in the Colombian Andes. Global Ecology and Biogeography 14:327-337.

Estades, C.F. 1997. Bird-habitat relationships in a vegetational gradient in the Andes of Central Chile. The Condor 99:719-727.

Ewert, D. 1982. Birds in isolated bogs in central Michigan. American Midland Naturalist 108:41-50.

Faeth, S.H. and D. Simberloff. 1981. Population regulation of a leaf-mining insect *Cameraria* sp. Nov., at increased field densities. Ecology 62:620-624.

Faith, D.P., P.R. Minchin and L. Belbin. 1987. Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 69:57-68.

Fasham, M.J.R. 1977. A comparison of nonmetric multidimensional scaling, principal components and reciprocal averaging for the ordination of simulated coenoclines, and coenoplanes. Ecology 58-551-561.

Finke, D.L. and R.F. Denno. 2002. Intraguild predation diminished in complex-structured vegetation: implications for prey suppression. Ecology 83:643-652.

Fournier, V., J.A. Rosenheim, J. Brodeur, L.O. Laney and M.W. Johnson. 2003. Herbivorous mites as ecological engineers: indirect effects on arthropods inhabiting papaya foliage. Oecologia 135:442-450.

Friend, G.R. and K.M. Cellier. 1990. Wetland herpetofauna of Kakadu National Park, Australia: seasonal richness trends, habitat preferences and the effects of feral ungulates. Journal of Tropical Ecology 6:131-152.

Garvey, J.E., R.A. Stein and H.M. Thomas. 1994. Assessing how fish predation and interspecific prey competition influence a crayfish assemblage. Ecology 75:532-547.

Gilinsky, E. 1984. The role of fish predation and spatial heterogeneity in determining benthic community structure. Ecology 65:455-468.

Gillespie, R.G. and T. Caraco. 1987. Risk-sensitive foraging strategies of 2 spider populations. Ecology 68:887-899.

Gleason, H.A. and A. Cronquist. 1991. Manual of vascular plants of northeastern United States and adjacent Canada. New York Botanical Garden, Bronx, New York.

Gorman, O.T. 1988. The dynamics of habitat use in a guild of Ozark minnows. Ecological Monographs 58:1-18.

Gotelli, N.J. 2000. Null model analysis of species co-occurrence patterns. Ecology 81:2606-2621.

Gotelli, N.J. and A.M. Ellison. 2002. Assembly rules for New England ant assemblages. Oikos 99:591-599.

Gotelli, N.J. and D.J. McCabe. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. Ecology 83:2091-2096.

Gotelli, N.J. and G.L. Entsminger. 2001. EcoSim: Null models software for ecology. Version 7.0. Acquired Intelligence Inc. & Kesey-Bear.

Gotelli, N.J. and G.L. Entsminger. 2003. Swap algorithms in null model analysis. Ecology 84:532-535.

Gotelli, N.J. and G.R. Graves. 1996. Null models in ecology. Smithsonian Institution Press, Washington DC, USA.

Gower, J.C. and D.J. Hand. 1996. Biplots. Chapman and Hall, London.

Gower, J.C. and R.F. Ngouenet. 2005. Nonlinearity effects in multidimensional scaling. Journal of Multivariate Analysis 94:344-365.

Grace, J.B. 2003. Examination of the relationship between environmental variables and ordination axes using latent variables and structural equation modeling. Pages 171-193 in B.H. Pugesek, A. Tomer and A. von Eye, editors. Structural equation modeling: applications in ecological and evolutionary biology. Cambridge University Press, Cambridge, UK.

Grace, J.B. and B.H. Pugesek. 1998. On the use of path analysis and related procedures for the investigation of ecological problems. American Naturalist 152:151-159.

Grace, J.B. in prep. Structural equation modeling and the study of natural systems. in prep for Cambridge University Press, Cambridge, UK.

Graham, M.H. 2003. Confronting multicollinearity in ecological multiple regression. Ecology 84:2809-2815.

Greenquist, E.A. and J.S. Rovner. 1976. Lycosid spiders on artificial foliage: stratum choice, orientation preferences, and prey-wrapping. Psyche 83:196-209.

Greenstone, M.H. 1984. Determinants of web spider species diversity: vegetation structural diversity vs. prey availability. Oecologia 62:299-304.

Gruner, D.S., A.D. Taylor and R.E. Forkner. 2005. The effects of foliar pubescence and nutrient enrichment on arthropod communities of *Metrosideros polymorpha* (Myrtaceae). Ecological Entomology 30:428-443.

Gunnarsson, B. 1988. Spruce-living spiders and forest decline: the importance of needle-loss. Biological Conservation 43:309-320.

Gunnarsson, B. 1990. Vegetation structure and the abundance and size distribution of spruce-living spiders. Journal of Animal Ecology 59:743-752.

Gunnarsson, B. 1992. Fractal dimension of plants and body size distribution in spiders. Functional Ecology 6:636-641.

Gunnarsson, B. 1996. Bird predation and vegetation structure affecting spruce-living arthropods in a temperate forest. Journal of Animal Ecology 65:389-397.

Gurevitch, J., J.A. Morrison, and L.V. Hedges. 2000. The interaction between competition and predation: a meta-analysis of field experiments. American Naturalist 155:435-453.

Hairston, N.G. 1986. Species packing in Desmognathus salamanders: experimental demonstration of predation and competition. American Naturalist 127:266-291.

Halaj, J., D.W. Ross, and A.R. Moldenke. 1998. Habitat structure and prey availability as predictors of the abundance and community organization of spiders in western Oregon forest canopies. Journal of Arachnology 26:203-220.

Halaj J., D.W. Ross, and A. R. Moldenke. 2000. Importance of habitat structure to the arthropod food-web in Douglas-fir canopies. Oikos 90:139-152.

Halaj, J. and D.H. Wise. 2001. Terrestrial trophic cascades: how much do they trickle? American Naturalist 157:262-281.

Hallander, H. 1970. Prey, cannibalism and microhabitat selection in the wolf spiders *Pardosa chelata* O. F. Müller and *P. pullata* Clerck. Oikos 21:37-340.

Harvey, P.H., R.K. Colwell, J.W. Silvertown and R.M. May. 1983. Null models in ecology. Annual Review of Ecology and Systematics 14:189-211.

Harwood, J.D., K.D. Sunderland and W.O.C. Symondson. 2003. Web-location by linyphiid spiders: prey-specific aggregation and foraging strategies. Journal of Animal Ecology 72:745-756.

Hastings, A. 1987. Can competition be detected using species co-occurrence data? Ecology 68:117-123.

Hatley, C.L. and J.A. MacMahon. 1980. Spider community organization - seasonal-variation and the role of vegetation architecture. Environmental Entomology 9:632-639.

Haukisalmi, V. and H. Henttonen. 1998. Analysing interspecific associations in parasites: alternative methods and effects of sampling heterogeneity. Oecologia 116:565-574.

Hawkins, B.A. and J.G. Pausas. 2004. Does plant richness influence animal richness?: the mammals of Catalonia (NE Spain). Diversity and Distributions 10:247-252.

Hawkins, B.A. and E.E. Porter. 2003. Does herbivore diversity depend on plant diversity? The case of california butterflies. American Naturalist 161:40-49.

Heck, K.L. and G.S. Wetstone. 1977. Habitat complexity and invertebrate species richness and abundance in tropical seagrass meadows. Journal of Biogeography 4:135-142.

Heikkinen, M.W. and J.A. MacMahon. 2004. Assemblages of spiders on models of semi-arid shrubs. Journal of Arachnology 32:313-323.

Hershberger, S.L., G.A. Marcoulides and M.M. Parramore. 2003. Structural equation modeling: an introduction. Pages 3-41 *in* B.H. Pugesek, A. Tomer and A. von Eye, editors. Structural equation modeling: applications in ecological and evolutionary biology. Cambridge University Press, Cambridge, UK.

Hilborn, R. and M. Mangel. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, New Jersey, USA.

Hill, A.M. and D.M. Lodge. 1994. Diel changes in resource demand: competition and predation in species replacement among crayfishes. Ecology 75:2118-2126.

Hodkinson, I.D., S.J. Coulson and J. Harrison. 2001. What a wonderful web they weave: spiders, nutrient capture and early ecosystem development in the high Arctic—some counter-intuitive ideas on community assembly. Oikos 95:349-352.

Horton, C.C. and D.H. Wise. 1983. The experimental analysis of competition between two syntopic species of orb-web spiders (Araneae: Araneidae). Ecology 64:929-944.

Hughes, A.R. and J.H. Grabowski. 2006. Habitat context influences predator interference interactions and the strength of resource partitioning. Oecologia 149:256-264.

Hurd L.E. and W.F. Fagan. 1992. Cursorial spiders and succession: age or habitat structure? Oecologia 92:215-221.

Huston, M.A. 1979. A general hypothesis of species diversity. American Naturalist 113:81-101.

Johnsingh, A.J.T. and J. Joshua. 1994. Avifauna in three vegetation types on Mundanthurai Plateau, South India. Journal of Tropical Ecology 10:323-335.

Jongman, R.H.G., C.J.F. ter Braak and O.F.R. van Tongeren. 1995. Data analysis in community and landscape ecology. Cambridge University Press, Cambridge.

Kareiva, P. 1994. Special feature - higher-order interactions as a foil to reductionist ecology. Ecology 75:1527-1528.

Kaston, B.J. 1981. Spiders of Connecticut. State Geological and Natural History Survey of Connecticut, Hartford, Connecticut, USA.

Kelaher, B.P. 2003. Changes in habitat complexity negatively affect diverse gastropod assemblages in coralline algal turf. Oecologia 135:431-441.

Koen, J.H. and T.M. Crowe. 1987. Animal-habitat relationships in the Knysna Forest, South Africa: discrimination between forest types by birds and invertebrates. Oecologia 72:414-422.

Kronk, A.E. and S.E. Riechert. 1979. Parameters affecting the habitat choice of a desert wolf spider, *Lycosa santrita* Chamberlin and Ivie. Journal of Arachnology 7:155-166.

Landres, P.B. and J.A. MacMahon. 1983. Community organization of arboreal birds in some oak woodlands of western North America. Ecological Monographs 53:183-208.

Langellotto, G.A. and R.F. Denno. 2004. Responses of invertebrate natural enemies to complex-structured habitats: a meta-analytical synthesis. Oecologia 139:1-10

Legendre, P. and E.D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129:271-280.

Legendre, P. and L. Legendre. 1998. Numerical ecology. Elsevier, Amsterdam.

Lowrie, D.C. 1948. The ecological succession of spiders of the Chicago area dunes. Ecology 29:334-351.

M'Closkey, R.T., and D.T. Lajoie. 1975. Determinants of local distribution and abundance in white-footed mice. Ecology 56:467-472.

MacArthur, R.H. and J.W. MacArthur. 1961. On bird species diversity. Ecology 42:594-598.

MacArthur, R.H., H. Recher and M. Cody. 1966. On the relation between habitat selection and species diversity. American Naturalist 100:319-332.

Malaeb, Z.A., J.K. Summers and B.H. Pugesek. 2000. Using structural equation modeling to investigate relationships among ecological variables. Environmental and Ecological Statistics 7:93-111.

Martínez-Vilalta J., A. Bertolero, D. Bigas, J-Y. Paquet and A. Martínez-Vilalta. 2002. Habitat selection of passerine birds nesting in the Ebro Delta reedbeds (NE Spain): management implications. Wetlands 22:318-325.

May, R.M. 1984. An overview: real and apparent patterns in community structure. Pages 3-16 *in* D. R. Strong, D. Simberloff, L. G. Abele and A. B. Thistle, editors. Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton, New Jersey, USA.

McCoy, E.D. and S.S. Bell. 1991. Habitat structure: the evolution and diversification of a complex topic. Pages 3-27 *in* S. S. Bell, E. D. McCoy, and H. R. Mushinsky, editors. Habitat structure: the physical arrangement of objects in space. Chapman & Hall, New York, New York, USA.

McCune, B. 1997. Influence of noisy environmental data on canonical correspondence analysis. Ecology 78:2617-2623.

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

McNett, B.J. and A.L. Rypstra. 2000. Habitat selection in a large orb-weaving spider: vegetational complexity determines site selection and distribution. Ecological Entomology 25:423-432.

McReynolds, C.N. 2000. The impact of habitat features on web features and prey capture of *Argiope aurantia* (Araneae, Araneidae). J Arachnology 28:169-179.

Miklós, I. and J. Podani. 2004. Randomization of presence-absence matrices: comments and new algorithms. Ecology 85:86-92.

Minchin, P.R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. Vegetatio 69:89-107.

Miyashita, T., T. Fukami and A. Shimazaki. 1998. One spider a leaf: competition for web-sites in *Nephila maculata* spiderlings. Acta Arachnologica 47:31-35.

Moermond, T.C. 1979. Habitat constraints on the behavior, morphology, and community structure of *Anolis* lizards. Ecology 60:152-164.

Moran, M.D. 2003. Arguments for rejecting the sequential Bonferroni in ecological studies. Oikos 100:403-405.

Moran, M.D. and A.R. Scheidler. 2002. Effects of nutrients and predators on an old-field food chain: interactions of top-down and bottom-up processes. Oikos 98:116-124.

Morris, D.W. 2003. Toward an ecological synthesis: a case for habitat selection. Oecologia 136:1-13.

Murdoch, W.W., F. C. Evans, and C. H. Peterson. 1972. Diversity and pattern in plants and insects. Ecology 53:819-829.

Neter, J., M.H. Kutner, W. Wasserman, and C.J. Nachtsheim. 1996. Applied linear regression models, 3rd ed. McGraw-Hill/Irwin, New York, NY, USA.

Olive, C.W. 1982. Behavioral-response of a sit-and-wait predator to spatial variation in foraging gain. Ecology 63:912-920.

Paine, R.T. 1981. Barnacle ecology: is competition important?: the forgotten roles of disturbance and predation. Paleobiology 7:553-560.

Panzer, R. and M.W. Schwartz. 1998. Effectiveness of a vegetation-based approach to insect conservation. Conservation Biology 12:693-702.

Parker, J.D., J.E. Duffy and R. J. Orth. 2001. Plant species diversity and composition: experimental effects on marine epifaunal assemblages. Marine Ecology Progress Series 224:55-67.

Peres-Neto, P.R. 2004. Patterns in the co-occurrence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. Oecologia 140:352-360.

Peres-Neto, P.R., J.D. Olden and D.A. Jackson. 2001. Environmentally constrained null models: site suitability as occupancy criterion. Oikos 93:110-120.

Persons, M.H., S.E. Walker, A.L. Rypstra and S.D. Marshall. 2001. Wolf spider predator avoidance tactics and survival in the presence of diet-associated predator cues. Animal Behaviour 61:43-51.

Peterson, C. H. 1982. The importance of predation and intra- and interspecific competition in the population biology of two infaunal suspension-feeding bivalves, *Protothaca staminea* and *Chione undatella*. Ecological Monographs 52:437-475.

Polis, G.A., C.A. Myers and R.D. Holt. 1989. The ecology and evolution of intraguild predation: potential competitors that eat each other. Annual Review of Ecology and Systematics 20:297-330.

Post, W.M. and S.E. Riechert. 1977. Initial investigation into the structure of spider communities. Journal of Animal Ecology 46:729-749.

Prentice, I.C. 1977. Non-metric ordination methods in ecology. Journal of Ecology 65:85-94

Pugesek, B.H. 2003a. Concepts of structural equation modeling in biological research. Pages 42-59 *in* B.H. Pugesek, A. Tomer and A. von Eye, editors. Structural equation modeling: applications in ecological and evolutionary biology. Cambridge University Press, Cambridge, UK.

Pugesek, B.H. 2003b. Modeling a complex theory of population change in the Shiras moose: history and recasting as a structural equation model. Pages 60-84 *in* B.H. Pugesek, A. Tomer and A. von Eye, editors. Structural equation modeling: applications in ecological and evolutionary biology. Cambridge University Press, Cambridge, UK.

Rand, A.S. 1964. Ecological distribution in anoline lizards of Puerto Rico. Ecology 45:745-752.

Real, L. and T. Caraco. 1986. Risk and foraging in stochastic environments. Annual Review Of Ecology And Systematics 17:371-390.

Ribas, C.R. and J.H. Schoereder. 2002. Are all ant mosaics caused by competition? Oecologia 131:606-611.

Riechert, S.E. 1981. The consequences of being territorial: spiders, a case study. American Naturalist 117:871-892.

Riechert, S.E. 1982. Spider interaction strategies: communication versus coercion. *in* P.N. Witt and J. Rovner, editors. Spider communication: mechanisms and ecological significance. Princeton University Press, Princeton, New Jersey, USA..

Riechert, S.E. and W.G. Reeder. 1970. Effects of fire on spider distribution in southwestern Wisconsin prairies. Pages 73-90 *in* J. H. Zimmerman, editor. Proceedings of the Second Midwest Prairie Conference.

Riechert, S.E. and A.B. Cady. 1983. Patterns of resource use and tests for competitive release in a spider community. Ecology 64(4):899-913.

Riechert, S.E. and C.R. Tracy. 1975. Thermal balance and prey availability: basis for a model relating web-site characteristics to spider reproductive success. Ecology 56:265-284.

Riechert, S.E. and R.G. Gillespie. 1986. Habitat choice and utilization in web-building spiders. Pages 23-48 *in* W. A. Shear, editor. Spiders: webs, behavior and evolution. Stanford University Press, Stanford, California, USA.

Riechert, S.E. and F.D. Singer. 1995. Investigation of potential male mate choice in a monogamous spider. Animal Behavior 49:715-723.

Rigby, C. and J.H. Lawton. 1981. Species-area relationships of arthropods on host plants: herbivores on bracken. Journal of Biogeography 8:125-133.

Robinson, J.V. 1981. The effects of architectural variation in habitat on a spider community: an experimental field study. Ecology 62:73-80.

Root, R.B. 1967. The niche exploitation pattern of the blue-grey gnatcatcher. Ecological Monographs 37:317-350.

Rosenheim, J.A. and A. Corbett. 2003. Omnivory and the indeterminacy of predator function: can a knowledge of foraging behavior help? Ecology 84:2538-2548.

Rosenheim, J.A., T.E. Glik, R.E. Goeriz and B. Rämert. 2004. Linking a predator's foraging behavior with its effects on herbivore population suppression. Ecology 85:3362-3372.

Rosenzweig, M.L. 1981. A theory of habitat selection. Ecology 62:327-335.

Rosenzweig, M.L. 1991. Habitat selection and population interactions: the search for mechanism. American Naturalist 137:S5-S28.

Rosenzweig, M.L. and J. Winakur. 1969. Population ecology of desert rodent communities: habitats and environmental complexity. Ecology 50:558-572.

Rotenberry, J.T. 1985. The role of habitat in avian community composition: physiognomy or floristics? Oecologia 67:213-217.

Rotenberry, J.T. and J.A. Wiens. 1980. Habitat structure, patchiness, and avian communities in North American steppe vegetation: a multivariate analysis. Ecology 61:1228-1250.

Rypstra, A.L. 1983. The importance of food and space in limiting web-spider densities: a test using field enclosures. Oecologia 59:312-316.

Rypstra, A.L. 1986. Web spiders in temperate and tropical forests: relative abundance and environmental correlates. American Midland Naturalist 115:42-51.

Rypstra, A.L., P.E. Carter, R.A. Balfour and S.D. Marshall. 1999. Architectural features of agricultural habitats and their impact on the spider inhabitants. Journal of Arachnology 27:371-377.

Sánchez-Piñero, F. and G.A. Polis. 2000. Bottom-up dynamics of allochthonous input: direct and indirect effects of seabirds on islands. Ecology 81:3117-3132.

Sanderson, R.A., S.P. Rushton, A.J. Cherril and J.P. Byrne. 1995. Soil, vegetation and space: an analysis of their effects on the invertebrate communities of a moorland in north-east England. Journal of Applied Ecology 32:506-518.

Schluter, D. 1982. Distributions of Galapagos ground finches along an altitudinal gradient: the importance of food supply. Ecology 63:1504-1517.

Schluter, D. 1984. A variance test for detecting species associations, with some example applications. Ecology 65:998-1005.

Schmitz, O.J. 1998. Direct and indirect effects of predation and predation risk in old-field interaction webs. American Naturalist 151:327-342.

Schmitz, O.J. 2003. Top predator control of plant biodiversity and productivity in an old-field ecosystem. Ecology Letters 6:156-163.

Schmitz, O.J. 2006. Predators have large effects on ecosystem properties by changing plant diversity, not plant biomass. Ecology 87:1432-1437.

Schmitz, O.J. and K.B. Suttle. 2001. Effects of top predator species on direct and indirect interactions in a food web. Ecology 82:2072-2081.

Schmitz, O.J., A. P. Beckerman and K.M. O'Brien. 1997. Behaviorally mediated trophic cascades: effects of predation risk on food web interactions. Ecology 78:1388-1399.

Schmitz, O.J., E.L. Kalies and M.G. Booth. 2006. Alternative dynamic regimes and trophic control of plant succession. Ecosystems 9:659-672.

Schmitz, O.J., V. Krivan and O. Ovadia. 2004. Trophic cascades: the primacy of trait-mediated interactions. Ecology Letters 7:153-163.

Schoener, T.W. 1983. Field experiments on interspecific competition. American Naturalist 122:240-285.

Schoener, T.W. and G.H. Adler. 1991. Greater resolution of distributional complementarities by controlling for habitat affinities: a study with Bahamian lizards and birds. American Naturalist 137:669-692.

Seamon, J.O. and G.H. Adler. 1996. Population performance of generalist and specialist rodents along habitat gradients. Canadian Journal of Zoology 74:1130-1139.

Searcy, L.E., A.L. Rypstra and M. H. Persons. 1999. Airborne chemical communication in the wolf spider *Pardosa milvina*. Journal of Chemical Ecology 25:2527-2533.

Sfenthourakis, S., E. Tzanatos and S. Giokas. 2005. Species co-occurrence: the case of congeneric species and a causal approach to patterns of species association. Global Ecology and Biogeography 15:39-49.

Shipley, B. 2000. Cause and correlation in biology: a user's guide to path analysis, structural equations, and causal inference. Cambridge University Press, Cambridge, UK.

Siemann, E. 1998. Experimental tests of effects of plant productivity and diversity on grassland arthropod diversity. Ecology 79:2057-2070.

Siemann, E., D. Tilman, J. Haarstad and M. Ritchie. 1998. Experimental tests of the dependence of arthropod diversity on plant diversity. American Naturalist 152:738-750.

Sih, A., P. Crowley, M. McPeek, J. Petranka, and K. Strohmeier. 1985. Predation, competition, and prey communities: a review of field experiments. Annual Review of Ecology and Systematics 16:269-311.

Simberloff, D. 1983. Competition theory, hypothesis-testing, and other community ecological buzzwords. American Naturalist 122:626-635.

Sinclair, A.R.E. 1985. Does interspecific competition or predation shape the African ungulate community? Journal of Animal Ecology 54:899-918.

Somershoe, S.G. and C.R. Chandler. 2004. Use of oak hammocks by neotropical migrant songbirds: the role of area and habitat. Wilson Bulletin 116:56-63

Spence, J.R. 1981. Experimental analysis of microhabitat selection in water-striders (Heteroptera: Gerridae). Ecology 62:1505-1514.

Spiller, D.A. 1984. Competition between two spider species: experimental field study. Ecology 65:909-919.

Spiller, D.A. and T.W. Schoener. 1988. An experimental study of the effects of lizards on web-spider communities. Ecological Monographs 58:57-77.

Stanton, M.L. 2003. Interacting guilds: moving beyond the pairwise perspective on mutualisms. American Naturalist (supplement) 162:S10-S23.

Stapp, P. 1997. Community structure of shortgrass-prairie rodents: competition or risk of intraguild predation? Ecology 78:1519-1530.

Stevens, R.D. 2004. Untangling latitudinal richness gradients at higher taxonomic levels: Familial perspectives on the diversity of new world bat communities. Journal of Biogeography 31:665-674.

Stiling, P.D. and D.R. Strong. 1984. Experimental density manipulation of stem-boring insects: some evidence for interspecific competition. Ecology 65:1683-1685.

Stone, L. and A. Roberts. 1992. Competitive exclusion, or species aggregation? An aid in deciding. Oecologia 91:419-424.

Stratton, G.E., G.W. Uetz and D.G. Dillery. 1979. A comparison of the spiders of three coniferous tree species. Journal of Arachnology 6:219-226.

Strauss, S.Y. and R.E. Irwin. 2004. Ecological and evolutionary consequences of multispecies plant-animal interactions. Annual Review of Ecology and Systematics 35:435-466.

Strong, D.R. 1982. Harmonious coexistence of hispine beetles on *Heliconia* in experimental and natural communities. Ecology 63:1039-1049.

Swihart, R.K. and N.A. Slade. 1990. Long-term dynamics of an early successional small mammal community. American Midland Naturalist 123:372-382.

Tews, J., U. Brose, V. Grimm, K. Tielbörger, M.C. Wichmann, M. Schwager and F. Jeltsch. 2004. Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. Journal of Biogeography 31:79-92.

Tomer, A. 2003. A short history of structural equation models. Pages 85-124 *in* B.H. Pugesek, A. Tomer and A. von Eye, editors. Structural equation modeling: applications in ecological and evolutionary biology. Cambridge University Press, Cambridge, UK.

Tomer, A. and B.H. Pugesek. 2003. Guidelines for the implementation and publication of structural equation models. Pages 125-140 *in* B.H. Pugesek, A. Tomer and A. von Eye, editors. Structural equation modeling: applications in ecological and evolutionary biology. Cambridge University Press, Cambridge, UK.

Tscharntke, T., A. Gathmann and I. Steffan-Dewenter. 1998. Bioindication using trap-nesting bees and wasps and their natural enemies: community structure and interactions. Journal of Applied Ecology 35:708-719.

Turner, M. and G.A. Polis. 1979. Patterns of co-existence in a guild of raptorial spiders. Journal of Animal Ecology 48:509-520.

Uetz, G.W. 1979. The influence of variation in litter habitats on spider communities. Oecologia 40:29-42.

Uetz, G.W. 1991. Habitat structure and spider foraging. Pages 325-348 *in* S. S. Bell, E. D. McCoy, and H. R. Mushinsky, editors. Habitat structure: the physical arrangement of objects in space. Chapman & Hall, New York, New York, USA.

Uetz, G.W., J. Halaj, and A.B. Cady. 1999. Guild structure of spiders in major crops. J Arachnology 27:270-280.

Ulrich, W. 2004. Species co-occurrences and neutral models: reassessing J. M. Diamond's assembly rules. Oikos 107:603-609.

van Veen, F.J.F., P.D. van Holland and H.C.J. Godfray. 2005. Stable coexistence in insect communities due to density- and trait-mediated indirect effects. Ecology 86:1382-1389.

Vazquez, D.P. and D. Simberloff. 2004. Indirect effects of an introduced ungulate on pollination and plant reproduction. Ecological Monographs 74:281-308.

Wagner, J.D. and D.H. Wise. 1996. Cannibalism regulates densities of young wolf spiders; evidence from field and laboratory experiments. Ecology 77:639-652.

Ward, D. and Y. Lubin. 1993. Habitat selection and life history of a desert spider, *Stegodyphus lineatus* (Eresidae). Journal of Animal Ecology 62:353-363.

Wellborn, G.A., D.K. Skelly and E.E. Werner. 1996. Mechanisms creating community structure across a freshwater habitat gradient. Annual Review of Ecology and Systematics 27:337-363.

Williamson, M.H. and J.H. Lawton. 1991. Fractal geometry of ecological habitats. Pages 69-86 *in* S. S. Bell, E. D. McCoy, and H. R. Mushinsky, editors. Habitat structure: the physical arrangement of objects in space. Chapman & Hall, New York, New York, USA.

Willis, S.C., K.O. Winemiller and H. Lopez-Fernandez. 2005. Habitat structural complexity and morphological diversity of fish assemblages in a Neotropical floodplain river. Oecologia 142:284-295.

Wimp, G.M., G.D. Martinsen, K.D. Floate, R.K. Bangert and T.G. Whitham. 2005. Plant genetic determinants of arthropod community structure and diversity. Evolution 59:61-69.

Wise, D.H. 1983. Competitive mechanisms in a food-limited species: relative importance of interference and exploitative interactions among labyrinth spiders (Araneae: Araneidae). Oecologia 58:1-9.

Wise, D.H. 1984. The role of competition in spider communities: insights from field experiments with a model organism. Pages 42-53 *in* D. R. Strong, D. Simberloff, L. G. Abele and A. B. Thistle, editors. Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton, New Jersey, USA.

Wise, D.H. 1993. Spiders in ecological webs. Cambridge University Press, New York, New York, USA.

Wise, D.H. and B. Chen. 1999. Impact of intraguild predators on survival of a forest floor wolf spider. Oecologia 121:129-137.

Wofford, B.E. 1989. Guide to the vascular plants of the Blue Ridge. University of Georgia Press, Athens, Georgia, USA.

Wootton, J.T. 1994a. The nature and consequences of indirect effects in ecological communities. Annual Review of Ecology and Systematics 25:443-466.

Wootton, J.T. 1994b. Predicting direct and indirect effects—An integrated approach using experiments and path analysis. Ecology 75:151-165.

Wootton, J.T. and M. Emmerson. 2005. Measurement of interaction strength in nature. Annual Review of Ecology and Systematics 36:419-444.

Wywialowski, A.P. 1987. Habitat structure and predators: choices and consequences for rodent specialists and generalists. Oecologia 72:39-45.

Zaman, A. and D. Simberloff. 2002. Random binary matrices in biogeographical ecology—Instituting a good neighbor policy. Environmental and Ecological Statistics 9:405-421.

APPENDICES

APPENDIX A

Tables and figures

Table 1.1 Spider species associated with NMS ordination axes and their correlations with NMS axes for each site (SF=sloped field, TGF=tussock grass field, DW=deciduous woodland). Guild classifications: W=web builders, H=hunters, A=ambushers/stalkers.

Site	Axis	Species	Family	Guild	τ
SF	S1	*Clubiona abbotti*	Clubionidae	H	0.45
		Rabidosa rabida	Lycosidae	H	0.36
		Neoantistea agilis	Hahniidae	W	0.33
	S2	*Bathyphantes pallida*	Linyphiidae	W	0.60
		Clubiona abbotti	Clubionidae	H	-0.37
		Gea heptagon	Araneidae	W	-0.36
	S3	*Pardosa milvina*	Lycosidae	H	0.56
		Gea heptagon	Araneidae	W	-0.32
TGF	S1	*Pardosa milvina*	Lycosidae	H	0.35
		Clubiona abbotti	Clubionidae	H	-0.35
		Pisaurina undulata	Pisauridae	H	-0.32
		Tetragnatha laboriosa	Tetragnathidae	W	-0.31
	S2	*Pardosa milvina*	Lycosidae	H	-0.48
		Hogna helluo	Lycosidae	H	-0.39
		Bathyphantes pallida	Linyphiidae	W	0.35
		Clubiona abbotti	Clubionidae	H	0.34
	S3	*Tetragnatha laboriosa*	Tetragnathidae	W	-0.47
DW	S1	*Theridion frondeum*	Theridiidae	W	0.59
	S2	*Habrocestum parvulum*	Salticidae	A	0.39
		Mangora maculata	Araneidae	W	-0.38
		Bathyphantes pallida	Linyphiidae	W	-0.35
		Lepthyphantes nebulosa	Linyphiidae	W	0.34
	S3	*Lepthyphantes nebulosa*	Linyphiidae	W	-0.48
		Theridion frondeum	Theridiidae	W	0.39
		Cicurina arcuata	Agelenidae	W	0.31

Table 1.2 Results of community-wide analyses and analyses of guilds and families within species co-occurrence matrices in the Sloped Field. Simulated values are means of all randomized matrices, prop ≥ obs is p-value, C is C-score, T is T-score, and CHECKER is the number of perfect checkerboards. C- and T-scores are given as totals rather than averages. SES is the standardized effect size. P-values less than or equal to 0.05 are highlighted in bold.

Matrix	Measure	Observed	Simulated	Prop ≥ obs	SES
Community-wide	C	83079	82375.77	0.11	1.23
	T	97833	97129.77	0.11	1.23
	CHECKER	1200	1189.56	0.24	0.72
Guilds					
Web-builder	C	13317	13458.70	0.62	-0.31
	T	15849	15927.35	0.51	-0.08
Hunter	C	12312	12413.11	0.58	-0.19
	T	14438	14465.33	0.50	-0.03
Ambusher	C	2346	2476.28	0.85	-1.02
	T	3842	2966.31	**0.05**	1.79
Families					
Araneidae	C	94	78.29	0.33	0.95
	T	104	93.61	0.30	0.12
Clubionidae	C	420	406.86	0.52	0.15
	T	448	478.21	0.67	-0.23
Linyphiidae	C	2198	2374.47	0.83	-0.94
	T	3258	2808.25	0.14	1.07
Lycosidae	C	4665	4905.46	0.75	-0.67
	T	6304	5673.07	0.10	1.33
Salticidae	C	552	597.47	0.96	-1.91
	T	1282	722.31	**0.03**	2.10
Tetragnathidae	C	24	30.80	0.95	-1.11
	T	96	37.70	0.34	1.03
Theridiidae	C	403	401.45	0.50	0.02
	T	394	477.63	0.73	-0.57
Thomisidae	C	63	65.77	0.79	-0.16
	T	86	79.95	0.64	0.08

Table 1.3 Results of community-wide analyses and analyses of guilds and families within species co-occurrence matrices in the Tussock Grass Field. Simulated values are means of all randomized matrices, prop ≥ obs is p-value, C is C-score, T is T-score, and CHECKER is the number of perfect checkerboards. C- and T-scores are given as totals rather than averages. SES is the standardized effect size. P-values less than or equal to 0.05 are highlighted in bold.

Matrix	Measure	Observed	Simulated	Prop ≥ obs	SES
Community-wide	C	51082	52031.12	0.99	-2.06
	T	72274	73223.12	0.99	-2.06
	CHECKER	1391	1354.10	**0.007**	2.50
Guilds					
Web-builder	C	9801	10075.39	0.78	-0.75
	T	14036	14253.03	0.57	-0.23
Hunter	C	7150	6749.39	0.16	1.01
	T	8379	9404.78	0.96	-1.62
Ambusher	C	1317	1391.63	0.78	-0.77
	T	2105	1967.60	0.33	0.37
Families					
Araneidae	C	61	54.66	0.39	0.93
	T	0	77.12	1	-0.96
Clubionidae	C	158	207.62	0.77	-0.87
	T	344	288.69	0.27	0.52
Linyphiidae	C	2012	1998.37	0.47	0.11
	T	2517	2835.50	0.75	-0.71
Lycosidae	C	2035	2152.00	0.69	-0.51
	T	3061	2973.00	0.37	0.28
Pisauridae	C	51	38.04	0.44	1.00
	T	0	53.35	1	-0.97
Salticidae	C	308	300.00	0.40	0.35
	T	242	423.10	0.89	-0.97
Tetragnathidae	C	72	64.25	0.60	0.32
	T	67	89.89	0.81	-0.36
Theridiidae	C	461	550.89	0.83	-0.96
	T	948	782.05	0.18	0.87
Thomisidae	C	40	45.33	0.86	-0.71
	T	178	63.35	**0.05**	1.62

Table 1.4 Results of community-wide analyses and analyses of guilds and families within species co-occurrence matrices in the Deciduous Woodland. Simulated values are means of all randomized matrices, prop ≥ obs is p-value, C is C-score, T is T-score, and CHECKER is the number of perfect checkerboards. C- and T-scores are given as totals rather than averages. SES is the standardized effect size. P-values less than or equal to 0.05 are highlighted in bold.

Matrix	Measure	Observed	Simulated	Prop ≥ obs	SES
Community-wide	C	28697	28781.30	0.63	-0.35
	T	31624	31708.30	0.63	-0.35
	Checker	1054	1041.00	0.09	1.39
Guilds					
Web-builder	C	10521	10712.69	0.76	-0.69
	T	13024	11793.13	**0.03**	1.99
Hunter	C	1425	1496.27	0.84	-0.99
	T	1948	1651.22	0.16	1.01
Ambusher	C	421	433.40	0.62	-0.25
	T	533	478.15	0.32	0.36
Families					
Agelenidae	C	16	39.80	0.99	-1.70
	T	120	43.23	0.14	1.74
Araneidae	C	21	53.65	0.99	-2.67
	T	192	58.62	**0.05**	2.34
Clubionidae	C	260	264.90	0.61	-0.22
	T	261	291.49	0.61	-0.23
Linyphiidae	C	1888	1926.23	0.67	-0.41
	T	2446	2126.16	0.16	1.00
Lycosidae	C	153	140.23	0.33	0.48
	T	117	155.77	0.65	-0.46
Salticidae	C	22	30.90	0.92	-0.66
	T	58	34.58	0.50	0.61
Tetragnathidae	C	9	11.49	0.99	-2.04
	T	76	13.14	0.16	2.01
Theridiidae	C	425	575.07	0.94	-1.55
	T	921	631.19	**0.03**	2.05
Thomisidae	C	60	71.77	0.91	-1.12
	T	134	79.48	0.32	0.79

Table 1.5 *P*-values generated by Monte Carlo resampling for differences in correlation coefficients for spider species pairs associated with NMS axes in each site (SF=sloped field, TGF=tussock grass field, DW=deciduous woodland). Values ≤0.05 are highlighted in bold. Guild classifications: W=web builders, H=hunters, A=ambushers/stalkers.

NMS Axis	Species Pair	Guild	Obs. difference	Prop. ≥ obs.
S1 (SF)	*Clubiona abbotti* v. *Rabidosa rabida*	H	0.09	0.80
S2 (SF)	*Bathyphantes pallida* v. *Gea heptagon*	W	0.96	**0.03**
S1 (TGF)	*Pardosa milvina* v. *C. abbotti*	H	0.70	**0.05**
	P. v. *Pisaurina undulata*	H	0.68	0.09
	C. abbotti v. *P. undulata*	H	0.03	0.95
S2 (TGF)	*P. milvina* v. *C. abbotti*	H	0.82	**0.05**
	P. milvina v. *Hogna helluo*	H	0.09	0.91
	C. abbotti v. *H. helluo*	H	0.73	0.09
S2 (DW)	*Mangora maculata* v. *B. pallida*	W	0.03	0.98
	M. maculata v. *Lepthyphantes nebulosa*	W	0.71	**0.02**
	B. pallida v. *L. nebulosa*	W	0.69	**0.03**
S3 (DW)	*L. nebulosa* v. *Theridion frondeum*	W	0.87	**0.02**
	L. nebulosa v. *Cicurina arcuata*	W	0.79	**0.02**
	T. frondeum v. *C. arcuata*	W	0.09	0.89
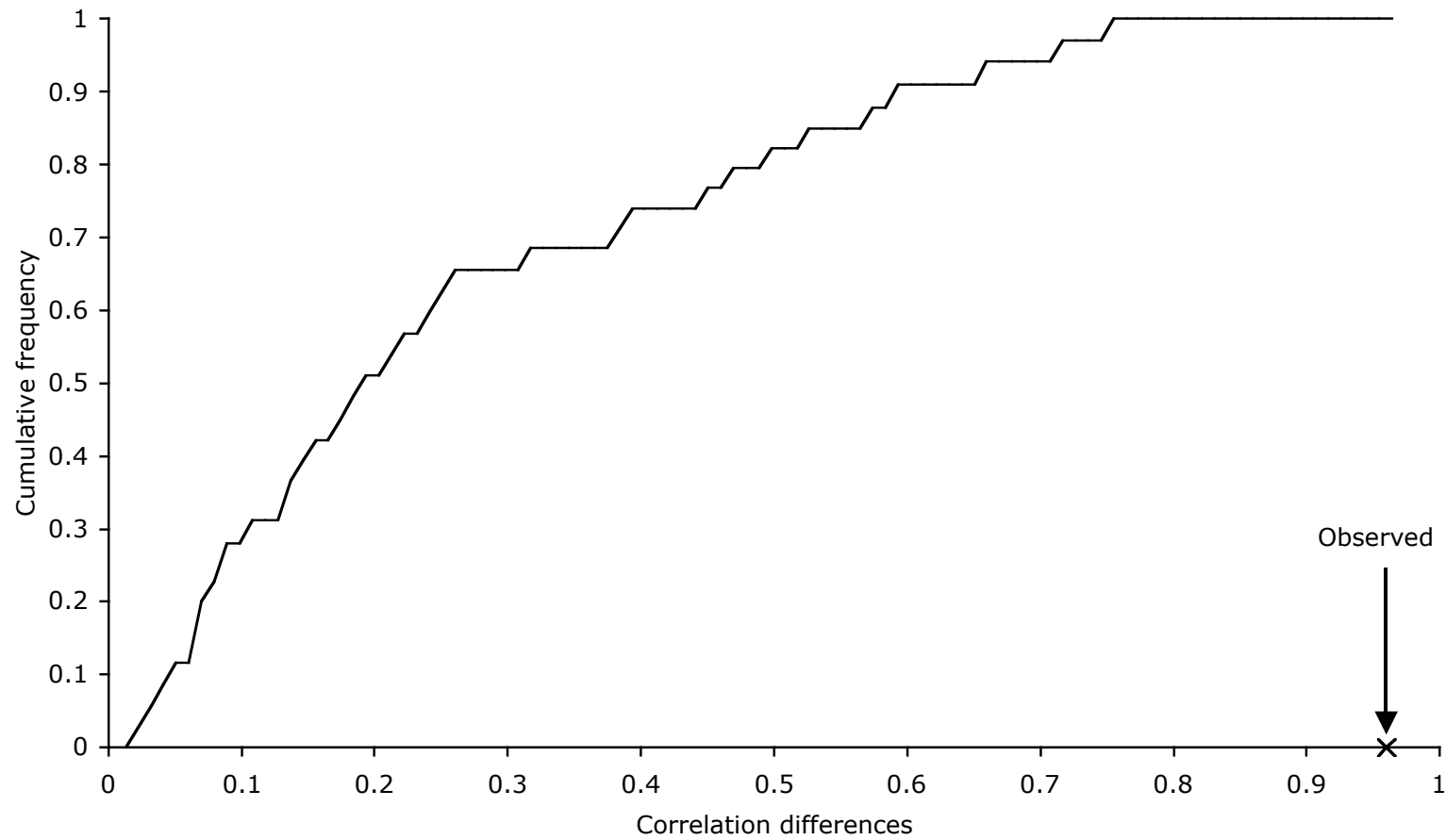
Figure 1.1 Example of null distribution of pairwise differences between correlation coefficients. "Observed" is the actual difference for the species pair *Bathyphantes pallida* and *Gea heptagon* associated with NMS axis 1 (S1) in the sloped field.

Table 2.1 Final regression models for general spider community structure in each site (number of individuals, species richness, and species diversity). Predictor variables are plant ordination axes (P_) and architecture ordination axes (A_).

Site	Response	F-value	P-value	R^2	Predictor	Estimate	T-value	P-value
Sloped Field	Individuals (no.)	8.22	<.0001	0.19	P1	-3.52	-2.97	0.004
N=108					P2	2.90	2.57	0.01
					A2	3.20	3.35	0.001
	Richness	10.18	0.002	0.09	A2	1.10	3.19	0.002
	Diversity	5.77	0.0042	0.10	P1	0.07	2.73	0.007
					A2	0.04	2.04	0.04
Tussock Grass Field	Individuals (no.)	NS						
N=100	Richness	4.52	0.04	0.04	P1	-1.00	-2.03	0.04
	Diversity	4.86	0.03	0.05	P1	-0.10	-2.20	0.03
Deciduous Woodland	Individuals (no.)	9.77	0.0002	0.22	P1	2.26	4.07	0.0001
N=80					P3	-1.14	-1.96	0.05
	Richness	7.54	0.008	0.10	P1	1.07	2.75	0.008
	Diversity	NS						

Table 2.2 Positive and negative responses of general community structure variables to original plant and habitat architecture variables based on regression models (Table 1). Correlation coefficients (Kendall's τ) for each variable are given in parentheses. Response signs were derived from signs of the correlation coefficients and regression parameters. Predictor variables are abbreviated as described in Table 2.1.

Response variables by site — Predictor NMS axes and associated variables

Sloped Field

	P1		P2		A2				
	Microstegium vimineum (0.57)	*Dactylis glomerata* (-0.47)	*Desmodium* sp. (-0.58)	*Dactylis glomerata* (0.43)	Vertical heterogeneity (1) (0.57)	Vegetation density (0.56)	Horizontal heterogeneity (-0.52)	Maximum height (0.44)	Herb cover (0.30)
Individuals (no.)	−	+	−	+	+	+	−	+	+
Richness					+	+	−	+	+
Diversity (D')	+	−			+	+	−	+	+

Tussock Grass Field

	P1				
	Lysimachia nummularia (0.69)	*Glechoma hederacea* (-0.68)	*Polygonum* sp. (0.39)	*Impatiens capensis* (-0.37)	*Solidago canadensis* (-0.33)
Richness	−	+	−	+	+
Diversity (D')	−	+	−	+	+

Deciduous Woodland

	P1			P3			
	Toxicodendron radicans (-0.56)	*Microstegium vimineum* (0.34)	*Pilea pumila* (0.32)	*Pilea pumila* (0.55)	*Waldsteinia fragarioides* (0.37)	*Euonymous* sp. (-0.35)	*Acer* sp. (0.31)
Individuals (no.)	−	+	+	−	−	+	−
Richness	−	+	+				

Table 2.3 Final regression models for spider species composition in each site. Response variables are spider ordination axes (S_). Predictor variables are as described in Table 1.

Site	Response	F-value	P-value	R^2	Predictor	Estimate	T-value	P-value
Sloped								
Field	S1	4.08	0.05	0.04	A1	0.15	2.02	0.05
N=106	S2	9.01	0.0002	0.15	P2	-0.24	-2.57	0.01
					P3	0.31	3.01	0.003
	S3	NS						
Tussock Grass								
Field	S1	6.71	0.01	0.07	A1	0.26	2.59	0.01
N=95	S2	20.28	<.0001	0.18	P1	-0.40	-4.50	<.0001
	S3	NS						
Deciduous								
Woodland	S1	9.46	0.003	0.11	A1	0.30	3.08	0.003
N=76	S2	6.97	0.01	0.09	P1	-0.28	-2.64	0.01
	S3	7.14	0.009	0.10	P3	0.32	2.67	0.009

Table 2.4 Responses of individual spider species to original plant and habitat architecture variables based on regression models for the Sloped Field (SF) (Table 3). Correlation coefficients (Kendall's τ) for each variable are given in parentheses. See Table 2 for explanation of response signs. Predictor variables are plant ordination axes (P_) and architecture ordination axes (A_); response variables are spider ordination axes (S_). Numbers in parentheses for Vertical heterogeneity indicate which of the two variables (NMS axes) were associated with the predictor.

Response NMS axes and associated variables by site	Predictor NMS axes and associated variables					
S1 (SF)	A1					
	Maximum height (-0.80)	Vertical heterogeneity (1) (-0.47)	Herb cover (-0.32)	Vertical heterogeneity (2) (-0.32)	Vegetation density (-0.30)	
Clubiona abboti (Clubionidae) (-0.45)	+	+	+	+	+	
Rabidosa rabida (Lycosidae) (-0.36)	+	+	+	+	+	
Neoantistea agilis (Hahniidae) (-0.33)	+	+	+	+	+	
S2 (SF)	P2		P3			
	Desmodium sp. (-0.58)	*Dactylis glomerata* (-0.43)	*Microstegium vimineum* (-0.44)	*Convolvulus arvensis* (-0.41)	*Diodia virginiana* (-0.33)	*Verbesina* sp. (-0.30)
Bathyphantes pallida (Linyphiidae) (-0.60)	+	–	–	–	+	+
Clubiona abboti (Clubionidae) (-0.37)	–	+	+	+	–	–
Gea heptagon (Araneidae) (-0.36)	–	+	+	+	–	–

Table 2.5 Responses of individual spider species to original plant and habitat architecture variables based on regression models for the Tussock Grass Field (TGF) (Table 3). Correlation coefficients (Kendall's τ) for each variable are given in parentheses. See Table 2 for explanation of response signs. Predictor variables are plant ordination axes (P_) and architecture ordination axes (A_); response variables are spider ordination axes (S_). Numbers in parentheses for Vertical heterogeneity indicate which of the two variables (NMS axes) were associated with the predictor.

Response NMS axes and associated variables by site	Predictor NMS axes and associated variables				
S1 (TGF)	A1				
	Horizontal heterogeneity (-0.73)	Vertical heterogeneity (1) (-0.41)	Vegetation density (-0.39)	Herb cover (-0.33)	
Pardosa milvina (Lycosidae) (-0.35)	+	−	−	−	
Clubiona abboti (Clubionidae) (-0.35)	−	+	+	+	
Pisaurina undulata (Pisauridae) (-0.32)	−	+	+	+	
Tetragnatha laboriosa (Tetragnathidae) (-0.31)	−	+	+	+	
S2 (TGF)	P1				
	Lysimachia nummularia (-0.69)	*Glechoma hederacea* (-0.68)	*Polygonum* sp. (-0.39)	*Impatiens capensis* (-0.37)	*Solidago canadensis* (-0.33)
Pardosa milvina (Lycosidae) (-0.48)	+	−	+	−	−
Hogna helluo (Lycosidae) (-0.39)	+	−	+	−	−
Bathyphantes pallida (Linyphiidae) (-0.35)	−	+	−	+	+
Clubiona abboti (Clubionidae) (-0.34)	−	+	−	+	+

Table 2.6 Responses of individual spider species to original plant and habitat architecture variables based on regression models for the Deciduous Woodland (DW) (Table 3). Correlation coefficients (Kendall's τ) for each variable are given in parentheses. See Table 2 for explanation of response signs. Predictor variables are plant ordination axes (P_) and architecture ordination axes (A_); response variables are spider ordination axes (S_). Numbers in parentheses for Vertical heterogeneity indicate which of the two variables (NMS axes) were associated with the predictor.

Response NMS axes and associated variables by site	Predictor NMS axes and associated variables			
S1 (DW)	**A1**			
	Horizontal heterogeneity (-0.91)	Vertical heterogeneity (2) (-0.37)		
Theridion frondeum (Theridiidae) (-0.59)	–	–		
S2 (DW)	**P1**			
	Toxicodendron radicans (-0.56)	*Microstegium vimineum* (-0.34)	*Pilea pumila* (-0.32)	
Habrocestum parvulum (Salticidae) (-0.39)	+	–	–	
Mangora maculata (Araneidae) (-0.38)	–	+	+	
Bathyphantes pallida (Linyphiidae) (-0.35)	–	+	+	
Lepthyphantes nebulosa (Linyphiidae) (-0.34)	+	–	–	
S3 (DW)	**P3**			
	Pilea pumila (-0.55)	*Waldsteinia fragaroides* (-0.37)	*Euonymous* sp. (-0.35)	*Acer* sp. (-0.31)
Lepthyphantes nebulosa (Linyphiidae) (-0.48)	–	–	+	–
Theridion frondeum (Theridiidae) (-0.39)	+	+	–	+
Cicurina arcuata (Agelenidae) (-0.31)	+	+	–	+
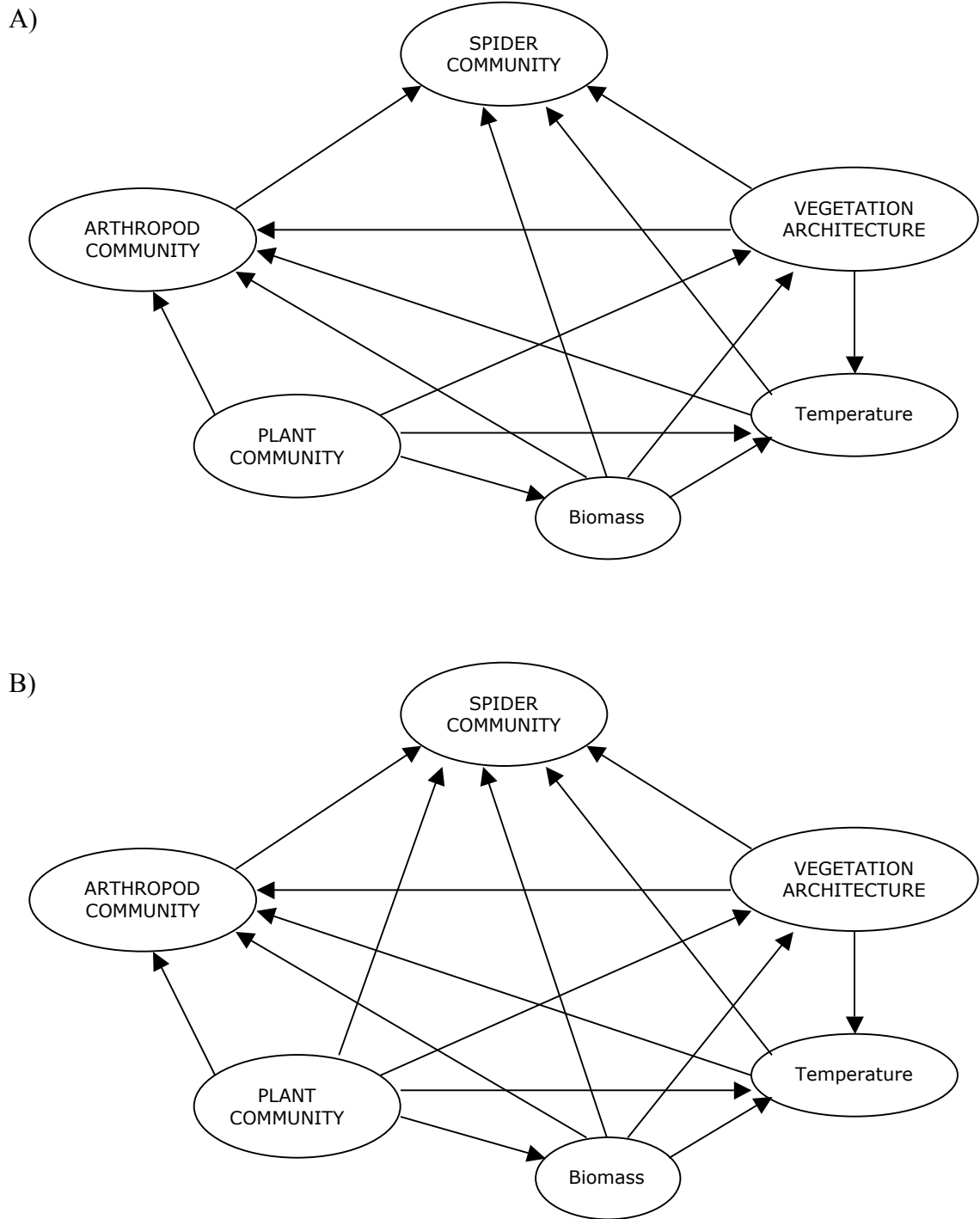
Figure 3.1 Graphical models of *a priori* hypotheses. Plants influence spiders indirectly (A) or directly (B). Variables in all capitals were represented by multiple NMS ordination axes in analyses (see text).

Table 3.1 Stress values for NMS configurations and proportion variation explained by each axis for spider, arthropod, plant, and vegetation structure data.

Data matrix	Stress	Variation explained			
		Axis 1	Axis 2	Axis 3	Cumulative
Spiders	20.00	9.60%	32.40%	23.40%	65.50%
Arthropods	18.31	22.20%	36.20%	14.40%	72.80%
Plants	13.03	19.00%	48.20%	21.50%	88.60%
Structure	14.54	19.80%	21.70%	44.30%	85.70%
Herbivore/scavenger	18.41	14.60%	31.20%	27.30%	73.10%
Predator/parasite	13.76	25.50%	28.10%	28.00%	81.60%

Table 3.2 Correlations between NMS axes and original variables for all data matrices. Variables with correlations (Kendall's τ) $\tau \geq |0.30|$ are listed. See Methods for explanation of variables associated with Structure NMS axes.

NMS axes and associated variables	Correlation with NMS axis	NMS axes and associated variables	Correlation with NMS axis
Spider 1		Plant 1	
Clubiona abbotii (Clubionidae)	0.56	*Calystegia sepium*	-0.38
Eperigone trilobata (Linyphiidae)	-0.38	*Dactylis glomerata*	0.35
Florinda coccinea (Linyphiidae)	-0.48	*Polygonum caespitosum*	-0.40
Theridion frondeum (Theridiidae)	-0.31	*Toxicodendron radicans*	0.32
Spider 2		*Solanum ptychanthemum*	0.40
Pisaurina undulata (Pisauridae)	0.31	*Stellaria media*	0.37
Theridion cheimatos (Theridiidae)	-0.36	Plant 2	
Spider 3		*Desmodium* sp.	0.47
Bathyphantes pallida (Linyphiidae)	0.37	*Calystegia sepium*	0.64
Eridantes erigonoides (Linyphiidae)	-0.45	*Dactylis glomerata*	-0.37
Oxyopes sp. (Oxyopidae)	-0.45	*Glechoma hederacea*	-0.58
Arthropod 1		*Toxicodendron radicans*	-0.32
Orthoptera	-0.61	*Pilea pumila*	-0.35
Diplura	0.31	Plant 3	
Opiliones	0.31	*Desmodium* sp.	-0.41
Arthropod 2		*Dactylis glomerata*	0.60
Diptera	-0.31	*Stellaria media*	0.30
Hymenoptera	-0.48	Structure 1	
Hemiptera	-0.37	Herb cover	-0.63
Acari	-0.33	Glabrous cover	-0.39
Isopoda	0.42	Structure 2	
Arthropod 3		Horizontal heterogeneity	-0.41
Hymenoptera	-0.35	Vine cover	-0.63
Collembola	-0.36	Structure 3	
Diplopoda	-0.39	Density	-0.39
Homoptera	-0.40	Grass cover	-0.76
Opiliones	-0.35	Glabrous cover	-0.50

Table 3.3 Model selection criteria for comparison of two *a priori* models of spider community structure for each spider NMS axis. Model A contains no direct paths from plants to spiders, while Model B incorporates these (see Figure 3.1). The specified model is the a priori model, the saturated model contains all possible paths between variables, and the independence model contains no paths. AIC and CAIC are discussed in text.

Spider axes	Model A		Model B	
	AIC	CAIC	AIC	CAIC
Spider 1				
Specified model	196.65	432.78	200.80	445.89
Saturated model	182.00	454.00	182.00	454.00
Independence model	327.20	366.05	327.20	366.05
Spider 2				
Specified model	196.14	432.27	200.80	445.89
Saturated model	182.00	454.00	182.00	454.00
Independence model	337.38	376.23	337.38	376.23
Spider 3				
Specified model	202.64	438.77	200.80	445.89
Saturated model	182.00	454.00	182.00	454.00
Independence model	347.63	386.49	347.63	386.49

Table 3.4 Measures of absolute fit for final path models. See text for explanations of CMIN/DF, CFI and RMSEA.

Final path model	χ^2	df	Global p-value	CMIN/DF	CFI	RMSEA
Spider 1	43.44	53	0.82	0.82	1.00	<0.01
Spider 2	46.16	54	0.77	0.86	1.00	<0.01
Spider 3	46.15	53	0.74	0.87	1.00	<0.01
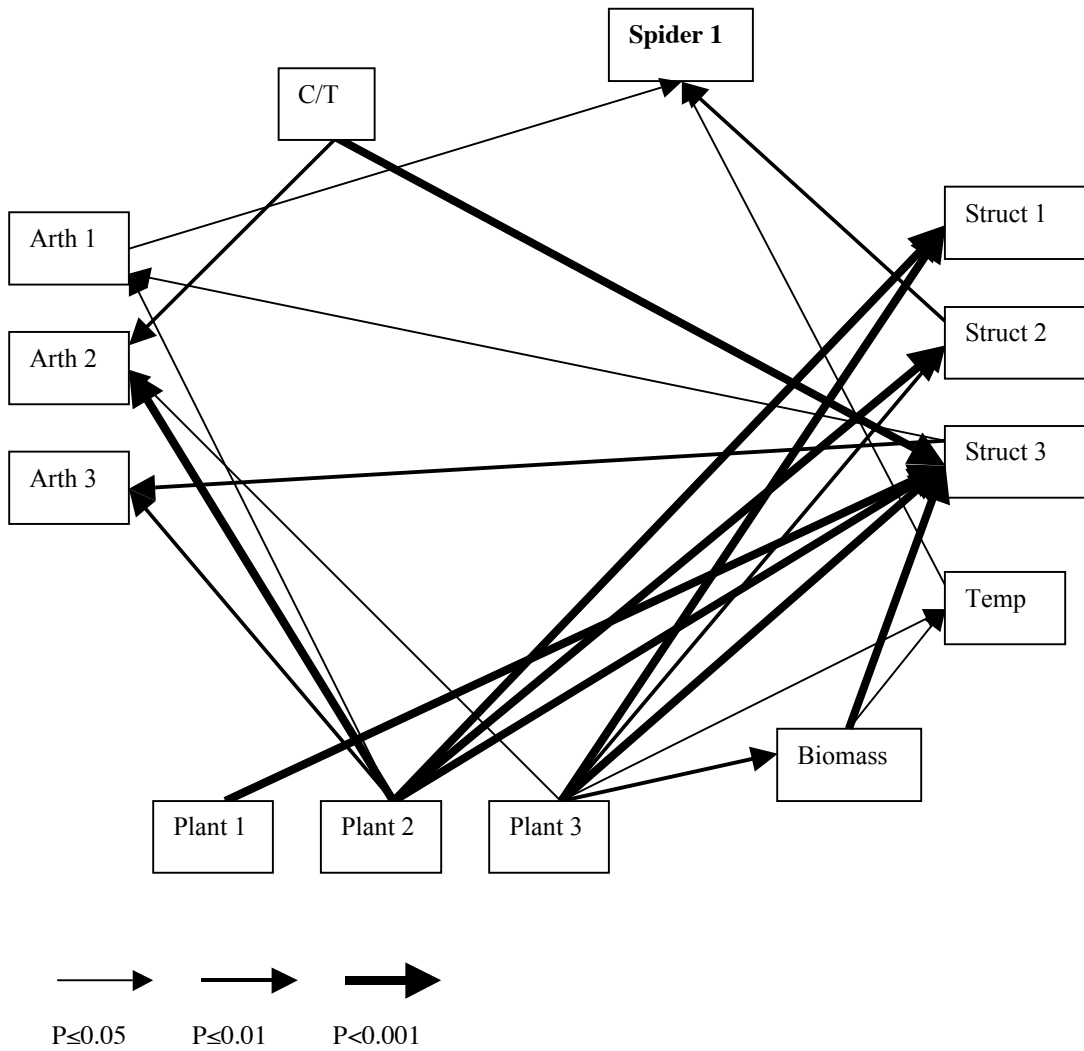
Figure 3.2 Final path model for Spider 1. Differing significance levels for each path coefficient are indicated by the thickness of the arrows. Unexplained error variances for each variable are omitted from the figure for clarity. Path coefficients are not included in this and following figures because the direction of their effects depends on the sign of the correlations between original variables and the NMS axes (see text).
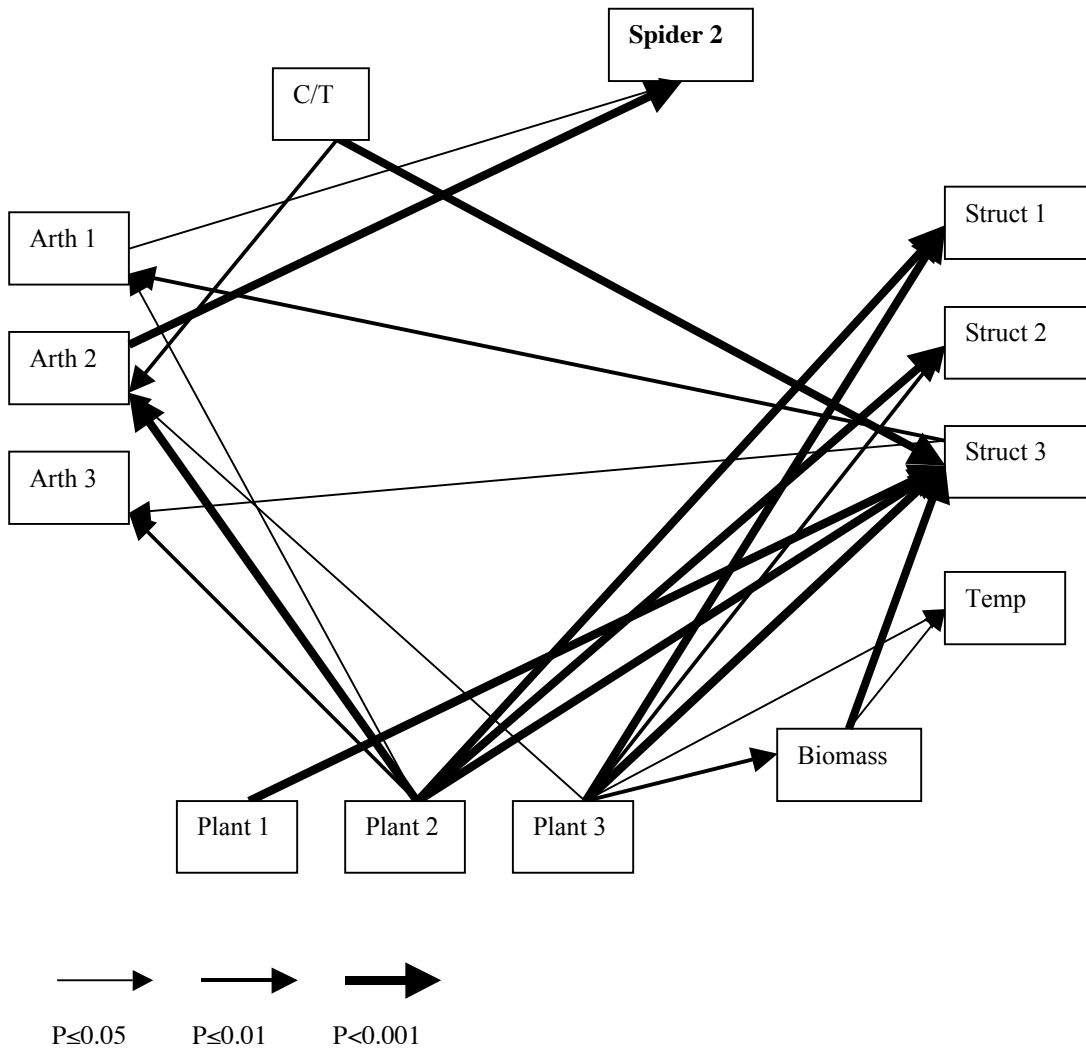
Figure 3.3 Final path model for Spider 2. Differing significance levels for each path coefficient are indicated by the thickness of the arrows. Unexplained error variances for each variable are omitted from the figure for clarity.
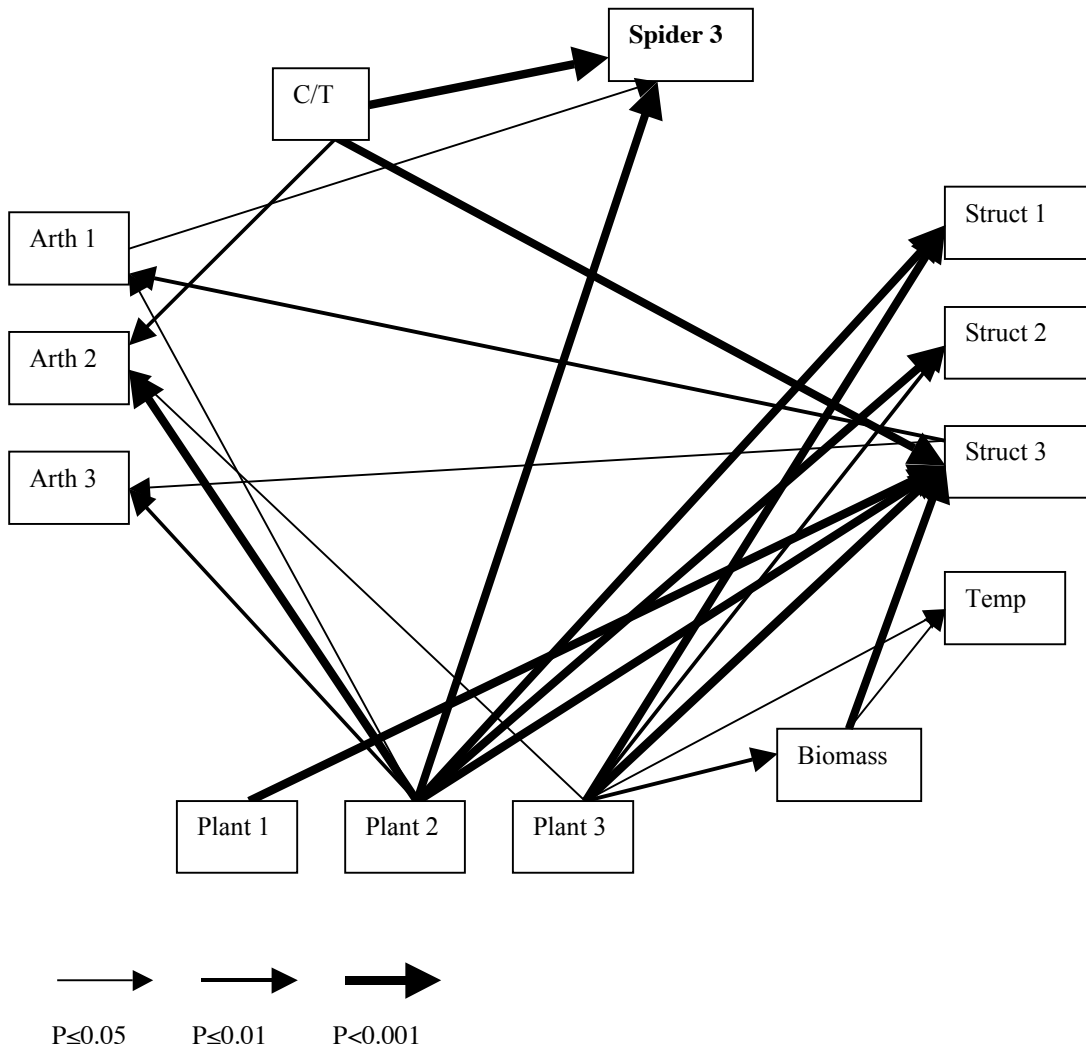
Figure 3.4 Final path model for Spider 3. Differing significance levels for each path coefficient are indicated by the thickness of the arrows. Unexplained error variances for each variable are omitted from the figure for clarity.

Table 3.5 Coefficients of multiple determination for variables from path models for all spider NMS axes.

Spider 1 model	R^2	Spider 2 model	R^2	Spider 3 model	R^2
Spider 1	0.20	Spider 2	0.31	Spider 3	0.46
Arth 1	0.33	Arth 1	0.33	Arth 1	0.33
Arth 2	0.38	Arth 2	0.38	Arth 2	0.38
Arth 3	0.16	Arth 3	0.16	Arth 3	0.16
Biomass	0.13	Biomass	0.13	Biomass	0.13
Struct 1	0.31	Struct 1	0.31	Struct 1	0.31
Struct 2	0.42	Struct 2	0.42	Struct 2	0.42
Struct 3	0.83	Struct 3	0.83	Struct 3	0.83
Temp	0.23	Temp	0.23	Temp	0.23

Table 3.6 Total effects for all predictor variables in path models for spider NMS axes (Figures 3.2-3.4). Effects of predictors on non-spider variables were identical across path models.

Variable	C/T	Biomass	Plant 1	Plant 2	Plant 3	Struct3
Arth1	0.15	-0.08	-0.10	0.44	-0.28	0.41
Arth2	-0.61	0	0	-0.50	-0.26	0
Arth3	0.11	-0.06	-0.07	-0.31	-0.20	0.30
Biomass	0	0	0	0	0.34	0
Struct1	0	0	0	0.42	0.40	0
Struct2	0	0	0	-0.60	-0.30	0
Struct3	0.36	-0.19	-0.25	0.36	-0.68	0
Temp	0	0.31	0	0	0.38	0

Table 3.7 Total, direct and indirect effects of predictor variables on spider NMS axes (Figures 3.2-3.4). C/T is the nominal variable for control versus treatment quadrats. Effects are unstandardized.

Effects	C/T	Arth 1	Arth 2	Biomass	Plant 1	Plant 2	Plant 3	Struct 2	Struct 3	Temp
Spider 1										
Total Effects	-0.04	-0.26	0	-0.07	0.03	0.09	0.07	-0.34	-0.11	-0.27
Direct Effects	0.00	-0.26	0	0.00	0.00	0.00	0.00	-0.34	0.00	-0.27
Indirect Effects	-0.04	0.00	0	-0.07	0.03	0.09	0.07	0.00	-0.11	0.00
Spider 2										
Total Effects	-0.31	-0.26	0.44	0.02	0.03	-0.34	-0.04	0	-0.11	0
Direct Effects	0.00	-0.26	0.44	0.00	0.00	0.00	0.00	0	0.00	0
Indirect Effects	-0.31	0.00	0.00	0.02	0.03	-0.34	-0.04	0	-0.11	0
Spider 3										
Total Effects	-0.80	0.26	0	-0.02	-0.03	-0.53	-0.07	0	0.11	0
Direct Effects	-0.84	0.26	0	0.00	0.00	-0.64	0.00	0	0.00	0
Indirect Effects	0.04	0.00	0	-0.02	-0.03	0.11	-0.07	0	0.11	0

Table 3.8 Responses of individual spider species associated with spider NMS axis 1 (S1). Original variables are grouped by the sign of their correlation with their respective NMS axes; in the case of Struct 2 and Struct 3, all associated variables had negative correlations. Response directions are based on total effects (Table 3.6). See Figure 3.2 for path model.

Predictors	Variables associated with NMS axes	S1			
		C. abbotti	E. trilobata	F. Coccinea	T. frondeum
Arth 1 (+)	Diplura Opiliones	−†	+†	+†	+†
Arth 1 (−)	Orthoptera	+†	−†	†	−†
Biomass		−	+	+	+
Plant 1 (+)	*D. glomerata* *T. radicans* *S. ptychanthemum* *S. media*	+	−	−	−
Plant 1 (−)	*C. sepium* *P. caespitosum*	−	+	+	+
Plant 2 (+)	*Desmodium* sp. *C. sepium*	+	−	−	−
Plant 2 (−)	*D. glomerata* *G. hederacea* *T. radicans* *P. pumila*	−	+	+	+
Plant 3 (+)	*D. glomerata* *S. media*	+	−	−	−
Plant 3 (−)	*Desmodium* sp.	−	+	+	+
Struct 2	Horizontal heterogeneity Vine cover	+†	−†	−†	−†
Struct 3	Density Grass cover Glabrous cover	+	−	−	−
Temp		−†	+†	+†	+†

† Effects greater than or equal to 0.20.

Table 3.9 Responses of individual spider species associated with spider NMS axis 2 (S2). Original variables are grouped by the sign of their correlation with their respective NMS axes; in the case of Struct 3, all associated variables had negative correlations. Response directions are based on total effects (Table 3.6). See Figure 3.3 for path model.

Predictors	Variables associated with NMS axes	S2	
		P. undulata	*T. cheimatos*
Arth 1 (+)	Diplura		
	Opiliones	$-^\dagger$	$+^\dagger$
Arth 1 (−)	Orthoptera	$+^\dagger$	$-^\dagger$
Arth 2 (+)	Isopoda	+	−
Arth 2 (−)	Diptera		
	Hymenoptera		
	Hemiptera	−	+
	Acari		
Biomass		+	−
Plant 1 (+)	*D. glomerata*		
	T. radicans		
	S. ptychanthemum	+	−
	S. media		
Plant 1 (−)	*C. sepium*		
	P. caespitosum	−	+
Plant 2 (+)	*Desmodium* sp.		
	C. sepium	$-^\dagger$	$+^\dagger$
Plant 2 (−)	*D. glomerata*		
	G. hederacea		
	T. radicans	$+^\dagger$	$-^\dagger$
	P. pumila		
Plant 3 (+)	*D. glomerata*		
	S. media	−	+
Plant 3 (−)	*Desmodium* sp.	+	−
Struct 3	Density		
	Grass cover	+	−
	Glabrous cover		

† Effects greater than or equal to 0.20.

Table 3.10 Responses of individual spider species associated with spider NMS axis 3 (S3). Original variables are grouped by the sign of their correlation with their respective NMS axes; in the case of Struct 3, all associated variables had negative correlations. Response directions are based on total effects (Table 3.6). See Figure 3.4 for path model.

Predictors	Variables associated with NMS axes	S3		
		B. pallida	*E. erigonoides*	*Oxyopes sp.*
C/T		$-^{\ddagger}$	$+^{\ddagger}$	$+^{\ddagger}$
Arth 1 (+)	Diplura Opiliones	$+^{\dagger}$	$-^{\dagger}$	$-^{\dagger}$
Arth 1 (−)	Orthoptera	$-^{\dagger}$	$+^{\dagger}$	$+^{\dagger}$
Biomass		−	+	+
Plant 1 (+)	*D. glomerata* *T. radicans* *S. ptychanthemum* *S. media*	−	+	+
Plant 1 (−)	*C. sepium* *P. caespitosum*	+	−	−
Plant 2 (+)	*Desmodium* sp. *C. sepium*	$-^{\ddagger}$	$+^{\ddagger}$	$+^{\ddagger}$
Plant 2 (−)	*D. glomerata* *G. hederacea* *T. radicans* *P. pumila*	$+^{\ddagger}$	$-^{\ddagger}$	$-^{\ddagger}$
Plant 3 (+)	*D. glomerata* *S. media*	−	+	+
Plant 3 (−)	*Desmodium* sp.	+	−	−
Struct 3	Density Grass cover Glabrous cover	−	+	+

[†] Effects greater than or equal to 0.20.
[‡] Effects greater than or equal to 0.50.

Table 3.11 Final regression models for spider NMS axes. Spider axes were regressed on herbivore/scavenger NMS axes (Herb_) and predator/parasite NMS axes (Pred_), with control/treatment (C/T) as a nominal variable.

Response	F-value	P-value	R^2 (adj)	Predictor	Estimate	T-value	P-value
S1	ns						
S2	17.67	<0.0001	0.39	Herb2	-0.68	-5.90	<0.0001
				C/T	-0.31	-2.70	0.009
S3	9.69	<0.0001	0.50	Herb2	-0.50	-4.27	<0.0001
				Herb3	0.25	2.29	0.03
				Pred3	0.13	1.05	0.30
				C/T	0.20	1.82	0.07
				C/T x Herb2	0.47	3.88	0.0003
				C/T x Pred3	-0.26	-2.19	0.03
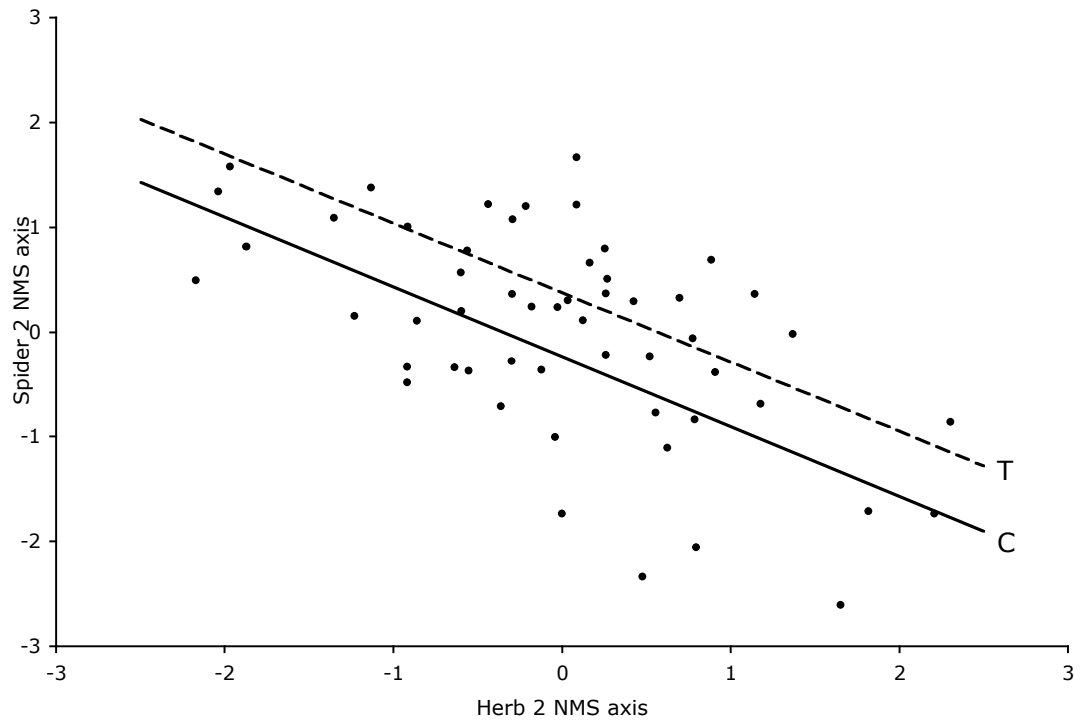
Figure 3.5 Final regression model for spider NMS axis 2 (S2). Regression lines illustrate the difference in intercepts, but not slopes, in spider response to the herbivore/scavenger NMS axis in control (C) and treatment (T) quadrats.

Table 3.12 Responses of individual spider species to herbivore/scavenger and predator/parasite orders based on regression models (Table 3.12).

NMS axes and associated variables	Species associated with S2		Species associated with S3		
	P. undulata	*T. cheimatos*	*B. pallida*	*E. erigonoides*	*Oxyopes sp.*
Herb 2					
Diptera	−	+	−[†]	+[†]	+[†]
Isopoda	+	−	+[†]	−[†]	−[†]
Herb 3					
Hemiptera			−	+	+
Opiliones			−	+	+
Pred 3					
Diptera			−[†]	+[†]	+[†]
Hemiptera			+[†]	−[†]	−[†]
Acari			−[†]	+[†]	+[†]
Diplura			−[†]	+[†]	+[†]

[†] Responses only significant in treatment quadrats

APPENDIX B

Spider species list by family

Agelenidae
  Cicurina arcuata
  Wadotes sp.
Amaurobiidae
  Amaurobius sp.
Antrodiaetidae
  Antrodiaetus unicolor
Anyphaenidae
  Anyphaena pectorosa
Araneidae
  Acanthepeira sp.
  Araneus sp.
  Argiope aurantia
  Argiope trifasciata
  Gea heptagon
  Mangora maculata
  Neoscona hentzii
  Nuctenea sp.
  Verrucosa arenata
Clubionidae
  Castianeira gertschi
  Chiracanthium sp.
  Clubiona abbotti
  Clubionoides sp.
  Phrurolithus fratrellus
  Phrurotimpus alarius
  P. borealis
  Trachelas similis
Gnaphosidae
  Drassylus eremitus
  Zelotes sp.
Hahniidae
  Neoantistea agilis
Leptonetidae
  Leptoneta gertschi
Linyphiidae
  Bathyphantes pallida
  Centromerus cornupalpis
  Ceraticelus fissiceps
  C. laetabilis
  Ceratinopsis laticeps
  Eperigone maculata
  E. tridentata
  E. trilobata
  Eridantes erigonoides
  Florinda coccinea
  Frontinella communis
  Grammonota ornata
  G. pictilis
  Graphomoa theridioides
  Islandia flaveola
  Lepthyphantes nebulosa
  Meioneta longipes

  Meioneta micaria
  Meioneta picta
  M. unimaculata
  M. zygia
  Microneta viaria
  Neriene clathrata
  N. maculata
  Origanates rostratus
  Walckenaeria spiralis
  species A (Erigoninae)
  species B (Erigoninae)
  species C (Erigoninae)
Lycosidae
  Allocosa funerea
  Hogna helluo
  Pardosa milvina
  Pirata arenicola
  P. insularis
  P. sylvanus
  Rabidosa rabida
  Rabidosa punctulata
  Schizocosa crassipes
  Trabea aurantiaca
  Trochosa sp.
Mimetidae
  Ero pensacolae
Oxyopidae
  Oxyopes sp.
Philodromidae
  Philodromus sp.
Pisauridae
  Dolomedes sp.
  Pisaurina undulata
  Pisaurina sp.
Salticidae
  Ballus sp.
  Eris militaris
  Habrocestum parvulum
  Habronattus coronatus
  Maevia sp.
  Marpissa lineata
  Phiddipus audax
  Sitticus floridanus
  Thiodina sylvana
  Zygoballus bettini
  Zygoballus nervosus
Tetragnathidae
  Leucauge venusta
  Pachygnatha autumnalis
  Tetragnatha sp.
Theridiidae
  Euryopis funebris
  Robertus sp.

  Steatoda sp.
  Theridion albidum
  T. cheimatos
  T. frondeum
  T. neshamini
  Theridula opulenta
  species A
  species B
Thomisidae
  Misumena vatia
  Misumenoides formosipes
  Misumenops sp.
  Ozyptila monroensis
  Synema sp.
  Tmarus sp.
  Xysticus ferox
Zoridae
  Zora pumilis

APPENDIX C

Use of multisample cluster analysis to examine relationships between three sites

based on three types of data

Introduction and problem statement

Traditional multiple comparison procedures are limited to pairwise comparisons of the K groups or samples that comprise a data set. The basis of these comparisons is typically an F-statistic chosen based on a level of alpha determined to represent an acceptable risk of Type I error by the researcher. While there are "rules of thumb" regarding this selection, the choice of alpha is ultimately an arbitrary one, and in the case of multiple comparisons it is difficult to be certain that the nominal alpha level actually corresponds to the overall error rate of the test (Bozdogan 1986). Moreover, because these procedures only examine relationships between two samples at a time, they preclude the possibility of uncovering potential homogeneous clusters of more than two samples. They also tend to assume homogeneity of variance (Bozdogan 1986; Cox and Cowpertwait 1992).

Multiple comparison procedures have also largely been limited to the univariate case, and few procedures are available for dealing with multivariate data (Bozdogan, 1986). Ecological data are rarely univariate; the complexity of natural systems and their interacting components often require that large numbers of variables be measured. However, we frequently wish to be able to compare samples or treatments to determine which can be reasonably assumed to have come from the same population. There is also considerable interest in determining whether *multiple* groups are homogeneous (rather than just comparing them two at a time), and a robust method is needed that takes into account the complexity (e.g., number of parameters or covariance structure) of the models.

146

Many authors have argued against the use of multiple comparison procedures in general (Nelder 1971; Plackett 1971) or against specific procedures (Saville 1990). Concerns include the utility of these procedures (Nelder 1971), the problems of testing multiple hypotheses simultaneously without being able to reliably control the experiment-wise error rate (Nelder 1971; Plackett 1971; Bozdogan 1986), and the lack of consistency of many of the procedures (Saville 1990). However, the comparison of groups or treatments is a critical question in many research programs. Plackett, in his discussion of O'Neill and Wetherill (1971), suggested abandoning traditional methods in favor of cluster analysis, and this approach was taken by Scott and Knott as early as 1974. However, until the work of Bozdogan (1986) these approaches still relied on more traditional methods of model selection (likelihood ratio test).

Statistical method

Multisample Cluster Analysis (Bozdogan 1986) clusters K groups or samples into k homogeneous groups (where $k \leq K$) using a clustering algorithm in which all possible partitions of the K groups are formed. The different partitioning schemes are compared using Bozdogan's information complexity criterion, ICOMP (Bozdogan 1988, 1990, 1993, 1994) and Akaike's information criterion, AIC (Akaike 1973, 1974). Three possible models are evaluated for each partition scheme: Model 1 (varying means and varying variance/covariance), Model 2 (varying means and common variance/covariance), and Model 3 (common means and common variance/covariance). These model selection criteria seek the most parsimonious model by penalizing models for either the number of parameters (AIC) or the complexity of the model (ICOMP).

147

Both criteria have two components: lack of fit of the model, and a penalty term. The model with the smallest value of either AIC and/or ICOMP is considered the best model. Unlike the frequentist approach, criteria using information theory allow one to compare multiple models simultaneously and don't rely on an arbitrarily chosen level of alpha. Analyses were performed using MATLAB code (MATLAB 6.5.1, 1984-2000, The Mathworks, Inc., Natick, Massachusetts) provided by H. Bozdogan of the University of Tennessee's Department of Statistics.

Data collection

The data used for this project are from Beals 2006. Three adjacent sites at the University of Tennessee's Cherokee Woodlot were sampled in 1997 and 1998. The Sloped Field (site 1) is a grassy area maintained by periodic mowing throughout the summer. The Tussock Grass Field (site 2) lies to the southeast of the Sloped Field, and is separated from it by a band of shrubs approximately five meters wide. This site is maintained as a field by flooding in the spring and early summer. The Deciduous Woodland (site 3) is a forested area located to the southwest of the Sloped Field. Circular quadrats (0.1 m^2) were located in each site using the random walk method based on Catana, 1955 ($n=108$ in Sloped Field, $n=100$ in Tussock Grass Field, $n=80$ in Deciduous Woodland, and total $n=288$). Within each quadrat, data were recorded for the plant species (percent of the quadrat covered by each species) and the architecture or structure of the plants/habitat. Following these measurements, spiders were collected from each quadrat and subsequently identified to species.

Eleven variables were measured for the habitat architecture, including the density

of the vegetation, the amount of open space in the quadrat, and the growth forms of the

plants. Sixty-eight plant species and 103 spider species were observed. To achieve

variable reduction I used nonmetric multidimensional scaling (NMS) as implemented by

the software program PC-ORD (PC-ORD 4, 1999, MjM Software Design, Gleneden

Beach, Oregon). PC-ORD's NMS performs multiple runs with real data followed by

multiple runs with randomized data to select the best dimensionality for the data based on

minimizing stress (stress is a measure of lack of fit [McCune and Grace 2002]). For both

the plant and spider species a three-dimensional solution was recommended and these

axes were then used as the variables for the Multisample Cluster Analysis.

Results

Results for the plant species data are presented in Table C.1. The lowest ICOMP

(428.2) and AIC (439.4) scores occurred for Model 1 (varying means and varying

variance/covariance) and the fifth partition scheme (k=3). The optimal decision tree

classifier is shown in Figure C.1. The next best score is for k=2, and the best partition

scheme combines sites 1 and 3 (Sloped Field and Deciduous Woodland) based on the

plant species data. The worst of these three partition schemes is the one in which k=1

(ICOMP=1389.4, AIC=1406.2).

Table C.1 ICOMP and AIC values for plant species data. Subscripts indicate different models: Model 1 (varying means and varying variance/covariance), Model 2 (varying means and common variance/covariance), and Model 3 (common means and common variance/covariance).

Partition Scheme	$ICOMP_1$	AIC_1	$ICOMP_2$	AIC_2	$ICOMP_3$	AIC_3
(1, 2, 3)	1389.4	1406.2	1389.4	1406.2	1389.4	1406.2
(1, 2) (3)	921.5	937.2	988.5	983.4	1382.3	977.4
(1, 3) (2)	739.0	751.2	922.5	918.2	1383.0	912.2
(1) (2, 3)	886.7	905.3	998.2	996.3	1383.1	990.2
(1) (2) (3)	428.2*	439.4*	556.3	521.9	1375.2	509.9

			k	ICOMP	AIC
Stage 1	(1,2,3)		1	1389.4	1406.2
Stage 2	(1, 3)	(2)	2	739	751.2
Stage 3	(1) (3)	(2)	3*	428.2*	439.4*

Figure C.1 Optimal decision tree classifier for plant species data. ICOMP and AIC values for stages two and three are based on the model with varying means and varying variance/covariance (Model 1).

For the habitat architecture data ICOMP and AIC did not choose the same partition scheme, although they did select the same model (Model 1 [varying means and varying variance/covariance]) (Table C.2). ICOMP (-2235.1) selected the partition scheme in which $k=2$, with the two grassy fields clustered together and the forested area separate. This makes intuitive sense. AIC's first choice (AIC=-5018.2) is the partition with $k=3$ groups, but the next smallest score (-4741.1) is for the same partition as ICOMP. The optimal decision tree classifier (Figure C.2) suggests that a partition of $k=1$ (i.e., no partitioning) is the next best choice (ICOMP=-1917.1, AIC=-3146.0), followed by $k=3$ (ICOMP=-1279.4, AIC=-5018.2).

150

Table C.2 ICOMP and AIC values for habitat architecture data. Subscripts indicate different models: Model 1 (varying means and varying variance/covariance), Model 2 (varying means and common variance/covariance), and Model 3 (common means and common variance/covariance).

Partition Scheme	$ICOMP_1$	AIC_1	$ICOMP_2$	AIC_2	$ICOMP_3$	AIC_3
(1, 2, 3)	-1917.1	-3146.0	-1917.1	-3146.0	-1917.1	-3146.0
(1, 2) (3)	-2235.1*	-4741.1	-1696.0	-3508.9	-1942.4	-3530.9
(1, 3) (2)	-1370.5	-3806.3	-1604.5	-3409.5	-1941.1	-3431.5
(1) (2, 3)	-1336.4	-3883.4	-1504.3	-3371.7	-1940.7	-3393.7
(1) (2) (3)	-1279.4	-5018.2*	-1300.9	-3730.5	-1968.9	-3774.5

		k	ICOMP	AIC
Stage 1	(1,2,3)	1	-1917.1	-3146.0
Stage 2	(1, 2) (3)	2*	-2235.1*	-4741.1
Stage 3	(1) (2) (3)	3	-1279.4	-5018.2*

Figure C.2 Optimal decision tree classifier for habitat architecture data. ICOMP and AIC values for stages two and three are based on the model with varying means and varying variance/covariance (Model 1).

For the spider species data, ICOMP and AIC again both selected Model 1 (Table C.3). The partition scheme chosen was $k=3$ (ICOMP=1080.6, AIC=1113.5). The optimal decision tree classifier (Figure C.3) suggests that the next best model is partition scheme two (Sloped Field and Tussock Grass Field together, with Deciduous Woodland separate) (ICOMP=1119.8, AIC=1143.8), and the worst of these three is $k=1$ (ICOMP=1429.5, AIC=1446.5).

Table C.3 ICOMP and AIC values for spider species data. Subscripts indicate different models: Model 1 (varying means and varying variance/covariance), Model 2 (varying means and common variance/covariance), and Model 3 (common means and common variance/covariance).

Partition Scheme	$ICOMP_1$	AIC_1	$ICOMP_2$	AIC_2	$ICOMP_3$	AIC_3
(1, 2, 3)	1429.5	1446.5	1429.5	1446.5	1429.5	1446.5
(1, 2) (3)	1119.8	1143.8	1139.9	1140.5	1422.6	1134.5
(1, 3) (2)	1315.5	1342.3	1377.8	1384.8	1422.9	1378.8
(1) (2, 3)	1295.2	1322.2	1364.0	1371.2	1423.1	1365.2
(1) (2) (3)	1080.6*	1113.5*	1125.7	1114.1	1415.4	1102.1

		k	ICOMP	AIC
Stage 1	(1,2,3)	1	1429.5	1446.5
Stage 2	(1, 2) (3)	2	1119.8	1143.8
Stage 3	(1) (2) (3)	3*	1080.6*	1113.5*

Figure C.3 Optimal decision tree classifier for spider species data. ICOMP and AIC values for stages two and three are based on the model with varying means and varying variance/covariance (Model 1).

Conclusions and discussion

The Multisample Cluster Analysis approach of Bozdogan (1986) is an effective method for elucidating parsimonious structure in multivariate data sets. The method produces similar results to those of traditional ordination analyses such as Bray-Curtis (or Polar) ordination, which has been used extensively in the plant ecology literature. However, the advantages of this new approach are that 1) the results provide information about the means and variance/covariance of the groups, 2) the model selection method (information criteria) can be relied on to choose the most parsimonious structure for the data, taking both lack of fit and model complexity into account, and 3) this approach

produces an easily interpretable decision tree classifier. For example, in the case where *k*=2, both the habitat architecture data and the spider species data cluster sites 1 and 2 (the two grassy fields) together, suggesting a possible relationship between spider species and the architecture of their environment and therefore potential future research directions. In the case of the plant species, the clustering for *k*=2 groups sites 1 and 3 (one of the grassy fields and the forested area) separately from the other grassy field. Although all three sites are adjacent, there is a gradation between sites 1 and 3 (the edge of the woods is distinct, but there are forest plant species in the Sloped Field near the edges, and some of the Sloped Field species are found a little ways into the Deciduous Woodland), while the Tussock Grass Field is more disjunct (it is separated from the Sloped Field by a band of shrubs about five to ten meters wide). The results therefore agree with empirical observations.

References

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csake (eds.), Second International Symposium on Information Theory. Budapest: Akademiai Kiado, pp. 267-281.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control, AC-19:716-723.

Beals, M..L. 2006. Understanding community structure: a data-driven multivariate approach. Oecologia DOI 10.1007/s00442-006-0551-8.

Bozdogan, H. 1986. Multi-sample cluster analysis as an alternative to multiple comparison procedures. Bulletin of Informatics and Cybernetics 22:95-130.

Bozdogan, H. 1988. ICOMP: A new model-selection criterion. In H. H .Bock (ed.), Classification and Related Methods of Data Analysis. Amsterdam: Elsevier Science Publishers B. V. (North-Holland), pp.599-608.

Bozdogan, H. 1990. On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. Communications in Statistics, Theory and Methods 19:221-278.

Bozdogan, H. 1993. Choosing the number of component clusters in the mixture-model using a new informational complexity criterion of the inverse-Fisher information matrix. In O. Opitz, B. Lausen, and R. Klar (eds.), Studies in Classification, Data Analysis, and Knowledge Organization. Springer-Verlag, Heidelberg, pp.40-54.

Bozdogan, H. 1994. Mixture-model cluster analysis using a new informational complexity and model selection criteria. In H. Bozdogan (ed.), Multivariate Statistical Modeling, Volume 2, Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling. Dordrecht: Kluwer Academic Publishers, the Netherlands, pp.69-113.

Catana, Jr., A.J. 1955. The wandering quadrant—a new ecological sampling method utilizing interspace measurements. Bulletin of the Ecological Society of America 36:88.

Cox, T.F. and P.S.P. Cowpertwait. 1992. Clustering population means under heterogeneity of variance. The Statistician 41:591-598.

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

Nelder, J.A. 1971. Discussion of the present state of multiple comparison methods by O'Neill and Wetherill. Journal of the Royal Statistical Society, Series B 33:244-246.

O'Neill, R. and G.B. Wetherill. 1971. The present state of multiple comparison methods. Journal of the Royal Statistical Society, Series B 33:218-241.

Plackett, R.L. 1971. Discussion of the present state of multiple comparison methods by O'Neill and Wetherill. Journal of the Royal Statistical Society, Series B 33:242-244.

Saville, D.J. 1990. Multiple comparison procedures: the practical solution. The American Statistician 44:174-180.

Scott, A.J. and M. Knott. 1974. A cluster analysis method for grouping means in the analysis of variance. Biometrics 30:507-512.

VITA

Monica Lynn Beals was born in Madison, Wisconsin on 24 September, 1970. She graduated from Earlham College in Richmond, Indiana in 1992 with a Bachelor of Arts in Biology. She received her Master of Science in Ecology and Evolutionary Biology at the University of Tennessee, Knoxville in 2001. In 2006, she received her doctorate in Ecology and Evolutionary Biology and a Master of Science in Statistics from the University of Tennessee.