2006

Spatially explicit null models in biogeography: Toward a multi-scale understanding of the niche.

Jeremy VanDerWal
University of Windsor

Recommended Citation

VanDerWal, Jeremy, "Spatially explicit null models in biogeography: Toward a multi-scale understanding of the niche." (2006). *Electronic Theses and Dissertations*. 3551.
https://scholar.uwindsor.ca/etd/3551

SPATIALLY EXPLICIT NULL MODELS IN BIOGEOGRAPHY: TOWARD A MULTI-SCALE UNDERSTANDING OF THE NICHE

by

Jeremy VanDerWal

A Dissertation
Submitted to the Faculty of Graduate Studies and Research
through Biological Sciences
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada
2006

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-17131-8
Our file Notre référence
ISBN: 978-0-494-17131-8

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

ABSTRACT

In general, most ecologists envision the “niche” as a central organizing tenet, and that particular parameters of the niche help structure biogeographic patterns of diversity, distribution and abundance. The major emergent alternative to the niche concept requires the inference of background stochasticity, and its application through null models. For example, rather than competitive interactions of species shaping the coexistence of species, “historical accidents of dispersal” have been suggested. In this thesis I explore, in some detail, the concept of niche using of null models.

In this thesis, two detailed and quite different null models are presented. The first, based on the “Mid-Domain Effect” (MDE), explores the influence of continental geometry on patterns in species richness and range size frequency distributions. I compared the MDE predictions first to observations on tree species richness in continental North America (n = 417 species), and then to amphibian, bird and mammal species richness across North and South America (n = 2216, 3771 and 1605 species, respectively) contrasting the relative contributions of null model results and environmental correlates. I have developed a novel null methodology to predict the niche of a species, or a group of species; I applied this at local and regional scales to examine null spatial distribution predictions for a single, endangered species at the local scale (*Opuntia humifusa* at Point Pelee National Park), and for groups of rare species at a regional scale (based on reported occurrences across south-western Ontario).

Results can be regarded as representing intermediate states between the extremes of continua of which niche and neutral models form the ends. With respect to the relative strengths of stochastic and deterministic processes, this thesis has characterized the attributes of groups of species. For example, large-ranged NA tree species are influenced by the MDE more than small-ranged species; moreover, regional, null species distribution models performed best for birds, insects, reptiles, sedges, as well as for aquatic and terrestrial plants. It seems most likely that real species distributions are the product of variation in relative strength of stochastic and deterministic processes.

ACKNOWLEDGEMENTS

I would like to thank:

Amazing “tag-team” supervisors – Jon and Lesley Lovett-Doust.

Committee – Jan Ciborowski and Phil Graniero of the Univeristy of Windsor, and Carl Freeman of Wayne State University.

Biology faculty and staff (expecially Nancy and Louise)

Funding sources – NSERC, OGSST, AERC Inc., Grad Studies (University of Windsor).

Data sources – Point Pelee National Park staff, Ontario Geospatial Data Exchange, Ministry of Agriculture, Food and Rural Affairs.

Labmates – Helen Murphy, Joe Thomas, Lisa Tulen, Kelly Thickett, Nolan Evans, Mirabella Hali, Nancy MacDonald, Moona Sahar Moeen, and the many workstudy students.

Friends – Jessica Skipper and Julien Samson, Tawnya Durrant, Carla Wytrykush and Adam Carriere, Jen and Steve Pedersen.

Family – Mom and Dad, Van and Linda, my sisters (Bridget, Barbara, Brenda, Becky) and their significant others, and Vanessa.

And most importantly, Andrea, Dylan and Emily, for their tolerance, encouragement and support throughout.

STATEMENT OF ORIGINALITY

The work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

This thesis incorporates some results of research undertaken in collaboration with Dr. Helen Murphy as part of a research project under the supervision of Dr. Jon Lovett-Doust. The collaboration related to research on the extent to which patterns in North American tree richness were influenced by the Mid-Domain Effect, covered in Chapter 3. In that chapter, I was the sole developer of the methodology, model creation and data extraction / manipulation; responsibility for data analysis and authoring the chapter was shared between Dr. Murphy and myself.

I certify that, with the above qualification, this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline. I acknowledge the helpful guidance and support of my supervisors, Dr. Lesley Lovett-Doust and Dr. Jon Lovett-Doust.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Chapter 1 - General Introduction

Background

A foundation of landscape ecology, if not ecology in general, is that environmental patterns influence ecological processes, and so influence species distribution and abundance (Turner 1989). Ecological research in general has focussed on understanding how the environment (both directly and indirectly) influences the diversity, abundance and geographic distribution of organisms (Krebs 1972, Ricklefs 1979, Kearney and Porter 2004). Indeed, spatial patterns in species distributions have long intrigued ecologists (Brown and Lomolino 1998, Gaston 2003).

A common approach to studying the distribution of species is to quantitatively describe the environmental conditions where a species occurs, and derive statistical models of the probability of occurrence that can be interpolated to other parts of the landscape (e.g., Lindenmayer et al. 1991, Sykes et al. 1996, Peterson et al. 1999, Stockwell and Peters 1999, Peterson 2001, Kearney and Moussalli 2003, Kearney et al. 2003). Such predictive modeling has gained prominence in conservation efforts as a decision support tool for assessing the impact of accelerated environmental change on the distribution of organisms across regional landscapes (e.g., Margules and Austin 1994).

The above approach (and indeed, most of the existing theories explaining patterns in species richness) is predicated upon a major assumption, that the patterns observed are due to *some* particular set of relevant biologically important drivers – typically omitting the possibility that these patterns are, at least in part, due to random, stochastic elements in the absence of biological mechanisms. There is an increasing awareness, albeit contentious, that these stochastic factors can play an important role in defining species distributions from global to local scales. At the very least, these stochastic elements, as null models, offer a baseline for comparison with ‘real world’ patterns (Harvey et al. 1983, Colwell and Winkler 1984, Gotelli and Graves 1996, Gotelli 2001). According to Colwell et al. (2004), it is the simple deviation from null predictions that is arguably most interesting, and worthy of investigation.

Ecologists tend to work on the premise that the niche is a central organizing concept in ecology and that niche attributes structure biogeographic patterns of diversity, abundance and distribution (Gaston and Chown 2005). In this thesis I explore the concept of niche, first by examining the role of random, non-biological aspects in conjunction with environmental correlates of species distributions at a continental scale, and second, by examining the non-random, spatial clustering (deviation from random) of species occurrences at the regional scale for the null prediction of niche. For each approach, I develop novel methodology (building models) and using different biotic groups as test cases, explore the concept of species “neutrality” – the assumption that all species can be considered as essentially equivalent (see below), as a kind of reciprocal, or “flip-side” of the requirements for the concept of niche.

Niche

> “most [ecologists] would agree that niche is a central concept of ecology, even though we do not know exactly what it means”…
>
> Real and Levin (1991)

The term ‘niche’ has evolved over time from representing the habitat in which an organism resides (Grinnell 1917, 1924, 1928), to representing the ecological role an organism fills within a community (Elton 1927), to the intersection of ranges of tolerances for a set of resources utilized by an organism (Hutchinson 1957). Ricklefs (2001) defined habitat as the local place where a species lives and grows, characterized by, at least, the physical environment that surrounds, influences, and is utilized by a population of a particular species. Niche describes the relationship between a species and its area of inhabitation, in both physical and ecological terms. More specifically, it describes the unique position of the species in terms of the range of conditions it can tolerate and the characteristics that relate the geophysical environment of a species to its ecological functions – its role in the ecological system (Ricklefs 2001).

Grinnell defined the niche as a spatial unit that represented the "concept of the ultimate distributional unit, within which each species is held by its structural and instinctive limitations." (Grinnell 1928); in other words, niche represented the actual physical distribution of an organism (Vandermeer 1972). He was interested in determining the physical or climatic factors of a species’ geographical distribution, ignoring relationships with other species such as predation and competition (Leibold 1995). Two important elements of Grinnell's niche were that he envisioned species as having evolved to fill a niche, and the assumption that no two species could occupy the same niche.

At nearly the same time, Elton (1927) proposed a definition of niche that differed from Grinnell's, in that niche was defined as an organism's "place in the biotic environment, its relations to food and enemies." In other words, an organism's niche was defined by its ecological role in the community rather than its geographic location. The Eltonian view of niche delimited niches based on the size of an organism (as this influenced the type and amount of resources consumed, and influenced the potential number of predators) and its food habits (Elton 1927, Leibold 1995).

Hutchinson (1951) defined niche as an abstract multi-dimensional space, an "n-dimensional hypervolume," defining the environmental limits within which an organism is able to survive and reproduce. The limits were in terms of abiotic environmental tolerances (e.g., climatic, geophysical) and biological functions, such as in competitive effects and predation (extending beyond Elton's more limited biotic criteria of organism size and food habit).

The key difference between Hutchinson's concept of niche and those of Elton and Grinnell was that Hutchinson (1957) used the niche to represent the environmental requirements of a species rather than a place or "recess" in the environment that has the potential to support a species (Schoener 1989, Colwell 1992) – thus Hutchinson (1957) emphasised attributes of the species, not the environment (Pulliam 2000).

Hutchinson's hypervolume could include any number of dimensions or environmental axes (Holt et al. 2005). As the response of an organism to all possible environmental factors is difficult to determine, most ecologists limit investigation to a smaller set of "major", plausible factors. The "fundamental" niche was thus defined as the hypervolume created in the absence of relations with other species, and so it represents a species' potential to use available resources (Holt et al. 2005). While some have suggested that a species' geographic range represents a spatial expression of its fundamental niche (Peterson et al. 1999, Peterson 2001), many other factors such as competition, predation or dispersal limitations may influence the "realized" niche (Hutchinson 1957, Holt et al. 2005). As the fundamental niche is determined largely by the species' physiological tolerances, it is the niche that would be observed in the absence of predators and competitors (a rare and unlikely event).

Competitive interactions between species can affect the breadth of a species' niche along one or several niche axes. For example, predation could decrease the breadth of a species' food niche axis if the probability of being preyed upon increased if the species were searching for certain kinds of food items. Realistically, a niche is limited in extent by the presence of interactions with other species; this is regarded as the realized niche. The realized niche of a species may vary from location to location because of the presence of different sets of predators, and competitors (Leibold 1995, Pulliam 2000).

The niche concept is often emphasized as a term in the domain of community ecology (see e.g., Begon et al. 1990, Pianka 1994, Ricklefs 2001). However, the concept is used for studies at most levels of ecological organisation (Liebold 1995). For example, identifying environmental conditions limiting an organism's performance or fitness is often done by physiologists (e.g., Anthony and Connolly 2004, Ochocinska and Taylor 2005, Welsh et al. 2005); population biologists examine the limiting factors that alter population dynamics (e.g., Halpern et al. 2005, Lõhmus and Remm 2005); biogeographers examine the environmental constraints limiting species distributions (e.g., Peterson et al. 1999, Peterson 2001); and ecosystem ecologists examine how the functional traits of taxa alter ecosystem structure and process (e.g., Hunter and Simons 2004). Thus, moving towards larger scales, one can examine an organism's niche, a population's niche, a species' niche, and the niche of a taxon or group of species.

Occurrence of a species at a location represents the expression of its ecology and evolutionary history (Brown 1995). It is a combination of different factors at different scales (Gaston 2003, Pearson and Dawson 2003). Soberón and Peterson (2005) suggested that there are four classes of factors influencing distribution of species:

1. Abiotic conditions that impose physiological limits on a species, such as climate, physical environment, etc.
2. Biotic factors that describe the interactions, either positive or negative, with other species that affect distributions.

3. Areas accessible to dispersal from currently occupied areas, distinguishing a species' actual distribution from its potential (fundamental) distribution based on dispersal abilities and landscape composition/configuration.
4. The evolutionary capacity of a species to adapt to new conditions.

There is an ongoing debate over the relative importance of niche-assembly vs. dispersal-assembly theories of species coexistence (Potts et al. 2004). Niche-assembly theories posit that biological interactions and environmental heterogeneity are the processes underlying species coexistence and community structure (Tilman 1982, Lieberman et al. 1985, Hubbell and Foster 1986, Denslow 1987, Kohyama 1994, Terborgh et al. 1996, Clark et al. 1998). It is believed that species can only coexist when they differ from each other in the resources they use most efficiently, or in their adaptation to the local environmental conditions (Ostling 2005). This theory makes the assumption that coexisting species must have different niches.

In contrast, chance, history, and dispersal explain species coexistence in dispersal-assembly theories (Hubbell 1997, Bell 2001, Hubbell 2001). These theories suggest that "historical accidents of dispersal," rather than competitive interactions of species shape the coexistence of species (Ostling 2005). Rather than being quickly out-competed, it is suggested that species that are less efficient at using a resource evolve to be as efficient as their competitors. Dispersal to the same habitable region is the main criterion for coexistence (Ostling 2005).

Classical niche-assembly theory implies that species diversity should vary geographically, in concert with biotic or abiotic gradients (Murphy et al. 2006). The assumption that species richness should be spatially uniform in the absence of these gradients has prevailed for a long time (Colwell et al. 2004). Conversely, under the assumptions of theories of dispersal limitation and chance, most species are able to grow at most sites, so that community composition is determined largely by the accidents of dispersal, and local diversity is strongly influenced by the composition of the regional species pool (Hubbell et al. 1999, Bell 2001). Null models have recently emerged from dispersal-assembly theories as a mathematical framework (Ostling 2005), that can generate patterns that resemble those of real distribution data, without a need to incorporate data on biologically important factors.

As ecological communities are undoubtedly a product of both niche- and dispersal-assembly rules (Hubbell 2001), the question that should be posed is what is the relative importance of each?

Null models

There is increasing recognition that stochasticity may play an important role in defining species distributions. Gotelli and Graves (1996) defined a null model as a "pattern-generating model that is based on randomization of ecological data or random sampling from a known or imagined distribution". Contrary to most other modeling approaches, null models attempt to predict patterns similar to those of empirical patterns by deliberately excluding a mechanism of interest (Gotelli 2001, Colwell et al. 2004) – in other words, with respect to ecological applications, they explore whether a simple, stochastic model can reproduce patterns in real data, without incorporating biologically important mechanisms.

There is a large body of literature that attempts to interpret patterns of species distribution, abundance and diversity in terms of ecological processes (Bell 2001). However, there is often a lack of well-understood expected values within which to interpret empirical studies. As it is often difficult to infer the process generating a pattern, the interpretation of observed patterns is also difficult (Cale et al. 1989, Moloney et al. 1991).

While analytical and simulation models attempt to mimic reality, null models exclude the mechanism or factor of interest, offering a baseline for comparison (Harvey et al. 1983, Colwell and Winkler 1984, Gotelli and Graves 1996, Gotelli 2001). The deviation from null thus describes the influence of the factor or mechanism of interest (Colwell et al. 2004). The 'ideal' null model only excludes the single factor of interest and incorporates, as realistically as possible, all other potential influences (Colwell and Winkler 1984, and see examples in Gotelli and Graves 1996). As with any model, null models are abstractions of reality. Such models should therefore not be taken to be free from bias; all models make assumptions and simplifications of reality and these assumptions must be considered when interpreting the model results and model utility (Gotelli 2001, Colwell et al. 2004).

Neutral theory

Neutral models are a specialized subset of null models. Null models often do not specify demographic processes of species, assuming demographic processes and parameters are randomly distributed among taxa. In contrast, neutral models assume demographic processes are equal for all individuals across taxa (Enquist et al. 2002). First proposed and debated within population genetics (e.g., King and Jukes 1969, Lewontin 1974), Hubbell (2001) proposed a unified neutral theory of biodiversity for community ecology and macroecology.

General patterns such as log-normal or geometric range size frequency distributions, relationship between range and abundance, etc. are often examined using complex niche-assembly models (Hubbell 2005). These models tend to incorporate as many details as possible about the species and their biotic and abiotic environment, assuming each species is unique. However, such patterns can actually be generated from neutral models that assume a *per capita* functional equivalence of all species (Hubbell 2001, 2005, Bell 2001, Ricklefs 2003). This idea of 'neutrality' or '*per capita* functional equivalence' means that while species can differ in many ways (size, shape, color, etc.), they are assumed to be demographically identical with respect to vital rates of birth, death, dispersal and speciation. While all species violate this assumption to some degree, the question posed by neutral models is, how good is this approximation? It is indeed the assumption of *per capita* functional equivalence that has proven controversial (see discussions in e.g., Zhang and Lin 1997, Yu et al. 1998, Hubbell 2001, Enquist et al.

2002, Chase and Leibold 2003, Hubbell and Lake 2003, Ricklefs 2003, Chave 2004, Poulin 2004, Hubbell 2005).

The fundamental question of neutral models (and null models in general) is: to what extent do these approaches (theories) capture the mechanism behind the structuring of the patterns being tested? Gaston and Chown (2005) suggest there are three possible conclusions to the controversy of neutral- vs. niche-theories. The first is that the neutral approach is wrong, in that it fails to account for the patterns observed. While the correlation between observed and predicted patterns may be quite similar, correlation does not represent causation. Thus these approaches often start with assumptions that are wrong but may end up with patterns that happen to match empirical patterns.

Alternatively, niche theory could be wrong. This would be contrary to many ecological beliefs. Indeed, ecologists tend to work on the premise that the niche is a central organizing concept in ecology and that attributes of niche structure biogeographic patterns of diversity, abundance and distribution (Gaston and Chown 2005). The third perspective would recognize some truth in both theories. Ecological systems may operate on gradients, or continua of stochastic (null or neutral) to non-stochastic forces. It is with this point of view that I approach this thesis. I would expect that both stochastic and deterministic forces drive the distribution of species and for that reason I will examine how the deviation from randomness may provide a starting point for investigation of our understanding of the ecological niche.

Mid-domain effect

Null models have been a useful tool for describing patterns in community ecology and biogeography (Gotelli and Graves 1996) and now are being applied in macroecology (see e.g., Colwell and Lees 2000, Gotelli 2001, Colwell et al. 2004). In macroecology, such approaches (in the form of theoretical null models that assume no direct effects of environmental gradients) have shown that observed geographic patterns of species richness can be produced through stochastic processes combined with geometric constraints on species ranges. In this thesis, I use a null model, the “mid-domain effect” (MDE) of Colwell and Hurtt (1994), to examine continental patterns in species richness. The results are interpreted, describing the relative influences of stochasticity and environmental correlates on empirical observations of patterns of species richness.

Colwell and Hurtt (1994) proposed that latitudinal gradients in species richness could be explained by stochastic processes acting within the geometric constraints (or edges) of a domain. Later, this came to be known as the mid-domain effect. The MDE is observed as a mid-domain peak or plateau in species richness, for theoretical or actual species ranges, when ranges are placed randomly within a bounded geographic domain (Colwell and Lees 2000). Predictions of MDE models have been powerful in elucidating patterns of species richness along gradients of latitude, longitude and altitude (reviewed in Zapata et al. 2003, Colwell et al. 2004; and see Chapters 3 and 4 of this thesis).

Mid-domain models are null models that simulate the random spatial arrangement of species ranges, within a bounded domain such as a continent, assuming no direct effects of environmental gradients on patterns of species richness (Colwell et al. 2004). Colwell and Hurtt (1994) first proposed MDE as an alternative hypothesis to latitudinal and altitudinal gradients as factors determining species richness. Previously Stevens (1989, 1992) had suggested these were due to a biological effect of environmental gradients, which he labeled Rapoport's rule (and see Rapoport 1982). The "rule" denotes an apparent latitudinal gradient in which species richness increased while mean range size decreased, with decreasing latitude; in other words, species richness decreases towards the poles. The MDE models of Colwell and Hurtt (1994) predicted gradients in species richness, similar to Rapoport's rule, peaking at mid latitude (or altitudes) simply from the random placement of ranges in a bounded domain, without any invocation of and given the absence of any biological effects.

To envision the mid-domain effect, consider the random placement of line segments of random length between domain limits of 0 to 1, as depicted in Fig. 1.1. The line segments represent species ranges placed randomly within the limits of some geographical uni-dimensional domain (latitude, longitude, altitude). At any point within the domain, the number of species (overlapping line segments) can be counted, (as initially proposed by Colwell and Hurtt, 1994). Mean range size (mean line segment size) and its associated variance and skewness in range size can also be calculated (Arita 2005). The random placement of ranges (niches) was first described with MacArthur's two-hit broken stick model (MacArthur 1957), a random, "overlapping niche" model. MacArthur was interested in the distribution of niche sizes, not the patterns of overlap among them. However, if the number of overlapping species ranges is plotted for all points within the limits of the domain, a mid-domain effect is observed as a parabolic or quasi-parabolic curve, peaking at 0.5 times the total number of species, over the center of the domain (Colwell and Hurtt 1994, Willig and Lyons 1998, Colwell and Lees 2000, Arita 2005).

Null models are controversial in that they take no consideration of bological features of the species (Connolly 2005) and MDE models are no exception (see e.g., Colwell et al. 2004, Hawkins et al. 2005, Zapata et al. 2005, Colwell et al. 2005). MDE effectively ignores spatial environmental gradients when placing species within the bounds of a domain; it assumes that "environmental conditions vary but that species' responses to environmental conditions would be sufficiently individualistic that, in the aggregate, no part of the domain would be more hospitable to species than any other part" (Connolly 2005). The main point of contention involves range cohesion. In the absence of environmental gradients, it would be reasonable to ask why all species are not distributed throughout the entire domain (Diniz-Filho et al. 2002, Hawkins and Diniz-Filho 2002, Zapata et al. 2003, Hawkins et al. 2005, Zapata et al. 2005).

Colwell et al. (2005) responded that this is not a shortcoming of MDE models. MDE modelers do believe actual range size frequency distributions (RSFD) are shaped and limited by environmental factors, historical effects, and dispersal limitation. They seek to determine what 'real world' patterns would arise in the absence of any direct effects of environmental gradients on species richness patterns. By randomly placing species ranges within a bounded domain and comparing predicted and actual patterns of species richness, the question being asked is not "what constrains the range of a species" but rather, "to what degree would stochastic processes be sufficient to explain actual species richness patterns?" The most complete studies of MDE models examine, in a multivariate context, the influence of both stochasticity under geometric constraints (MDE), and the direct effects of both biotic and abiotic environmental gradients – exploring the covariation of the two sets of factors (Colwell et al. 2004).

The major focus of MDE models has been on predicting species richness gradients as a function of a single dimension, primarily latitude and altitude (Arita 2005). Colwell et al. (2004) reviewed the MDE literature and showed that 19 of the 21 empirical MDE studies were one-dimensional, with 47% examining latitudinal gradients, 10% longitudinal gradients, and 52% altitudinal (or depth) gradients (some studies evaluated multiple single dimensions). While uni-dimensional models explained, on average, 54% (± 4.9 SE, range 0 - 96%, based on the R^2 for ordinary least squares regression) of the variance in empirical patterns of species richness, it has been suggested the lack of consistent model fit may be due to: (1) the method for selection of range sizes (the RSFD of ranges being placed)(Willig and Lyons 1998, Colwell and Lees 2000, Colwell et al. 2004); (2) a failure to control for spatial autocorrelation, as measures of fit were based on simple correlations between predicted and empirical results (see too Jetz and Rahbek 2002, Diniz-Filho et al. 2002, Colwell et al. 2004); and/or (3) the variation in relative strength of stochasticity (MDE) and the direct effects of both biotic and abiotic environmental gradients on range location and overlap (Colwell et al. 2004, 2005, Rangel and Diniz-Filho 2005).

Bokma and Mönkkönen (2000) suggested MDE models may be limited in predictive power and applicability by their lack of multi-dimensionality, and that a move from one-dimensional to two-dimensional MDE models was the next logical step. A few studies have proposed two-dimensional models (longitude x latitude or latitude x altitude; see e.g. Bokma et al. 2001, Arita 2005, Rangel and Diniz-Filho 2005) but none have yet proposed a three-dimensional model. In this thesis, a method for the stochastic creation and placement of ranges within a three-dimensional (longitude, latitude and altitude), irregularly shaped domain is developed (Chapter 2). The predicted patterns in species richness and range-size frequency distribution (RSFD) in each of a series of one-dimensional, two-dimensional, and three-dimensional MDE models are presented, followed by discussion of the effects of domain shape on the predicted patterns.

There are two general types of MDE models; those which randomly create and place species ranges within a domain (a "fully stochastic" model) and those that place ranges "re-sampled" from a real world RSFDs. While fully stochastic MDE models produce theoretical RSFDs that are biologically realistic (a log-normal curve; Colwell and Lees 2000), it may differ from that of the actual RSFD. Thus, differences between actual and MDE-predicted species richness patterns may be a product of either the non-random placement of species within the domain, or the differences in RSFD, or a combination of both. All these potentially underestimate the MDE (McCain 2003, Colwell et al. 2004).

Random placement of ranges resampled from a range-size frequency distribution based on empirical data may incorporate taxon-specific biological characteristics, which are independent of patterns of species richness (Lees et al. 1999, Colwell and Lees 2000, Hawkins and Diniz-Filho 2002, Jetz and Rahbek 2001, 2002, McCain 2004, Colwell et al. 2004). Such niche-based attributes could include speciation and extinction potential, population density, body size, etc. As others have noted, these may not be independent of spatial patterns of species richness; moreover the use of a purely theoretical RSFD is less subject to biological assumptions, and thus more representative of a null conceptual model (see e.g., Kollef and Gaston 2001, Hawkins and Diniz-Filho 2002, Laurie and Silander 2002). The use of a theoretical RSFD permits the comparison of richness patterns of multiple groups of species to that predicted by one MDE, allowing one to probe the "neutrality" (equivalence) of groups of species with respect to the variation in relative strength of stochastic (MDE) forces and environmental correlates on range location and overlap (as in Chapter 4).

Mapping the niche

Understanding that a species' range is not homogeneously suitable throughout, finer-scale habitat suitability mapping, quantitative habitat models and predictive distribution maps (of a species' niche) all are potentially important tools to guide management and restoration as approaches to conservation of rare species (Guisan and Zimmerman 2000, Johnson et al. 2004). Application of such tools is a function of advances in geographical information systems (GIS), availability of geo-referenced databases (such as species distributions, topographical, climatic and landuse information) and the ability to utilize computationally intensive numerical techniques (Rushton et al. 2004).

The concept of the niche provides a useful starting point for understanding the distribution of species (Kearney and Porter 2004). With GIS technology, a common approach to defining a species' geographic distribution is first to characterize, quantitatively, a suite of environmental conditions for known occurrences of the species (Peterson 2001, Kearney and Porter 2004). Then statistical models of the probability of occurrence are derived, that can be interpolated to other parts of the landscape (e.g., Lindenmayer et al. 1991, Sykes et al. 1996, Peterson et al. 1999, Peterson 2001, Kearney and Moussalli 2003, Kearney et al. 2003).

There are two general methods whereby the fundamental niche of a species has been estimated (Soberón and Peterson 2005). The first is by the direct measurement of responses of organisms to abiotic environmental conditions (e.g., gradients of temperature, humidity) and inferring fitness from combinations of these conditions. Then the investigator can map areas of positive fitness using GIS technology (see e.g., Porter et al. 2000, Porter et al. 2002).

It can be argued that, in nature, only the realized niche can be observed. However it has been suggested that, if the distribution of a species in the context of varied biotic backgrounds can be observed, the composite fundamental niche can be described (Peterson et al. 1999, Peterson 2001). It is upon this premise that the second method for estimating a species' fundamental niche relates species' occurrences with geo-referenced databases (e.g., GIS layers of climate, topography, soil characteristics). Combinations of abiotic conditions that best describe occurrences are projected across the landscape to describe the limits of the fundamental niche of a species (e.g., Lindenmayer et al. 1991, Sykes et al. 1996, Peterson et al. 1999, Peterson 2001, Kearney and Moussalli 2003, Kearney et al. 2003). This simple 'correlative approach' provides insight into the fundamental niche of a species (Peterson et al. 1999, Peterson 2001, Peterson and Holt 2003, Soberón and Peterson 2005). The inclusion of biotic environmental variables (landuse, dominant vegetation, etc.) with abiotic variables provides a powerful tool for predicting species occurrences elsewhere on a landscape (Kearney and Porter 2004, Sanchez-Cordero et al. 2005, Soberón and Peterson 2005).

Information about species distributions is often based on known occurrences of the species – "presence only" data (Pearce and Boyce 2005). Different approaches for statistically modeling species occurrences have been proposed that can interpolate the probability of occurrence in other parts of the landscape. When using presence-only data, the most common and simplest approach involves tallying the limits of each environmental dimension and defining the suitability of an area based on the intersection of the apparent environmental tolerances (see e.g., Pearce and Lindenmayer 1998, Walther et al 2004). BIOCLIM (Nix 1986) is a program that uses such "environmental envelopes" (i.e., If X > x1 and X <= x2 then ...), trimming arbitrarily, 5% off the limits to map either the fundamental niche (if only abiotic environmental envelopes are use), or the realized niche (areas of suitable habitat if both biotic and abiotic envelopes are used) of a species or group of species.

Many studies have applied techniques that can be applied to presence-absence data by generating pseudo-absences, representing the background areas where species data are missing (Pearce and Boyce 2005). Algorithms such as logistic regression (see Keating and Cherry 2004), generalized linear models (e.g. Ferrier et al 2002), classification and regression trees (e.g., Breiman et al. 1984, Bourg et al. 2005), genetic algorithms (e.g., Stockwell and Peters 1999, Peterson 2001) and Bayesian logic (e.g., Bayliss et al. 2005, Mac Nally 2005) have all been used to determine the unique attributes of locations where a species occurs, relative to places where it has not been reported (pseudo-absences).

The method for selecting pseudo-absences is often to choose sites at random from within the study area (Stockwell and Peterson 2002). Such spatial randomness is a null hypothesis used in many tests to detect spatial patterns (e.g., point clustering, spatial autocorrelation). Often, such a null hypothesis is not really relevant for complex ecological systems (Sokal et al. 1998, Fortin and Jacquez 2000), so rejection of the null hypothesis may represent little scientific value. For example, when testing spatial patterns of fish occurrences, it would be unrealistic and unhelpful to randomly place pseudo-absences in the terrestrial environment. The appropriate null model is that which captures the notion of a plausible system state (as in Chapter 5).

Two general types of error are generated with predictive niche models: errors of omission and commission (Fielding and Bell 1997) akin to false negatives and false positives of a contingency table. The first, a false negative, or the omission of areas where the species occurs represents a failure of the model to include the full realized ecological niche of a species. The second, false positive or commission, represents areas that are recognized as suitable but where there is no occurrence. There are two aspects to the error of commission: the model may have failed, and incorrectly predicted areas that are not part of the species' niche; or the area is part of the species' niche but either the species is there but has not been reported or the species does not occur there due to historical factors (dispersal limitation, local extinction, etc.) or interspecific interactions (competition, predation, etc.) (Peterson et al. 1999).

Analysis of spatial datasets

Within ecology, many theories and models assumed environmental homogeneity; this is not a valid assumption (Weins 1989). In acknowledging heterogeneity, the importance of spatial pattern and scale change the way studies are designed, analyzed and interpreted (Allen and Hoekstra 1992). Over the last decade, analysis of landscape to regional level correlates of species distribution and abundance patterns has increased with increased availability of spatially explicit data from geographical information systems (Turner et al. 2001, Bullock et al. 2002). With such tools, vast amounts of spatially explicit data can be extracted; the limitation is determining which data should be considered, and selecting the appropriate analysis tools.

Spatial autocorrelation of ecological data is common, with many ecological theories assuming an underlying spatial pattern in species distributions and their environments (Legendre and Fortin 1989). Typically, nearby points sampled for species abundance or occurrence have similar values, more frequently than expected by chance (Lichstein et al.

2002). While this is ecologically significant (in terms of demography, dispersal limitation, competition, etc.), it is problematic for standard parametric tests, such as analysis of variance (ANOVA) and ordinary least squares (OLS) regression (Anselin 1992, Legendre 1993). These tests assume independently distributed errors; however, the assumption of independence is invalid when the response and the effect of the covariates are spatially autocorrelated (Gumpertz et al. 1997).

Spatial autocorrelation estimates the similarity between samples of a given variable at varying spatial distances (Sokal and Oden 1978a,b, Legendre 1993, Rossi and Quénéhervé 1998). While many methods have been proposed, it is most frequently assessed in univariate analysis using Moran's I (a standardized measure of correlation between neighboring observations; Diniz-Filho et al. 2003). Non-zero values of Moran's I indicate that pairs of values at a given geographical distance are more similar (positive autocorrelation, max of 1) or less similar (negative autocorrelation, max of -1) than expected for randomly associated pairs (Diniz-Filho et al. 2003). The geographical distances are often "binned" or categorized into discrete classes for which different Moran's I -values are calculated for the same variable, permitting the evaluation of autocorrelation as a function of spatial distance. The values are plotted against distance to produce a plot termed a "spatial correlogram" (Diniz-Filho et al. 2003).

While many different correlogram profiles are possible (Legendre and Fortin 1989, Rossi and Quénéhervé 1998), Diniz-Filho et al. 2003 describe three basic correlogram profiles usually found in ecological data. The first shows positive autocorrelation in short distance classes, shifting to negative autocorrelation at larger distance classes. This profile can be interpreted as characterizing a linear gradient at macro-scales. The second profile occurs when only short distance autocorrelation is found, indicating that spatial variation is structured in patches. The distance up to which spatial autocorrelation is observed has been interpreted as the average patch size for that specific variable (see Diniz- Filho and Telles 2002). The third profile is flat with Moran's I coefficients at or near 0 for all bin classes indicating no spatial pattern in the data. If no Moran's I coefficients significantly differ from 0, there is no spatial pattern in the data.

If Moran's I, or other similar statistics such as Geary's c (Geary, 1954), G statistics (Getis and Ord 1992), etc., show spatial autocorrelation, Type I errors may be inflated when testing statistical hypotheses using standard methods (ANOVA, correlation or regression; Legendre 1993). OLS regression is commonly used to assess the distribution of species without assessing or discussing spatial autocorrelation (Stohlgren et al. 1999, Harrison et al. 2000, McKinney 2002, Dark 2004). However, spatial autoregressive (SAR) models modify OLS to incorporate spatial dependence (Anselin 1998, Dark 2004)

Spatial autoregressive models

There are different SAR models that depend upon the nature of the spatial dependence (for example, error or lag dependence; Anselin 1995, Dark 2004). The OLS regression model may be written as:

$$y = X\beta + \epsilon$$

where y is the dependent variable, X is the independent variable, β is the regression coefficient, and ϵ is the error term assumed to be normally and independently distributed (Anselin and Rey 1991). When errors can no longer be assumed independent and identically distributed, the error SAR model is used. The spatial dependence influences the error term only. This model is expressed as:

$$y = X\beta + \lambda W \epsilon + \tau$$

where λ is the spatial autoregression coefficient, W is a spatial weights matrix, and τ is the spatially-dependent error term.

If the dependent variable at a particular location is autocorrelated with values at other locations, OLS estimation is no longer consistent and a spatial lag SAR model should be used (Anselin & Rey 1991). In the spatial lag model, the standard regression equation may be rewritten as:

$$y = \gamma W y + X\beta + \tau$$

where γ is the spatial autoregression coefficient. Maximum likelihood estimations are used for both the lag and error models (Anselin 1995).

The SAR model includes a spatially dependent variable in the regression equation; thus, a spatial weights matrix is necessary. The simplest is one where adjacent values of empirical species diversity are given a value of 1 in the weight matrix and a value of 0 is applied to values that are not adjacent. Each non-zero element in the matrix represents potential spatial interaction between two observations.

The traditional R^2 measure of fit, based on the decomposition of total sum of squares into explained and residual sum of squares, is not applicable to the SAR model. Instead, a "pseudo R^{2}" measure (ratio of the variance of the predicted values over the variance of the observed values for the dependent variable) is often reported (Anselin 1995). However, to compare measures of fit between OLS and SAR, it is inappropriate to compare the traditional R^2 with the pseudo R^2 (Anselin 1995). SAR models are based on maximum likelihood estimations; the proper measures for goodness-of-fit for the SAR are based on the likelihood function, which can also be calculated for OLS. In order to correct the log likelihood for overfitting due to unequal number of variables among regressions, information criteria (IC) such as the Akaike Information Criterion (AIC) and the Schwartz Criterion (SC) have been proposed (Anselin 1988). These information criteria adjust the log likelihood values such that, in general they can be described by:

$$IC = -2LL + f(K,N)$$

where LL is the maximized log likelihood, K is the number of variables and N is the number of observations in the model; the function f(K,N) represents 2K for AIC and K ln(N) for SC (Anselin 1995). The best model is the one with the lowest AIC and/or SC value.

With respect to null model predictions, there are well-documented problems with assessing the fit of models using correlation coefficients, due to the fact that absolute differences in magnitude are obscured (Zapata et al. 2003, Colwell et al. 2004, Romdal et al. 2005). As recommended by several authors (Colwell et al. 2004, Zapata et al. 2005), t-tests can be used to examine deviation of the slope from unity and intercept from zero, for regressions between predicted and observed patterns. These tests provide an indication of deviation in both shape and magnitude between predicted and actual data.

Thesis objectives

In this thesis, I investigate observed deviations from randomness, using spatial analysis at both regional and continental scales, to clarify the concept of niche within the context of biogeography. I also explore applications of the assumption that species can be described by models that are ecologically neutral, assuming that all species have equivalent life histories.

Following a review of the literature, I derive, first, a GIS-based null model, specifically a MDE model exploring the influence of continental geometry on patterns in species richness and range size frequency distributions. I compare the MDE predictions first to observations on tree species richness in continental North America, and then to amphibian, bird and mammal species richness across the continents of North and South

America contrasting the relative inputs of null and niche-assembly theories. I then develop a novel methodology for a null model predicting the niche of a species or a group of species. I apply this approach at local to regional scales, using environmental correlates of non-random clusters of species occurrences. I used the methodology for null prediction of the niche of a single, endangered species at the local scale (*Opuntia humifusa* at Point Pelee National Park, ON) and for groups of species at a regional scale (groups of rare and common species, based on reported occurrences across South-Western Ontario).

Throughout the thesis, I examine the applicability of the model and methodology with respect to particular species or groups of species, using the "null" approach that niche-based attributes of a species are excluded from the analysis or are assumed to be ecologically equivalent. I discuss the implications of these results and show how use of the null model as a baseline allows development of a better understanding of niche, with the opportunity to quantify niche more precisely across multiple scales.

Thesis structure

This chapter (**Chapter 1**), provides a general introduction and overview of the thesis.

Chapter 2 presents a null model based upon the 'mid-domain effect', first described by Colwell and Hurtt (1994). The model is used to predict patterns of species richness, and range size frequency distributions of species. It is based on a stochastic placement of ranges (niche limits in latitude, longitude and/or altitude), within a bounded domain. I review current models and discuss their advantages and limitations, and then argue for a modification of these models to extend predictions beyond one- and two- dimensional space, to three-dimensional space, while accounting for the fact that domains are often irregularly shaped. I introduce this model as a "six-hit", three dimensional model (based upon existing "two-hit", uni-dimensional models); I discuss the predicted species richness and RSFD patterns with respect to domain shape. The homogeneity and concordance of predictions are clarified in regard to varying domain shape by comparing predictions of the model for three different domains (a square-based pyramid, continental North America and Australia).

In **Chapter 3**, predicted patterns of species richness were created for continental North America, using the "six-hit" mid-domain model. These predictions, along with GIS-based environmental gradients in climate and topography were compared to observed tree species richness, using spatial autoregressive models. In this analysis the relative contributions of each of the stochastic range placement (i.e., the null, mid-domain effect) and environmental gradients (e.g., niche-assembly theory) are separately examined and discussed with respect to tree species richness in continental North America. Similarly, in **Chapter 4**, I determine the relative explanatory power of stochastic range placement and environmental gradients separately and combined to account for patterns of species richness and range-size frequency distributions for each of three species groups (amphibian, avian and mammal) across North and South America, using spatial autoregressive models. In this chapter, I discuss the assumption that each group of species may be considered as equivalent with respect to the predictions of the mid-domain model; in other words, “does one model fit all?”

In **Chapter 5,** I introduce a methodology for null prediction of the niche at local to regional scales. I first review the literature and outline the theoretical background to this approach. It links theories and frameworks of landscape ecology with computational sciences and eco-informatics. The procedure first determines the scale of each environmental parameter that best distinguishes the clustering of species occurrences from what would be expected based on spatial randomness; then it uses the minimum number of environmental parameters necessary to develop an accurate predictive model; and finally it applies the model over a selected landscape. This assesses the probability that a specific location is part of the species’ niche. As this prediction of niche ignores species-specific attributes, it can be regarded as a null model in which deviation of observed occurrences from the predicted areas are discussed with respect to species-specific attributes and potential biotic interactions. I apply this method spatially, from local to regional scales, and ecologically, from individuals to populations and using single species to groups of species. I demonstrate the approach using *Opuntia humifusa* in Point Pelee National Park, ON. In this site, plant distribution, habitat attributes and habitat qualities have been closely studied.

I then extend the approach to the larger grouping of rare species across south-western Ontario. Neutral approaches assume species are equivalent with respect to probabilities of birth, death, dispersal and speciation, and that species are able to grow at all sites (i.e. that they have no habitat preference)(Gaston and Chown 2005). I test these assumptions using null models of the species' niche and different subsets of species occurrence data for southern Ontario, comparing outcomes using spatial correlations. In **Chapter 6**, I develop a gradient-based niche model first, for each of four groups of rare biota (birds, insects, plants and reptiles), based on occurrences of rare species (S-rank 1-3) in SW Ontario, and report patterns of landcover, soil types, and elevation and climate. I then further model distributions of rare rare plants by grouping them into: trees, shrubs, herbs, grasses, and sedges; S1, S2, S3 plants and herbs separately; and into terrestrial and aquatic classifications. The models for each of the various groups of rare species were used to test the efficacy of the ecological equivalence assumption. Further, spatial predictions of groups of species were examined with respect to determining if some groups of species can be more accurately modelled than others. Finally, the implications in an applied conservation perspective are discussed, as this approach could be used for potential prioritization of areas of potential niche for conservation / restoration efforts for groups of rare species rather than for single species.

Chapter 7 is a general discussion, where the major results of the thesis are discussed in the context of traditional and current theories that seek to explain patterns of species distributions across scales from local to continental. It also addresses the contribution of null models toward a better understanding of niche. I propose a framework that allows current null and niche-assembly theories to be used in combination to further our understanding of the stochastic and biological processes influencing species niche and distributions.

References

Allen, T.F.H. & Hoekstra, T.W., 1992. Toward a unified ecology. Columbia University Press, New York, NY.

Anselin, L., 1992. Spacestat tutorial: a workbook for using Spacestat in the analysis of spatial data. Technical Report S-92-1, NCGIA, Santa Barbara, CA.

Anselin, L., 1995. SpaceStat, A software programme for the analysis of spatial data. Version 1.80. Regional Research Institute, West Virginia University, Morgantown, WV.

Anselin, L., 1998. SpaceStat Version 1.90. Regional Research Institute, West Virginia University, Morgantown, WV.

Anselin, L. & Rey, S., 1991. The performance of tests for spatial dependence in a linear regression. Report, 91-13. National Center for Geographic Information and Analysis, University of California Santa Barbara, CA.

Anthony, K.R.N. & Connolly, S.R., 2004. Environmental limits to growth: physiological niche boundaries of corals along turbidity-light gradients. Oecologia 141: 373-384.

Arita, H.T., 2005. Range size in mid-domain models of species diversity. Journal of Theoretical Biology 232: 119-126.

Bayliss, J.L., Simonite, V. & Thompson, S., 2005. The use of probabilistic habitat suitability models for biodiversity action planning. Agriculture, Ecosystems and Environment 108: 228-250.

Begon, M., Harper, J.L. & Townsend, C. R., 1990. Ecology: individuals, populations and communities. Blackwell Scientific, Boston, MA.

Bell, G., 2001. Neutral macroecology. Science 293: 2413-2418.

Bell, G., 2005. The co-distribution of species in relation to the neutral theory of community ecology. Ecology 86: 757-770.

Bokma, F., Bokma, J. & Mönkkönen, M., 2001. Random processes and geographic species richness patterns: why so few species in the north? Ecography 24: 43-49.

Bokma, J. & Mönkkönen, M., 2000. The mid-domain effect and the longitudinal dimension of continents. Trends in Ecology and Evolution 15: 288-289.

Bourg, N.A., McShea, W.J. & Gilla, D.E., 2005. Putting a cart before the search: successful habitat prediction for a rare forest herb. Ecology 86: 2793-2804.

Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. G., 1984. Classification and regression trees. The Wadsworth statistics/probability series. Chapman and Hall, New York, NY.

Brown, J.H., 1995. Macroecology. University of Chicago Press, Chicago, IL.

Brown, J.H. & Lomolino, M.V., 1998. Biogeography. Sinauer Associates, Sunderland, MA.

Cale, W.G., Henebry, G.M. & Yeakley, J.A., 1989. Inferring process from pattern in natural communities. BioScience 39: 600-605.

Chase, J.M. & Leibold, M.A., 2003. Ecological Niches: Linking Classical and Contemporary Approaches. University of Chicago Press, Chicago.

Chave, J., 2004. Neutral theory and community ecology. Ecology Letters 7: 241-253.

Clark, D.B., Clark, D.A. & Read, J.M., 1998. Edaphic variation and the mesoscale distribution of tree species in a neotropical rain forest. Journal of Ecology 86: 101-112.

Colwell, R.K., 1992. Niche: a bifurcation in the conceptual lineage of the term. pg 241-248 in E. Fox-Keller and E. A. Lloyd, eds. Keywords in evolutionary biology. Harvard University Press, Cambridge, MA.

Colwell, R.K. & Hurtt, G.C., 1994. Nonbiological gradients in species richness and a spurious Rapoport effect. The American Naturalist 144: 570-595.

Colwell, R.K. & Lees, D.C. 2000. The mid-domain effect: geometric constraints on the geography of species richness. Trends in Ecology and Evolution 15: 70-76.

Colwell, R.K. & Winkler, D.W., 1984. A null model for null models in biogeography. Pages 344359 in D. R. Strong, Jr., D. Simberloff, L. G. Abele, and A. B Thistle, eds. Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton, N.J.

Colwell, R.K., Rahbek, C. & Gotelli, N.J., 2004. The mid-domain effect and species richness patterns: What have we learned so far? The American Naturalist 163: E1-E23.

Colwell, R.K., Rahbek, C. & Gotelli, N.J., 2005. The mid-domain effect: there's a baby in the bathwater. The American Naturalist 166: E149-E154.

Connolly, S.R., 2005. Process-based models of species distributions and the mid-domain effect. The American Naturalist 166: 1-11.

Dark, S.J., 2004.The biogeography of invasive alien plants in California: an application of GIS and spatial regression analysis. Diversity & Distributions 10: 1-9.

Denslow, J.S., 1987. Tropical rainforest gaps and tree species diversity. Annual Review of Ecology and Systematics 18: 431-451

Diniz-Filho, J.A.F. & Telles, M.P.C., 2002. Spatial autocorrelation analysis and the identification of operational units for conservation in continuous populations. Conservation Biology 16: 924-935.

Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A., 2003. Spatial autocorrelation and redd herrings in geographical ecology. Global Ecology and Biogeography 12: 53-64.

Diniz-Filho, J.A.F., de Sant'Ana, C.E.R., de Souza, M.C. & Rangel, T.F.L.V.B., 2002. Null models and spatial patterns of species richness in South American birds of prey. Ecology Letters 5: 47-55.

Elton, C., 1927. Animal Ecology. Sidgwick and Jackson, London.

Enquist, B.J., Sanderson, J. & Weiser, M.D., 2002. Modeling macroscopic patterns in ecology. Science 295: 1835-1837.

Ferrier, S., Watson, G., Pearce, J. & Drielsma, M., 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. Biodiversity and Conservation 11: 2275-2307.

Fielding, A.H. & Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24: 38-49.

Fortin, M.J. & Jacquez, G.M., 2000. Randomization tests and spatially autocorrelated data. Bulletin of the Ecological Society of America 81: 201-205.

Gaston, K.J., 2003. The Structure and Dynamics of Geographic Ranges. Oxford University Press, Oxford, UK.

Gaston, K.J. & Chown, S.L., 2005. Neutrality and the niche. Functional Ecology 19 (1), 1-6.

Geary, R., 1954. The contiguity ratio and statistical mapping, The Incorporated statistician 5: 115-45.

Getis, A. & Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geographical Analysis 24: 189-206.

Gotelli, N.J., 2001. Research frontiers in null model analysis. Global Ecology & Biogeography 10: 337-343.

Gotelli, N.J. & Graves, G.R., 1996. Null Models in Ecology. Smithsonian Institution Press, Washington, DC.

Grinnell, J., 1917. The niche-relationships of the California Thrasher. Auk 34: 427-433.

Grinnell, J., 1924. Geography and evolution. Ecology 5: 225-229.

Grinnell, J., 1928. A distributional summation of the ornithology of Lower California. University of California Publications in Zoology 32: 1-30.

Guisan, A. & Zimmerman, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135: 147-186.

Gumpertz, M. L., Graham, J.M. & Ristaino, J.B., 1997. Autologistic model of spatial pattern of Phytophthora epidemic in bell pepper: effects of soil variables on disease presence. Journal of Agricultural, Biological, and Environmental Statistics 2: 131-156.

Halpern, B.S., Gaines, S.D. & Warner, R.R., 2005. Habitat size, recruitment, and longevity as factors limiting population size in stage-structured species. The American Naturalist 165: 82-94.

Harrison, S., Viers, J.H. & Quinn, J.F., 2000. Climatic and spatial patterns of diversity in the serpentine plants of California. Diversity and Distributions 6: 153-161.

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

Hawkins, B.A. & Diniz-Filho, J.A.F., 2002. The mid-domain effect cannot explain the diversity gradient of Nearctic birds. Global Ecology & Biogeography 11: 419-426.

Hawkins, B.A., Diniz-Filho, J.A.F. & Weis, A.E., 2005. The mid-domain effect and diversity gradients: is there anything to learn? The American Naturalist 166: E140-E143.

Holt, R.D., Keitt, T.H., Lewis, M.A., Maurere, B.A. and Taper, M.L., 2005. Theoretical models of species' borders: single species approaches. Oikos 108: 18-27.

Hubbell, S.P., 1997. A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. Coral Reefs 16: S9-S21.

Hubbell, S.P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton, NJ.

Hubbell, S.P., 2005. Neutral theory in community ecology and the hypothesis of functional equivalence. Functional Ecology 19: 166-172.

Hubbell, S.P. & Foster, R.B., 1986. Biology, chance and history and the structure of tropical rain forest tree communities. pp 314-329 In: J. Diamond & T.J. Case (eds). Community ecology. Harper and Row, New York, NY.

Hubbell, S.P. & Lake, J.K., 2003. The neutral theory of biodiversity and biogeography, and beyond. pp. 45-63 In T.M. Blackburn & K.J. Gaston (eds). Macroecology: Concepts and Consequences. Blackwell Publishing, Oxford, UK.

Hubbell, S.P., Foster, R.B., O'brien, S.T., Harms, K.E., Condit, R., Wechsler, B., Wright, S.J. & De Lao, S.L., 1999. Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. Science 283: 554-557.

Hunter, R.D. & Simons, K.A., 2004. Dreissenids in Lake St. Clair in 2001: evidence for population regulation. Journal of Great Lakes Research 30: 528-537.

Hutchinson, G.E., 1957. Concluding remarks. Cold Spring Harbor Symposium on Quantitative Biology 22: 415-427.

Jetz, W. & Rahbek, C., 2001. Geometric constraints explain much of the species richness patterns in African birds. Proceedings of the National Academy of Science USA 98: 5661-5666.

Jetz, W. & Rahbek, C., 2002. Geographic range size and determinants of avian species richness. Science 297: 1548-1551.

Johnson, C.J., Seip, D.R. & Boyce, M.S., 2004. A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales. Journal of Applied Ecology 41: 238-251.

Kearney, M. & Moussalli, A., 2003. Geographic parthenogenesis in the Australian arid zone. II. Climatic analyses of orthopteroid insects of the genus *Warramaba* and *Sipyloidea*. Evolutionary Ecology Research 5: 977-997.

Kearney, M. & Porter, W.P., 2004. Mapping the fundamental niche, physiology, climate and the distribution of a nocturnal lizard. Ecology 85: 3119-3131.

Kearney, M., Moussalli, A., Strasburg, J., Lindenmayer, D. & Moritz, C., 2003. Geographic parthenogenesis in the Australian arid zone. I. A climatic analysis of the *Heteronotia binoei* complex (Gekkonidae). Evolutionary Ecology Research 5: 953-976.

Keating, K.A. & Cherry, S., 2004. Use and interpretation of logistic regression in habitat-selection studies. Journal of Wildlife Management 68: 774-789.

King, J.L. & Jukes, T.H., 1969. Non-Darwinian Evolution. Science 164: 788-798.

Kohyama T., 1994. Size-structure-based models of forest dynamics to interpret population-level and community-level mechanisms. Journal of Plant Research 107: 107-116

Koleff, P. & Gaston, K.J., 2001. Latitudinal gradients in diversity: real patterns and random models. Ecography 24: 341-351.

Krebs, C.J., 1972. Ecology: The experimental analysis of distribution and abundance. Harper & Row, Publ., New York, NY.

Laurie, H. & Silander Jr., J.A., 2002. Geometric constraints and spatial patterns of species richness: critique of range-based null models. Diversity and Disturbance 8: 351-364.

Lees, D.C., Kremen, C. & Andriamampianina, L., 1999. A null model for species richness gradients: bounded range overlap of butterflies and other rainforest endemics in Madagascar. Biological Journal of the Linnean Society 67: 529-584.

Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? Ecology 74: 1659-1673.

Legendre, P. & Fortin, M.J., 1989. Spatial pattern and ecological analysis. Vegetatio 80: 107-138.

Leibold, M.A., 1995. The niche concept revisited: mechanistic models and community context. Ecology 76: 1371-1382.

Lewontin, R.C., 1974. The Genetic Basis of Evolutionary Change. Columbia University Press, New York, NY.

Lichstein, J.W., Simons, T.R., Shriner, S.A. & Franzerb, K.E., 2002. Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs 72: 445-463.

Lieberman, M., Lieberman, D., Hartshorn, G.S. & Peralta, R., 1985. Smallscale altitudinal variation in low-land wet tropical forest vegetation. Journal of Ecology 73: 505-516.

Lindenmayer, D., Nix, H., McMahon, J.P., Hutchinson, G.E. & Tanton, M.T., 1991. The conservation of Leadbeater's possum, *Gymnobelideus leadbeateri* (McCoy): a case study of the use of bioclimatic modeling. Journal of Biogeography 18: 371-383.

Lõhmus, A. & Remm, J., 2005. Nest quality limits the number of hole-nesting passerines in their natural cavity-rich habitat. Acta Oecologica 27: 125-128.

Mac Nally, R., 2005. Ecological boundary detection using Carlin-Chib Bayesian model selection. Diversity and Distributions 11: 499-508.

Margules, C.R. & Austin, M.P., 1994. Biological models for monitoring species decline: the construction and use of data bases. Philosophical Transactions of the Royal Society London series B. 344: 69-75.

McCain, C.M., 2003. North American desert rodents: a test of the mid-domain effect in species richness. Journal of Mammalology 84: 967-980.

McCain, C.M., 2004. The mid-domain effect applied to elevational gradients: species richness of small mammals in Costa Rica. Journal of Biogeography 31: 19-31.

McKinney, M.L., 2002. Influence of settlement time, human population, park shape and age, visitation and roads on the number of alien plant species in protected areas in the USA. Diversity and Distributions 8: 311-318.

Moloney, K.A., Morin, A. & Levin, S.A., 1991. Interpreting ecological patterns generated through simple stochastic processes. Landscape Ecology 5: 163-174.

Murphy, H. T., VanDerWal, J., Khalatkhar, N. & Lovett-Doust, J. 2006. Incorporating 3-dimensional geographic range geometry in mid-domain models: geometric constraints and environmental correlates of North American tree species richness. Ecography. In Review.

Nix, H.A., 1986. A biogeographic analysis of Australian elapid snakes, p. 4-15. In Bureau of Flora and Fauna [eds.], Atlas of Australian elapid snakes. Bureau Flora Fauna, Canberra, Australia.

Ochocinska, D. & Taylor, J.R.E., 2005. Living at the physiological limits: field and maximum metabolic rates of the common shrew (Sorex araneus). Physiological and Biochemical Zoology 78: 808-818.

Ostling, A., 2005. Ecology – neutral theory tested by birds. Nature 436: 635-636.

Pearce, J. & Lindenmayer, D., 1998. Bioclimatic analysis to enhance reintroduction biology of the endangered Helmeted Honeyeater (*Lichenostomus melanops cassidix*) in southeastern Australia. Restoration Ecology 6: 238-243.

Pearce, J.L. & Boyce, M.S., 2005. Modelling distribution and abundance with presence-only data. Journal of Applied Ecology, In Press.

Pearson, R.G. & Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography 12: 361-371.

Peterson, A.T., 2001. Predicting species' geographic distributions based on ecological niche modeling. Condor 103: 599-605.

Peterson, A.T. & Holt, R. D., 2003. Niche differentiation in Mexican birds: Using point occurrences to detect ecological innovation. Ecology Letters 6: 774-782.

Peterson, A.T., Soberón, J. & Sánchesz-Cordero, V., 1999. Conservatism of ecological niches in evolutionary time. Science 285:1265-1267.

Pianka, E.R., 1994. Comparative ecology of *Varanus* in the Great Victoria desert. Australian Journal of Ecology 19: 395-408.

Porter, W.P., Budaraju, S., Stewart, W. E. & Ramankutty, N., 2000. Calculating climate effects on birds and mammals: Impacts on biodiversity, conservation, population parameters, and global community structure. American Zoologist 40: 597-630.

Porter, W.P., Sabo, J.L., Tracy, C.R., Reichman, O.J. & N. Ramankutty. 2002. Physiology on a landscape scale: Plant-animal interactions. Integrative and Comparative Biology 42: 431-453.

Potts, M.D., Davies, S.J., Bossert, W.H., Tan, S. & Supardi, M.N.N. 2004. Habitat heterogenity and niche structure of trees in two tropical rain forests. Oceologia 139: 446-453.

Poulin, R., 2004. Parasites and the neutral theory of bio-diversity. Ecography 27: 119-123.

Pulliam, H.R., 2000. On the relationship between niche and distribution. Ecology Letters 3: 349-361.

Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F., 2005. An evolutionary tolerance model explaining spatial patterns in species richness under environmental gradients and geometric constraints. Ecography 28: 253-263.

Rapoport, E.H., 1982. Areography: geographical strategies of species. B. Drausel, Trans. Pergamon, Oxford, UK.

Real, L.A. & Levin, S.A., 1991. The role of theory in the rise of modern ecology. In: Real LA, Brown JH (eds) Foundations of ecology: classic papers with commentaries. The university of Chicago Press, Chicago, IL.

Ricklefs, R.E., 1979. Ecology, second edition. Chiron Press, New York, NY.

Ricklefs, R.E., 2001. The Economy of Nature, 5th ed. WH Freeman, New York, NY.

Ricklefs, R.E., 2003. A comment on Hubbell's zero-sum ecological drift model. Oikos 100: 185-192.

Romdal, T.S., Colwell, R.K. & Rahbek, C., 2005. The influence of band sum area, domain extent, and range sizes on the latitudinal mid-domain effect. Ecology 86: 235-244.

Rossi, J.P. & Quénéhervé, P., 1998. Relating species density to environmental variables in presence of spatial autocorrelation: a study case on soil nematodes distribution. Ecography 21: 117-123.

Rushton, S.P., Ormerod, S.J. & Kerby, G., 2004. New paradigms for modeling species distributions? Journal of Applied Ecology 41: 193-200.

Sánchez-Cordero, V., Cirelli, V., Munguía, M. & Sarkar, S., 2005. Place prioritization for biodiversity representation using species' ecological niche modeling. Biodiversity informatics 2: 11-23.

Schoener, T.W., 1989. The ecological niche. pg 79-114 in J. M. Cherrett, ed. Ecological concepts: the contribution of ecology to an understanding of the natural world. Blackwell Scientific, Oxford, UK.

Soberón, J. & Peterson, A.T., 2005. Interpretation of models of fundamental ecological niches and species' distributions. Biodiversity Informatics 2: 1-10.

Sokal, R.R. & Oden, N.L., 1978a. Spatial autocorrelation in biology. 1. Methodology. Biological Journal of the Linnean Society 10: 199-228.

Sokal, R.R. & Oden, N.L., 1978b. Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. Biological Journal of the Linnean Society 10: 229-249.

Sokal, R.R., Oden, N.L. & Thomson, B.A., 1998. Local spatial autocorrelation in a biological model. Geographical Analysis 30: 331-354.

Stevens, G.C., 1989. The latitudinal gradient in geographical range: How so many species coexist in the tropics. The American Naturalist 133: 240-256.

Stevens, G.C., 1992. The elevational gradient in geographical range: an extension of Rapoport's latitudinal rule to altitude. The American Naturalist 140: 893-911.

Stockwell, D.R.B. & Peters, D.P., 1999. The GARP modelling system: problems and solutions to automated spatial prediction. International Journal of Geographic Information Science 13: 143-58.

Stockwell, D.R.B. & Peterson, A.T., 2002. Controlling bias in biodiversity data. Predicting Species Occurrences: Issues of Accuracy and Scale. pp. 537-546 In. J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall & F.B. Samson (eds). Island Press, Washington, DC.

Stohlgren, T.J., Binkley, D., Chong, G.W., Kalkhan, M.A., Schell, L.D., Bull, K.A., Otsuki, Y., Newman, G., Bashkin, , M. & Son, Y., 1999. Exotic plant species invade hot spots of native plant diversity. Ecological Monographs 69: 25-46.

Sykes, M.T., Prentice, I.C. & Cramer, W., 1996. A bioclimatic model for the potential distributions of north European tree species present and future climates. Journal of Biogeography 23: 203-233.

Terborgh, J., Foster, R.B. & Nunez, P.V., 1996. Tropical tree communities: A test of the non-equilibrium hypothesis. Ecology 77: 561-567.

Tilman, D., 1982. Resource competition and community structure. Princeton University Press, Princeton, N.J.

Turner, M.G. 1989 Landscape Ecology: the effect of pattern on process. Annual Review of Ecology and Systematics 20: 171-197.

Turner, M.G., Gardner, G.H. & O'Neill, R.V., 2001. Landscape ecology in theory and practice: pattern and process. Springer-Verlag, New York, NY.

Vandermeer, J.H., 1972. Niche Theory. Annual Review of Ecology and Systemematics 3: 107-132.

Walther, B., Wisz, M. & Rahbek, C., 2004. Known and predicted African winter distributions and habitat use of the endangered Basra reed warbler (*Acrocephalus griseldis*) and the near-threatened cinereous bunting (*Emberiza cineracea*). Journal of Ornithology 145: 287-299.

Welsh, H.H. Jr., Hodgson, G.R. & Lind, A.J., 2005. Ecography of the herpetofauna of a northern California watershed: linking species patterns to landscape processes. Ecography 28: 521-536.

Wiens, J.A., 1989. Spatial scaling in ecology. Functional Ecology 3: 385-397.

Willig, M.R. & Lyons, S.K., 1998. An analytical model of latitudinal gradients of species richness with an empirical test for marsupials and bats in the New World. Oikos 81: 93-98.

Yu, D.W., Terborgh, J.W. & Potts, M.D., 1998. Can high tree species richness be explained by Hubbell's null model? Ecology Letters 1: 193-199.

Zapata, F.A., Gaston, K.J. & Chown, S.L., 2003. Mid-domain models of species richness gradients: assumptions, methods and evidence. Journal of Animal Ecology 72: 677-690.

Zapata, F.A., Gaston, K.J. & Chown, S.L., 2005. The mid-domain effect revisited. The American Naturalist 166: E144-E148.

Zhang, D.-Y. & Lin, K., 1997. The effects of competitive asymmetry on the rate of competitive displacement: how robust is Hubbell's community drift model? Journal of Theoretical Biology 188: 361-367.

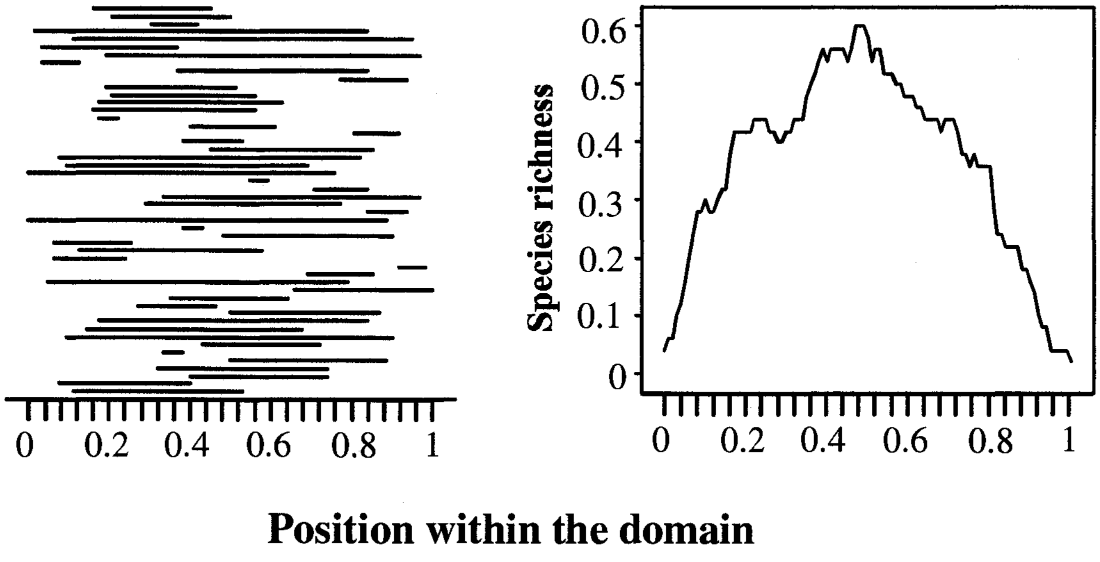


Figure 1.1. Illustration of the rationale of the mid-domain models. Fifty species ranges were randomly placed within the domain limits (0 - 1) based on the random selection of range endpoints from a uniform distribution of all potential points with the limits of the domain. The horizontal lines in the left panel depict the size and placement within the domain for the 50 theoretical species. The overlapping species ranges produce a pattern of species richness shown in the right panel. While this single simulation produces a mid-domain peak of 0.6 x number of species, replicating the simulation multiple times would produce a mean peak of approx. 0.5 as predicted by Colwell and Hurtt (1994), Willig and Lyons (1998) and Colwell and Lees (2000).

Chapter 2 - Species richness and range size predictions of mid-domain models – now in 3-D!

Summary

The mid-domain effect (MDE) proposes that gradients in species richness arise due to stochasticity and geometric domain boundaries. Thus far null patterns in species richness have been predicted mostly using random range placements within the limits of a single dimension (e.g., latitude), with a few models exploring two-dimensional or irregularly shaped domains. Here I present a modified, fully stochastic, three-dimensional MDE model, using a "six-hit" approach that builds from an established two-hit, one-dimensional model. Thus, the model randomly selects the extent (minimum and maximum range endpoints) for each of the three dimensions of longitude (X), latitude (Y) and altitude (Z), such that the size of a theoretical species' range, R, is the area represented by $R_X \cap R_Y \cap R_Z$. Existing one- and two-dimensional MDE models are improved upon by having range endpoints drawn from a distribution which takes into account the proportion of the domain in each unit of a particular dimension. This chapter explores predictions for species richness and range-size frequency distribution (RSFD) patterns, examining effects of domain shape using three different domains (a square-based pyramid, continental North America, and Australia) within the simulations.

Several general results occurred independent of domain shape: (1) mean range size at any point in the domain was relatively constant at nearly 50% of the domain for single dimensions individually, at *c.* 25% for two-dimension combinations, and at *c.* 13% for

the three-dimension combination; (2) variance in range size decreased from domain edges (areas of lowest species richness) toward the center of the domain (highest richness); and (3) the RSFD was always right-skewed, with skewness increasing with the number of dimensions being resolved. Effects of domain shape were greatest in regard to elevation, with a shift of peak richness to lower altitudes rather than mid altitudes. The model provides explicit one-, two- and three-dimensional baselines against which empirical data can be compared, deriving contributions of multi-dimensional geometric factors upon the stochastic arrangement of species ranges.

Introduction

Colwell and Hurtt (1994) proposed that latitudinal gradients in species richness could be explained by stochastic processes acting within the geometric constraints (or edges) of a domain; later this came to be known as the mid-domain effect (MDE). The MDE is observed as a mid-domain peak or plateau in species richness, for theoretical or empirical ranges, when placed randomly within a bounded geographic domain (Colwell and Lees 2000). Predictions of MDE models have been used to examine patterns of species richness in gradients of latitude, longitude and altitude (reviewed in Zapata et al. 2003, Colwell et al. 2004).

Mid-domain models are null models that simulate the random spatial arrangement of species ranges, within a bounded domain (such as a continent), and assuming no direct effects of environmental gradients on species richness (Colwell et al. 2004). Colwell and Hurtt (1994) first proposed MDE as an alternate hypothesis explaining latitudinal and altitudinal gradients in species richness. Previously, Stevens (1989, 1992) had suggested such richness gradients were due to environmental gradients, which he labeled Rapoport's rule (and see Rapoport 1982: the "rule" denotes an apparent latitudinal gradient in which species richness increased and mean range size decreased, with decreasing latitude). MDE models of Colwell and Hurtt (1994) predicted gradients in species diversity, similar to Rapoport's rule, peaking at mid latitude (or altitudes) simply from the random placement of ranges in a bounded domain, and in the absence of any direct biological effects.

Null models have proven controversial for their non-biological basis (Connolly 2005), and MDE models are no exception (see e.g., Colwell et al. 2004, Hawkins et al. 2005, Zapata et al. 2005, Colwell et al. 2005). This chapter will not reiterate the arguments, but will present, with a MDE model, the effects of geometric constraints (domain limits) on species ranges as a driver of spatial richness patterns. Elsewhere, in Chapters 2 and 3, I show that actual patterns in species richness are influenced by factors of both geometric constraints and environment (including topography, climatic patterns and biotic factors). My interest in this chapter is to extend the theoretical predictions of MDE to three dimensions, and to probe the geometry of real domains.

To date, the major focus of MDE models has been on predicting gradients in species richness as a function of a single dimension, primarily latitude (Arita 2005). Bokma and Mönkkönen (2000) suggested MDE models may be limited in predictive power and applicability by their lack of multi-dimensionality. Here I propose a method for the stochastic creation and placement of ranges within a bounded domain. Patterns are predicted in species richness and range-size frequency distribution (RSFD) in each of a series of one-dimensional, two-dimensional, and three-dimensional MDE models. Finally, effects of domain shape are tested by applying the model first to a square-based, pyramidal domain, and then to continental North American- and Australian-shaped domains, respectively.

Mid-domain models

Consider the random placement of line segments of random length between domain limits of 0 to 1. The line segments represent species ranges placed randomly within the limits of a geographic uni-dimensional domain (latitude, longitude, altitude). At any point within the domain the number of species (overlapping line segments) can be counted, along with mean range size (mean line segment size) and its associated variance and skewness in range size can be calculated (Arita 2005). The random placement of ecological ranges (niches) was first described with MacArthur's two-hit broken stick model (MacArthur 1957), a random, "overlapping niche" model. MacArthur was interested in the distribution of niche sizes, rather than the patterns of overlap among them. However, if number of overlapping species ranges is plotted for all points within the limits of the domain, a mid-domain effect is observed as a parabolic or quasi-parabolic curve, peaking at 0.5 times the total number of species, over the center of the domain (Colwell and Hurtt 1994, Willig and Lyons 1998, Colwell and Lees 2000, Arita 2005).

There are two general types of MDE models: those which randomly create and place species ranges within a domain ("fully stochastic" models) and those which place ranges "re-sampled" from real world RSFDs (Zapata et al. 2003, Colwell et al. 2004). While fully stochastic MDE models produce a theoretical RSFD that is biologically realistic (Colwell and Lees 2000), it may differ from that of the actual RSFD. Thus, differences between actual and MDE-predicted species richness patterns may be a product of either the non-random placement of species within the domain, or the differences in RSFD, or a combination of both. All this potentially underestimates the MDE (McCain 2003, Colwell et al. 2004).

Random placement of ranges re-sampled from a RSFD based on observed data could incorporate taxon-specific biological characteristics, and these may be independent of patterns of species richness (Lees et al. 1999, Colwell 2000, Hawkins and Diniz-Filho 2002, Jetz and Rahbek 2001, 2002, McCain 2004, Colwell et al. 2004). Such niche-based attributes could include speciation and extinction potential, population density, body size, etc. However, as others have noted, it may be that use of a purely theoretical RSFD is less subject to biological assumptions, and thus more "null," (see e.g., Koleff and Gaston 2001, Hawkins and Diniz-Filho 2002, Laurie and Silander 2002).

Further, the random placement of ranges (based on either theoretical or empirical RSFDs) permits the tracking and hence prediction of RSFD characteristics at particular locations or points within the domain. Arita (2005) explored exact RSFDs for species at different latitudinal positions, under assumptions of a fully stochastic one-dimensional MDE model. Arita (2005) found three major RSFD generalizations: 1) MDE models predict no gradient in mean range size (≈ 0.5 x the extent of the domain); 2) variance in range size decreases from the edge to the center of the domain; and 3) while the RSFD is right-skewed at any point in the domain, the skewness increases with decreasing species diversity (moving from center to edge of the domain).

Modeling in 3-D – true domain geometry

Here I extend predictions of MDE to three dimensions using a modified, fully stochastic MDE model. Mid-domain models assume that spatial richness patterns are constrained by the geometry of the domain. Yet the true geometry of any domain is not solely in a single dimension or even in two, but in three – latitude, longitude and altitude.

With one-dimensional MDE models, two methods dominate creation and placement of randomly generated ranges within a domain. The first, proposed by Colwell and Hurtt (1994), selects range midpoints and range size randomly from all possible combinations of values. The presumption was, because the domain is bounded, as range midpoints approach the domain limits, ranges are constrained to be progressively smaller. Thus while small-range midpoints may be located throughout the domain, large-range midpoints are constrained to be nearer the center (Zapata et al. 2003). The alternative approach is to define a range by selecting endpoints from a uniform distribution of possible points within the domain (e.g., MacArthur's 1957 two-hit broken stick model, the binomial model of Willig and Lyons 1998, and the fully stochastic model of Colwell and Lees 2000).

Underlying all one-dimensional MDE models is the assumption that the domain of interest is linear (i.e., shapeless beyond the focal dimension), with the only limits to species placement being the extent of the domain. Thus, for a single dimension, it is assumed a species cannot occur outside the "hard" limits (0 and 1) of any domain, and so a species range can be defined by its limits L_1 and L_2 where $0 \leq L_1, L_2 \leq 1$ and $L_2 \geq L_1$ (Willig and Lyons 1998, Arita 2005). To generate a species range, two random points within the range are randomly placed within the domain (i.e., for longitude X_1 and X_2) where $L_1 = \text{MIN}(X_1, X_2)$ and $L_2 = \text{MAX}(X_1, X_2)$. The range size R_X is a function of X_1 and X_2 and can be represented as $R_X = |X_1 - X_2|$, such that $0 \leq R \leq 1$.

Arita (2005) extended this to two dimensions (longitude x latitude). Thus, for a square-shaped continent, the range could be defined as a rectangle with limits in longitude of X_1 and X_2, and for latitude, Y_1 and Y_2. The values X_1, X_2, Y_1 and Y_2 are selected at random within the limits of the domain. For each single dimension, the size of the range can be represented by $R_X = |X_1 - X_2|$ and $R_Y = |Y_1 - Y_2|$. The area of the rectangle that represents the range in two-dimensions can be defined as $R = R_X * R_Y$ or $= |X_1 - X_2| * |Y_1 - Y_2|$.

I propose we can further extend Arita's (2005) two-dimensional concept to three dimensions; I suggest that a range can be defined by its limits in longitude (X_1, X_2), latitude (Y_1, Y_2) and altitude (Z_1, Z_2). As the universe of possible values for two dimensions is a unit square, the universe of possible values for all three dimensions can be represented by a unit cube. In this case, the species range, R, could be defined as a volume $R_X * R_Y * R_Z$ or $|X_1 - X_2| * |Y_1 - Y_2| * |Z_1 - Z_2|$. While this would be suitable for developing a model for three-dimensional domains such as ocean or lake basins, such volumetric measurements are not suitable for terrestrial models.

In terrestrial environments, the range limits represent the surface of the domain in longitude and latitude between some altitudinal range; the species range, R, is the area delimited in longitude and latitude that can be represented by $R_X \cap R_Y \cap R_Z$ where R_X, R_Y and R_Z are defined by their limits $X_1$and X_2, Y_1 and Y_2, and Z_1 and Z_2, respectively. Given that the universe of possible values for X_1, X_2, Y_1, Y_2, Z_1 and Z_2 is represented by a unit cube, the random selection of endpoints may create a range where $R = 0$ (i.e., when $R_X \cap R_Y$ but does not intersect R_Z). As actual domains are irregularly shaped surfaces (rather than a cube), the probability of creating theoretical species ranges where $R= 0$ increased when calculating the size of the range in three dimensions, where R is defined as the area represented by $R_X \cap R_Y \cap R_Z$. For this model, where $R = 0$ the theoretical species was discarded and a new one was created, with six new, randomly selected endpoints.

While the move from one-dimensional to two-dimensional MDE models is the next logical step (Bokma and Mönkkönen 2000), some of the appealing simplicity of one-dimensional MDE models is lost. Indeed, accounting for three dimensions is even more complex. Beyond the single dimension, irregularly shaped domains have been problematic as MDE models often assume a rectangular or square domain (Bokma et al. 2001, Hawkins and Diniz-Filho 2002, Arita 2005; but see e.g., Bokma et al. 2001, Jetz and Rahbek 2001, 2002, Diniz-Filho et al. 2002, Hawkins and Diniz-Filho 2002, Rangel and Diniz-Filho 2005a, b).

Rather than selecting endpoints from a uniform distribution of potential points within the domain (as done by e.g., Willig and Lyons 1998, Colwell and Lees 2000, Arita 2005), I suggest the irregular shapes of domains may be better accounted for by selecting endpoints from non-uniform distributions, based on the proportion of the domain represented by single points in (or for each measured unit of) one dimension. Thus, with respect to latitude and longitude, and given that peninsular proportions of a domain are typically small compared to the entire domain, the probability of selecting range endpoints within the peninsula is relatively small.

Simulations

To examine predictions of this modified, fully stochastic MDE model, 100 simulations were run in which 500 theoretical species ranges (defined as the intersection of $R_X \cap R_Y \cap R_Z$) were randomly placed within each of three domains (Fig. 2.1). The first domain was geometrically regular, a square-based pyramid, used for comparisons with predictions of other models with respect to longitude x latitude interactions. The pyramidal shape also permits examination of effects of elevation. The other two domains were irregularly shaped domains, namely continental North America and Australia. Perimeters of these domains were extracted from "World Countries 1992" base map, supplied with ArcView 3.2 (ESRI, California, USA); altitudes were obtained from the GLOBE project (GLOBE 1999). Australia was used as an approximately "regular" shaped domain with respect to longitude and latitude, with irregularity in elevation; while North America is a domain that is irregularly shaped in all three dimensions.

Each domain was gridded with 20 x 20 km cells in longitude and latitude, while mean altitude was estimated in metres above sea level for each cell. Thus the pyramid domain was 300 x 300 cells in longitude and latitude, and peaked at 300m above sea level in altitude. The maximum extents of the North American domain were 309 x 366 cells in longitude and latitude, and peaked at 3644m above sea level. The extents of the Australian domain were 243 x 197 cells in longitude and latitude, and peaked at 1441m ASL.

From the 100 simulations, mean values of predicted species richness and range size were used, along with measures of associated standard deviation and skewness in range size, for further analysis. In effect, 50,000 species were placed within each domain whose ranges were randomly created using the “six-hit” approach. Range limits for each of longitude, latitude and altitude (i.e., two endpoints for each dimension) were selected at random from the non-uniform distributions of possible endpoints. Distributions of possible endpoints were, again, based on the proportion of the particular domain represented by individual points in any one dimension. For example, based on the proportionate area within the square-based pyramid domain, the probability of selecting any point for latitude and longitude is uniform and equally likely (≈ 0.02; Fig. 2.2a,b); however for elevation, the probability of selecting any point decreases with increasing elevation (from ≈ 0.04 at low elevations, approaching 0 at higher elevations; Fig. 2.2c).

The theoretical ranges created were often irregularly shaped and could be discontinuous, similar to actual ranges. Additionally, the extents of the theoretical ranges in longitude, latitude and altitude were often smaller than those defined by randomly selected endpoints. For example, X_{max} for range R, where R is a function of $R_X \cap R_Y \cap R_Z$, was less than the randomly selected L_{X2} where $L_{X2} = MAX(X_1, X_2)$, and thus $R_X < |X_1 - X_2|$. When this occurred, the maximum and minimum extents of the range were defined by the limits of R rather than the randomly selected endpoints.

To examine RSFD for the theoretical species, range sizes for each of the single dimensions and combinations of dimensions were calculated as proportions of the domain within the limits of the range. In other words the proportion of the domain within the extent of the range limits, in each of the one-, two- and three-dimensions, was calculated. From the species extents and corresponding range sizes, species richness, mean range size, and standard deviation and skewness in range size were calculated for each 20 x 20 km cell, in each individual dimension and combination of dimensions, within each domain.

Results

Theoretical range-size frequency distribution

Proportionate area frequencies for the pyramidal-, the North American-, and the Australian domains are shown in Fig. 2.2a,b,c with respect to longitude, latitude and altitude, respectively. The square-based pyramid domain shows a nearly uniform distribution of proportionate frequencies for longitude and latitude. However, for elevation lower altitudes comprise a greater proportion of the domain, and thus this should be reflected in the distribution of potential endpoints. With respect to the North American and Australian domains, in both cases the proportion of the domain represented by longitude and latitude provide a normal distribution (Fig. 2.2d,e), while elevation shows a significantly right-skewed distribution (Fig. 2.2f).

The RSFD of the theoretical species shows, in general, a monotonically decreasing frequency of range sizes, with the maximum range size equaling that of the full breadth of the single domain, in each of the unitary dimensions (Fig. 2.3, top panels). Furthermore, this model shows decreased mean range size and increased skewness with increasing dimensionality (Fig. 2.3, middle and lower panels).

Patterns in species richness and range-size frequency distribution

An average peak of 47.3 % (± 0.82 SE) was predicted for individual dimensions, similar to that of other one-dimensional models (Fig. 2.2 d,e,f; and see Colwell and Hurtt 1994, Willig and Lyons 1998, Colwell and Lees 2000, Arita 2005). In all three domains, the predicted species richness peaks ranged from 0.22 to 0.28 x number of species for the two-dimension combinations, and ranged between 0.10 and 0.14 x number of species, when accounting for all three dimensions.

With respect to the patterns in RSFD, and with the exception of very low species richness, the following general trends were noted. First, all dimensions and combinations of dimensions predict a relatively constant mean range size of 0.450 (± 0.0001), 0.244 (± 0.0001) and 0.130 (± 0.0001) for one-, two- and three-dimensions respectively, independent of location within the domain (Fig. 2.4, top panels). Second, variance in mean range size decreased from edge to center (i.e., low-to-high predicted species richness; Fig. 2.4, middle panels). Finally, while the RSFD is right-skewed at any point of the domain, skewness was in general relatively constant across the domain, increasing with increasing dimensionality (Fig. 2.4, lower panels). The exception for skewness was with respect to longitude and latitude of the square-based pyramid, where skewness increased with decreasing species diversity (moving from center to edge).

Domain shape

Patterns in predicted species richness are shown in Fig. 2.5, becoming increasingly complex with increasing number of dimensions, and differing greatly depending on the dimensions examined. This is exemplified with respect to the pyramidal domain, where there was a shift in the richness center, in comparing longitude x latitude predictions with those of longitude x altitude, or latitude x altitude. While the single dimensions of longitude and latitude both show a distinct mid-domain peak in species richness, the elevational peak was shifted toward lower altitudes (Fig. 2.2 d,e,f). This was evidently a function of the greater proportionate representation of lower altitudes within the domain (see above).

Extending the MDE beyond one-dimension increases the complexity of both the model and its predictions. For the simplest domain, the pyramid, species richness peaks at 0.5 in each dimension for longitude and latitude, i.e., at the center of the square domain (Fig. 2.5, left panels). However, the two-dimensional predictions of longitude x altitude or latitude x altitude show species richness peaks located where longitude (or latitude) is at position 0.5 of the domain and altitude is at 0.34 – two major and two minor peaks in species richness. With respect to the three-dimensional predictions, four peaks occurred (Fig. 2.5, lower left panel).

Within the empirical domains, the two-dimensional prediction of species richness for longitude x latitude appears in the center of the domain, with a single peak. By including altitude in the mid-domain model presented here, multiple peaks in species richness become apparent, especially with respect to combinations including altitude. For both the North American and Australian domains, there was a single dominant peak in species richness. However, there was no gradual decrease in species richness from those peaks, unlike the situation for longitude x latitude predictions (Fig. 2.5, center and right panels).

Discussion

Theoretical range-size frequency distribution

The predictions presented here are most appropriately tested against empirical species having RSFD matching that of the theoretical species RSFD. Theoretical RSFDs created by the present model are similar to those of other fully stochastic models, monotonically decreasing the frequency of range size, and with the maximum range size equaling the full breadth of the domain in each dimension (Fig. 2.3, top panels). This is similar to what was observed for the fully stochastic one-dimensional models of Colwell and Hurtt

(1994), Willig and Lyons (1998) and Colwell and Lees (2000). One could expect that as conditions of range size became more stringent (increasing the number of dimensions to be accounted for), mean range size should decrease and become more right-skewed. For example, if a species range is defined as 0.2 to 0.8 for each of the longitude and latitude dimensions within a domain having limits of 0 and 1, the range size for that species in either single dimension should be limited to (0.8-0.2)/1 = 0.6 of the domain. When accounting for both dimensions, range size is limited to 60% of each dimension, or just 36% of the total domain. Indeed this model demonstrates decreased mean range size and increased skewness, with increasing dimensionality (Fig. 2.3).

The log-normal RSFD curve is biologically realistic (Anderson 1985, Brown et al. 1996, Colwell and Lees 2000), and is a theoretical standard to which empirical RSFD patterns have been compared (Gaston et al. 2005; and see e.g., Pagel et al. 1991, Blackburn and Gaston 1996, Gaston 1998, Macpherson 2003). The differences between empirical and MDE-predicted patterns, in species richness and in RSFD, may be due to the non-random placement of species within the domain or to the differences in species RSFD, or a combination of both – potentially underestimating, or overestimating, the MDE (McCain 2003, Colwell et al. 2004, Arita 2005).

Patterns in species richness and range-size frequency distribution

Independent of domain shape, the stochastic model presented here further confirmed previously reported predictions of species richness patterns (suggested by Colwell and Hurtt 1994, Willig and Lyons 1998, Colwell and Lees 2000, Arita 2005) and patterns of

RSFD (Arita 2005). The models presented here confirmed the predictions of MDE models showing peak richness to be 0.5 x the number of species in one-dimensional predictions, 0.25 (= 0.5^2) for two-dimensional prediction (Colwell and Hurtt 1994, Willig and Lyons 1998, Colwell and Lees 2000, Arita 2005) and then, extending this to three dimensions, 0.13 (= 0.5^3) x the number of species.

Arita (2005) was the first to address patterns in RSFD at different latitudes, under the assumption of a fully stochastic one-dimensional MDE model, and extended the predictions into two-dimensions. Because Arita's (2005) theoretical domain was square, in order to make direct comparison with my irregularly-shaped domains, I attempted to remove the influence of shape by using the assumption that the predicted species richness was greatest at mid-domain and lowest at the limits of the domain.

Arita's first generalization was that MDE models predict no gradient in mean range size, with a mean of 0.5 x the extent of the domain for any single dimension or 0.25 x the extents for two dimensional combinations. Extending this to three dimensions, mean range size should be 0.13 x the extent. For the model presented here, all predictions fit this generalization, except in areas of very low species richness. All dimensions and combinations of dimensions predicted a relatively constant mean range size for the one-, two- and three-dimensional predictions, in line with Arita (2005), across all domains (Fig. 2.4, top panels). Small ranges can be located throughout the domain and large ranges are constrained to be nearer the center of the domain (Zapata et al. 2003), so it seems not unrealistic to expect that species at the limits of the domain exhibit smaller mean range sizes.

Arita's second MDE generalization was "variance in range size is lowest at the middle of the continent and highest near the corners of a square-shaped continent". Put simply, variance in range size should decrease from low species richness (at the limits of the domain) to high species richness (center of the domain). My results also support this (Fig. 2.4, middle panels). With the exception of extremely low species richnesses, variance in mean range size decreased from edge to center of domains. Arita's third generalization was that, while RSFD is right-skewed at any point of the domain, the skewness increases with decreasing species diversity (moving from the center to the edge of the domain). This was only observed with respect to longitude and latitude of the square-based pyramid – a square domain as used by Arita (2005), suggesting domain shape is important here (Fig. 2.4, lower panels). With the exception of the above-mentioned square-based pyramid domain, the predictions suggested that skewness was relatively constant at any point within the domains and that skewness increased with increasing dimensionality.

Domain shape

Few have attempted to modify MDE models to account for irregularly-shaped domains. The basic Bokma et al. (2001) model has been modified by using proportionate distance from domain edge for the exact latitudinal and longitudinal extent at each point in the domain (Diniz-Filho et al. 2002, Hawkins and Diniz-Filho 2002) or by taking the distance to the absolute edge of the domain for any point and deriving the proportionate distance as a measure of the maximum domain size in each dimension (Hawkins and Diniz-Filho 2002). However, these methods can still show anomalous secondary peaks in peninsulas (Hawkins and Diniz-Filho 2002).

While attempting to account for irregularly-shaped domains, the developments of Bokma et al. (2001) models do not speak to the RSFD patterns. Two further methods exist, based again on the random placement of ranges within the domain, for prediction of both species richness and location-specific RSFD within irregularly shaped domains. The first uses a "spreading dye" algorithm (Jetz and Rahbek 2001, 2002). The model re-samples range sizes in two dimensions (area) from an empirical RSFD. These are randomly placed by first selecting a cell within a gridded domain and "growing" the species range to the selected size by adding adjacent cells chosen at random – creating irregularly shaped ranges. While range size in area is maintained from an empirical RSFD, range size in either single dimension is not (Zapata et al. 2003).

The second method is a process-based, evolutionary MDE model. Initially proposed by Bokma et al. (2001), this approach differs from others in that it considers the evolutionary and biotic processes of speciation, extinction and dispersal in defining species ranges – suggesting an ecological basis for MDE. Species are randomly placed in a single cell of a gridded domain and with successive generations they may speciate, go extinct or disperse. If rates of speciation and extinction are equal, species richness becomes driven by dispersal, and since dispersal at the edge of the domain is lower (due to fewer shared cell edges), a mid-domain peak in species richness is again predicted. Recently, several other process-based approaches have been proposed that further support MDE predictions (see e.g., Connolly 2005, Rangel and Diniz-Filho 2005a, b). These models are limited by the need to estimate parameters such as rates of speciation, extinction and dispersal.

While the model presented here makes predictions of species richness peaks and patterns of RSFD independent of shape similar to other models, it predicts complex gradients in species richness that increase with increasing complexity of the domain being examined. The procedure used for stochastically creating and placing theoretical species ranges generates a pattern of richness values that peaks mid-domain for longitude and latitude, as previously reported (Colwell and Hurtt 1994, Willig and Lyons 1998, Colwell and Lees 2000). However, the richness peak for elevation is shifted toward lower altitudes. In general, species richness tends to decrease with elevation (e.g., Stevens 1992, Brown and Lomolino 1998). Indeed Rahbek (1995) reviewed 97 studies on elevational gradients in species diversity and found most studies showed greatest diversity at lower altitudes, but nearly half the studies exhibited a mid-domain peak. It is unclear where within the domain the peak occurred (shifted from the middle of the domain toward higher or lower elevations) and to what degree the richness peak corresponded to the proportional area of the domain (per unit of elevation).

Sanders (2002) showed that peaks in species richness corresponded with the area in elevational bands, for ant richness in three U.S. states. Similarly, Carpenter (2005) showed that for trees and understory plants in the Himalayas, peaks in species richness were not centered over the mid-domain for an elevational gradient, but were shifted to the left. This corresponds to a greater proportion of the domain being represented by lower elevations. The "mid-domain" elevational peaks Rahbek (1995) noted may indeed be shifted to the left (not immediately over the middle of the domain), as with my observations. Sanders (2002) stated that "species richness peaked at mid-elevations", yet in one of that paper's figures (Sanders 2002: Fig. 1), the mid-domain peak was not immediately over the mid-domain but shifted toward the lower elevations, for each of three domains examined.

While the shift in elevational richness toward lower altitudes seems empirically justifiable, it can also be demonstrated geometrically. Fig. 2.6 depicts each possible two-dimensional combination for the pyramid domain. In looking explicitly at longitude x latitude (Fig. 2.6, left panel), a "mid-domain" peak is expected at the geometric center of the square domain, thus predicting richness peaks for both longitude and latitude at the midpoint of each dimension. However, when examining longitude x altitude or latitude x altitude (Fig. 2.6, center and right panels), the domain becomes triangular in shape, not square. Geometric centers can be found by connecting the vertices of the triangle to the opposite midpoints. The center of the triangle (peak in species richness) continues to correspond to the 0.5 x domain extent for longitude and latitude, but the peak is shifted toward lower altitudes, to 0.34 (within domain extents of 0 and 1) – the same as what was predicted for the pyramid-shaped domain by selecting range endpoints based on the proportionate area frequencies.

The novelty of the model proposed here is that it can be applied to irregularly-shaped domains, in the three dimensions that make up a domain. However, this increases the complexity of both the model and the predictions. Single-dimension models predicted that in any of the three dimensions, mid-domain peaks at 0.5 (in a domain of 0 to 1) should be observed (Colwell and Hurtt 1994, Willig and Lyons 1998, Colwell and Lees 2000, Arita 2005), with the shifting of the peak in elevation toward lower altitudes already discussed with respect to domain shape and proportionate area. Given that peak richness occurs at position 0.5 for both latitude and longitude, and at 0.34 for altitude, peak domain richness should occur where all conditions are met in any of the two- or three-dimension combinations. Indeed, with the simplest of domains, the pyramid (Fig. 2.5, left panels), the longitude x latitude pattern of species richness, peaking at the center of the square domain, is arguably the most intuitive and recognizable pattern, extending MDE beyond a single dimension (see e.g., Willig and Lyons 1998, Arita 2005, Rangel and Diniz-Filho 2005a).

Application of two-dimensional MDE models has been limited, due to the limited ability to account for the irregular shape of empirical domains. Within the domains examined here, the two-dimensional prediction of species richness for longitude x latitude is intuitively in the center of the domain with a single peak. However, by including altitude in the mid-domain model presented here, multiple peaks in species richness become apparent, especially with respect to longitude x altitude and latitude x altitude combinations.

The two-dimensional predictions of longitude x altitude, and latitude x altitude for the pyramid domain, although not as intuitive, resulted in two major and two minor peaks in species richness, located where longitude (or latitude) occurred at position 0.5 on the domain and where altitude occurred at 0.34. Extending this to include all three dimensions, peak species richness should be seen at the position where longitude and latitude equal 0.5 and altitude is 0.34. As these conditions do not occur within the pyramidal domain, four peaks in species richness are observed, each with a peak height lower than expected.

With respect to the two- and three-dimensional predictions for both the North American and Australian domains, a single dominant peak in species richness occurred; however, there was no gradual decrease in species richness from that peak (unlike longitude x latitude predictions; Fig. 2.5, center and right panels). Secondary peaks in the two- and three- dimensional models are not anomalies as found by Hawkins and Diniz-Filho (2002) but rather show the major effect of including elevation in the model. Indeed, multiple richness peaks and non-uniform gradients in species richness are seen in empirical data (see e.g., Hawkins and Diniz-Filho 2002, Rangel and Diniz-Filho 2005a, Murphy et al. 2006) and may be due to multi-dimensional MDE rather than environmental gradients.

Hawkins and Diniz-Filho (2002) concluded that a two-dimensional MDE model (longitude x latitude) could not be used to explain the diversity gradient in North American Nearctic birds. Colwell et al. (2004) later noted there is a clear latitudinal gradient in bird richness, and suggested the lack of a longitudinal gradient may be due to pooling of avian data from longitudinally distinct biomes – i.e., an environmental gradient. While an environmental gradient may influence species richness patterns, I suggest the observed species richness pattern in Nearctic birds resembles the predicted gradient for latitude x altitude, and support Colwell et al.'s (2004) suggestion that all possible combinations of dimensions should be used to examine MDE.

In recognizing that irregular shape of domains may be important, some caution is needed when defining a domain. As maps are abstract representations of the earth, the transition from a sphere to flat surface distorts the true dimensions of a domain in shape, area and/or direction (Brainerd and Pang 2001). The problem of conversion of a sphere to a planar surface has interested cartographers, mathematicians and navigators for nearly 2500 years (Snyder and Voxland 1989); hundreds of such projections are used – each with tradeoffs in maintaining distance, direction, size, area, etc. Given the basic assumption of MDE that geometry, or shape of a domain, drives distributions, domains should use projections that best maintain the original shape of the domain. In doing so, they should also minimize distortions in area (Bokma et al. 2001). This becomes more important with larger domains as distortion generally increases with increasing domain size (Brainerd and Pang 2001).

References

Anderson, S., 1985. The theory of range-size (RS) distributions. American Museum Novitates 2833: 1-20.

Arita, H.T., 2005. Range size in mid-domain models of species diversity. Journal of Theoretical Biology 232: 119-126.

Blackburn, T.M. & Gaston, K.J., 1996. Spatial patterns in the geographic range sizes of bird species in the New World. Philosophical Transactions of the Royal Society of London, Series B 351: 897-912.

Bokma, F., Bokma, J. & Mönkkönen, M., 2001. Random processes and geographic species richness patterns: why so few species in the north? Ecography 24: 43-49.

Bokma, J. & Mönkkönen, M., 2000. The mid-domain effect and the longitudinal dimension of continents. Trends in Ecology and Evolution 15: 288-289.

Brainerd, J. & Pang, A., 2001. Interactive map projections and distortion. Computers & Geosciences 27: 299-314.

Brown, J.H. & Lomolino, M.V., 1998. Biogeography. Sinauer Associates, Sunderland, MA.

Brown, J.H., Stevens, G.C. & Kaufman, D.M., 1996. The geographic range: size, shape, boundaries, and internal structure. Annual Review of Ecology and Systematics 27: 597-623.

Carpenter, C., 2005. The environmental control of plant species density on a Himilayan elevation gradient. Journal of Biogeography 32: 999-1018.

Colwell, R.K., 2000. Rangemodel: A Monte Carlo Simulation Tool for Assessing Geometric Constraints on Species Richness, Version 3; User's Guide and Application. [WWW document]. URL http://viceroy.eeb.uconn.edu/rangemodel.

Colwell, R.K. & Hurtt, G.C., 1994. Nonbiological gradients in species richness and a spurious Rapoport effect. American Naturalist 144: 570-595.

Colwell, R.K. & Lees, D.C., 2000. The mid-domain effect: geometric constraints on the geography of species richness. Trends in Ecology and Evolution 15: 70-76.

Colwell, R.K., Rahbek, C. & Gotelli, N.J., 2004. The mid-domain effect and species richness patterns: What have we learned so far? American Naturalist 163: E1-E23.

Colwell, R.K., Rahbek, C. & Gotelli, N.J., 2005. The mid-domain effect: there's a baby in the bathwater. American Naturalist 166: E149-E154.

Connolly, S.R., 2005. Process-based models of species distributions and the mid-domain effect. American Naturalist 166: 1-11.

Diniz-Filho, J.A.F., de Sant'Ana, C.E.R., de Souza, M.C. & Rangel, T.F.L.V.B., 2002. Null models and spatial patterns of species richness in South American birds of prey. Ecology Letters 5: 47-55.

Gaston, K.J., 1998. Species–range size distributions: products of speciation, extinction and transformation. Philosophical Transactions of the Royal Society of London, Series B 353: 219-230.

Gaston, K.J., Davies, R.G., Gascoigne, C.E. & Williamson, M., 2005. The structure of global species-range size distributions: raptors and owls. Global Ecology and Biogeography 14: 67-76.

GLOBE Task Team 1999. The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0. National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, CO, U.S.A. Digital data base on the World Wide Web (URL: http://www.ngdc.noaa.gov/mgg/topo/globe.html) and CD-ROMs.

Hawkins, B.A. & Diniz-Filho, J.A.F., 2002. The mid-domain effect cannot explain the diversity gradient of Nearctic birds. Global Ecology and Biogeography 11: 419-426.

Hawkins, B.A., Diniz-Filho, J.A.F. & Weis, A.E., 2005. The mid-domain effect and diversity gradients: is there anything to learn? American Naturalist 166: E140-E143.

Hubbell, S.P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton, NJ.

Jetz, W. & Rahbek, C., 2001. Geometric constraints explain much of the species richness patterns in African birds. Proceedings of the National Academy of Sciences U.S.A. 98: 5661-5666.

Jetz, W. & Rahbek, C., 2002. Geographic range size and determinants of avian species richness. Science 297: 1548-1551.

Koleff, P. & Gaston, K.J., 2001. Latitudinal gradients in diversity: real patterns and random models. Ecography 24: 341-351.

Laurie, H. & Silander Jr., J.A., 2002. Geometric constraints and spatial patterns of species richness: critique of range-based null models. Diversity and Distributions 8: 351-364.

Lees, D.C., Kremen, C. & Andriamampianina, L., 1999. A null model for species richness gradients: bounded range overlap of butterflies and other rainforest endemics in Madagascar. Biological Journal of the Linnean Society 67: 529-584.

MacArthur, R.H., 1957. On the relative abundance of bird species. Proceedings of the National Academy of Sciences U.S.A. 43: 293-295.

Macpherson, E., 2003. Species range size distributions for some marine taxa in the Atlantic Ocean: effect of latitude and depth. Biological Journal of the Linnean Society 80: 437-455.

McCain, C. M., 2003. North American desert rodents: a test of the mid-domain effect in species richness. Journal of Mammalogy 84: 967-980.

McCain, C.M., 2004. The mid-domain effect applied to elevational gradients: species richness of small mammals in Costa Rica. Journal of Biogeography 31: 19-31.

Mora, C. & Robertson, D.R., 2005. Causes of latitudinal gradients in species richness: a test with fishes of the tropical Eastern Pacific. Ecology 86: 1771-1782.

Murphy, H.T., VanDerWal, J., Khalatkhar, N. & Lovett-Doust, J., 2006. Incorporating 3-dimensional geographic range geometry in mid-domain models: geometric constraints and environmental correlates of North American tree species richness. Ecography. In Review.

Pagel, M.P., May, R.M. & Collie, A.R., 1991. Ecological aspects of the geographic distribution and diversity of mammalian species. American Naturalist 137: 791-815.

Rahbek, C., 1995. The elevational gradient of species richness: a uniform pattern? Ecography 18: 200-205.

Rahbek, C. & Graves, G.R., 2001. Multiscale assessment of patterns of avian species richness. Proceedings of the National Academy of Sciences U.S.A. 98: 4534-4539.

Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F., 2005a. An evolutionary tolerance model explaining spatial patterns in species richness under environmental gradients and geometric constraints. Ecography 28: 253-263.

Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F., 2005b. Neutral community dynamics, the mid-domain effect and spatial patterns in species richness. Ecology Letters 8: 783-790.

Rapoport, E.H., 1982. Areography: geographical strategies of species. B. Drausel, Trans. Pergamon, Oxford.

Sanders, N.J., 2002. Elevational gradients in ant species richness: area, geometry, and Rapoport's rule. Ecography 25: 25-32.

Snyder, J.P. & Voxland, P.M., 1989. An Album of Map Projections. U.S. Geological Survey Professional Paper 1453. United States Government Printing Office, Denver, CO.

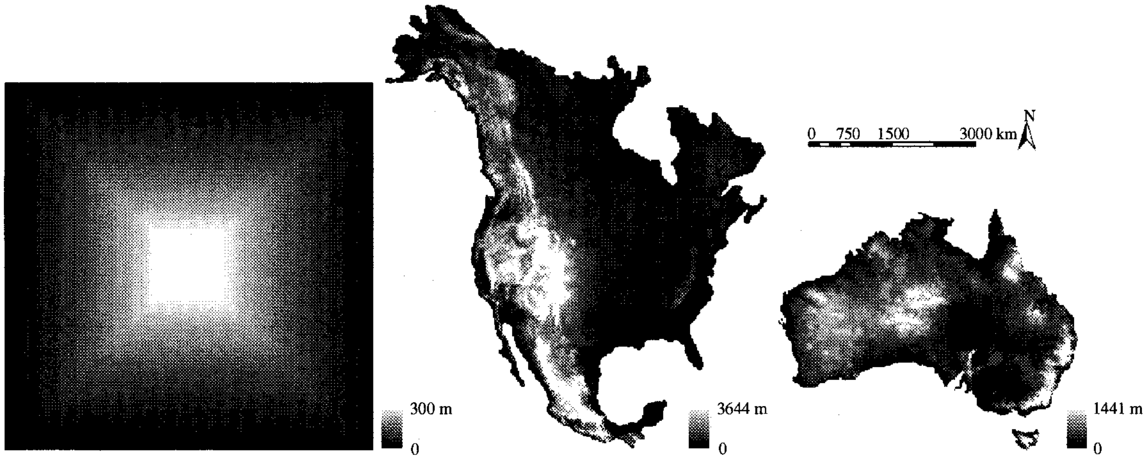
Stevens, G.C., 1989. The latitudinal gradient in geographical range: How so many species coexist in the tropics. American Naturalist 133: 240-256.

Stevens, G.C., 1992. The elevational gradient in geographical range: an extension of Rapoport's latitudinal rule to altitude. American Naturalist 140: 893-911.

Willig, M.R. & Lyons, S.K., 1998. An analytical model of latitudinal gradients of species richness with an empirical test for marsupials and bats in the New World. Oikos 81: 93-98.

Zapata, F.A., Gaston, K.J. & Chown, S.L., 2003. Mid-domain models of species richness gradients: assumptions, methods and evidence. Journal of Animal Ecology 72: 677-690.

Zapata, F.A., Gaston, K.J. & Chown, S.L., 2005. The mid-domain effect revisited. American Naturalist 166: E144-E148.



Square-based pyramid North America Australia

Figure 2.1. Illustration of the three domains: a square-based pyramid, continental North America, and Australia. Perimeters of the continental domains were extracted from "World Countries 1992" base map supplied with ArcView 3.2 (ESRI, California, USA) and altitudes were obtained from the GLOBE project (GLOBE 1999). Lighter areas represent higher elevations, reported in metres above sea level.

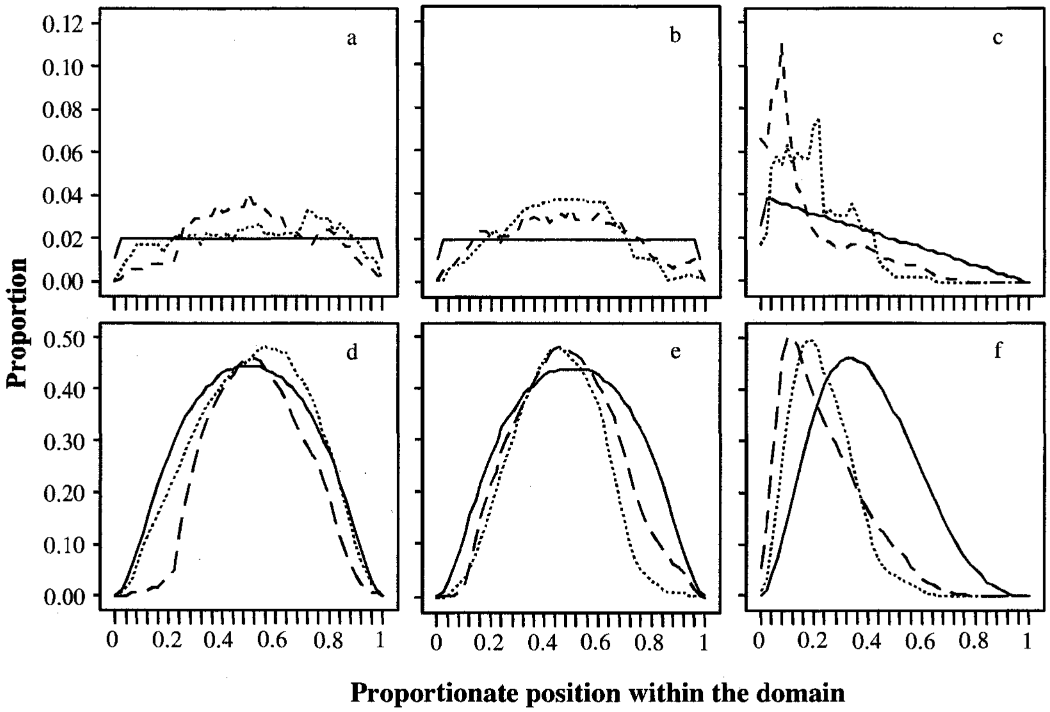


Figure 2.2. For each single dimension, longitude (a,d), latitude (b,e) and altitude (c,f), the relationship between proportionate area (a,b,c), predicted species richness (d,e,f) and position within the domain is depicted for a square-based pyramid (——), North American (- - -) and Australian (· · ·) continental domains.

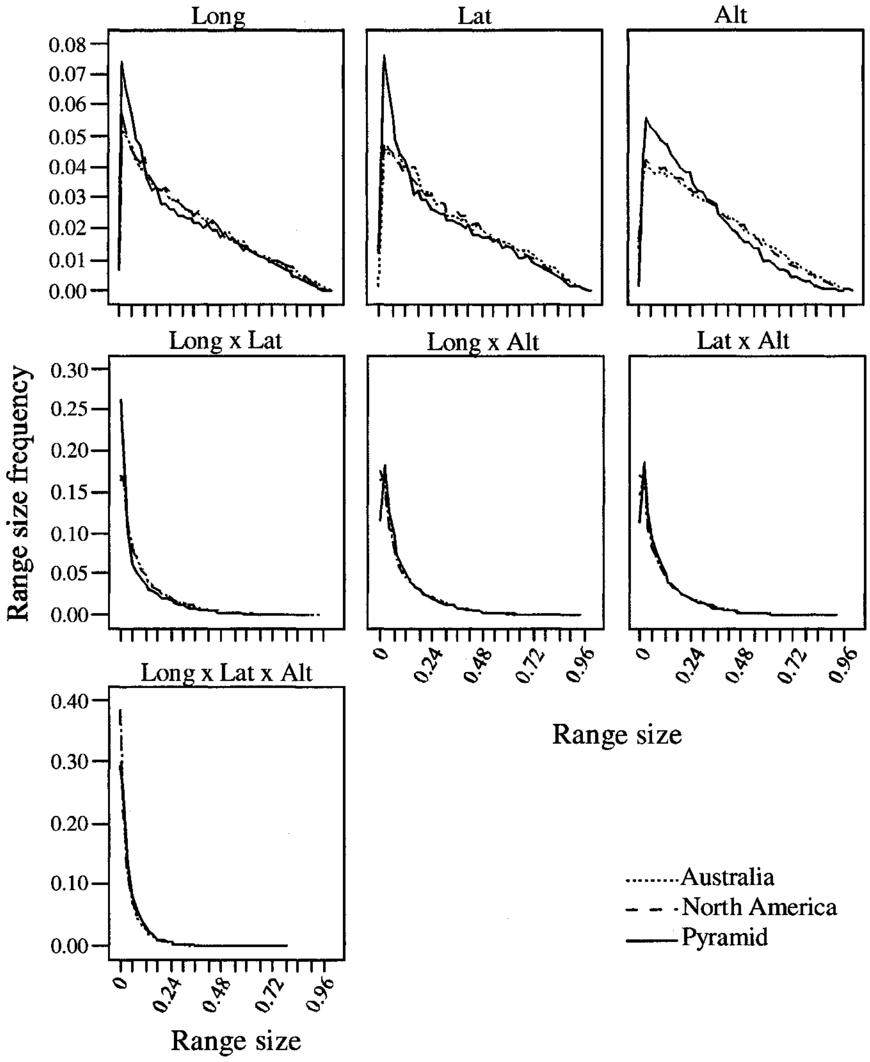


Figure 2.3. Range-size frequency distribution of theoretical species, stochastically placed within each domain, for a square-based pyramid (——), North American (- - -) and Australian (· · ·) continental domains.

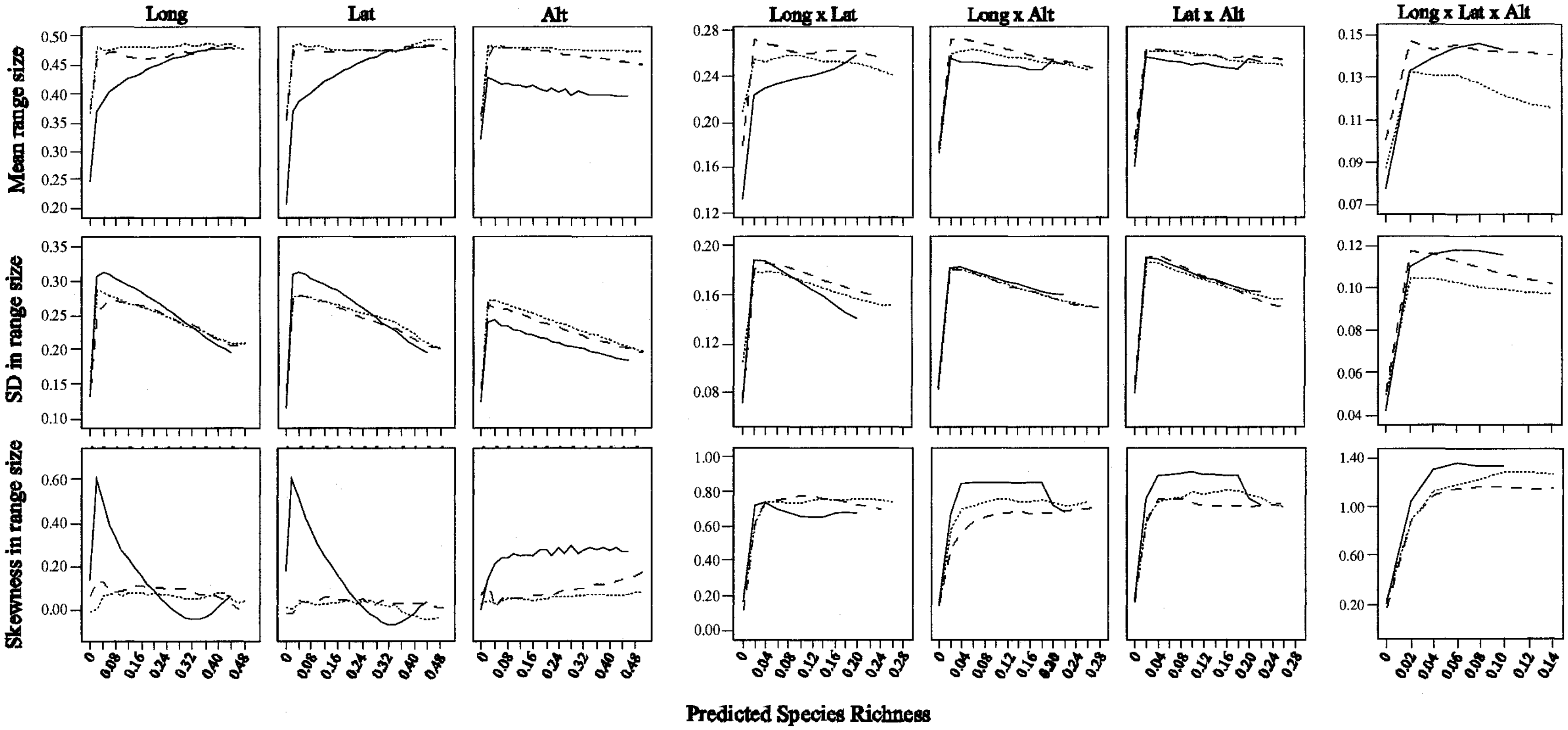


Figure 2.4. The relationship between predicted patterns in range-size frequency distributions and species richness for individual dimensions and multi-dimensional combinations. Lines represent the mean values of mean range size (top panels), standard deviation in range size (middle panels) and skewness in range size (lower panels), for each domain. The three domains consisted of a square-based pyramid (——), North American (- - -) and Australian (· · ·) continental domains.

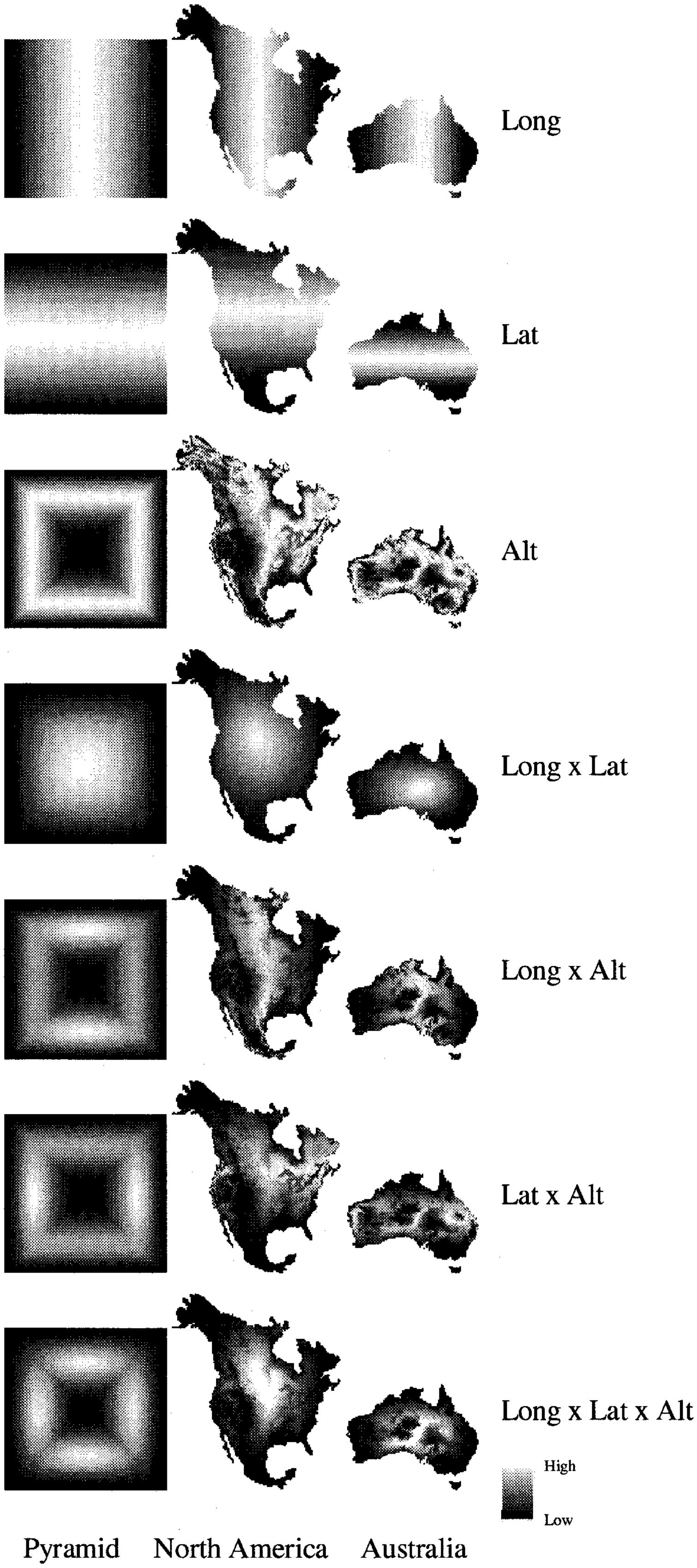

Long
Lat
Alt
Long x Lat
Long x Alt
Lat x Alt
Long x Lat x Alt
High
Low
Pyramid
North America
Australia

Figure 2.5. Patterns of species richness as predicted by a modified, fully stochastic, three dimensional MDE model for each of a square-based pyramid, North American and Australian continental domains. One-dimensional (Longitude, Latitude and Altitude) and multi-dimensional (Long x Lat, Long x Alt, Lat x Alt and Long x Lat x Alt) predictions are shown. Equal interval classification is shown with color ramps indicating minimum (dark, bottom of legend) to maximum (light, top of legend) predicted species richness. The color ramp represents 0 - 0.5 x total number of species for the uni-dimensional predictions, 0 - 0.28 for the two-dimensional, and 0 - 0.14 for the three-dimensional model.

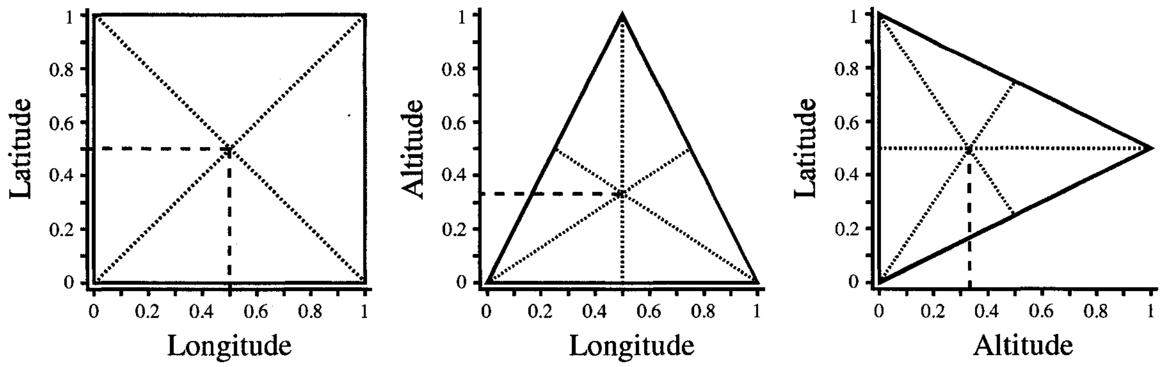


Figure 2.6. Illustration of the rationale of predicting location of peak species richness. Mid-domain models predict a peak in species richness at the geometric center of a domain. The intersection of the dotted lines represents the center of the domain for each two-dimensional combination for a square-based pyramid domain. The axis represents the proportionate distance across the domain in each of longitude, latitude and altitude. For both longitude and latitude, species richness peaks at the domain mid point (0.5) while for altitude, species richness peaks at lower altitudes at position 0.34.

Chapter 3 - Incorporating three-dimensional geographic range geometry in mid-domain models: geometric constraints and environmental correlates of North American tree species richness.

Summary

The 'mid-domain effect' (MDE) has received much attention recently as a candidate explanation for patterns in species richness over large geographic areas. Mid-domain models generate a central peak in richness, in the absence of environmental gradients, when species ranges are randomly placed within a bounded geographic area (i.e., the domain). Until now, domain limits have been described mostly in one dimension, usually latitude or altitude, and only occasionally in two dimensions. Here, one-, two- and, for the first time, three-dimensional mid-domain models are tested and the independent and concurrent effects of geometric constraints and environmental variables on species richness of 417 North American tree species are assessed.

For species with large ranges, MDE alone explained between 69% and 98% of the variation in species richness. For these species, MDE also substantially improved the fit (by up to 21%) of multiple regression models which also included environmental variables. The largest difference in fit between models with and without MDE was for the two-dimensional (latitude x altitude) model for all ranges, with an improvement of 23% when MDE was included in the model. For medium- and small-range species, geometric

constraints have little influence on patterns of species richness and environmental parameters are much more important.

The analysis addresses many of the recent methodological criticisms directed at studies testing the MDE, and the results support the hypothesis that species richness patterns may be influenced by geometric constraints. I suggest few good arguments remain for why geometric constraints should not be considered alongside more traditional environmental correlates in understanding patterns of species richness.

Introduction

Ecologists and biogeographers for over two centuries have hotly debated the mechanisms underlying geographic gradients in species richness, without an emergent consensus (e.g., Rhode 1992; Willig et al. 2003; Hawkins et al. 2003). Biogeographic patterns in species richness have traditionally been seen as reflecting underlying geographical gradients in a variety of ecological and evolutionary factors (see e.g., Currie and Paquin 1987; Stevens 1989; Grytnes and Vetaas 2002; Hawkins et al. 2003; Willig et al. 2003; Zapata et al. 2003). In principle, if classical niche-assembly theory prevails, species richness should vary geographically in concert with biotic or abiotic gradients. For the longest time, biologists worked under the assumption that species richness should be spatially uniform in the absence of these gradients (Colwell et al. 2004).

Recently, using null models, Colwell and Hurtt (1994) showed that 'nonbiological' gradients in species richness arise from the assumption of a random one-dimensional association between the size and placement of species' ranges. A central peak in species richness is generated in the absence of a direct effect of environmental gradients, when ranges (empirical or theoretical) are randomly placed within a geographic area bounded by features that limit dispersal (i.e., the domain). Colwell and Lees (2000) labeled the geometric theory of species richness gradients the "mid-domain effect" (MDE) and defined it as "the increasing overlap of species ranges towards the centre of a shared geographic domain due to geometric boundary constraints in relation to the distribution of species' range sizes and midpoints." Mid-domain models view the real-world distribution of species within a geographic domain as a statistical aggregate of deterministic factors acting at more local scales (Colwell et al. 2004). The MDE focuses on the emergent, macroecological pattern that geometric constraints predict, rather than on the effects of constraints on individual ranges (Colwell and Lees 2000).

Most studies of the MDE have used one-dimensional models to study patterns of species richness, in either a latitudinal (e.g., Willig and Lyons 1998; Koleff and Gaston 2001; McCain 2003; Mora and Robertson 2005) or altitudinal dimension (e.g., Bachman et al. 2004; Gryntes and Vetaas 2002; Sanders 2002). Only a few studies have used two-dimensional models (e.g., Bokma et al. 2001; Jetz and Rahbek 2001; Diniz-Filho et al. 2002; Hawkins and Diniz-Filho 2002), and none that have modeled species richness with an MDE model in three dimensions (but see Chapter 2). Three-dimensional models in terrestrial systems would explicitly include the vertical altitudinal dimension as well as the two horizontal dimensions of latitude and longitude. One-dimensional (usually latitudinal) mid-domain models are criticized as simplistic and unrealistic because natural ranges are always constrained in at least two horizontal axes (Bokma et al. 2001; Zapata et al. 2003; Hawkins 2004). Moreover, analysis solely of one-dimensional patterns restricts explanations to mechanisms that vary in that dimension and ignores mechanisms having more complex spatial patterns (Hawkins 2004). Colwell et al. (2004) counter these arguments by noting that neither is the world two-dimensional, yet a reduction in dimensionality is a commonly used strategy for studying complex patterns in statistics and science. Yet, as Colwell et al. suggest, where possible, presenting both one- and two-dimensional analyses (in the absence at that time of any methodology for evaluating three-dimensional models) of the same data set seems the strongest approach.

Small-ranged species are less likely to experience 'hard' boundaries than large-range species, and thus the impact of boundaries on their richness patterns should be smaller (Jetz and Rahbek 2002), and more likely to reflect environmental and historical factors (Colwell and Lees 2000). Larger-ranged species are much more likely to be affected by continental geometry and are also more likely to occupy the centre of the bounded area (Colwell and Lees 2000). As predicted by MDE, where ranges are small relative to the extent of the domain, MDE tends to be weaker (Laurie and Silander 2002), and where ranges are large relative to the extent of the domain, MDE is usually stronger (McCain 2003). Similarly, studies that have partitioned datasets into range size categories have consistently found stronger support for MDE among large-ranged species than among the smaller range species in the dataset (Hawkins and Diniz-Filho 2002; Jetz and Rahbek 2002; Vetaas and Grytnes 2002; Mora and Robertson 2005). Mora and Robertson (2005) found habitat features had a much greater influence on gradients in species richness of narrow-ranging tropical eastern Pacific fishes, whereas the mid-domain effect had a significant influence on species richness of large-ranged species (though they only tested a one-dimensional MDE model).

Several recent reviews provide a good background to the development of various types of MDE models and their progression from one- to two-dimensional (see in particular Zapata et al. 2003; Colwell et al. 2004). These will not be summarized in this chapter. Provided here is the first empirical test of a novel methodology for incorporating a third dimension to a modified, fully stochastic MDE model. The aim of this chapter is to determine whether patterns of North American tree species richness are driven primarily by 'non-biological' geometric constraints (whether one-, two-or three-dimensional), or by environmental drivers. Spatial linear regression models are used to determine the relative contribution of a number of potential casual agents (MDE, climate, topography, habitat diversity) to observed patterns of richness of 417 tree species. Furthermore, how these relationships vary for species having small-, medium- and large-range sizes (relative to the domain) are examined.

Methods

Species ranges

Between 1971 and 1977, Elbert Little, Chief Dendrologist with the U.S.D.A. Forest Service published a series of maps of tree species ranges based on inventory lists, detailed forest surveys, field notes and herbarium specimens (Prasad and Iverson 2003). These published (and now digitized) maps have become the standard reference for most

U.S. and Canadian tree species ranges. The database currently includes range maps for 431 species (see Appendix 3A). Fourteen species ranges were excluded from the analysis, either because their ranges extended beyond the hard boundary of the domain or their range was smaller than one 20 km x 20 km cell. Empirical range sizes were defined by their limits in each of the dimensions, latitude, longitude and altitude, as measured from their range maps.

Mid-domain models

Continental North America was divided into 20 km x 20 km grid cells. Cells not part of the continental land mass or which formed small peninsulas (size = 1 cell) were excluded from the analysis, giving a total of 49,323 cells within the domain. The relationships of the main effects (one-dimensional) and interaction effects (two- and three-dimensional) of latitude, longitude and altitude were explored on species richness across the domain (a total of seven models) using a modified, fully stochastic mid-domain model (described in Chapter 2). For this type of model, ranges are usually generated by selecting range endpoints or midpoints randomly from a uniform distribution of possible values, or by sampling from permissible range-size mid-point location pairs (Colwell and Lees 2000; Arita 2005). As a result, fully stochastic models are sometimes criticized because the range-size frequency distribution (RSFD) of the simulated species differs from the empirical data and thus may produce different patterns of richness when ranges are placed randomly on a domain.

Building on the 'two-hit' model of Arita (2005) I created 417 theoretical species ranges by randomly selecting the endpoints of a range in each of the three dimensions, employing in effect a 'six-hit' model (Chapter 2). For any single dimension, it is assumed a species cannot occur outside the "hard" limits (0 and 1) of the domain and thus, a species range can be defined by its limits L_1 and L_2 where $0 \leq L_1, L_2 \geq 1$ and $L_2 \geq L_1$ (Willig and Lyons 1998; Arita 2005). To generate a species range, two random points within the range are randomly placed within the domain (i.e., for longitude X_1 and X_2) where $L_1 = \text{MIN}(X_1, X_2)$ and $L_2 = \text{MAX}(X_1, X_2)$. The range size R is a function of X_1 and X_2 and can be represented as $R = |X_1 - X_2|$, such that $0 \leq R \leq 1$. Extending this beyond the single dimension, a range can be defined by its limits in longitude (X_1, X_2), latitude (Y_1, Y_2) and altitude (Z_1, Z_2).

For the three-dimensional model, the extent of possible values for X_1, X_2, Y_1, Y_2, Z_1 and Z_2 is represented by a unit cube. However, because of the irregularly shaped surface of North America, the random selection of endpoints may create a range where $R = 0$ (i.e. when $R_X \cap R_Y$ but does not intersect R_Z) in the three-dimensional model. When this occurred the theoretical species was discarded and a new one was created, with six new endpoints. In addition, the extents of the theoretical ranges in longitude, latitude and altitude may be smaller than that defined by its randomly selected endpoints. For example, if X_{max} for range R, where R is a function of $R_X \cap R_Y \cap R_Z$, was less than the randomly selected L_{X2} where $L_{X2} = \text{MAX}(X_1, X_2)$ then $R_X < |X_1 - X_2|$). In this case, the maximum and minimum extents of the range were defined by the limits of R rather than the randomly selected endpoints.

In Chapter 2 I have shown how domain shape can influence MDE model predictions and proposed modifying the fully stochastic model to account for the proportion of the domain in each unit of a particular dimension. For the irregularly-shaped North American domain, the proportion of the domain represented by latitude and longitude represent a nearly normal distribution while altitude shows a right-skewed distribution (Chapter 2). Thus, much more of the domain in the altitudinal dimension occurs at low altitudes than at high altitudes. In the model used here, range limits for each of longitude, latitude and altitude (i.e., two endpoints for each dimension – six hits) were selected at random from the non-uniform distributions of possible endpoints based on the proportion of the domain represented by a single point (or for each measured unit) in each dimension.

All analyses shown here compare the mean species richness generated by 100 runs (each run creating a new set of 417 theoretical species ranges) of the mid-domain model with the empirical species richness across the domain. The theoretical and empirical species ranges were mapped and species richness in each cell across the domain was calculated. Both empirical and theoretical species range sizes were further broken into small (<10% of the domain, n = 240), medium (between 10-30% of the domain, n = 137) and large range sizes (>30% of the domain, n = 40), and mapped for comparisons of species richness within each group.

Environmental variables

Climatic data layers included in the analysis were sourced from WorldClim Version 1.3 (Hijmans et al. 2004). WorldClim includes 19 interpolated global climate layers on a square kilometer grid. Many of these layers are highly correlated; variables included in the analysis were ones with greatest correlation to the dependent variable while any variables correlated at >75% to the independent variable were exclued. Three layers remained in the analysis: (1) mean annual temperature, (2) mean temperature of the wettest annual quarter, (3) precipitation seasonality (coefficient of variation). Landcover data was extracted from the AVHRR Global Landcover Classification (Hansen et al. 1998). The diversity of landcover was calculated in each 20 x 20 km cell of the domain as

$$\text{Landcover diversity} = -\sum(P_i * \ln P_i)$$

where P_i = proportion of the landscape occupied by patch type (class) i.

Topographic parameters were extracted from the USGS HYDRO1k dataset (USGS 2005) and the mean for each 20 x 20 km cell in the domain was calculated. Variables extracted from this dataset were: (1) flow accumulation – the amount of upstream area flowing into each cell; (2) slope; and (3) aspect.

Statistical analysis

Spatial linear regression analyses using spatial autoregressive (SAR) models was used to determine the effect of geometric constraints predicted by the MDE models and environmental variables on the empirical pattern of richness (Lichstein et al. 2002). In each regression, the dependent variable was the rounded empirical value (x 1000) of species richness, while the independent variables represent the mean value of that variable for each value of the dependent. This methodology avoids collapsing the results of two- and three-dimensional MDE models into one dimension for analysis (e.g., using latitudinal bands). Single SAR models were first applied to each of the eight independent variables. Each of the variables was then included in a multiple SAR model and, finally, ran the multiple SAR model again but without the MDE variable.

The SAR model includes a spatially lagged dependent variable in the regression equation; thus, a spatial weights matrix was generated where adjacent values of empirical species richness are given a value of 1 in the weight matrix and a value of 0 is applied to values that are not adjacent. Each non-zero element in the matrix represents potential spatial interaction between two observations. The traditional R^2 measure of fit, based on the decomposition of total sum of squares into explained and residual sum of squares, is not applicable to the SAR model. Instead, a pseudo R^2 measure (ratio of the variance of the predicted values over the variance of the observed values for the dependent variable) is reported here (Anselin 1995).

There are well documented problems with assessing the fit of MDE models using correlation coefficients, due to the fact that absolute differences in magnitude are obscured (Zapata et al. 2003; Colwell et al. 2004; Romdal et al. 2005). As recommended by several authors (Colwell et al. 2004; Zapata et al. 2005), t-tests were used to examine deviation of the slope from unity and intercept from zero, for the regressions between predicted MDE species richness and empirical species richness. Together, these tests provide an indication of deviation in both shape and magnitude between predicted and empirical species richness.

Spatial statistics were done using SpaceStat (Anselin 1995). All data extraction, modeling and mapping was done in ArcGIS 9.0 (ESRI, Redlands CA).

Results

One-, two- and three-dimensional patterns of species richness

As expected, the latitudinal MDE models generally predict species richness to peak in the centre of the domain (see Fig. 3.1), at between 45°N and 58°N. The empirical species richness peak was higher (65% of species compared with 48% for the MDE model) and offset to the south of the centre of the domain, at around 35°N. The peak of empirical species richness for large-range species alone was close to the MDE peak, at 43°N, but becomes more offset (moving southerly) from the MDE peak with decreasing range size.

MDE model species richness is predicted generally down the longitudinal centre of the continent, between 103°W and 105° W (Fig. 3.2). However, empirical species richness across longitude shows a peak to the east of the centre at 85°W. For small-range species, there is a second, smaller peak in richness at 133°W.

All MDE models show a peak in species richness at low altitudes, between approximately 400-500 metres (Fig. 3.3). Empirical species richness for all ranges, large- and medium-ranges also show a peak in species richness at low altitudes. Small-range empirical species show a bimodal distribution of species richness with a strong peak at the lowest altitudes (0 - 300 m) and a second, comparably-sized peak at mid-altitudes (1700-2000 m) (Fig. 3.3).

Figure 3.4 illustrates geographically where the MDE model for different range sizes predicts richness peaks for the combinations of mid-domain for latitude and low altitude. Similarly, Figure 3.5 illustrates the MDE model predictions for mid-domain longitude and low altitude, the intersection of which is predicted to result in high species richness. The three-dimensional pattern of species richness predicted by the MDE model is shown in Figure 3.6. As expected, the three-dimensional MDE model generally predicts a peak in species richness at mid latitudes and longitudes in the domain. The two-dimensional (lat-long) predicted patterns are relatively similar (and not shown here) but show a stronger 'bull's-eye' effect; whereas, the addition of the third (altitudinal) dimension to the model adds complexity to the predicted three-dimensional pattern. When all species were combined, the empirical North American tree species show a peak in species richness in the south-eastern region of the continent. However species with large ranges (>30% of the domain) peak at the approximate centre of the domain.

One-, two- and three-dimensional predictors of species richness

The one-dimensional latitudinal model shows MDE as a significant predictor of species richness for all species, and for large-range species only (Table 3.1). The fit of the multiple regressions for all species, large- and small-range species was improved by the inclusion of MDE in the model (Table 3.1). In the one-dimensional longitudinal dimension, MDE was also independently significant in predicting species richness for all species and for large-range species but only improved the fit of the multiple regression model for large-range species (Table 3.1). MDE was a significant predictor of species richness in the one-dimensional altitude model for all range size classifications, however, the maximum improvement in fit of the multiple regression models was only 3% for the small-range species.

As an independent predictor of species richness, MDE was significant ($p < 0.05$) in 18 of the 28 models (Table 3.1). MDE was the *most* significant single predictor of species richness in 11 of the 28 models (the highest of any of the variables), including all seven of the large-range models (explaining between 69% [one-dimensional altitude] and 98% [two-dimensional lat-long] of the variation, Table 3.2) and two of the all-range models (two-dimensional lat-alt and one-dimensional altitude). MDE was also the most significant predictor of species richness for the three-dimensional and two-dimensional (lat-long) small-range models, however, the relationship was negative for these models. Table 3.1 clearly demonstrates the contrasting influence of MDE on species with different range sizes. Species richness generated by the mid-domain models is correlated significantly and positively for species with large ranges; the correlation is generally weak but positive for species with medium ranges; and for species with small ranges the correlation is generally negative and sometimes significantly so (except for in the one-dimensional altitudinal dimension where the relationship is strongly positive). Figures 3.4, 3.5 and 3.6 ([d] and [h]) also clearly show the contrasting pattern in MDE small-range species richness versus the empirical pattern.

The three-dimensional MDE model was not a significant predictor of species richness when all ranges were combined, probably because the strong positive effect of the large-range species and the strong negative effect of the small-range species serve to cancel each other out (Table 3.1). In the multiple SAR models, MDE substantially improved the fit of the regressions for large range species in all but the one-dimensional altitude model, by between 6% (one-dimensional latitude) and 21% (two-dimensional lat-long). The largest difference in fit between models with and without MDE was for the two-dimensional (lat-alt) model for all ranges, with an improvement of 23% (Table 3.1) with MDE included in the model.

The relative importance of the environmental factors varied between models. In the single spatial regressions, slope was a significant predictor of species richness in 26 of the 28 models (Table 3.1). In 21 of the 28 single spatial regression models, precipitation seasonality was significant. Of the environmental variables, precipitation seasonality and slope appeared most often as the strongest single predictors of species richness (8 of 28 models and 5 of 28 models, respectively) (Table 3.1).

MDE models consistently under-predicted empirical species richness (Table 3.2 and see Figs. 3.1 through 3.6), particularly for large-range species. Only one model, the two-dimensional lat-alt model for all species, predicted both shape and magnitude of the empirical species richness pattern well. Three of the one-dimensional altitudinal models predicted shape (i.e., slope not significantly different from 1) for the empirical species richness curve, but magnitude was under-predicted; the exception was the model for small-range species. The two-dimensional lat-alt model for all species and the one-dimensional longitude model for large-range species also had significant regressions, and slope not different from unity (Table 3.2).

Discussion

Latitudinal gradients in species richness have been widely documented, and have usually been explained by a monotonic relationship with climatic, environmental and/or energy-related factors (Currie and Paquin 1987; Grytnes and Vetaas 2002; Hawkins et al. 2003). When all range sizes are considered, the results presented here show latitudinal gradients in species richness were best explained by geometric constraints, precipitation seasonality and slope, none of which showed a monotonic decline with latitude (Table 3.1). For example, precipitation seasonality had a U-shaped pattern across the latitudinal extent of the domain (see Appendix B), with the greatest values at both low and high latitudes. However, with decreasing range size, geometric constraints became unimportant and monotonically declining mean temperatures became significant predictors of species richness.

Very few studies have examined patterns of species richness across the longitudinal dimension in North America. Though there are not necessarily expected linear gradients of environmental variables across the longitudinal dimension, in theory, species richness can be expected to vary in concert with the same environmental variables that influence its variation in the latitudinal dimension, and the results presented here generally support this. Geometric constraints, mean temperatures, precipitation seasonality and slope were the most significant predictors of species richness across the longitudinal dimension (Table 3.1). Precipitation seasonality and slope tended to be higher in western longitudes where species richness was lower; mean temperatures tended to be higher in these areas. Again geometric constraints became less important with decreasing range size.

The relationship between species richness and altitude is often assumed to mirror the latitudinal gradient, resulting in a monotonic decline with increasing altitude (Rahbek 1995; Givnish 1999; Grytnes and Vetaas 2002) and several studies have reported such a pattern for plants (Leathwick et al. 1998; Ohlemüller and Wilson 2003). However, Rahbek (1995; 2005) reviewed species richness patterns in relation to altitude and found that the most common pattern, representing approximately half of the studies, was a mid-altitudinal peak in richness. As described in Chapter 2, the low altitude peak in species richness shown here is due to the non-uniform distribution of altitude across the domain; disproportionately more of the North American continent occurs at low to mid altitudes than at high. Thus when the MDE model draws randomly from allowable range endpoints in the altitudinal dimension, low altitudes are much more likely to be drawn. This results in a MDE model that predicts a peak at low- rather than mid-altitudes (Fig. 3.3).

Bachman et al. (2004) found that area alone accounted for around half of the variance in observed species richness patterns of New Guinea palms along an elevational gradient. Without accounting for area, species and genus richness of New Guinea palms decreased monotonically with elevation. However, when the effect of area was removed (by using equal-area bands instead of equal-elevation bands) a mid-elevational peak in both species and genus richness was observed. Other researchers have also shown the strong effects of area in mid-domain analyses of species richness with altitude (Lawton et al. 1987; Sanders 2002).

Grytnes and Vetaas (2002) showed that the observed asymmetric hump in plant species richness along a Himalayan altitude gradient could arise from the effect of geometric constraints, in combination with an underlying linearly decreasing trend in species richness with altitude. Thus, geometric constraints and monotonically decreasing (or increasing) environmental variables in combination were needed to best predict changes in species richness with altitude. The results of this study suggest geometric constraints alone can explain much of the variation in species richness with altitude, however added little explanatory power to a model which included environmental variables.

As expected, geometric constraints have greatest predictive power on the distribution of species richness for species having ranges that are large relative to the domain (Table 3.1). For species with large ranges, MDE alone often explained more of the variation in species richness than all of the environmental variables combined (Tables 3.1 and 3.2). For medium- and small-range species, geometric constraints clearly have little influence on patterns of species richness and, as expected, environmental parameters are much more important (Table 3.1). None of the MDE models alone was able to capture the empirical peak in tree species richness in the southeastern part of the US. When MDE was combined with environmental variables, particularly precipitation seasonality and slope, the percent of variation in species richness explained was often greater than 90%, and did capture this natural peak. Species richness is generally higher in regions where precipitation seasonality and slope are low. The fact that these two environmental parameters are consistently significant in predicting species richness, in both one-dimensional latitudinal, longitudinal and altitudinal models as well as two- and three-dimensional models, points to their considerable importance in structuring tree species richness. However, alone these two variables were also not able to capture the peak in species richness in the southeastern corner of the continent (see Appendix 3B).

Inclusion of longitude and altitude in analyses of patterns in species richness clearly better reflects real patterns, particularly over large geographic areas. In a large and heterogeneous domain such as North America, analysis of patterns of species richness and environmental gradients as mean values in one-dimensional latitudinal, longitudinal or altitudinal bands obscures much of the detail in spatial structure of these variables. For example, latitudinal patterns of species richness are strongly influenced by the peak between 30 and 40°N; but this peak occurs primarily in a 10° longitudinal band (between 70-80°W) and species richness declines significantly to the west between 80-90°W.

Jetz and Rahbek (2001) found that the predictive power of geometric constraints was weaker in their two-dimensional model of species richness in African birds compared with the one-dimensional model, and suggested that perhaps ecological and physiographical factors manifest themselves more in two dimensions than simply in latitude or longitude. In this chapter, the results in this respect varied. Inclusion of elevation in the two-dimensional models generally improved the match in shape of the predicted species richness curve over one-dimensional models that included latitude or longitude alone (Table 3.2). However, of all the models, elevation by itself (in the one-dimensional model) proved closest to matching both shape and magnitude of the empirical species richness curve (Table 3.2). The complexity in the patterns of species richness predicted with the three-dimensional model also demonstrates the major effect of including elevation in the model. Rather than the single peak in species richness in the centre of the domain, predicted by the two-dimensional lat-long model, inclusion of elevation results in multiple richness peaks and non-uniform gradients in species richness due to the non-uniform distribution of elevation across the domain. Future development and analysis of MDE models will benefit from the inclusion of multidimensional effects that take into consideration the proportion of each dimension in the domain of interest.

There is no doubt that geometric constraints result in non-uniform patterns of species richness in the absence of environmental gradients. Large-range tree species richness patterns generated by geometric constraints alone often have as much, if not more, predictive power than those generated by environmental gradients alone. However, as Colwell et al. (2005; and see Colwell and Lees 2000) state, the question is not *whether* geometry affects patterns of species richness, but what the magnitude of the contribution is. The results presented here show that the magnitude of the contribution can be quite substantial. However, they further demonstrate the importance of considering a number of statistical tests when evaluating the predictions of MDE models. Although MDE models often explained much of the variation in species richness across the domain, they generally did not predict either the magnitude or shape of the empirical species richness curve well. In 21 of the 28 models curve shapes were displaced on both the ordinate and the abscissa (Table 3.2). Currently, use of correlation values are the only way of assessing the contribution of MDE to species richness patterns on an equal footing with environmental variables (since environmental variables do not yield predictions in units of species, and so cannot be tested using the slope and intercept approach) (Zapata et al. 2005). Thus, though the limitations should be noted, the approach is valid.

Many of the methodological criticisms directed at studies testing MDE models recently have been avoided in the chapter; namely the lack of realistic geometry in range shapes (Bokma et al. 2001; Zapata et al. 2003) and of the domain itself (Bokma and Mönkkönen 2000), use of unrealistic RSFDs (Hawkins et al. 2005), and uncontrolled spatial autocorrelation in statistical analyses (Zapata et al. 2003; Colwell et al. 2004; Zapata et al. 2005). Some authors (Zapata et al. 2003; Connolly 2005) have noted that MDE randomization models that use a RSFD that mimics that of the empirical RSFD may inadvertently 'smuggle in' effects of environmental gradients and overestimate the contribution of MDE. In this study the direct effects of geometric constraints in addition to, and concurrently with, the indirect effects of environmental gradients expressed through the use of the empirically realistic RSFD have been analyzed. However, as noted recently by Colwell et al. (2005) and Connolly (2005), in the future a process-based framework, which mechanistically integrates geometric constraints with environmental gradients as well as evolutionary processes, may usefully complement traditional regression and randomization approaches for modeling species distributions in a bounded domain.

References

Anselin, L., 1995. SpaceStat, A software programme for the analysis of spatial data. Version 1.80. Regional Research Institute, West Virginia University, Morgantown, WV.

Arita, H.T., 2005. Range size in mid-domain models of species diversity. Journal of Theoretical Biology 232: 119-126.

Bachman, S., Baker, W.J., Brummitt, N., Dransfield, J., & Moat, J., 2004. Elevational gradients, area and tropical island diversity: an example from the palms of New Guinea. Ecography 27: 299-310.

Bokma, F., & Mönkkönen, M., 2000. The mid-domain effect and the longitudinal dimension of continents. Trends In Ecology and Evolution 15: 288-289.

Bokma, F., Bokma, J., & Mönkkönen, M., 2001. Random processes and geographic species richness patterns: Why so few species in the north? Ecography 24: 43-49.

Colwell, R.K. & Hurtt, G.C., 1994. Nonbiological gradients in species richness and a spurious Rapoport effect. American Naturalist 144: 570-595.

Colwell, R.K. & Lees, D.C., 2000. The mid-domain effect: geometric constraints on the geography of species richness. Trends in Ecology and Evolution 15: 70-76.

Colwell, R.K., Rahbek, C., & Gotelli, N.J., 2004. The mid-domain effect and species richness patterns: what have we learned so far? American Naturalist 163: E1-E23.

Colwell, R.K., Rahbek, C., & Gotelli, N.J., 2005. The mid-domain effect: there's a baby in the bathwater. American Naturalist 166: E150-E154.

Connolly, S.R., 2005. Process-based models of species distribution and the mid-domain effect. American Naturalist 166: E1-E11.

Currie, D.J. & Paquin, V., 1987. Large-scale biogeographical patterns of species richness of trees. Nature 329: 326-327.

Diniz-Filho, J.A.F., de Sant'Ana, C.E.R., de Souza, M.C. & Rangel, T.F.L.V.B., 2002. Null models and spatial patterns of species richness in South American birds of prey. Ecology Letters 5: 47-55.

Givnish, T.J., 1999. On the causes of gradients in tropical tree diversity. Journal of Ecology 87: 193-210.

Grytnes, J.A. & Vetaas, O.R., 2002. Species richness and altitude: a comparison between null models and interpolated plant species richness along the Himalayan altitudinal gradient, Nepal. American Naturalist 159:294-304.

Hansen, M., DeFries, R., Townsend, J.R.G. & Sohlberg, R., 1998. 1 Km Land Cover Classification Derived from AVHRR. College Park, Maryland: The Global Land Cover Facility.

Hawkins, B.A., 2004. “Latitude” and geographic patterns in species richness. Ecography 27: 268-272.

Hawkins, B.A. & Diniz-Filho, J.A.F., 2002. The mid-domain effect cannot explain the diversity gradient of nearctic birds. Global Ecology and Biogeography 11: 419-426.

Hawkins, B.A. & Diniz-Filho, J.A.F., 2004. ‘Latitude’ and geographic patterns in species richness. Ecography 27: 268-272.

Hawkins, B.A., Diniz-Filho, J.A.F. & Weis, A.E., 2005. The mid-domain effect and diversity gradients: is there anything to learn? American Naturalist 166: E140-E143.

Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guegan, J.F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'brien, E.M., Porter, E.E., & Turner, J.R.G., 2003. Energy, water, and broad-scale geographic patterns of species richness. Ecology 84: 3105-3117.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A., 2004. The WorldClim interpolated global terrestrial climate surfaces. Version 1.3. Available at http://biogeo.berkeley.edu/

Jetz, W. & Rahbek, C., 2001. Geographic constraints explain much of the species richness pattern in African birds. Proceedings of the National Academy of Sciences U.S.A. 98: 5661-5666.

Jetz, W. & Rahbek, C., 2002. Geographic range size and determinants of avian species richness. Science 297: 1548-1551.

Koleff, P. & Gaston, K.J., 2001. Latitudinal gradients in diversity: real patterns and random models. Ecography 24: 341-351.

Laurie, H. & Silander, J.A., 2002. Geometric constraints and spatial pattern of species richness: critique of range-based null models. Diversity and Distributions 8: 351-364.

Lawton, J.H., Macgarvin, M., & Heads, P.A., 1987. Effects of altitude on the abundance and species richness of insect herbivores on bracken. Journal of Animal Ecology 56: 147-160.

Leathwick, J.R., Burns, B.R., & Clarkson, B.D., 1998. Environmental correlates of tree alpha-diversity in New Zealand primary forests. Ecography 21: 235-246.

Lichstein, J.W., Simons, T.R., Shriner, S.A. & Franzreb, K.E., 2002. Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs 72: 445-463.

McCain, C.M., 2003. North American desert rodents: a test of the mid-domain effect in species richness. Journal of Mammology 84: 967-980.

Mora, C. & Robertson, D.R., 2005. Causes of latitudinal gradients in species richness: a test with fishes of the tropical Eastern Pacific. Ecology 86: 1771-1782.

Ohlemüller, R. & Wilson, J.B., 2000. Vascular plant species richness along latitudinal and altitudinal gradients: a contribution from New Zealand temperate rainforests. Ecology Letters 3: 262-266.

Prasad, A.M. & Iverson, L.R., 2003. *Little's Range and FIA Importance Value Database for 135 Eastern US Tree Species*, U.S.D.A. Forest Service, Delaware, URL http://www.fs.fed.us/ne/delaware/4153/global/littlefia/index.html.

Rahbek, C., 1995. The elevational gradient of species richness - a uniform pattern. Ecography 18: 200-205.

Rahbek, C., 2005. The role of spatial-scale and the perception of large-scale species-richness patterns. Ecology Letters 8:224-239.

Rhode, K., 1992. Latitudinal gradients in species diversity: the search for the primary cause. Oikos 65: 514-527.

Romdal, T.S., Colwell, R.K. & Rahbek, C., 2005. The influence of band sum area, domain extent, and range sizes on the latitudinal mid-domain effect. Ecology 86: 235-244.

Sanders, N.J., 2002. Elevational gradients in ant species richness: area, geometry, and Rapoport's rule. Ecography 25: 25-32.

Stevens, G.C., 1989. The latitudinal gradient in geographical range: How so many species coexist in the tropics. American Naturalist 133: 240-256.

U.S.G.S. 2005. HYDRO1K. URL: http://edc.usgs.gov//products/elevation/hydro1k.html

Vetaas, O.R. & Grytnes, J.A., 2002. Distribution of vascular plant species richness and endemic richness along the Himalayan elevation gradient in Nepal. Global Ecology and Biogeography 11: 291-301.

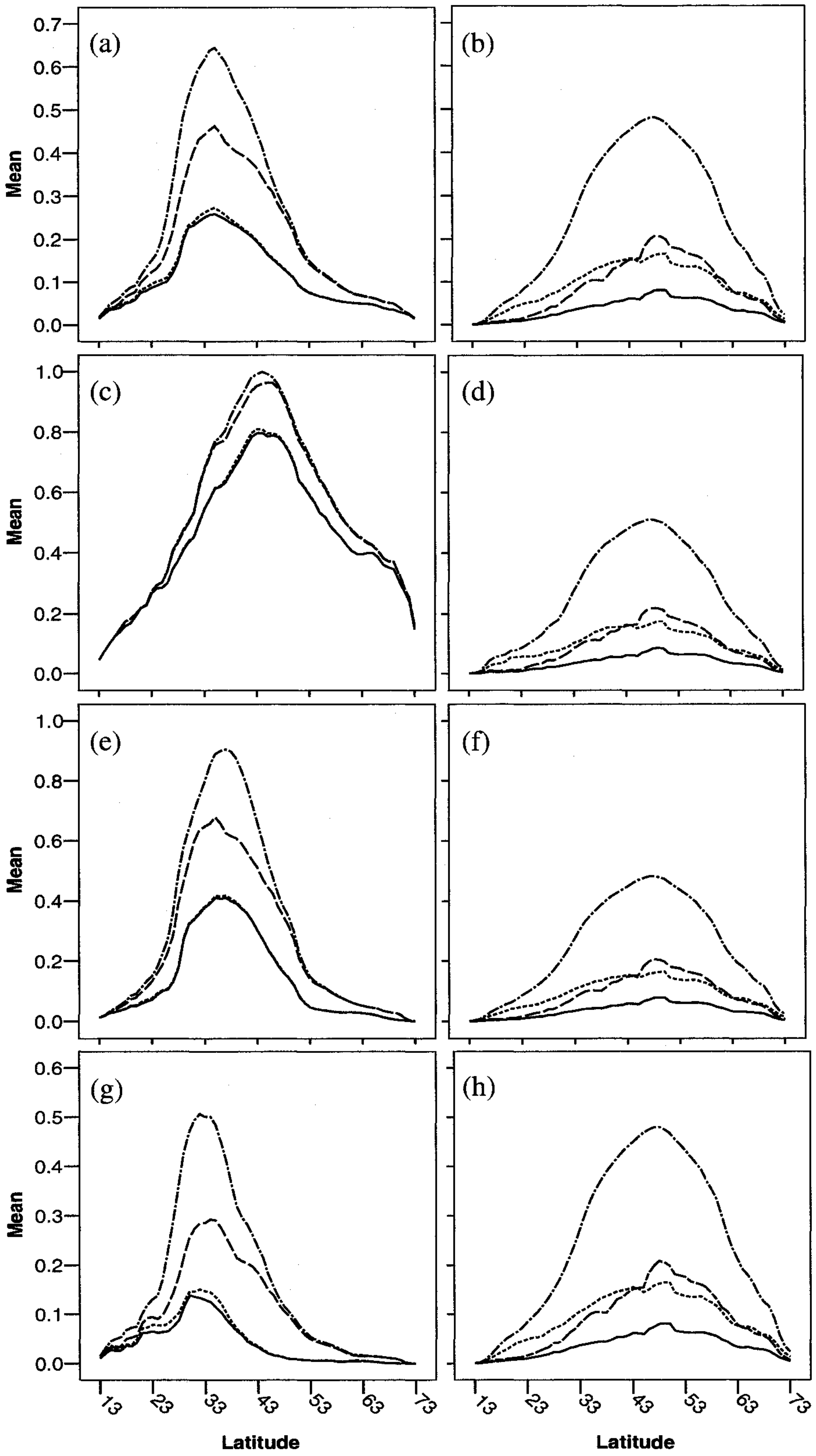
Willig, M.R. & Lyons, S.K., 1998. An analytical model of latitudinal gradients of species: richness with an empirical test for marsupials and bats in the New World. Oikos

83: 93-98.

Willig, M.R., Kaufman, D.M. & Stevens, R.D., 2003. Latitudinal gradients of biodiversity: pattern, process, scale and synthesis. Annual Review of Ecology, Evolution and Systematics 34: 273-309.

Zapata, F.A., Gaston, K.J., & Chown, S.L., 2003. Mid-domain models of species richness gradients: assumptions, methods and evidence. Journal of Animal Ecology 72: 677-690.

Zapata, F.A., Gaston, K.J. & Chown, S.L., 2005. The mid-domain effect revisited. American Naturalist 166: E144-E148.



(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Mean
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
1.0
0.8
13
23
33
43
53
63
73
Latitude

Figure 3.1. Species richness (mean proportion of species) at 1°latitudinal bands across the domain for (a) all empirical species, (b) all species MDE, (c) large-range empirical, (d) large-range MDE, (e) medium-range empirical, (f) medium-range MDE, (g) small-range empirical and (h) small-range MDE. Lines represent MDE models for three-dimensional (———), two-dimensional lat-long (···········), two-dimensional lat-alt (– – –), and one-dimensional latitude (– · – · – ·).

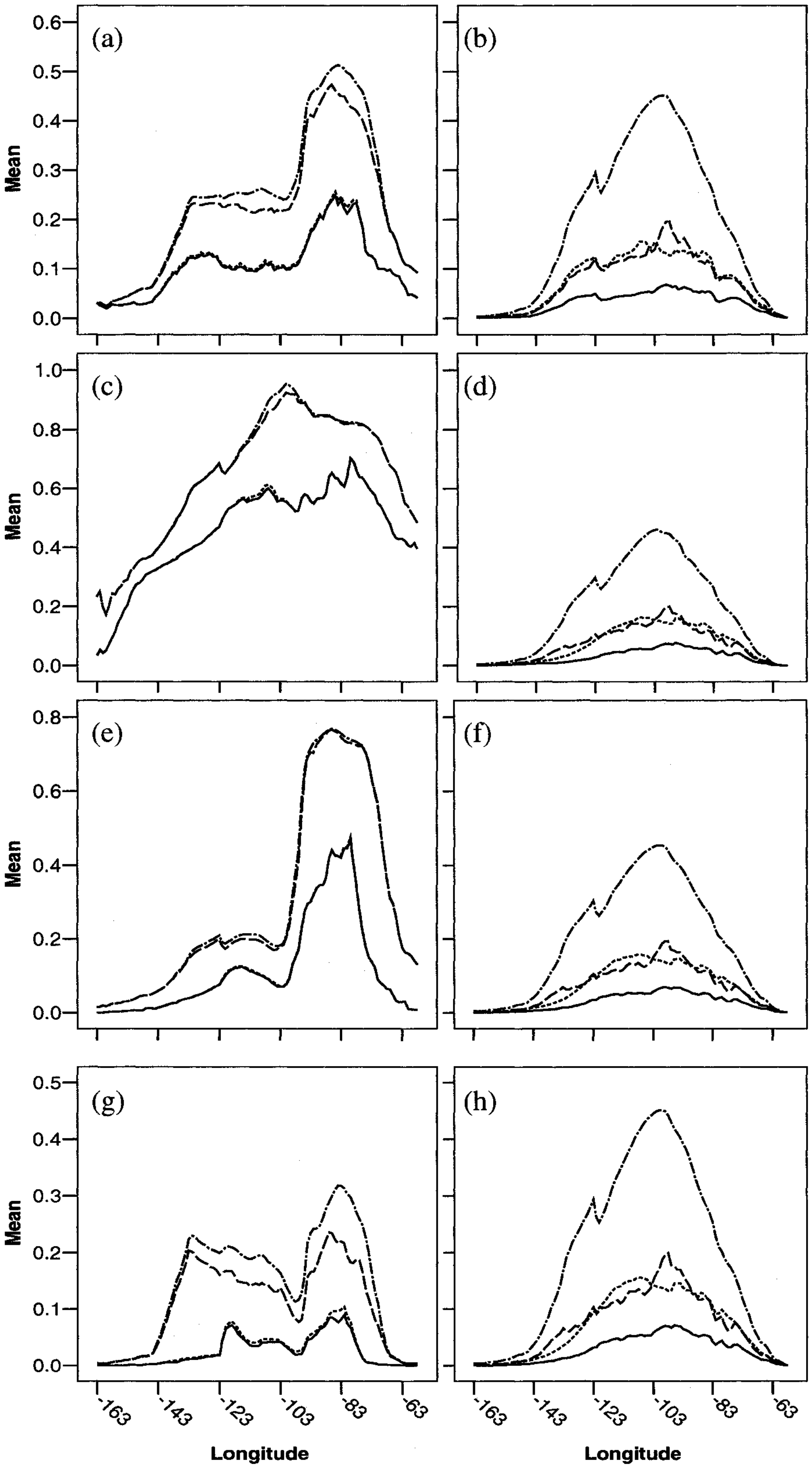

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Mean
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.8
1.0
-163
-143
-123
-103
-83
-63
Longitude

Figure 3.2. Species richness (mean proportion of species) at 1°longitudinal bands across the domain for (a) all empirical species, (b) all species MDE, (c) large-range empirical, (d) large-range MDE, (e) medium-range empirical, (f) medium-range MDE, (g) small-range empirical and (h) small-range MDE. Lines represent MDE models for three-dimensional (———), two-dimensional lat-long (··········), two-dimensional lat-alt (– – –), and one-dimensional latitude (– · – · – ·).

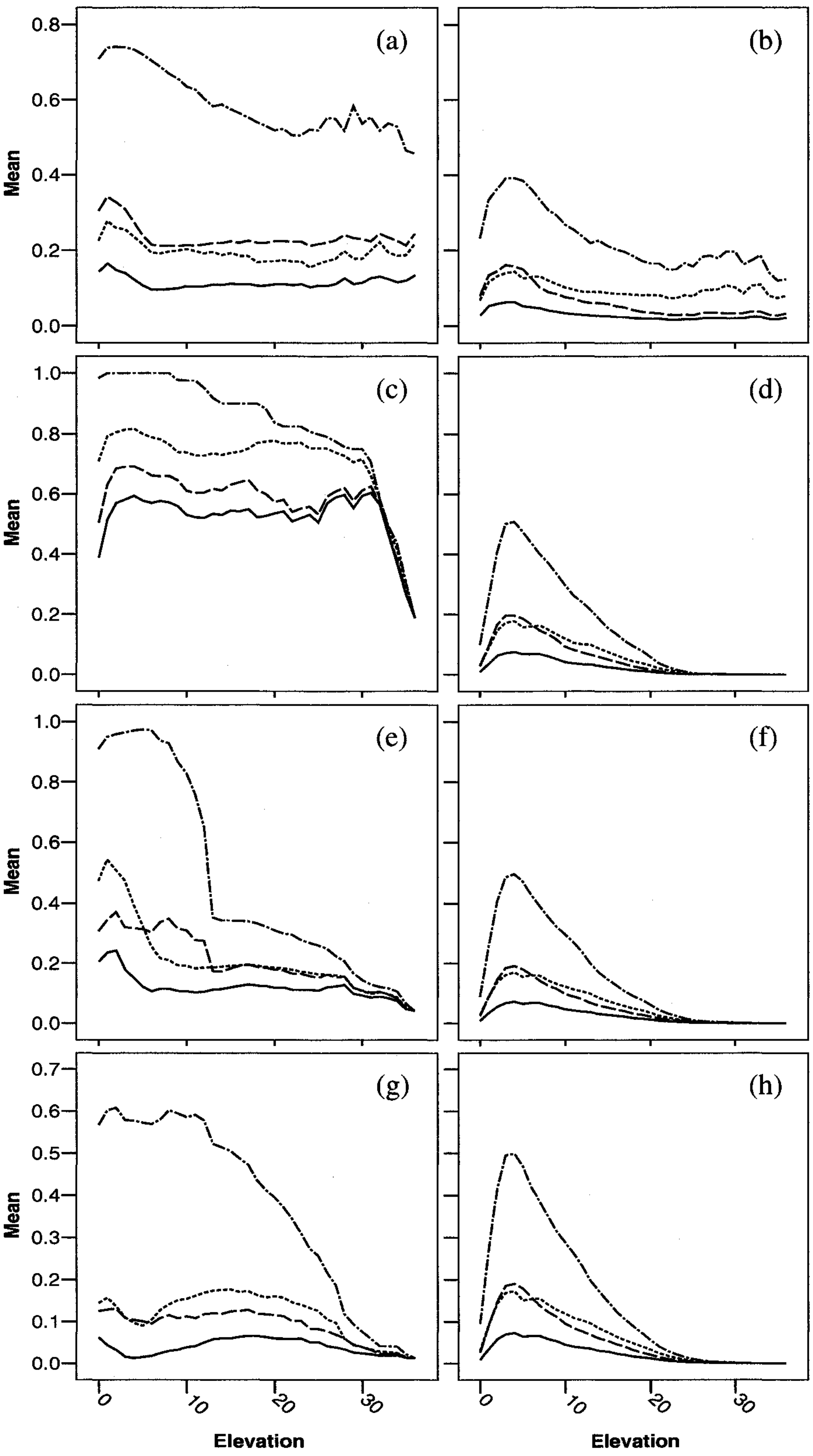

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Mean
Mean
Mean
Mean
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
10
20
30
0
10
20
30
Elevation
Elevation

Figure 3.3. Species richness (mean proportion of species) at 100 metre altitudinal bands across the domain for (a) all empirical species, (b) all species MDE, (c) large-range empirical, (d) large-range MDE, (e) medium-range empirical, (f) medium-range MDE, (g) small-range empirical and (h) small-range MDE. Lines represent MDE models for three-dimensional (———), two-dimensional lat-long (··········), two-dimensional lat-alt (– – –), and one-dimensional latitude (– · – · – ·).

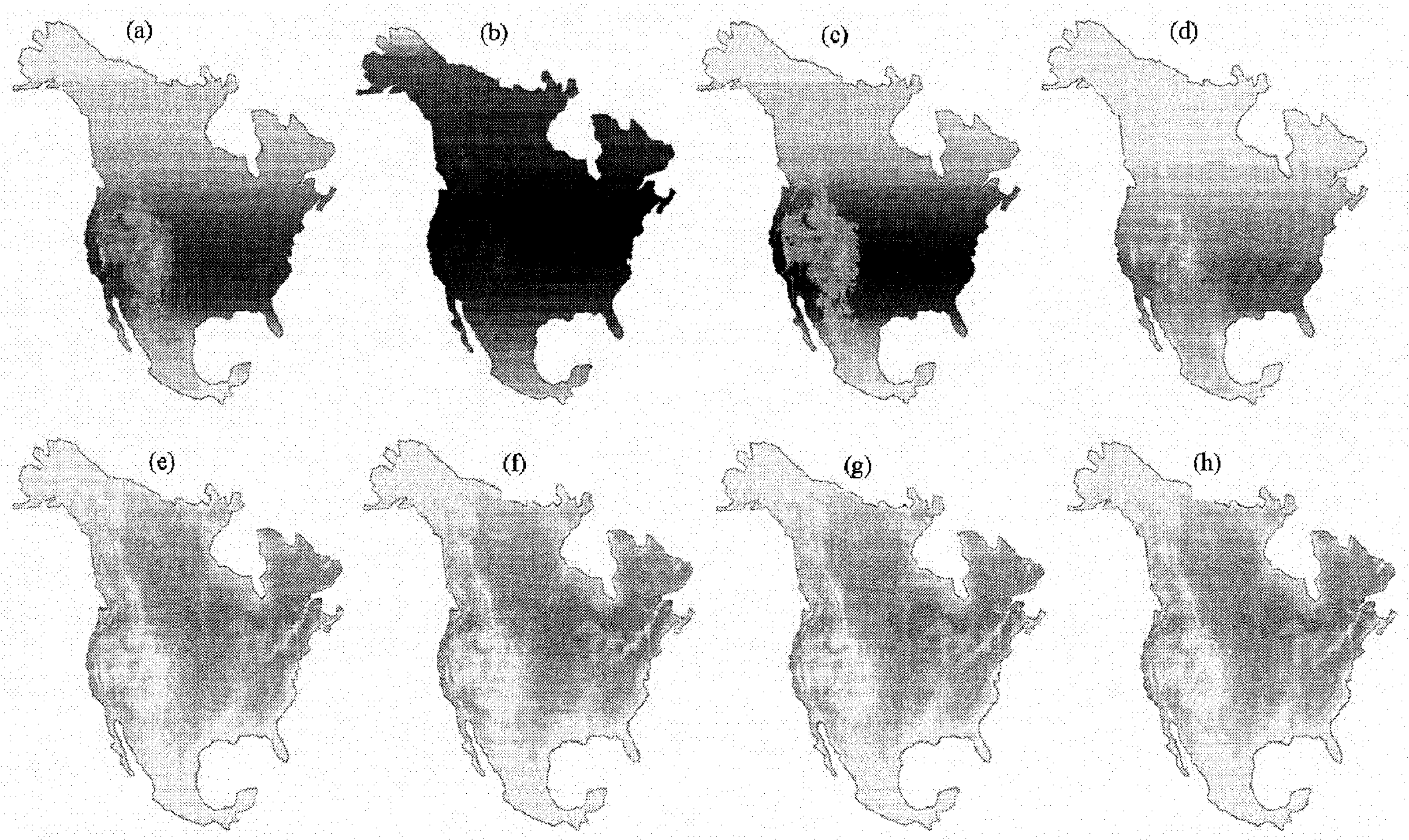


Figure 3.4. Two-dimensional (latitude-altitude) patterns of species richness (proportion of total species) (a) all species empirical, (b) large-ranges empirical, (c) medium-ranges empirical, (d) small-ranges empirical, (e) all species MDE, (f) large-ranges MDE, (g) medium-ranges MDE, and (h) small-ranges MDE. Equal interval classification is shown, with darker shading indicating areas of higher species richness.

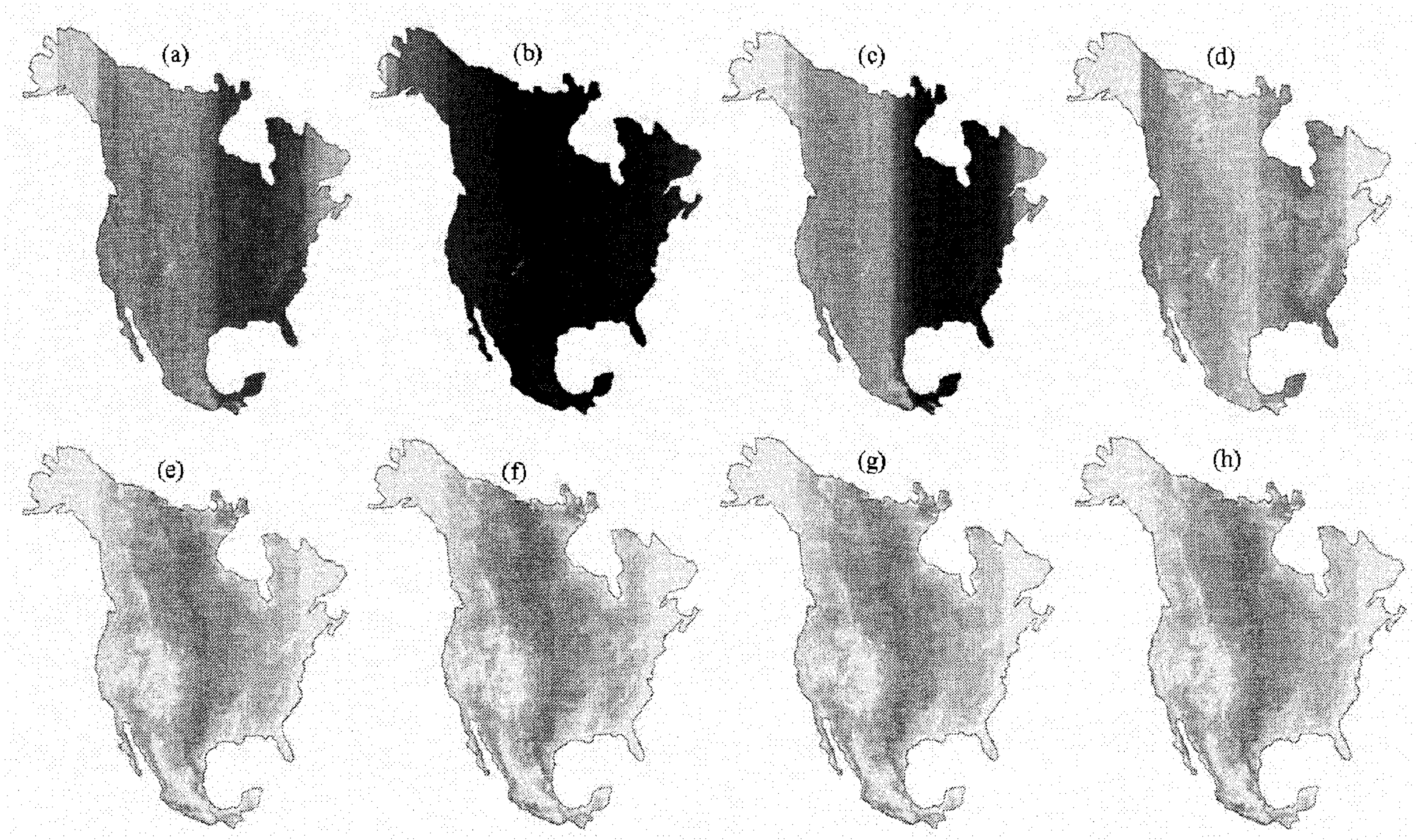


Figure 3.5. Two-dimensional (longitude-altitude) patterns of species richness (proportion of total species) (a) all species empirical, (b) large-ranges empirical, (c) medium-ranges empirical, (d) small-ranges empirical, (e) all species MDE, (f) large-ranges MDE, (g) medium-ranges MDE, and (h) small-ranges MDE. Equal interval classification is shown, with darker shading indicating areas of higher species richness.

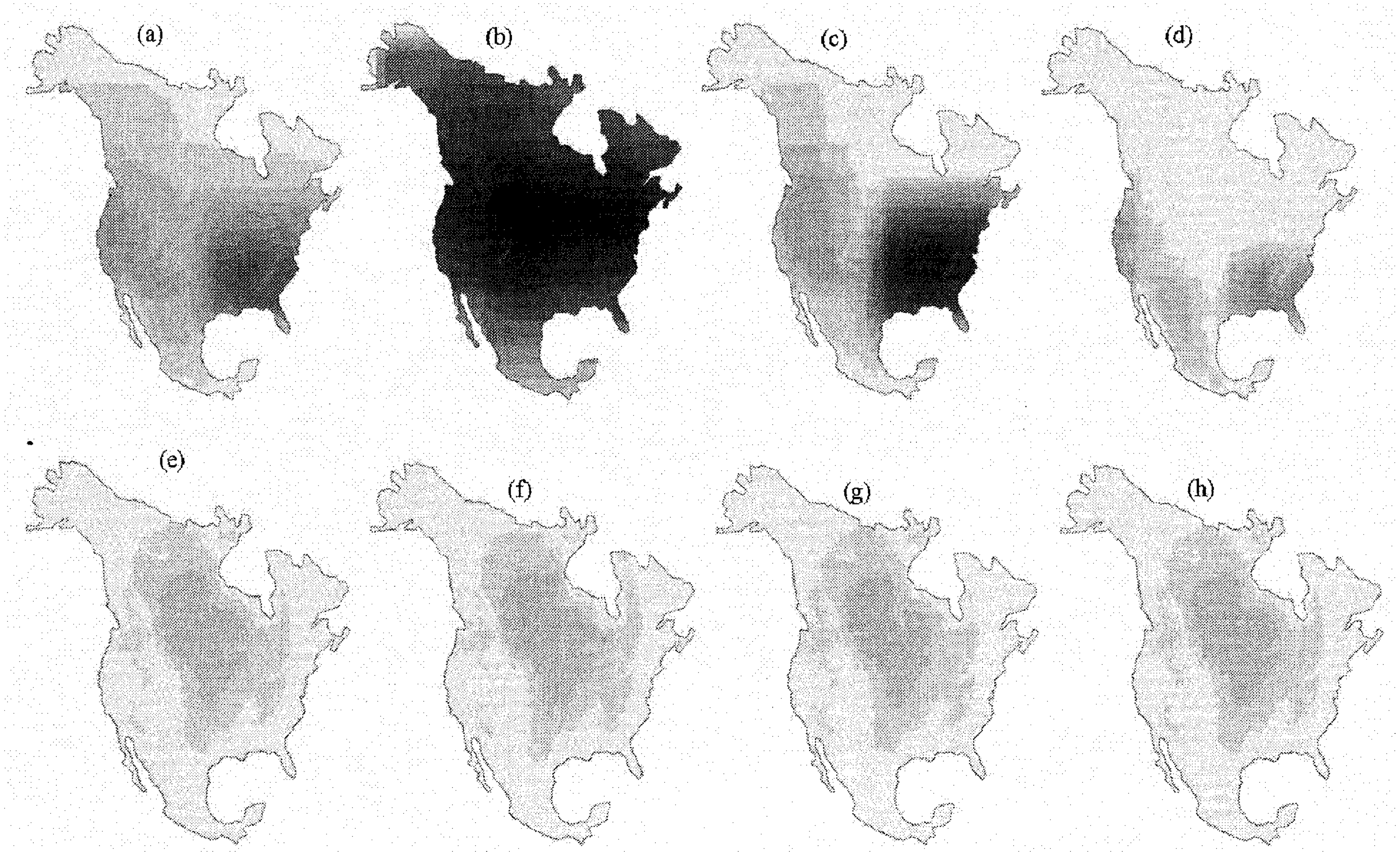


Figure 3.6. Three-dimensional patterns of species richness (proportion of total species) (a) all species empirical, (b) large-ranges empirical, (c) medium-ranges empirical, (d) small-ranges empirical, (e) all species MDE, (f) large-ranges MDE, (g) medium-ranges MDE, and (h) small-ranges MDE. Equal interval classification is shown, with darker shading indicating areas of higher species richness.

Table 3.1. Results of spatial linear regression analyses for predictors of empirical species richness for all species, and species having large-, medium- and small-ranges. Shown are the z-values (bold represents the strongest single predictor) of the single spatial linear regressions and significance (* p<0.05, ** p <0.01, *** p<0.001 after sequential Bonferroni adjustment). Also shown are the pseudo R^2 values for multiple spatial linear regression (MSL) with MDE and without MDE as a predictor variable.

Y = latitude, X = longitude, Z = altitude, 1-D = one-dimensional, 2-D = two-dimensional, 3-D = three-dimensional

	1-D models			**2-D models**			**3-D model**
All ranges	**Y**	**X**	**Z**	**XY**	**YZ**	**XZ**	**XYZ**
MDE	4.96***	5.07***	**14.57*****	-0.80	6.72***	**9.38*****	0.79
Mean temp.	1.01	**8.25*****	-11.78***	11.30***	1.22	2.20*	8.71***
Mean temp. (wet)	1.51	8.20***	-5.16***	6.79***	5.62***	-5.16***	6.76***
Precip. seasonality	-4.41***	-3.59***	-14.36***	**-18.36*****	**-12.89*****	-2.71**	**-23.07*****
Landcover diversity	-0.30	-0.28	-2.50*	8.33***	3.27***	3.26**	4.88***
Flow accumulation	3.35**	2.50*	7.49***	2.50*	5.52***	-2.84**	4.68***
Slope	**-5.17*****	-7.21***	-8.29***	-6.48***	-11.57***	3.13**	-10.86***
Aspect	-1.91	-2.12	-4.54***	4.39***	-0.24	9.20***	4.46***
MSL R^2	0.77	0.85	0.94	0.93	0.84	0.85	0.93
MSL w/o MDE R^2	0.72	0.84	0.93	0.91	0.84	0.62	0.93
Large ranges							
MDE	**13.56*****	**6.81*****	**5.17*****	**23.72*****	**11.27*****	**10.01*****	**21.60*****
Mean temp.	-0.61	4.52***	0.41	-0.02	1.02	2.26	-0.01
Mean temp. (wet)	-1.49	3.56***	1.13	0.71	0.52	2.86*	1.95
Precip. seasonality	-6.76***	0.05	1.87	-0.80	-5.82***	1.46	-0.86
Landcover diversity	0.91	-0.83	-5.08***	-3.24**	-0.06	-1.80	-4.90***
Flow accumulation	1.47	-0.09	3.99***	1.36	2.20*	0.18	1.02
Slope	-2.62*	-2.53	-3.44**	-4.04***	-3.19**	-2.35	-5.86***
Aspect	-1.04	-1.55	-1.87	-2.21	-1.52	-2.00	0.01
MSL R^2	0.97	0.92	0.93	0.99	0.95	0.93	0.98
MSL w/o MDE R^2	0.91	0.84	0.93	0.78	0.89	0.75	0.87
Medium ranges							
MDE	1.94	2.14	6.18***	-1.81	6.70***	2.07	-0.28
Mean temp.	3.75***	3.10**	-0.91	7.62***	7.17***	1.78	6.38***
Mean temp. (wet)	2.06	4.25***	3.93***	0.52	**9.32*****	4.02***	1.47
Precip. seasonality	**-4.69*****	-3.58***	0.47	**-13.33*****	-7.04***	-5.62***	**-12.89*****
Landcover	2.87*	-0.59	-3.10**	10.90***	1.74	-1.93	7.50***

diversity							
Flow accumulation	3.00 *	1.43	4.81 ***	6.52 ***	3.86 ***	0.82	5.75 ***
Slope	-4.31 ***	**-7.01** ***	**-6.26** ***	-3.13 **	-9.08 ***	**-6.99** ***	-4.94 ***
Aspect	0.69	-1.79	-2.15	6.84 ***	-0.66	-2.02	3.08 **
MSL R^2	0.93	0.86	0.94	0.93	0.88	0.81	0.94
MSL w/o MDE R^2	0.92	0.86	0.92	0.93	0.87	0.79	0.94
Small ranges							
MDE	-0.34	1.21	11.49 ***	**-8.91** ***	-1.70	-3.28 **	**-14.01** ***
Mean temp.	**5.96** ***	**6.03** ***	0.24	5.25 ***	10.32 ***	7.03 ***	10.09 ***
Mean temp. (wet)	5.02 ***	2.59 *	6.23 ***	-0.53	4.96 ***	2.14	7.19 ***
Precip. seasonality	-2.71 *	-5.12 ***	-0.33	-7.73 ***	**-17.94** ***	**-8.99** ***	-8.66 ***
Landcover diversity	3.44 **	-1.03	-10.07 ***	7.37 ***	13.86 ***	2.77 *	5.47 ***
Flow accumulation	4.70 ***	0.55	8.46 ***	0.25	3.73 ***	-0.65	0.10
Slope	-5.23 ***	-4.56 ***	**-15.73** ***	-8.21 ***	-10.86 ***	-4.05 ***	-11.39 ***
Aspect	0.49	-1.73	-2.34	3.01 **	7.77 ***	0.67	1.73
MSL R^2	0.97	0.81	0.96	0.97	0.97	0.85	0.96
MSL w/o MDE R^2	0.92	0.81	0.93	0.95	0.96	0.85	0.92

Table 3.2. Results for t-test of the hypotheses (a) slope = 1 and (b) intercept = 0, for spatial linear regression models of MDE predicted species richness against empirical species richness. t-values in bold are not significant at the $p<0.05$ level. Also shown are the pseudo R^2 values for the single spatial regression of MDE and empirical species richness.

Y = latitude, X = longitude, Z = altitude, 1-D = one-dimensional, 2-D = two-dimensional, 3-D = three-dimensional

	1-D models			**2-D models**			**3-D model**
	y	**x**	**z**	**xy**	**yz**	**xz**	**xyz**
All ranges							
slope	-13.94	-9.75	**0.67**	-7.94	**-0.04**	2.17	-2.11
intercept	2.59	6.47	13.45	4.31	**0.88**	5.38	3.08
R^2	62.70	58.70	80.60	63.40	59.20	41.60	56.00
Large ranges							
slope	2.90	**-0.22**	**1.28**	16.38	7.70	5.53	17.97
intercept	4.95	5.24	6.31	6.72	5.83	7.75	7.09
R^2	94.50	85.60	69.20	97.80	91.00	90.20	96.10
Medium ranges							
slope	-6.92	-7.83	**1.34**	-3.33	3.58	-2.52	**-1.31**
intercept	2.29	2.76	3.45	3.43	**1.64**	2.81	2.50
R^2	60.00	58.90	79.70	44.10	56.90	53.80	57.90
Small ranges							
slope	-13.72	-12.14	-3.96	-13.20	-6.09	-9.05	-16.62
intercept	3.82	3.94	5.96	9.92	2.83	5.34	14.35
R^2	55.50	32.60	77.90	78.90	50.20	22.00	82.20

Chapter 4 - Geometric constraints and environmental correlates of amphibian, avian and mammal richness across the Americas.

Summary

The mid-domain effect (MDE) proposes that geographic gradients in species richness can be accounted for by stochastic processes and geometric constraints of a domain. Predicting a mid-domain peak in species richness that arises simply from the random placement of ranges within a domain, MDE has been used as a null model to accurately predict species richness gradients in one-, two- and, more recently, three-dimensional domains. In this chapter, one-, two- and three-dimensional predictions of a modified, fully-stochastic, three-dimensional MDE model are compared to the natural species richness patterns for amphibians, birds and mammals across continental North, Central and South America. The model was used initially as an operational approach to generate null expectations of stochastic processes under geometric constraints; then, the predictions were invoked as an additional explanatory variable to account for patterns of species richness, separately with climatic variables. Species richness for each group of species was measured for each 20 x 20 km cell within the domain and compared using spatial autoregressive models, with mean species richness of 100 simulations in which 2000 theoretical species ranges were randomly placed within the domain.

The MDE alone accounted for a great proportion of the variability in empirical species richness, explaining some 73%, 77% and 84% of the variation in amphibian, mammal and avian species richness gradients, respectively. However, it generally underestimated

species richness for avian and mammal species, while over-estimating richness for amphibians.

As an additional variable together with climatic variables, MDE added little to the explanatory power of the analysis. On average, MDE explained an additional 1.1%, 0.5% and 6.9% of the variation in amphibian, avian and mammal species richness, respectively. Although environmental correlates provide equal or better explanatory value than that from MDE predictions, the principle of parsimony suggests that random placement of ranges within the limits of a domain should be used. This correlative approach suggests, with respect to the predictive power of MDE, species are randomly placed with the domain limits, or the richness pattern appears random but is a consequence of some larger set of interacting factors. Most likely, richness patterns are the product of variation in relative strength of stochastic and ecological processes.

Introduction

One of the newest and more controversial theories explaining broad-scale patterns in biological richness is the mid-domain effect (MDE), proposed by Colwell and Hurtt (1994). It proposes that stochastic processes and geometric constraints of a domain produce species richness gradients that are similar to 'real world' gradients. Mid-domain models are null models that simulate the random spatial arrangement of species ranges, within a bounded domain such as a continent, assuming no direct effects of environmental gradients (Colwell et al. 2004). Beyond offering a null expectation, MDE predictions could be used with other environmental or historical factors as explanatory variables for patterns in species richness (Colwell et al. 2005, Rangel and Diniz-Filho 2005).

MDE models have been the subject of some controversy (see e.g., Colwell et al. 2004, Hawkins et al. 2005, Zapata et al. 2005, Colwell et al. 2005). A major point of contention involves range cohesion. In the absence of environmental gradients, why are not all species distributed throughout the entire domain (Diniz-Filho et al. 2002, Hawkins and Diniz-Filho 2002, Zapata et al. 2003, 2005, Hawkins et al. 2005).

Colwell et al. (2005) proposes that the range cohesion argument is not a shortcoming of MDE models. While it is generally accepted that a species range size is shaped and limited by environmental factors, historical effects, and dispersal limitation, MDE models seek to determine what 'real world' patterns would arise in the absence of any direct effects of environmental gradients on species richness patterns. By randomly placing species ranges within a bounded domain and comparing predicted and actual patterns of species richness, the question being asked is not "what constrains the range of a species" but rather, to what degree do actual species richness patterns arise from stochastic processes? The most complete studies of MDE models examine, in a multivariate context, the influence of both stochasticity under geometric constraints (MDE), and the direct effects of both biotic and abiotic environmental gradients. Such studies explore the covariation of the two sets of factors (Colwell et al. 2004).

Fully stochastic MDE models have been criticized because the range size frequency distributions (RSFD) of the theoretical species may differ from the empirical, ‘real world’ data and thus potentially underestimating or overestimating the MDE (McCain 2003, Colwell et al. 2004). The alternative is to randomly place ranges resampled from an actual RSFD. This may produce species richness peaks similar in magnitude to that of actual richness peaks (Zapata et al. 2003, Colwell et al. 2004), and can incorporate taxon-specific biological characteristics (e.g., speciation and extinction potential, population density, body size) which may be independent of species richness patterns (Lees et al. 1999, Colwell and Lees 2000, Hawkins and Diniz-Filho 2002, Jetz and Rahbek 2001, 2002, McCain 2004, Colwell et al. 2004). As others have noted, these may not be independent of spatial patterns of species richness; moreover the use of a purely theoretical RSFD is less subject to biological assumptions, and thus more representative of a null conceptual model (see e.g., Kollef and Gaston 2001, Hawkins and Diniz-Filho 2002, Laurie and Silander 2002).

In this chapter, I examine the applicability of using a modified, fully stochastic, three dimensional MDE model, comparing model predictions with empirical data for each of three groups of biota (amphibians, birds, mammals) across the continental landmass of North, Central and South America. I first use the predicted patterns, operationally, to generate null expectations of stochastic patterns under geometric constraints. I then use the predictions as an additional explanatory variable to explain patterns of species richness together with climatic variables. Finally, I discuss the relative contributions of stochastic processes and environmental correlates influencing the observed patterns of species richness in these animal groups.

Methods

Species ranges

Digital distribution maps of the terrestrial ranges of birds and mammals for the Western Hemisphere (continental North, Central, and South America and associated islands) have been developed by a group of conservation organizations (NatureServe, the Center for Applied Biodiversity Science within Conservation International, the Migratory Bird Program within The Nature Conservancy, the US branch of the World Wildlife Fund and the WILDSPACE program within Environment Canada) to aid conservation planners and other interested users (Patterson et al. 2003, Ridgely et al. 2003). IUCN et al. (2004) further produced a series of digital distribution maps of the amphibians of the Western Hemisphere as part of the Global Amphibian Assessment project. All of these maps have been made available through NatureServe (www.natureserve.org). These digital distribution maps include 3019 amphibian species, 4247 avian species and 1786 mammal species. Of these, only 2216 amphibian, 3771 avian and 1605 mammal distributions were used, as I excluded non-indigenous species (MDE assumes species are indigenous), or the species ranges extended beyond the boundary of the domain, or were smaller than one 20 x 20 km cell.

Mid-domain models

Continental North America, Central, and South America was the domain used in this study (Fig. 4.1). The perimeter of the domain was extracted from "World Countries 1992" base map, supplied with ArcView 3.2 (ESRI, California, USA) and altitudes were obtained from the GLOBE project (GLOBE 1999). The domain was gridded with 20 x 20 km cells in longitude and latitude, while mean altitude was estimated in metres above sea level for each cell. Cells not part of the continental land mass (i.e., islands) or which formed small peninsulas (size = 1 cell) were excluded from the analysis, giving a total of 94,078 cells within the domain.

A modified, fully stochastic, three dimensional MDE model as described in Chapter 2 was implemented to explore the relationships of the main effects (one-dimensional) and interaction effects (two- and three dimensional) of longitude, latitude and altitude on species richness across the domain. For any single dimension, it is assumed a species cannot occur outside the "hard" limits (0 and 1) of the domain and thus, a species range can be defined by its limits L_1 and L_2 where $0 \leq L_1$, $L_2 \geq 1$ and $L_2 \geq L_1$ (Willig and Lyons 1998; Arita 2005). To generate a species range, two random points within the range are randomly placed within the domain (i.e., for longitude X_1 and X_2) where $L_1 = \mathrm{MIN}(X_1, X_2)$ and $L_2 = \mathrm{MAX}(X_1, X_2)$. The range size R_X is a function of X_1 and X_2 and can be represented as $R_X = |X_1 - X_2|$, such that $0 \leq R_X \leq 1$. Extending this beyond the single dimension, a range can be defined by its limits in longitude (X_1, X_2), latitude (Y_1, Y_2) and altitude (Z_1, Z_2).

Thus, a theoretical species range, R, is the area delimited in longitude and latitude that can be represented by $R_X \cap R_Y \cap R_Z$ where R_X, R_Y and R_Z are defined by their limits in longitude (X), latitude (Y) and altitude (Z), respectively. Because of the irregularly shaped surface of the domain used here, the random selection of endpoints may create a range where $R = 0$ (e.g., when $R_X \cap R_Y$ but does not intersect R_Z). When this occurred the theoretical species was discarded and a new one was created, with six new endpoints. The theoretical ranges created were often irregularly shaped and could be discontinuous, similar to actual ranges. Additionally, the extents of the theoretical ranges in longitude, latitude and altitude were often smaller than those defined by randomly selected endpoints. For example, X_{max} for range R, where R is a function of $R_X \cap R_Y \cap R_Z$, was less than the randomly selected L_{X2} where $L_{X2} = MAX(X_1, X_2)$, and thus $R_X < |X_1 - X_2|$. When this occurred, the maximum and minimum extents of the range were defined by the limits of R rather than the randomly selected endpoints.

I have shown, in Chapter 2, how domain shape can influence MDE model predictions, and proposed modifying the fully stochastic model to account for the proportion of the domain in each unit of a particular dimension. For this model, range limits for each of longitude, latitude and altitude were selected at random from the non-uniform distributions of possible endpoints based on the proportion of the domain represented by a single point (or for each measured unit) in each dimension (see Chapter 2). For this domain, there is a gradual increase in proportion of the domain represented in longitude, peaking at approximately 0.75 within domain limits of 0 to 1 (western limit to eastern limit; Fig. 2). Thus the probability of selecting a range endpoint representing 0.75 of the domain is ≈ 0.025, while the probability of selecting a point representing 0 or 1 is nearly 0. There is a bimodal peak in latitude (from north to south) with the valley representing the smaller proportion of the domain that is Central America. Altitude shows a right-skewed distribution (from low to high), in that much more of the domain occurs at lower altitudes than at high.

The null model consisted of the mean species richness of 100 simulations in which 2000 theoretical species ranges were randomly placed within the domain. In effect, 200,000 theoretical species were placed within the domain whose ranges were randomly created using the modified, fully-stochastic, three dimensional MDE model. Actual species ranges were defined by their limits in each dimension, as measured from their range maps to permit direct comparison of the theoretical and empirical species. The theoretical and empirical species ranges were mapped and species richness in each cell was calculated across the domain.

Statistical analysis

All data extraction, modeling and mapping was done in Visual Basic (VB.NET, Microsoft Corporation, Redmond, WA) and ArcGIS 9.0 (ESRI, Redlands, CA). Spatial statistics were done using SpaceStat (Anselin 1995). I used spatial autoregressive (SAR) models rather than ordinary least squared regression to account for spatial autocorrelation. SAR models were used to determine the effect of geometric constraints predicted by the MDE models and environmental variables on empirical richness patterns. As few software packages can store and handle analysis for such large amounts of data (e.g., 94078^2 as would be needed for the weight matrix), the data was summarized such that, for each SAR model, the dependent variable was the rounded empirical value (x 1000) of species richness, while the independent variables represent the mean value of that variable for each value of the dependent values. This avoids the alternative methodology of collapsing the results of two- and three-dimension MDE models into one dimension for analysis (e.g., using latitudinal bands).

WorldClim Version 1.3 (Hijmans et al. 2005) consists of 19 interpolated global climate layers on a square kilometer grid (listed in Appendix 4A). Upon examining the correlations of the environmental layers with each other, for each dependent variable, it was noted that many of these layers were highly correlated (using spatial correlation). Thus I excluded non-independent variables (correlated at >75%) from the analysis. All SAR models used a unique subset (5-6 variables) of the 19 climatic variables; the subsets are listed in Appendix 4B. The most commonly used climatic variables were related to precipitation seasonality (coefficient of variation) and isothermality (mean diurnal temperature range / mean annual temperature range).

A spatial weights matrix was generated for the SAR model as they included a spatially lagged dependent variable in the regression equation. The spatial weights matrix was generated such that adjacent values of empirical species richness were given a value of 1, and a value of 0 was applied to values that are not adjacent. Each value of 1 represents a potential spatial interaction between the two observations.

SAR models were first applied to MDE predictions alone, then to the environmental variables alone and lastly, a multiple SAR was run including both the environmental variables and MDE independent variables. The traditional R^2 measure of fit is not applicable to the SAR model. Instead, a pseudo R^2 measure (ratio of the variance of the predicted values over the variance of the observed values for the dependent variable) is reported here (Anselin 1995).

Assessing the fit of MDE models using correlation coefficients is problematic since absolute differences in magnitude are obscured (Zapata et al. 2003, Colwell et al. 2004, Romdal et al. 2005). t-tests were used to examine deviation of the slope from unity and intercept from zero for the regressions between predicted MDE species richness and empirical species richness (as suggested by Colwell et al. 2004, Zapata et al. 2005). These tests provide an indication of deviation in both shape and magnitude between predicted and empirical species richness.

Results

Range size frequency distribution

There was approximately a two-fold increase in mean range size for each group of species examined when ranges were defined by their extents in longitude, latitude and altitude rather than by polygons defined by NatureServe and IUCN et al. (2004). For amphibians, mean range size increased from 1.1 (± 0.07 SE) to 2.1 (± 0.12 SE) percent of the domain area; avian mean range size increased from 6.7 (± 0.18 SE) to 14.3 (± 0.23 SE) percent; and mammal mean range size increased from 5.0 (± 0.23 SE) to 8.6 (± 0.34

SE) percent. The theoretical species had a mean range size of 4.1 (± 0.02 SE) percent of the domain area.

The RSFD for each of the three empirical biotic, and the theoretical species are presented in Figure 3. RSFD of the theoretical species most closely represents the middle of a gradient between the amphibian and mammal RSFDs; it is most dissimilar from that of the birds. Amphibians have the greatest proportion of small ranges, with 83% of the ranges < 2.5% of the domain area. Mammals and the theoretical species had 53% and 59%, respectively, of the ranges < 2.5% of the domain area, while the avian species had only 34%.

Species richness patterns

Richness peaks of the theoretical species were from 0.46 to 0.50 x the number of theoretical species for individual dimensions, 0.24 to 0.28 for the two dimension combinations and 0.14 for the three dimension combination. However, the richness peaks were shifted from over the middle of the domain (0.5 in a domain with limits of 0 and 1), toward more eastern longitudes (0.6 to 0.75, with western edge of the domain being 0 and the eastern edge 1), more northern latitudes (0.2 to 0.6, with the northern edge being 0 and the southern edge 1), and towards lower altitudes (≈ 0.06, with 0 m ASL represented by 0 and 6499m ASL by 1; Fig. 4.4).

For longitude, biotic groups showed a shift in species richness toward the more eastern longitudes, similar to that of the null predictions (Fig. 4.4, top panel). However, only the birds had a similar peak height of 0.5 x the number of avian species (Table 4.1). The mammals peaked at 0.34 x the number of mammal species and the amphibians peaked at 0.19 x the number of amphibian species. Likewise, in altitude the three groups of species showed a similar shift in species richness toward lower elevations (Fig. 4.4, bottom panel). The null model predicted a species richness that was lower than that of all three groups; the theoretical species peak was 0.5 x the number of species whereas the peak for amphibian, avian and mammal richness was 0.62, 0.94, 0.86, respectively. Indeed the predicted elevational richness was lower than that of the three groups of species across the entire altitudinal range.

In regard to latitude, all three groups of species showed similar patterns in species richness, that differed from the predicted. While the null model predicted a peak that ranged from 0.2 to 0.6, the three biotic groups had peaks ranging from 0.5 to 0.7 (Fig. 4.4, middle panel). As for longitude-based results, the peak heights of the theoretical and avian species were approximately equal (0.46 and 0.43 x the number of species, respectively), with mammal peak height being lower at 0.26, and amphibians at 0.14.

Figure 4.5 illustrates geographically where the predicted and empirical species richness peaks were for each one-, two- and three-dimensional model. The two-dimensional longitude x latitude pattern shows, as expected, a 'bull's-eye' effect centred over the open water between North and South America. Adding effects of altitude adds complexity to the three-dimensional pattern. The three-dimensional MDE model predicts greatest species richness at lower elevations across northern South American and the southeast portion of North America.

Independent of the particular dimension or combination of dimensions being accounted for, similar patterns of species richness are evident for the amphibians, birds and mammals such that species richness peaked in the north western portion of South America, in accounting for all three dimensions. Only the predicted species richness patterns in longitude and longitude x altitude MDE models appear visually similar to the empirical patterns.

Deviations from the three-dimensional models for each group of species are shown in Figures 4.6, 4.7 and 4.8. For all groups lowest to negative deviation (observed - expected) in terms of the proportion of total species richness was observed in SE North America; greatest positive deviation was in general over NW South America.

Predictors of species richness

SAR showed that MDE was a significant ($p < 0.05$) independent predictor of species richness in general for all models, but in the one-dimensional model of altitude for both amphibians and mammals; only the two-dimension model of longitude x latitude, and the three-dimensional model were significant for avian species (Table 4.2). For amphibians, MDE accounted for between 54.3% (one-dimensional latitude) and 86.3% (two-dimensional long x lat) of the variation in species richness (mean pseudo $R^2 = 0.727 \pm 0.0476$ SE). A similar range was observed for mammals (52.6% in the single dimension of latitude to 87.1% in the single dimension of longitude; mean pseudo $R^2 = 0.767 \pm 0.0529$ SE). Moreover, MDE explained 88.5% and 80.1% (two-dimensional longitude x latitude, and three-dimensional models, respectively) of variation in avian species richness (mean pseudo $R^2 = 0.843 \pm 0.0476$ SE).

Only two models for mammals (the multi-dimensional models of longitude x altitude and longitude x latitude x altitude) predicted the shape of the empirical richness pattern (i.e., slope did not differ significantly from 1; Table 4.2).

In the multiple SAR models, environmental parameters could significantly ($p < 0.05$) explain between 88.3 and 98.3% of the variation in amphibian species richness, between 87.6 and 98.8% of the variation in avian species richness, and between 74.5 and 95.9 % of the variation in mammal species richness (Table 4.3). Little improvement in model fit was evident following the inclusion of MDE into the multiple SAR models for all groups. The one-dimension model of longitude and two-dimension model of longitude x latitude for mammals were the exceptions in that MDE improved the fit of the regressions by 23.1 and 12.3%, respectively.

Discussion

The general applicability of a modified, fully stochastic, three dimensional MDE model, both as a null prediction, and then as a separate explanatory variable of species richness for three different groups of biota were examined in this chapter. Results indicate that stochastic processes and geometric constraints can explain much of the variation in amphibian and mammal species richness patterns, independent of the number of dimensions being accounted for, and in avian richness patterns for the two-dimensional (longitude x latitude), and three dimensional models (see Table 4.1). This, in principle, suggests that either these species are randomly placed with the limits of the domain or that the richness pattern appears random but is a consequence of some set of interacting factors (Bokma et al. 2001) – reflecting the multitude of ecological and evolutionary processes that have adapted species to respond to the environment in many ways (Lyons and Willig 1997).

The MDE tends to be weaker where ranges are small relative to the domain extent (Laurie and Silander 2002), and stronger where ranges are large relative to the extent of the domain (McCain 2003). Small-ranged species are less likely to experience 'hard' boundaries than large-range species, and thus the impact of boundaries on their richness patterns should be smaller (Jetz and Rahbek 2002), and more likely to reflect environmental and historical factors (Colwell and Lees 2000). Colwell and Lees (2000) further noted that large-ranged species are more likely to be affected by continental geometry and to occupy the centre of the bounded area. Here, a similar gradient was observed with MDE accounting for some 73%, 77% and 84% of the variation in amphibian (smallest average range size), mammal (mid average range size) and avian (largest species range size) species richness gradients, respectively (see Table 4.1).

Studies that have partitioned datasets into range size categories have consistently found stronger support for MDE among large-ranged species than among the smaller range species in the dataset (Hawkins and Diniz-Filho 2002; Jetz and Rahbek 2002; Vetaas and Grytnes 2002; Mora and Robertson 2005, Murphy et al. 2006). In Chapter 3, it was shown that geometric constraints had greatest predictive power on the richness patterns of North America tree species having ranges that are large relative to the domain. For these large-ranged species, MDE alone often explained more of the variation in species richness than all of the environmental variables combined. However, for medium- and small-range species, environmental parameters were much more important. Furthermore, Mora and Robertson (2005) found habitat features had a much greater influence on gradients in species richness of small-range tropical eastern Pacific fishes, whereas a one-dimension mid-domain effect had a significant influence on species richness of large-ranged species.

While the use of modified, fully stochastic MDE models, as used here, have been criticized because the RSFD of the theoretical species may differ from the actual data, altering the fit of the MDE (McCain 2003, Colwell et al. 2004), the alternative (random placement of ranges resampled from an actual RSFD) may inadvertently ‘smuggle in’ effects of environmental gradients and overestimate the contribution of MDE (Zapata et al. 2003, Connolly 2005). Here, the theoretical RSFD produced by the modified, fully stochastic, three dimensional MDE model represents a log-normal RSFD. The log-normal RSFD curve is biologically realistic (Anderson 1985, Brown et al. 1996, Colwell and Lees 2000), and is a theoretical standard to which observed RSFD patterns have been compared (Gaston et al. 2005; and see e.g., Pagel et al. 1991, Blackburn and Gaston 1996, Gaston 1998, Macpherson 2003). While the theoretical RSFD is most similar to that of the amphibian and mammal species and least similar to that of the avian RSFD, the RSFD did differ from all three groups in both shape and mean range size. The differences in RSFD may account the inability of the model to accurately predict the shape and magnitude of species richness (see Tables 4.1 and 4.2). In general, amphibian richness was less than predicted (corresponding to lower mean range size than the theoretical species) and avian and mammal richness were greater than predicted (corresponding to greater mean range size). Even if the theoretical and actual RSFDs differ, MDE may show a high coefficient of determination (R^2), but the differences may be seen in the shape and magnitude of the richness prediction; if RSFD are resampled from actual data or the theoretical and actual RSFD are do not differ, MDE is more likely to accurately predict fit and magnitude of the actual species richness (Zapata et al. 2003, 2005, Colwell et al. 2004, 2005)

The lack of significant fit in five of the seven MDE models for avian species richness may be due to the differences in RSFD (Colwell et al. 2004), or may be due to the variation in relative strength of stochastic (MDE) and the effects of both biotic and abiotic environmental gradients (Colwell et al. 2005, Rangel and Diniz-Filho 2005). Climatic variables significantly predicted avian richness five of the seven SAR models. Indeed, for amphibian and mammals, the climatic variables significantly accounted for much of the variation in species richness. For all groups, precipitation seasonality and isothermality were the most commonly used predictors of species richness. It could be hypothesized that the patterns of species richness examined here were due, to some degree, to environmental stability or predictability (see e.g., Thiery 1982, Begon et al.

1986). These hypotheses are generally accepted as unsupported (Rohde 1992, Willig et al. 2003) and testing these potential hypotheses is beyond the scope of the chapter.

Currie et al. (2004) note that the statistical relationship between climate and broad-scale species richness is one of the strongest patterns in ecology. Climate-based theories of species richness abound, and offer a large body of tests of model fit, assumptions and secondary predictions (Zapata et al. 2005), however, these still have far to go to identifying causality of species richness gradients (Currie et al. 2004).

Strong correlations between dependent and independent variables do not imply a causal link. While correlative studies often lead to the best descriptor of an observed pattern (not necessarily the primary cause), the falsification of potential predictors is valuable (Bokma et al. 2001). Although I have shown here that environmental correlates provide equal or better explanatory value than that of the MDE predictions (see Table 4.3), the principle of parsimony (Occam's razor) suggests that the alternative with the least assumptions should be selected. In this study, random placement of ranges within the limits of a domain would be the simplest explanation. It seems most likely that real species distributions are the product of variation in relative strength of stochastic and ecological processes (Bokma et al. 2001, Colwell et al. 2005, Rangel and Diniz-Filho 2005).

References

Anderson, S., 1985. The theory of range-size (RS) distributions. American Museum Novitates 2833: 1-20.

Anselin, L., 1995. SpaceStat, A software programme for the analysis of spatial data. Version 1.80. Regional Research Institute, West Virginia University, Morgantown, WV.

Arita, H.T., 2005. Range size in mid-domain models of species diversity. Journal of Theoretical Biology 232: 119-126.

Blackburn, T.M. & Gaston, K.J., 1996. Spatial patterns in the geographic range sizes of bird species in the New World. Philosophical Transactions of the Royal Society of London, Series B 351: 897-912.

Begon, M., Harper, J.L. & Townsend, C.R., 1990. Ecology: individuals, populations and communities. Blackwell Scientific, Boston, MA.

Bokma, F., Bokma, J. & Mönkkönen, M., 2001. Random processes and geographic species richness patterns: why so few species in the north? Ecography 24: 43-49.

Brown, J.H., Stevens, G.C. & Kaufman, D.M., 1996. The geographic range: size, shape, boundaries, and internal structure. Annual Review of Ecology and Systematics 27: 597-623.

Colwell, R.K. & Hurtt, G.C., 1994. Nonbiological gradients in species richness and a spurious Rapoport effect. American Naturalist 144: 570-595.

Colwell, R.K. & Lees, D.C., 2000. The mid-domain effect: geometric constraints on the geography of species richness. Trends in Ecology and Evolution 15: 70-76.

Colwell, R.K., Rahbek, C. & Gotelli, N.J., 2004. The mid-domain effect and species richness patterns: What have we learned so far? American Naturalist 163: E1-E23.

Colwell, R.K., Rahbek, C. & Gotelli, N.J., 2005. The mid-domain effect: there's a baby in the bathwater. American Naturalist 166: E149-E154.

Connolly, S.R., 2005. Process-based models of species distributions and the mid-domain effect. American Naturalist 166: 1-11.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J., Hawkins, B.A., Kaufman, D.M., Kerr, J.T., Oberdorff, T., O'Brien, E. & Turner, J.R.G., 2004. Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. Ecology Letters 7: 1121-1134.

Diniz-Filho, J.A.F., de Sant'Ana, C.E.R., de Souza, M.C. & Rangel, T.F.L.V.B., 2002. Null models and spatial patterns of species richness in South American birds of prey. Ecology Letters 5: 47-55.

Gaston, K.J., 1998. Species–range size distributions: products of speciation, extinction and transformation. Philosophical Transactions of the Royal Society of London, Series B 353: 219-230.

Gaston, K.J., Davies, R.G., Gascoigne, C.E. & Williamson, M., 2005. The structure of global species-range size distributions: raptors and owls. Global Ecology and Biogeography 14: 67-76.

GLOBE Task Team, 1999. The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0. National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, CO, U.S.A. Digital data base on the World Wide Web (URL: http://www.ngdc.noaa.gov/mgg/topo/globe.html) and CD-ROMs.

Hawkins, B.A. & Diniz-Filho, J.A.F., 2002. The mid-domain effect cannot explain the diversity gradient of Nearctic birds. Global Ecology and Biogeography 11: 419-426.

Hawkins, B.A., Diniz-Filho, J.A.F. & Weis, A.E., 2005. The mid-domain effect and diversity gradients: is there anything to learn? American Naturalist 166: E140-E143.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

IUCN, Conservation International & NatureServe, 2004. Global Amphibian Assessment. IUCN,Conservation International, and NatureServe, Washington, DC and Arlington, Virginia, USA.

Jetz, W. & Rahbek, C., 2001. Geometric constraints explain much of the species richness patterns in African birds. Proceedings of the National Academy of Sciences U.S.A. 98: 5661-5666.

Jetz, W. & Rahbek, C., 2002. Geographic range size and determinants of avian species richness. Science 297: 1548-1551.

Koleff, P. & Gaston, K.J., 2001. Latitudinal gradients in diversity: real patterns and random models. Ecography 24: 341-351.

Laurie, H. & Silander Jr., J., A. 2002. Geometric constraints and spatial patterns of species richness: critique of range-based null models. Diversity and Distributions 8: 351-364.

Lees, D.C., Kremen, C. & Andriamampianina, L., 1999. A null model for species richness gradients: bounded range overlap of butterflies and other rainforest endemics in Madagascar. Biological Journal of the Linnean Society 67: 529-584.

Lyons, S.K. & Willig, M.R., 1997. Latitudinal patterns of range size: methodological concerns and empirical evaluations for New World bats and marsupials. Oikos 79: 568-580.

Macpherson, E., 2003. Species range size distributions for some marine taxa in the Atlantic Ocean: effect of latitude and depth. Biological Journal of the Linnean Society 80: 437-455.

McCain, C.M., 2003. North American desert rodents: a test of the mid-domain effect in species richness. Journal of Mammalogy 84: 967-980.

McCain, C.M., 2004. The mid-domain effect applied to elevational gradients: species richness of small mammals in Costa Rica. Journal of Biogeography 31: 19-31.

Mora, C. & Robertson, D.R., 2005. Causes of latitudinal gradients in species richness: a test with fishes of the tropical Eastern Pacific. Ecology 86: 1771-1782.

Murphy, H.T., VanDerWal, J., Khalatkhar, N. & Lovett-Doust, J., 2006. Geometric and environmental correlates of tree species richness in North America. Ecography. In Review.

Pagel, M.P., May, R.M. & Collie, A.R., 1991. Ecological aspects of the geographic distribution and diversity of mammalian species. American Naturalist 137: 791-815.

Patterson, B.D., Ceballos, G., Sechrest, W., Tognelli, M.F., Brooks, T., Luna, L., Ortega, P., Salazar, I. & Young, B.E., 2003. Digital Distribution Maps of the Mammals of the Western Hemisphere, version 1.0. NatureServe, Arlington, Virginia, USA.

Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F., 2005. An evolutionary tolerance model explaining spatial patterns in species richness under environmental gradients and geometric constraints. Ecography 28: 253-263.

Ridgely, R.S., Allnutt, T.F., Brooks, T., McNicol, D.K., Mehlman, D.W., Young, B.E. & Zook, J.R., 2003. Digital Distribution Maps of the Birds of the Western Hemisphere, version 1.0. NatureServe, Arlington, Virginia, USA.

Romdal, T.S., Colwell, R.K. & Rahbek, C., 2005. The influence of band sum area, domain extent, and range sizes on the latitudinal mid-domain effect. Ecology 86: 235-244.

Rohde, K., 1992. Latitudinal gradients in species diversity: the search for the primary cause. Oikos 65: 514-527.

Thiery, R.G., 1982. Environmental instability and community diversity. Biol. Rev. 57: 671-710.

Vetaas, O.R. & Grytnes, J.A., 2002. Distribution of vascular plant species richness and endemic richness along the Himalayan elevation gradient in Nepal. Global Ecology and Biogeography 11: 291-301.

Willig, M.R. & Lyons, S.K., 1998. An analytical model of latitudinal gradients of species: richness with an empirical test for marsupials and bats in the New World. Oikos 83: 93-98.

Willig, M.R., Kaufman, D.M. & Stevens, R.D., 2005. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. Annual Review of Ecology, Evolution and Systematics 34: 273-309.

Zapata, F.A., Gaston, K.J. & Chown, S.L., 2003. Mid-domain models of species richness gradients: assumptions, methods and evidence. Journal of Animal Ecology 72: 677-690.

Zapata, F.A., Gaston, K.J. & Chown, S.L., 2005. The mid-domain effect revisited. American Naturalist 166: E144-E148.

Figure 4.1. Illustration of the domain, Continental North America, Central, and South America, used in this study. The perimeter of the domain was extracted from "World Countries 1992" base map, supplied with ArcView 3.2 (ESRI, California, USA) and altitudes were obtained from the GLOBE project (GLOBE, 1999). Lighter areas represent higher elevations, reported in metres above sea level.

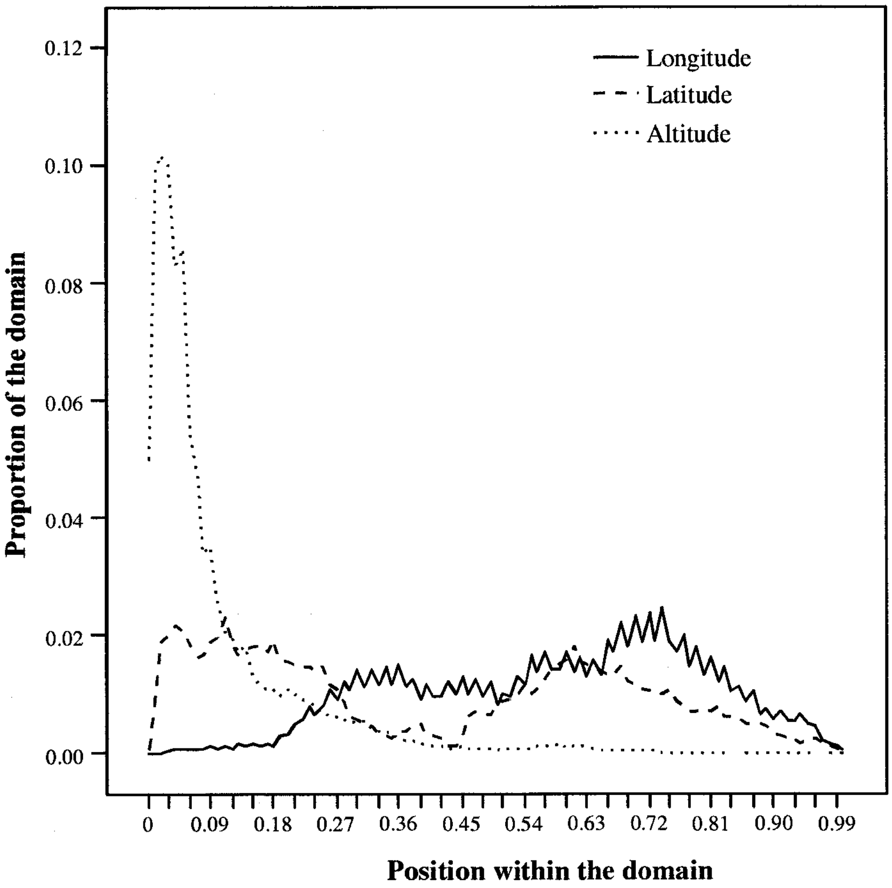


Figure 4.2. For each single dimension, longitude, latitude and altitude, the relationship between proportionate area and position within the domain is depicted.

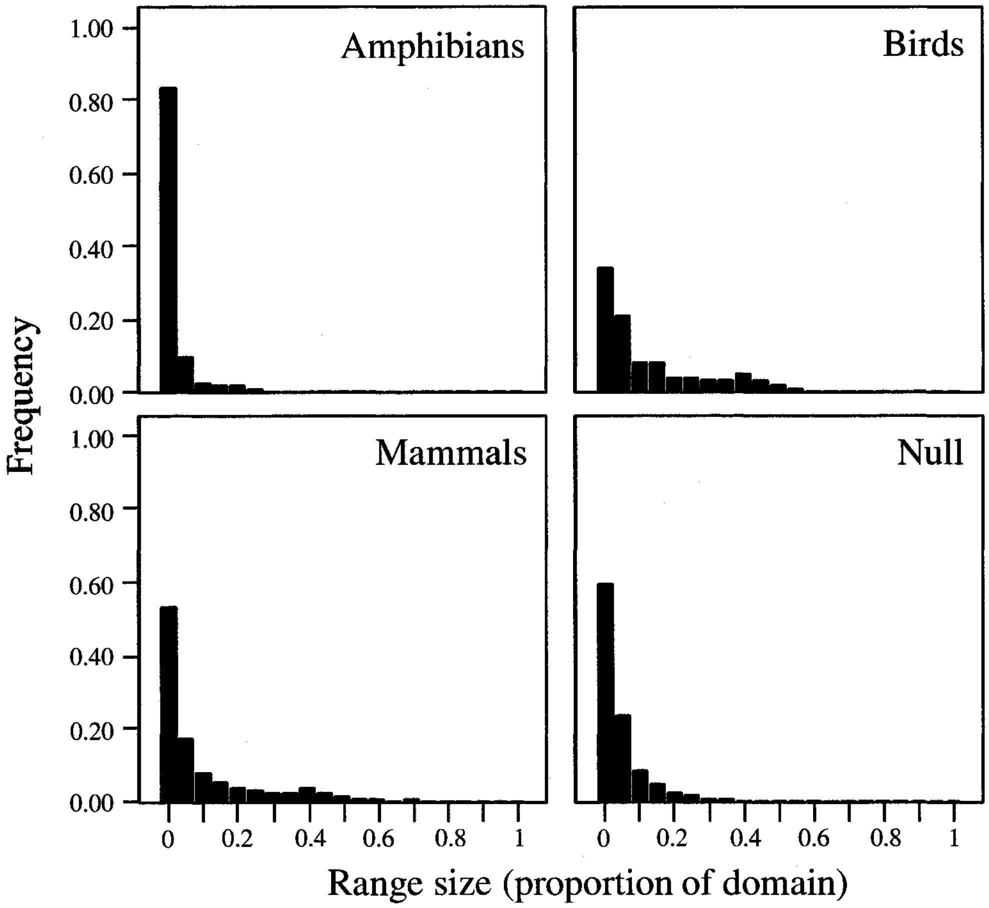


Figure 4.3. Range size frequency distributions for each of three biotic groups (amphibian, bird, mammal) and for the theoretical species created by a modified, fully stochastic MDE model.

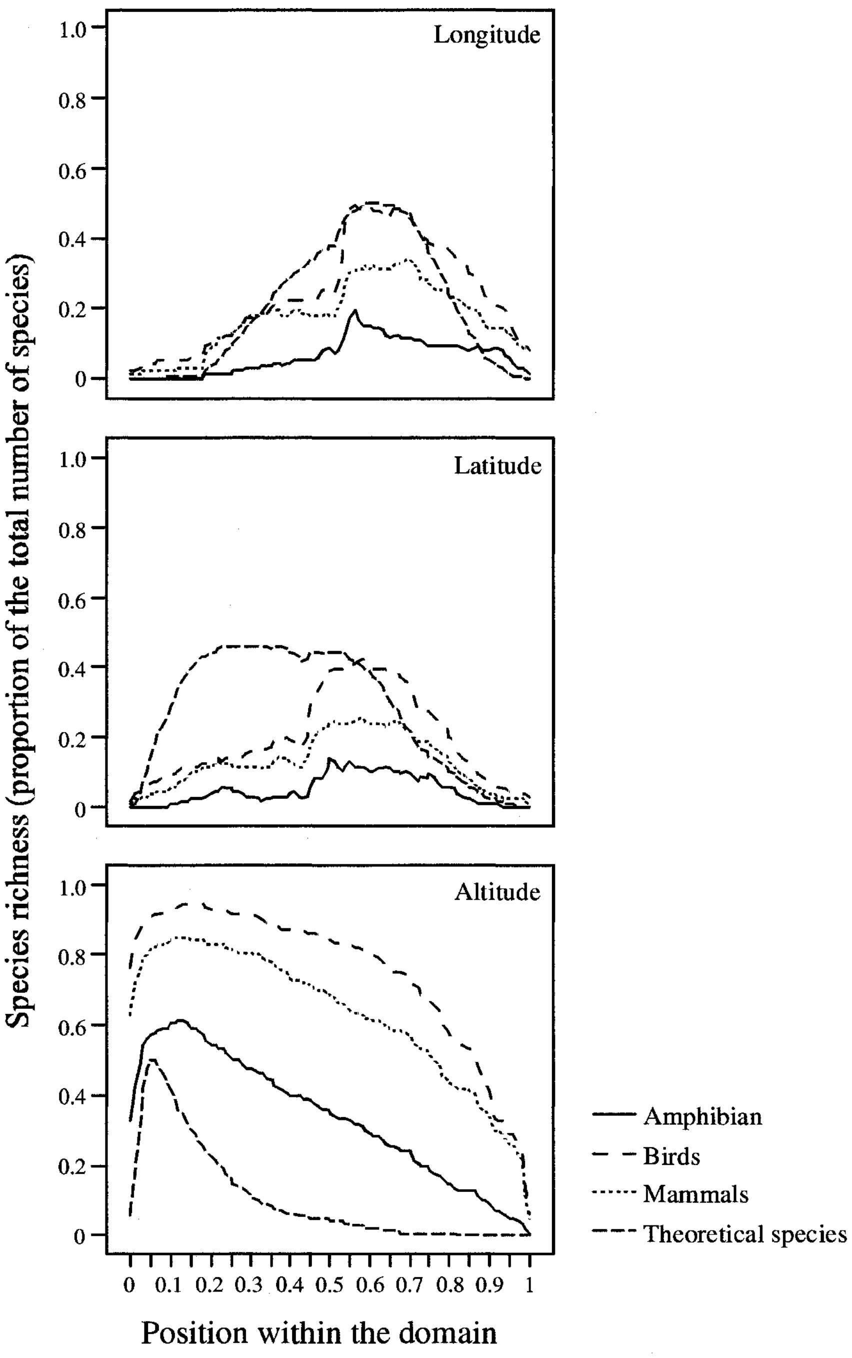


Figure 4.4. For each single dimension, species richness curves for each of three biotic groups (amphibian, bird, mammal) and for the theoretical species created by a modified, fully stochastic MDE model, with respect to position within the domain extents.

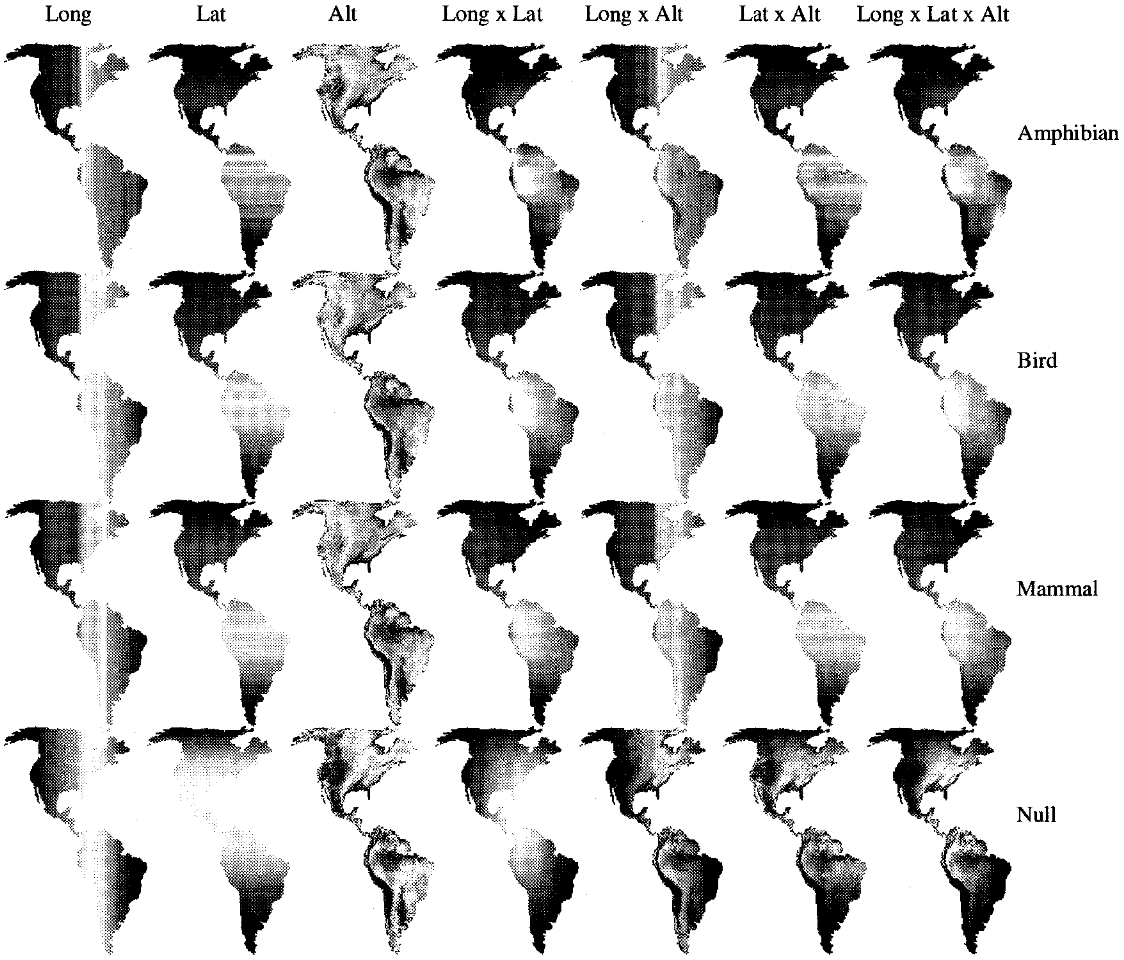


Figure 4.5. Patterns of species richness for each of three biotic groups (amphibians, birds and mammals) as predicted by a modified, fully stochastic, three dimensional MDE model. One-dimensional (Longitude, Latitude and Altitude) and multi-dimensional (Long x Lat, Long x Alt, Lat x Alt and Long x Lat x Alt) predictions are shown. Equal interval classification is shown with color ramps indicating minimum (dark, bottom of legend) to maximum (light, top of legend) predicted species richness.

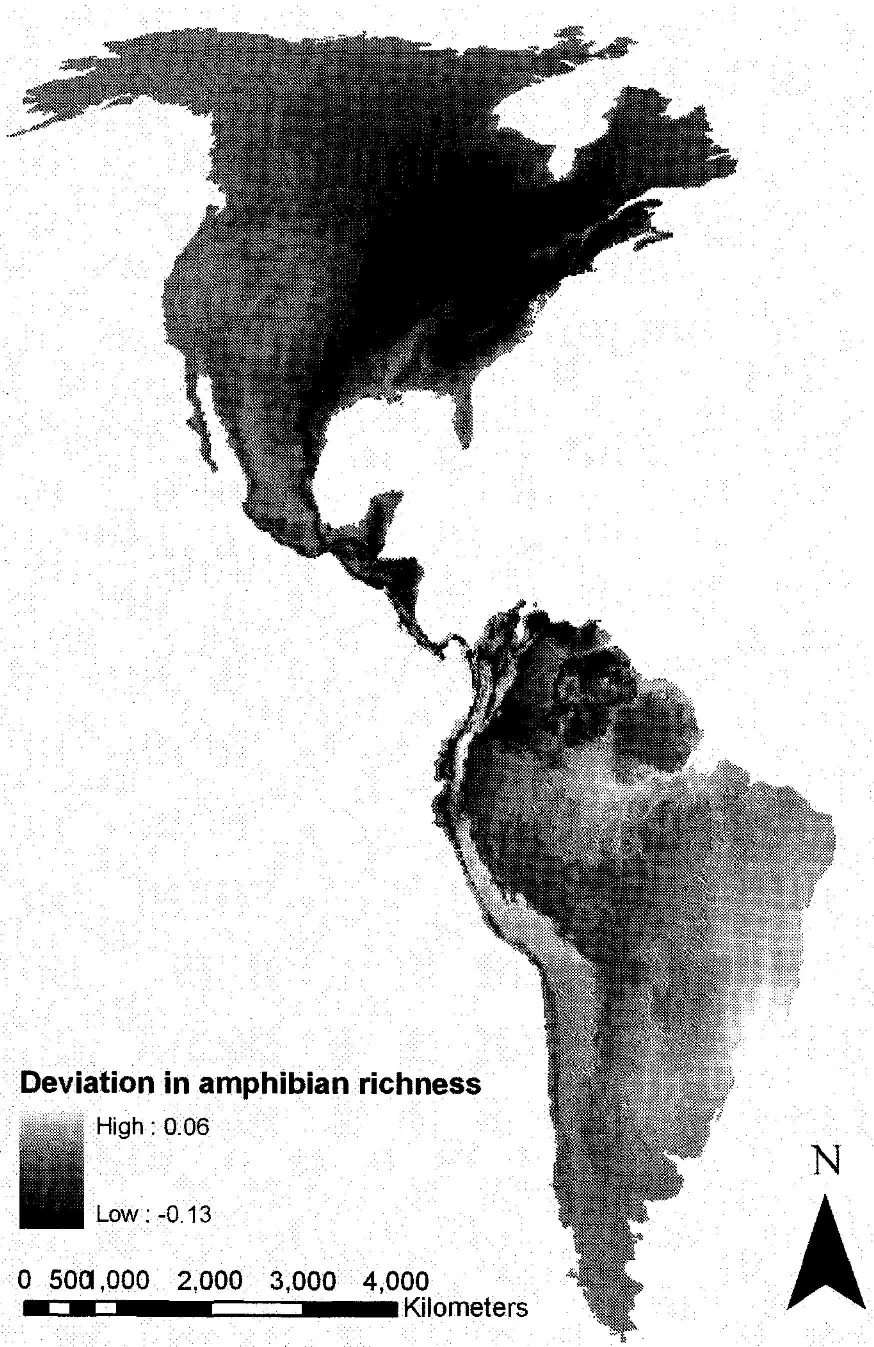


Figure 4.6. Deviation (observed - expected) from three dimensional MDE predictions for empirical amphibian richness. Values represent deviation in proportion of total species richness.

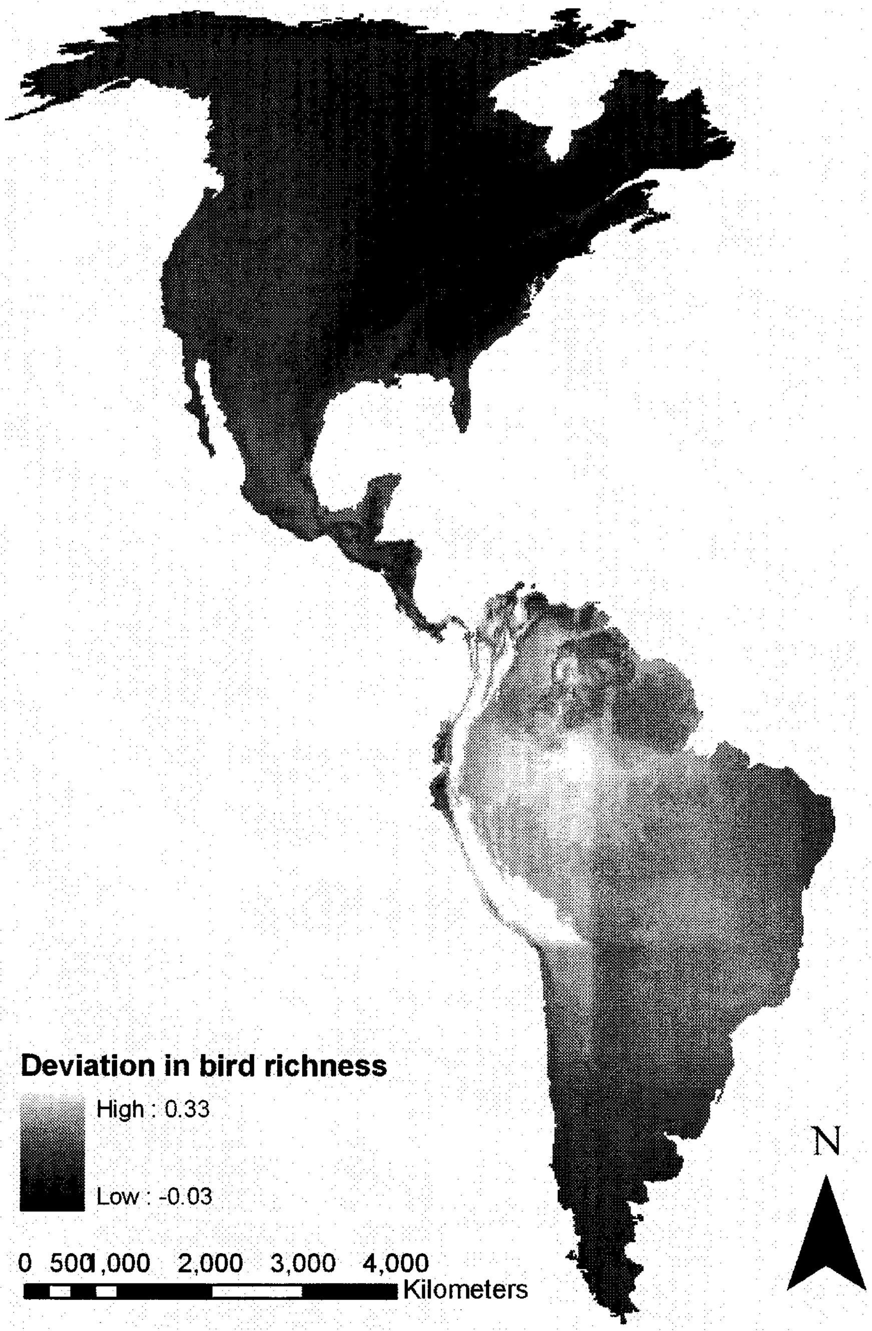


Figure 4.7. Deviation (observed - expected) from three dimensional MDE predictions for empirical bird richness. Values represent deviation in proportion of total species richness.

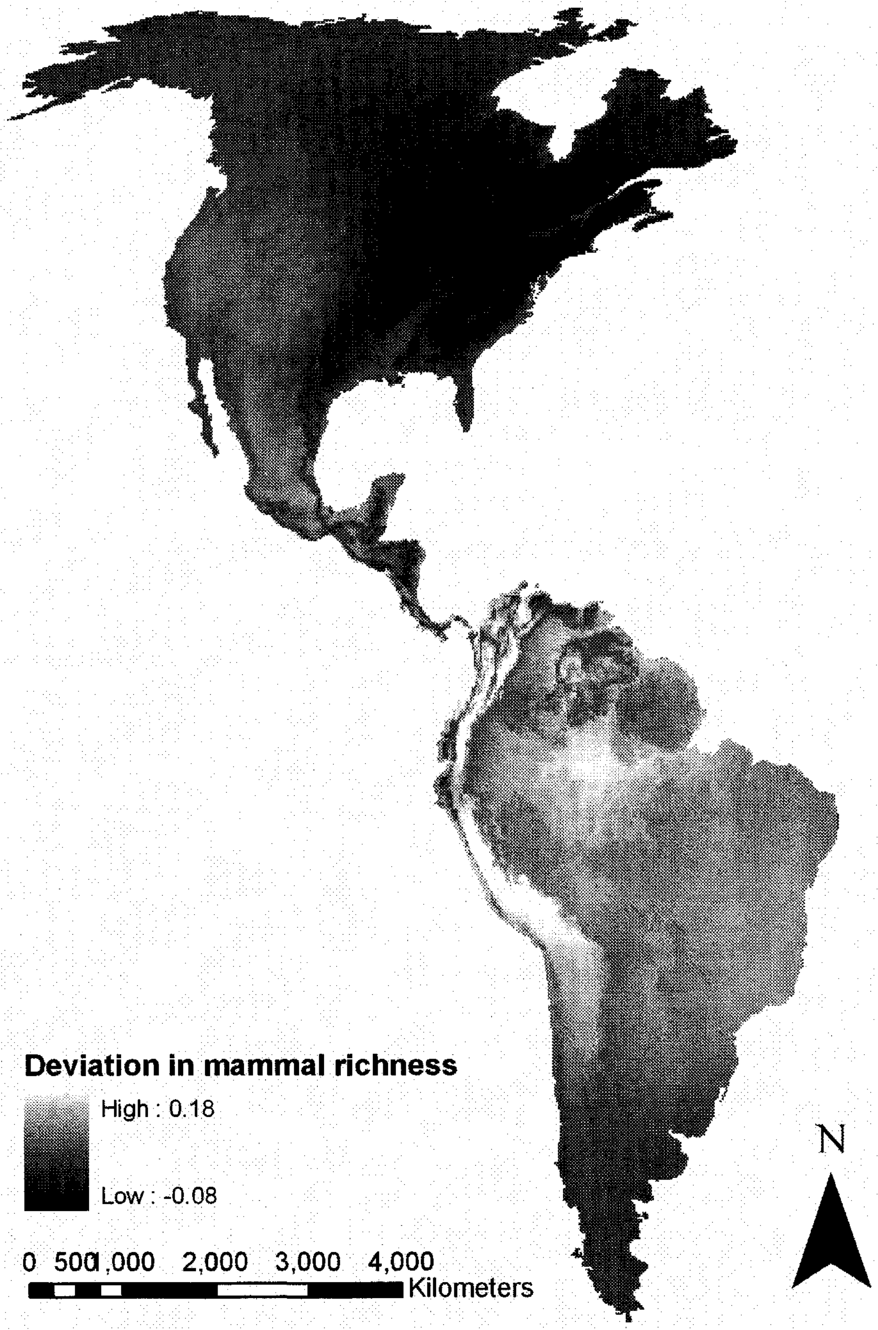


Figure 4.8. Deviation (observed - expected) from three dimensional MDE predictions for empirical mammal richness. Values represent deviation in proportion of total species richness.

Table 4.1. Maximum species richness noted for each of three biotic groups (amphibians, birds and mammals) and as predicted by a modified, fully stochastic, three dimensional MDE model. One-dimensional (Longitude, Latitude and Altitude) and multi-dimensional (Long x Lat, Long x Alt, Lat x Alt and Long x Lat x Alt) richness values are shown.

	Longitude	Latitude	Altitude	Long x Lat	Long x Alt	Lat x Alt	Long x Lat x Alt
Amphibians	0.19	0.14	0.62	0.08	0.17	0.12	0.08
Birds	0.50	0.43	0.94	0.35	0.49	0.42	0.35
Mammals	0.34	0.26	0.86	0.20	0.33	0.25	0.20
Null	0.50	0.46	0.50	0.28	0.26	0.24	0.14

Table 4.2. Results for spatial autoregressive models of MDE predicted species richness against empirical species richness of amphibians, birds and mammals. Significant ($p < 0.05$) pseudo R^2 values are bolded. Also shown are the t-test results of the hypotheses that slope = 1 and intercept = 0; bolded values are not significant at the $p < 0.05$ level.

		Longitude	Latitude	Altitude	Long x Lat	Long x Alt	Lat x Alt	Long x Lat x Alt
Amphibian								
	intercept	**1.414**	**1.814**	32.532	-6.126	2.839	-2.339	-4.345
	slope	-57.702	-48.196	-9.690	-36.602	-18.559	-13.526	-4.355
	R^2	**0.752**	**0.543**	0.759	**0.863**	**0.808**	**0.638**	**0.756**
Bird								
	intercept	13.264	6.797	38.180	3.304	31.248	14.067	7.503
	slope	-13.069	-25.740	-8.225	-8.378	-4.398	-9.646	3.010
	R^2	0.813	0.519	0.536	**0.885**	0.761	0.789	**0.801**
Mammal								
	intercept	18.510	6.480	44.881	2.706	27.153	6.675	2.192
	slope	-27.203	-33.033	-9.190	-23.963	**-0.323**	-13.887	**-1.356**
	R^2	**0.871**	**0.526**	0.617	**0.837**	**0.865**	**0.740**	**0.763**

Table 4.3. Results of spatial linear regression analysis for predictors of amphibian, avian and mammal richness. Shown are the pseudo R^2 values for separate analyses of MDE and the environmental variables, and then the analysis with MDE and environmental variables together. Significant ($p < 0.05$) values are bolded.

		Longitude	Latitude	Altitude	Long x Lat	Long x Alt	Lat x Alt	Long x Lat x Alt
Amphibian								
	MDE	**0.752**	**0.543**	0.759	**0.863**	**0.808**	**0.638**	**0.756**
	Environmental	**0.894**	**0.933**	0.775	**0.983**	**0.883**	**0.967**	**0.987**
	MDE + Environmental	**0.912**	**0.937**	0.867	**0.984**	**0.918**	**0.972**	**0.989**
Bird								
	MDE	0.813	0.519	0.536	**0.885**	0.761	0.789	**0.801**
	Environmental	0.824	**0.876**	0.669	**0.982**	**0.881**	**0.895**	**0.988**
	MDE + Environmental	**0.968**	**0.877**	0.707	**0.990**	**0.927**	**0.913**	**0.991**
Mammal								
	MDE	**0.871**	**0.526**	0.617	**0.837**	**0.865**	**0.740**	**0.763**
	Environmental	**0.745**	**0.907**	0.678	**0.919**	**0.768**	**0.959**	**0.915**
	MDE + Environmental	**0.975**	**0.909**	0.753	**0.944**	**0.891**	**0.961**	**0.944**

Chapter 5 – A Bayesian decision approach for the null prediction of a species' ecological niche

Summary

Species distribution models attempt to determine the non-random associations between environmental attributes associated with known occurrences and absences. Since most species surveys only record occurrences, pseudo-absences have been used in place of true absences. Pseudo-absences are randomly chosen locations where the species has not been reported and thus maybe assumed to be absent. Here, a methodological framework is presented whereby the species-environment relationship can be examined, in essence, providing a null model of a species niche. Recognizing that the species-environment relationship may be scale-dependent, logistic regressions are conducted between the environmental variable and occurrence / pseudo-absence data using spatially independent subsets, to determine the spatial scale at which the species-environment relationship is best fitted. The most significant (best fit) independent environmental attributes and 100 spatially independent subsets of a training dataset (a randomly drawn 50% of the occurrence / pseudo-absence data) were used to train a naïve Bayesian algorithm. The *a priori* and *a posteriori* probabilities from the Bayesian algorithm were used as a ruleset that was first evaluated using the training, testing (the remaining data not used for training) and entire occurrence / pseudo-absence data datasets. The algorithm was then applied throughout a landscape, generating predictions as to the probability that any locality is part of a species' niche.

The methodology was applied to occurrences of *Opuntia humifusa* at Point Pelee National Park, SW Ontario. The derived model accurately predicted species occurrences 97.1% of the time using the landuse category of the location of interest and the proportions of the surrounding landscape with respect to beach (at 50m radius), human use (at 500m), ponds (at 700m) and roads/parking (at 700m). This technique should be useful for a wide variety of studies including those examining the distribution of multiple species.

Introduction

The study of how organisms are distributed in space and time has long interested ecologists and inspired many modeling approaches that quantify the species-environment relationship (e.g., Rushton et al. 2004, Guisan and Thuiller 2005). Such species distribution models (SDMs) have also been referred to as habitat suitability mapping, quantitative habitat models, ecological niche models and predictive distribution maps, and all are potentially important tools to guide management and restoration approaches with respect to conservation of rare species and/or environmental change (land fragmentation, climate change, etc.) (Guisan and Zimmerman 2000, Johnson et al. 2004, Guisan and Thuiller 2005). Application of such tools is a function of advances in geographical information systems (GIS), availability of geo-referenced databases (such as species distributions, topographical, climatic and landuse information), and the ability to utilize computationally intensive numerical techniques (Rushton et al. 2004).

Kearney and Porter (2004) suggested that the concept of the niche provides a useful starting point for understanding the distribution of species. Indeed, many SDMs assume an underlying niche concept (Guisan and Zimmermann 2000, Guisan and Thuiller 2005, Soberón and Peterson 2005). The unique attributes of locations where a species occurs are often inferred as characteristics of a species niche, and thus a common approach used to determine the extent of a species' geographic distribution(definine limits of a species' niche) is to characterize, quantitatively, a suite of environmental conditions for known occurrences of the species (Peterson 2001, Kearney and Porter 2004).

Surveys of species often report only occurrences of species; rarely are absences recorded and thus the methods for modelling species distributions are limited (Pearce and Boyce 2005). One way to model presence-only data is to use pseudo-absences. Pseudo-absences are assumed to be true absences but may represent presence locations that have not been reported (Graham et al. 2004). Often the pseudo-absences are chosen at random to represent the background set of environmental attributes of the study area (Pearce and Boyce 2005, Soberón and Peterson 2005). Statistical models of the probability of occurrence can be derived for the species-environment relationship using occurrence / pseudo-absence data and then these models can be interpolated to other parts of the landscape, potentially under differing environmental conditions (see e.g., Lindenmayer et al. 1991, Sykes et al. 1996, Peterson et al. 1999, Peterson 2001, Kearney and Moussalli 2003, Kearney et al. 2003).

In this chapter I develop a method, and associated software, that uses species occurrences (ignoring species-specific life history attributes) such that the deviation from randomly placed pseudo-absences with respect to environmental attributes, effectively parameterizes a null model for a species' distribution. Thus the model created is a more appropriate null model (than the null assumption of complete spatial randomness) describing the niche of a single species. Deviations from this null niche model would represent greater scientific value than a null hypothesis of complete spatial randomness (Goovaerts et al. 2004, Goovaerts and Jacquez 2004) and suggest which species-specific attributes or biotic interactions are important factors defining a species' distributions.

Species niche modelling

Species distribution models have been used in a wide variety of applications from quantifying the niche of a species, to identification of sites with high potential of occurrence or high value for restoration, to assessing impact of environmental change (e.g., climate or landuse change) on species distributions (see reviews in Rushton et al. 2004, Guisan and Thuiller 2005). While such modelling efforts have been useful tools with respect to applications in ecology and conservation biology, common to many of these models is a pseudo-equilibrium assumption (Guisan and Theurillat 2000) and a reliance on the niche concept (Guisan and Zimmermann 2000).

Species and environmental data are often sampled during a limited time and thus the models produced often represent only a snapshot of the species-environment relationship. Assuming pseudo-equilibrium is a convenient postulate for projecting future distributions but the validity of the assumption is questioned in situations where systems are not in equilibrium (e.g., dispersal-limited range expansion, invasive species spread) (Guisan and Thuiller 2005).

Nearly all species distribution models rely on assumptions underlying the niche concept; it is assumed that a species is affected by its environmental requirements and/or itself affects the environment in which it occurs (Liebold 1995, Guisan and Zimmermann 2000). By quantifying where a species occurs, it is typically assumed that the distribution is constrained by biotic interactions; such models are *de facto* describing the realized niche (*sensu* Hutchinson 1957). In modelling a species' niche and its distribution with respect to potential change in climate, landuse, etc., one must assume a pseudo-equilibrium and that the realized niche, in terms of the abiotic characteristics, is a good approximation of the fundamental niche (*sensu* Hutchinson 1957) (Vetaas 2002). It has been suggested that, if the distribution of a species in the context of varied biotic backgrounds can be observed, then the composite fundamental niche can be described (Peterson et al. 1999, Peterson 2001). However, many species are able to survive outside their current distribution provided dispersal barriers are overcome (i.e., outside their realized niche; Vetaas 2002). One only has to examine the invasive species literature to see examples of species thriving outside their "realized niche" (Jesche and Strayer 2005, Murphy et al. 2006). Ackerly (2003) proposed an alternative to the classical concepts of niche by proposing the potential niche. The potential niche is that part of the fundamental niche, not yet recognized as part of the realized niche, but condition mimic that of the *realized environment* (the environmental attributes of the realized niche); hence predictive SDM's are, indeed, often exploring the potential niche of a species.

The species distribution modelling program

Three basic inputs are needed for the methodology described here: species occurrence data, environmental data and a specified set of buffer distances. The environmental data must be in raster format, with a minimum cell size less than or equal to the minimum buffer distance. Using the input data, the program goes through a series of calculations and user interactions to create the species distribution model describing the realized niche of a species, predicting the potential niche.

The first step involves the creation of pseudo-absences. Pseudo-absences, equal in number to that of occurrences, are randomly placed within the domain bounded by the extent of the environmental data. From this, a presence-absence (PA) database is created. The attributes of the environment at each PA locality are recorded. In adition to the environmental data of the specific raster cell in which the PA is located, environmental data is also recorded at a number of buffer distances; in other words, the environmental conditions are quantified at the specific PA location and for the surrounding landscape at the specified buffered sizes. The surrounding landscapes are treated as square delineations where the size is defined by the buffer distance (e.g., a buffer size of 1m on a grid with a resolution of 1m would describe a 3x3 cell area in which the edges are defined as 1m from the focal cell in the cardinal directions).

Quantifying the environment at the various buffer distances differs depending on the type of data used. For categorical environmental datasets (e.g., landcover, soil type), the point value at the PA location and the proportion of the surrounding landscape at each buffer distance for each category is recorded. With respect to continuous data, the point value at the PA location and the mean and standard error of the values representing the surrounding landscape at each buffer distance is recorded.

To determine the scale (buffer distance) that is the best predictor of a species occurrence, multiple single logistic regressions are performed on the PA data with each environmental parameter. At each buffer distance and for each environmental parameter, spatially independent subsets of the data are used. The subsets are selected by:

1. First, a PA location, n_1, is randomly selected from all PA locations, n_i.
2. A second site, n_2, is selected that satisfies the criteria of spatial independence. The condition of spatial independence here refers to the criteria set that areas within which the predictor variable is measured may not overlap; in other words, for the case considered here, the distance between n_1 and n_2 must be greater than two times the radius (buffer size).
3. Random PA locations are added to the subset until the spatial independence criteria can no longer be met.
4. Since the dependent variable is binary (presence or absence), logistic regression is used to determine the fit of the environmental data with the selected PA points. The significance and Nagelkerke's R^2, among other statistics are recorded.

5. Steps 1-4 are repeated for 100 different spatially independent subsets of the data such that the distribution of the model fit and effect size can be examined.

The sampling includes replacement of data between subsets. However, the spatial independence criterion excludes replacement within subsets and thus this technique differs from a bootstrapping procedure. The mean and SE of the regression statistics (log-likelihood and Nagelkerke's R^2) for each environmental attribute at each buffer distance are used such that the optimal buffer size for each environmental parameter can be determined based on significant ($p<0.05$ level) model fit.

A list of the environmental predictors at the optimal buffer size, based on greatest model fit, is created; often these are not independent of each other. By examining a correlation matrix that reports r (Pearson correlation coefficient), one can select the environmental predictors based on logistic regression model fit and independence of predictors. These predictors are then input into a naïve Bayesian algorithm for modelling the species niche.

Naïve Bayesian classifiers are simple Bayesian networks (Bayesian classification systems) in which there is only one class variable (dependent variable) with all other variables considered as attributes of the class variable (independent variables) (Porwal et al. 2006). The class variable is the root of the network with each attribute as a child of the parent (the class variable). Although the attribute variables can act as parents to other variables, here we assume attribute variables do not act as parents and thus, the environmental attributes are assumed to be independent of each other.

Class variables are categorical variables with two or more states. In this application of a naïve Bayesian classifier, the class variable of occurrence data has two states, species present or absent. The naïve Bayesian classifier here uses the following algorithms (Porwal 2006). Given C as a class variable in this case with two states = {0, 1} (representing species absent or present) and X is a vector of attributes in which X = {x_1, x_2, x_3,...x_d}, then the Bayesian algorithm can be written as:

$$p(C_i| \mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \ldots \mathrm{x}_\mathrm{d}) = p(\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \ldots \mathrm{x}_\mathrm{d}| C_i) p(C_i)$$

where $p(C_i| \mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \ldots \mathrm{x}_\mathrm{d})$ is the *a posteriori* probability of class membership.

Naïve Bayesian algorithms assume the conditional probabilities of the attribute variables are independent, thus:

$$p(X| C_i) = \prod_{k=1}^{d} p(x_k|C_i)$$

This means that *a posteriori* probability of class membership can be estimated as:

$$p(C_i|X) = p(C_i| \mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \ldots \mathrm{x}_\mathrm{d}) = p(C_i)\prod_{k=1}^{d} p(x_k|C_i)$$

Given that $p(C=1|X) + p(C=0|X)$ must equal 1, the probability of an occurrence can be estimated as:

$$p(C=1|X) = \frac{p(C=1)\prod_{k=1}^{d} p(x_k|C=1)}{p(C=1)\prod_{k=1}^{d} p(x_k|C=1) + p(C=0)\prod_{k=1}^{d} p(x_k|C=0)}$$

As one of the advantages of using naïve Bayesian classifiers is the ability to include both continuous and categorical variables, assuming a normal distribution, the *a posteriori* probabilities are calculated for continuous attributes as:

$$p(x_k|C_i)=\frac{1}{\sigma_{ki}\sqrt{2\pi}}\exp\frac{-(x-\mu_{ki})^2}{2\sigma_{ki}},-\infty<x<\infty,-\infty<\mu_{ki}<\infty,\sigma_{ki}>0$$

where μ_{ki} is the mean and σ_{ki} is the standard deviation of the attribute; and for categorical attributes as:

$$p(x_k|C_i)=\frac{count[x|C_i]}{count[x_k]}$$

in other words, as the proportion of each category of the attribute, with respect to species presence or absence.

Input into the naïve Bayesian classifier is 50% of the PA data (a training dataset selected at random from all occurrence and pseudo-absence points) and the environmental data at the scale that best distinguishes occurrences from pseudo-absences (again, using multiple single logistic regressions). One hundred spatially independent subsets of the training data (created using the procedure described above) are used to train the naïve Bayesian classifier. The subsets are spatially independent at the largest scale (buffer size) of the environmental attributes input into the Bayesian classifier. The output of the training is the mean and associated variance of a set of rules describing the *a priori* probability of an occurrence or pseudo-absence, the *a posteriori* probability that x belongs to C_i (= {0, 1} where 0 is an absence and 1 is an occurrence) for categorical variables and the mean and standard deviation to calculate the *a posterior* probability for continuous variables.

The training, testing and overall datasets (testing dataset is the 50% of the data not used to train the naïve Bayesian classifier, and overall is both training and testing datasets together) are then used to numerically validate the model. The output is a contingency table whereby the number of true positives (predicted and actual occurrence), true negatives (predicted absence and actual pseudo-absence), false positives (predicted occurrence and actual pseudo-absence) and false negatives (predicted absence and actual occurrence) are tallied and the percent accuracy is calculated (see e.g., Brown 1994, Guisan and Zimmerman 2000). The ruleset is then applied to the environmental databases in a GIS environment to predict the probability that any location within the landscape is part of the species potential niche. This is done such that each location (cell) within the landscape is treated as an occurrence or pseudo-absence in the PA database, calculating the environmental attributes at their "optimal" scale and then applying the naïve Bayesian ruleset to calculate the probability.

Methods

Environmental attributes and species occurrence data

To validate the model with empirical species data, occurrences of individuals of *Opuntia humifus*a, eastern prickly pear cactus at Point Pelee National Park (PPNP) in SW Ontario were used. *Opuntia humifusa* is listed by COSEWIC as Endangered in Canada (Klinkenberg and Klinkenberg 1985). It is a low, spreading succulent plant having jointed, circular-to-subovate flattened green stems. This species occurs in prairie and deciduous forest habitats, ranging from southwestern Ontario and Wisconsin, south to eastern Texas and along the gulf and Atlantic coasts from Florida to Massachusetts

(Whitehead 1995). Many of the individual plants at PPNP are being subjected to intense biotic and abiotic selection pressures, including competition from exotic introduced weeds, successional competition from native grasses, forbs and shrubs which may displace the cactus, physical wave and storm action along coastal areas, as well as impacts from trampling, collecting, and other human activities. All of this threatens the existence of *O. humifusa* at PPNP (Kraus 1991).

An intensive field survey for *O. humifusa* individuals was undertaken in July-August 2004, in which GPS locations of the individuals were recorded using a Trimble AG132 backpack GPS (see VanDerWal et al. 2005a). In total, 345 individuals were located. The environmental attribute data sets were made available by PPNP staff (landuse and aerial photos; both from 2000) and through the Ontario Geospatial Data Exchange (http://www.lio.mnr.gov.on.ca/ogdedescription.cfm; digital elevation model or DEM). All environmental data was rasterized or resampled to 1m resolution. Although climatic variables are often included in species distribution modelling, it was not considered that climate was a limiting factor in the relatively small region being modelled.

Species distribution model

Models of *O. humifusa*'s potential niche were developed based on 345 occurrences and an equal number of pseudo-absences randomly placed within the limits of PPNP. The environmental attributes used were landuse, DEM and aerial photo geospatial databases. The landuse database included 11 classes: beach, primary and secondary successional savanna, wet meadow, marsh, pond, juniper stand, forest, roads / parking, trails and human use / park infrastructure (Fig. 5.1). The aerial photo was included as a surrogate for canopy openness (since lighter greyscale values typically represented more open areas). The buffer zones used ranged from 10 to 2500m (10, 20, 30, 40, 50, 60, 80, 100, 120, 150, 200, 300, 400, 500, 700, 900, 1100, 1500, 2000, 2500).

The modelling software was developed using Visual Basic (VB.NET, Microsoft Corporation, Redmond, WA) and the libraries / functionality of ArcGIS 9.0 (ESRI, Redlands, CA). The occurrence / pseudo-absences, environmental data and buffer distances were put into the model. The first set of outputs was a list of spatially independent subsets and analysis of the logistic regressions of each subset with respect to each environmental attribute at the different buffer distances. A Pearson correlation matrix was also output. Included in the matrix was each environmental attribute at the scale with the greatest significant model fit (based on the multiple single logistic regressions). Five of the "best fitted", independent environmental attributes were included in the naïve Bayesian algorithm. The naïve Bayesian algorithm was trained using 100 subsets of a training dataset (a randomly drawn 50% of the entire occurrence / pseudo-absence dataset). A composite ruleset was derived, tested with the training, testing and overall occurrence / pseudo-absence dataset, and then applied to PPNP as a whole to map potential niche of the species.

Results

The modelling method proposed here first used logistic regression to examine the optimal scale of each environmental parameter by quantifying the environmental parameter at each of a set number of buffer distances that best distinguishes between occurrences and pseudo-absences. The optimal buffer distances ranged from 0m (or point values, as with landuse and elevation) to 700m (as with the proportion of the buffer zone that was ponds and roads/parking) (Table 5.1). The model fit ranged from 1.3% (Nagelkerke's R^2; proportion of the buffer zone that is juniper stand) to 50.6% (proportion of the buffer zone that is roads or parking). The landuse point values (the landuse category in which the occurrence or pseudo-absence occurred) explained only 42.5% of the variation, however it was the dominant classifier in the naïve Bayesian algorithm.

Results of the ruleset derived by training the naïve Bayesian algorithm with 100 spatially independent subsets are shown in Table 5.2. Only landuse attributes, point values and the proportion of the surrounding buffer zone with respect to beach, human use, ponds and roads/parking, were used to train the naïve Bayesian classifier as these were independent environmental attributes with the greatest model fit (as determined by examining Pearson correlation coefficients and Nagelkerke's R^2). The *a posteriori* probability of occurrence was greatest when locations were within primary or secondary successional savannas ($p = 0.568$ and 0.425, respectively; Table 5.2). The *a posteriori* probability of an absence was greatest when locations were within marsh, pond or forest classifications ($p = 0.502$, 0.198 and 0.163, respectively).

The accuracy of the model with the training dataset was 93.1% overall, with 3.3% false negatives and 11.5% false positives (Table 5.3). The testing dataset had a little higher accuracy, 94% overall, 3.5% false negatives and 8.5% false positives. As expected, when both training and testing datasets are put together, the composite contingency table shows an overall accuracy of 93.6%, with 3.9% false negatives and 9.8% false positives.

The model was applied to the entirety of PPNP (Fig. 5.2). The highest probability that any location was part of the species' potential niche occurred in the primary and secondary successional savannas. Within these landuse classifications, the probability of potential niche decreased with increasing proportion of the surrounding landscape (attribute specific buffer zone) that was classified as ponds and decreasing proportion comprised of beach, human use and roads / parking (Table 5.2).

Discussion

The framework for modelling species distributions presented here, in essence, describes aspects of a species' distribution in the Hutchinsonian n-dimensional niche space (Hutchinson 1957). Here, environmental attributes of the realized niche were quantitatively characterized, and then for any location within PPNP, the probability that the location is part of the species potential niche was estimated.

In modelling a species' niche and its distribution with respect to future environmental change (climate, habitat fragmentation, etc.), besides assuming a pseudo-equilibrium, it is often assumed that the realized niche, in terms of its abiotic characteristics, is a good approximation of the fundamental niche (Vetaas 2002). This assumption may not be valid. Indeed, it has been suggested that the "true" way to determine a species fundamental niche is through direct measurement of physiological responses of organisms to abiotic environmental conditions (e.g., gradients of temperature, humidity) and inferring fitness from combinations of these conditions (Soberón and Peterson 2005). Then the investigator can map areas of positive fitness using GIS technology (see e.g., Porter et al. 2000, Porter et al. 2002).

Soberón and Peterson (2005) proposed that a second method for estimation of a species' fundamental niche is to relate species occurrences with geo-referenced databases (e.g., GIS layers of climate, topography, soil characteristics). Combinations of abiotic conditions that best describe occurrences are projected across the landscape to describe the limits of the fundamental niche of a species (e.g., Lindenmayer et al. 1991, Sykes et al. 1996, Peterson et al. 1999, Peterson 2001, Kearney and Moussalli 2003, Kearney et al. 2003). This simple 'correlative approach' is argued to provide insight into the fundamental niche of a species (Peterson et al. 1999, Peterson 2001, Peterson and Holt 2003, Soberón and Peterson 2005). It has been suggested that, if the distribution of a species in the context of varied biotic backgrounds can be observed, the composite fundamental niche can thereby be described (Peterson et al. 1999, Peterson 2001).

Opuntia humifusa is used here as an example to demonstrate and validate the applicability of the SDM methodology presented. While this method could be used to examine the fundamental niche of a species, that was not the intended purpose here. Rather, this application was used to describe the realized niche and delineate the potential niche, suggesting unsurveyed sites with high potential of occurrence and identifying suitable sites for re-introduction. The potential niche of *O. humifusa* was accurately predicted (Table 5.3) using single logistic regressions to identify the most significant predictors of species occurrences (based on regression fit), followed by modelling using a naïve Bayesian classifier.

Many different methods have been proposed to model species distributions: expert opinion models (e.g., Burgman et al. 2001, Kuhnert et al. 2005), bioclimatic envelopes (e.g., Nix 1986, Walther et al. 2004), algorithms such as logistic regression (see Keating and Cherry 2004), generalized linear models (e.g., Ferrier et al. 2002), classification and regression trees (e.g., Breiman et al. 1984, Bourg et al. 2005), genetic algorithms (e.g., Stockwell and Peters 1999, Peterson 2001) and Bayesian logic (e.g., Bayliss et al. 2005, Mac Nally 2005). They have all been used to determine the unique attributes of locations where a species occurs and have particular benefits and limitations. Many of the statistical methods (e.g., generalized linear models) are restricted by parametric assumptions and linear interactions (Bourg et al. 2005), while bioclimatic envelopes and machine learning techniques (e.g., Bayesian classifiers, genetic algorithms) can overestimate a species niche (Peterson 2001).

Logistic regression has become increasingly popular for modelling species distributions (Keating and Cherry 2004). However, Keating and Cherry (2004) note that such models may be limited in appropriateness of the interpretations due to the choice of sampling design, underlying probability model and associated assumptions. The methodology proposed here first uses logistic regression to examine the optimal scale of each environmental parameter. It does this by quantifying the environmental attributes of a location at each of a set number of buffer distances; logistice regression was used to determine the buffer distance that best distinguishes the environmental attributes of an occurrence from a pseudo-absence. Logistic regression was not used to examine the probability of occurrence (or of potential niche) with respect to the multiple environmental attributes (each at the optimal buffer size) as the assumption of randomness may not be valid, and also due to the difficulty of integrating the results of multiple logistic regression analyses based on different subsets of the data. Many species occurrences are gathered in areas traversed by people; the distribution of occurrences may not be a random sample of all possible habitats. Furthermore, using multiple logistic regressions with spatially independent subsets of the data will yield a unique set of models equal in number to that of the spatially independent subsets used. Integration of the models into one would be difficult to justify. Thus, a naïve Bayesian algorithm was used such that the prediction of the probability of a location being part of a species' niche could be made without the limitations of the logistic regression.

Naïve Bayesian classifiers have been used in a wide variety of ecological / environmental applications, such as species niche modelling (e.g., Milne et al. 1989, Kuhnert et al. 2005), landcover and soil classifications (e.g., Mac Nally 2005, Porwal et al. 2006), reconstructing ancestral genealogies (e.g., Smith et al. 2005, Hardy 2006) and in other applications such as spam filtering (e.g., Delany et al 2005). A major advantage of Bayesian classifiers is that the attribute variables may be binary or multistate categorical or continuous, essentially including all quantitative or qualitative information (Ames 2002, Sadoddin et al. 2005). This is especially useful for resource management and species distribution models in that quantitative and qualitative (e.g., expert opinion, social views) can be incorporated into the classifier in its application as a decision support system (Sadoddin et al. 2005).

Many SDMs, especially bioclimatic envelope models, have no formal means of representing uncertainty (Burgman et al. 2005). However, Bayesian classifiers are capable of representing and considering uncertainty (Saddodin et al. 2005). Bayesian classifiers provide a formal reasoning about partial beliefs under conditions of uncertainty and conditional probabilities between different knowledge components represent uncertainty (Varis and Kuikka 1999). In the present application, the model provides the probability that a specified location is within the potential niche of *O. humifusa.* By applying probability theory (in using a Bayesian classifier), uncertainty is represented by the probability.

Recognition that species perceive different environmental cues at different resolutions and that the species-environment relationship is best observed at different scales has become increasingly important concept in much of landscape ecology (Turner et al. 2001, Holland et al. 2004, Guisan and Thuiller 2005). The method presented here is not limited to examining the species-environment relationship at a single scale. Many SDMs are limited in that quantification of the species-environment relationship is done at the scale of the patch, or some arbitrary buffer distance from where the species occurs (Holland et al. 2004). It is difficult to determine the appropriate scale at which the species-environment relationship occurs; different environmental attributes of species occurrences may be best observed at different scales (at different grain or extent) (Turner et al. 2001). For example, Holland et al. (2004) examined the species-environment relationship with respect to 12 cerambycid beetle species and found that abundance was correlated to the amount of forest cover at different buffer distances from 20 to 2000m, depending on the species. The method presented here examines the species-environment relationship at multiple scales (grains or buffer distance) to determine the "optimal" scale distinguishing a species occurrence from absence.

The methodology presented here further uses multiple subsets of spatially independent data in both the logistic regression and naïve Bayesian training. Holland et al. (2004) noted that this type of procedure (multiple single analyses of subsets of the data) have, in general, four advantages over random selection of a single set of independent sites (in this application, occurrence / pseudo-absence locations). These include:

1. due to the multiple regressions, the sites included in the subset are not affected by the first site chosen;
2. because different buffer sizes use different subsets of the possible sites, sites at different buffer sizes are not nested;
3. it increases the power of the analysis since this method uses multiple estimates of the regression (using different subsets of the data); and
4. it maximizes the use of the available data.

Two general types of error are generated with predictive niche models: errors of omission and commission (Fielding and Bell 1997), akin to false positive and false negatives of a contingency table. The first, a false negative, or the omission of areas where the species occurs represents a failure of the model to include the full realized ecological niche of a species. Here, the contingency table showed 3.9% of the occurrences were predicted to be absent (false negatives; Table 5.3). The areas in which these 10 (out of 345) individual plants are found are secondary successional savannas close to the ponds, or with few roads. The deviation from the model potentially represents a uniqueness of these individuals with respect to some biotic interactions. Indeed, six of the 10 individuals, although believed to be natural, are located in an area that was previously an orchard.

The second type of error, false positives or errors of commission, represents areas that are recognized as suitable but where there is no occurrence. There are two aspects to the error of commission: the model may have failed, and incorrectly predicted areas that are not part of the species' niche; or the area is part of the species' niche (within its potential niche) but either the species is there and has not been reported, or the species does not occur there due to historical factors (dispersal limitation, local extinction, etc.) or interspecific interactions (competition, predation, etc.) (Peterson et al. 1999). With respect to the *O. humifusa* model, 34 of the 345 pseudo-absences were predicted as part of the species niche. They all occurred within primary and secondary successional savannas. Ideally, these false positives are potential habitat of the species and thus the accuracy of the model could be reported as 97.1% (rather than 93.6% accounting for false positives). Indeed, some of the predicted niche is presently used at PPNP to "farm" *O. humifusa* for restoration projects (VanDerWal et al. 2005a,b).

The modeling methodology presented here allows one to examine aspects of a species niche by creating a null distribution model. The framework searches for non-random associations between environmental attributes (at different scales) of known occurrences compared with those of the study areas (pseudo-absences). The deviation from the null predictions (based on occurrences in the absence of life history information) may represent important factors (biotic interactions, niche limitations, etc.) defining species distributions. Inference with respect to deviation from the null may be more useful in multi-species applications.

Guissan and Thuiller (2005) suggest that one future challenge for SDM is to examine applications with respect to groups or guilds of species. Limiting such research may be the niche assumption underlying many SDMs. Assuming each species has its own "niche", models are often only built for single species, not groups or guilds of species since there is no ecologically justifiable hypothesis for modeling groups (Guissan and Thuiller 2005; but see Hubbel 2001, 2005). Each species is assumed to have its own "niche". Spatial predictions of groups of species could be examined with respect to determining whether one group of species can be better modelled than others (Boone and Krohn 2002, Huntly et al. 2004) and the likely ecological reasons for such patterns (Guissan and Thuiller 2005). From an applied conservation perspective, this approach could be used for potential prioritization of areas (potential niche) for groups of rare species rather than simply for single species.

References

Ackerly, D.D., 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. International Journal of Plant Science 164: S165-S184.

Ames, D. 2002. Bayesian decision networks for watershed management, Ph.D. Thesis, Civil and Environmental Engineering Department, Utah State University, Logan, UT.

Bayliss, J.L., Simonite, V. & Thompson, S., 2005. The use of probabilistic habitat suitability models for biodiversity action planning. Agriculture, Ecosystems and Environment 108: 228-250.

Boone, R.B. & Krohn, W.B., 2002. Modeling tools and accuracy assessment. In: Predicting Species Occurrences: Issues of Accuracy and Scale. pp. 265-270 In Scott, J.M., Heglund, P.J., Samson, F., Haufler, J., Morrison, M., Raphael, M. & Wall, B. (eds). Island Press, Covelo, California.

Bourg, N.A., McShea, W.J. & Gilla, D.E., 2005. Putting a cart before the search: successful habitat prediction for a rare forest herb. Ecology 86: 2793-2804.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.G., 1984. Classification and Regression Trees. The Wadsworth Statistics/Probability Series. Chapman and Hall, New York, NY.

Brown, D.G., 1994. Predicting vegetation types at treeline using topography and biophysical disturbance variables. Journal of Vegetation Science 5: 641-656.

Burgman, M.A., Breininger, D.R., Duncan, B.W. & Ferson, S., 2001. Setting reliability bounds on habitat suitability indices. Ecological Applications 11: 70-78.

Burgman, M.A., Lindenmayer, D.B. & Elith, J., 2005. Managing landscapes for conservation under uncertainty. Ecology 86: 2007-2017.

Delany, S.J., Cunningham, P. & Coyle, L., 2005. An assessment of case-based reasoning for spam filtering. Artificial Intelligence Review 24: 359-378.

Ferrier, S., Watson, G., Pearce, J. & Drielsma, M., 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. Biodiversity and Conservation 11: 2275-2307.

Fielding, A.H. & Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24: 38-49.

Goovaerts, P. & Jacquez, G., 2004. Accounting for regional background and population size in the detection of spatial clusters and outliers using geostatistical filtering and spatial neutral models: the case of lung cancer in Long Island, New York. International Journal of Health Geographics 3: 14-37.

Goovaerts, P., Jacquez, G.M., & Greiling, D., 2005. Exploring scale-dependent correlations between cancer mortality rates using factorial kriging and population-weighted semivariograms. Geographical Analysis 37: 152-182.

Guisan, A. & Theurillat, J.-P., 2000. Equilibrium modeling of alpine plant distribution: how far can we go? Phytocoenologia 30: 353-384.

Guisan, A. & Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8: 993-1009.

Guisan, A. & Zimmerman, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135: 147-186.

Hardy, C.R., 2006. Reconstructing ancestral ecologies: challenges and possible solutions. Diversity and Distributions 12: 7-19.

Holland, J.D. Bert, D.G. & Fahrig, L., 2004. Determining the spatial scale of species' response to habitat. Bioscience 54: 227-233.

Hubbell, S.P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton, NJ.

Hubbell, S.P., 2005. Neutral theory in community ecology and the hypothesis of functional equivalence. Functional Ecology 19: 166-172.

Huntley, B., Green, R.E., Collingham, Y.C., Hill, J.K., Willis, S.G., Bartlein, P.J. , Cramer, W., Hagemeijer, W.J.M. & Thomas, C.J., 2004. The performance of models relating species geographical distributions to climate is independent of trophic level. Ecology Letters 7: 417-426.

Hutchinson, G.E., 1957. Concluding remarks. Cold Spring Harbor Symposium on Quantitative Biology 22: 415-427.

Johnson, C.J., Seip, D.R. & Boyce, M.S., 2004. A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales. Journal of Applied Ecology 41: 238-251.

Kearney, M. & Moussalli, A., 2003. Geographic parthenogenesis in the Australian arid zone. II. Climatic analyses of orthopteroid insects of the genus *Warramaba* and *Sipyloidea*. Evolutionary Ecology Research 5: 977-997.

Kearney, M., Moussalli, A., Strasburg, J., Lindenmayer, D. & Moritz, C., 2003. Geographic parthenogenesis in the Australian arid zone. I. A climatic analysis of the *Heteronotia binoei* complex (Gekkonidae). Evolutionary Ecology Research 5: 953-976.

Kearney, M. & Porter, W.P., 2004. Mapping the fundamental niche, physiology, climate and the distribution of a nocturnal lizard. Ecology 85: 3119-3131.

Keating, K.A. & Cherry, S., 2004. Use and interpretation of logistic regression in habitat-selection studies. Journal of Wildlife Management 68: 774-789.

Klinkenberg, B. & Klinkenberg, R., 1985. Status on the Eastern Prickly Pear Cactus *Opuntia humifusa* in Canada. Committee on the Status of Endangered Wildlife in Canada, Ottawa, ON.

Kraus, D., 1991. Eastern Prickly Pear Cactus management plan. Canada Parks Service: Point Pelee National Park, Leamington, ON.

Kuhnert, P.M., Martin, T.G., Mengersen, K. & Possingham, H.P., 2005. Assessing the impacts of grazing levels on bird density in woodland habitat: a Bayesian approach using expert opinion. Environmetrics 16: 717-747.

Leibold, M.A., 1995. The niche concept revisited: mechanistic models and community context. Ecology 76: 1371-1382.

Lindenmayer, D., Nix, H., McMahon, J.P., Hutchinson, G.E. & Tanton, M.T., 1991. The conservation of Leadbeater's possum, *Gymnobelideus leadbeateri* (McCoy): a case study of the use of bioclimatic modeling. Journal of Biogeography 18: 371-383.

Mac Nally, R., 2005. Ecological boundary detection using Carlin-Chib Bayesian model selection. Diversity and Distributions 11: 499-508.

Milne, B.T., Johnston, K.M. & Forman, R.T.T, 1989. Scale-dependent proximity of wildlife habitat in a spatially-neutral Bayesian model. Landscape Ecology 2: 101-110.

Nix, H.A., 1986. A biogeographic analysis of Australian elapid snakes, p. 4-15. In Bureau of Flora and Fauna [eds.], Atlas of Australian elapid snakes. Bureau Flora Fauna, Canberra, Australia.

Peterson, A.T., 2001. Predicting species' geographic distributions based on ecological niche modeling. Condor 103: 599-605.

Peterson, A.T. & Holt, R.D., 2003. Niche differentiation in Mexican birds: Using point occurrences to detect ecological innovation. Ecology Letters 6: 774-782.

Peterson, A.T., Soberón, J. & Sánchesz-Cordero, V., 1999. Conservatism of ecological niches in evolutionary time. Science 285: 1265-1267.

Porter, W.P., Budaraju, S., Stewart, W.E. & Ramankutty, N., 2000. Calculating climate effects on birds and mammals: Impacts on biodiversity, conservation, population parameters, and global community structure. American Zoologist 40: 597-630.

Porter, W.P., Sabo, J.L., Tracy, C.R., Reichman, O.J. & Ramankutty, N., 2002. Physiology on a landscape scale: Plant-animal interactions. Integrative and Comparative Biology 42: 431-453.

Porwal, A., Carranza, E.J.M. & Hale, M., 2006. Bayesian network classifiers for mineral potential mapping. Computers & geosciences 32: 1-16.

Rushton, S.P., Ormerod, S.J. & Kerby, G., 2004. New paradigms for modeling species distributions? Journal of Applied Ecology 41: 193-200.

Sadoddin, A., Letcher, R.A., Jakeman, A.J. & Newham, L.T.H., 2005. A bayesian decision network approach for assessing the ecological impacts of salinity management. Mathematics and Computers in Simulation 69: 162-176.

Smith, S.A., Stephens, P.R. & Wiens, J.J., 2005. Replicate patterns of species richness, historical bigeography, and phylogeny in holarctic treefrogs. Evolution 59: 2433-2450.

Soberón, J. & Peterson, A.T., 2005. Interpretation of models of fundamental ecological niches and species' distributions. Biodiversity Informatics 2: 1-10.

Stockwell, D.R.B. & Peters, D.P., 1999. The GARP modelling system: problems and solutions to automated spatial prediction. International Journal of Geographic Information Science 13: 143-58.

Sykes, M.T., Prentice, I.C. & Cramer, W., 1996. A bioclimatic model for the potential distributions of north European tree species present and future climates. Journal of Biogeography 23: 203-233.

Turner, M.G., Gardner, R.H. & O'Neill, R.V., 2001. Landscape ecology: in theory and practice. Springer-Verlag New York, Inc. New York, NY.

VanDerWal, J.J., Wozniczka, I. & Lovett-Doust, L., 2005a. Report to Parks Canada: Demography of *Opuntia humifusa* at Point Pelee National Park. Canada Parks Service: Point Pelee National Park, Leamington, ON.

VanDerWal, J.J., Sahar Moeen, M. & Lovett-Doust, L., 2005b. Report to Parks Canada: Defining habitat requirements of *Opuntia humifusa* as determined by field and lab studies. Canada Parks Service: Point Pelee National Park, Leamington, ON.

Varis, O. & Kuikka, S., 1999. Learning Bayesian decision analysis by doing: lessons from environmental and natural resources management. Ecological Modelling 119: 177-195.

Vetaas, O.R., 2002. Realized and potential climate niches: a comparison of four Rhododendron tree species. Journal of Biogeography 29: 545-554.

Walther, B., Wisz, M. & Rahbek, C., 2004. Known and predicted African winter distributions and habitat use of the endangered Basra reed warbler (*Acrocephalus griseldis*) and the near-threatened cinereous bunting (*Emberiza cineracea*). Journal of Ornithology 145: 287-299.

Whitehead, G., 1995. 1995 eastern prickly pear cactus monitoring report: prepared for Natural Resource Conservation Point Pelee National Park. Department of Environment and Resource Studies, University of Waterloo. Waterloo, ON, CA.

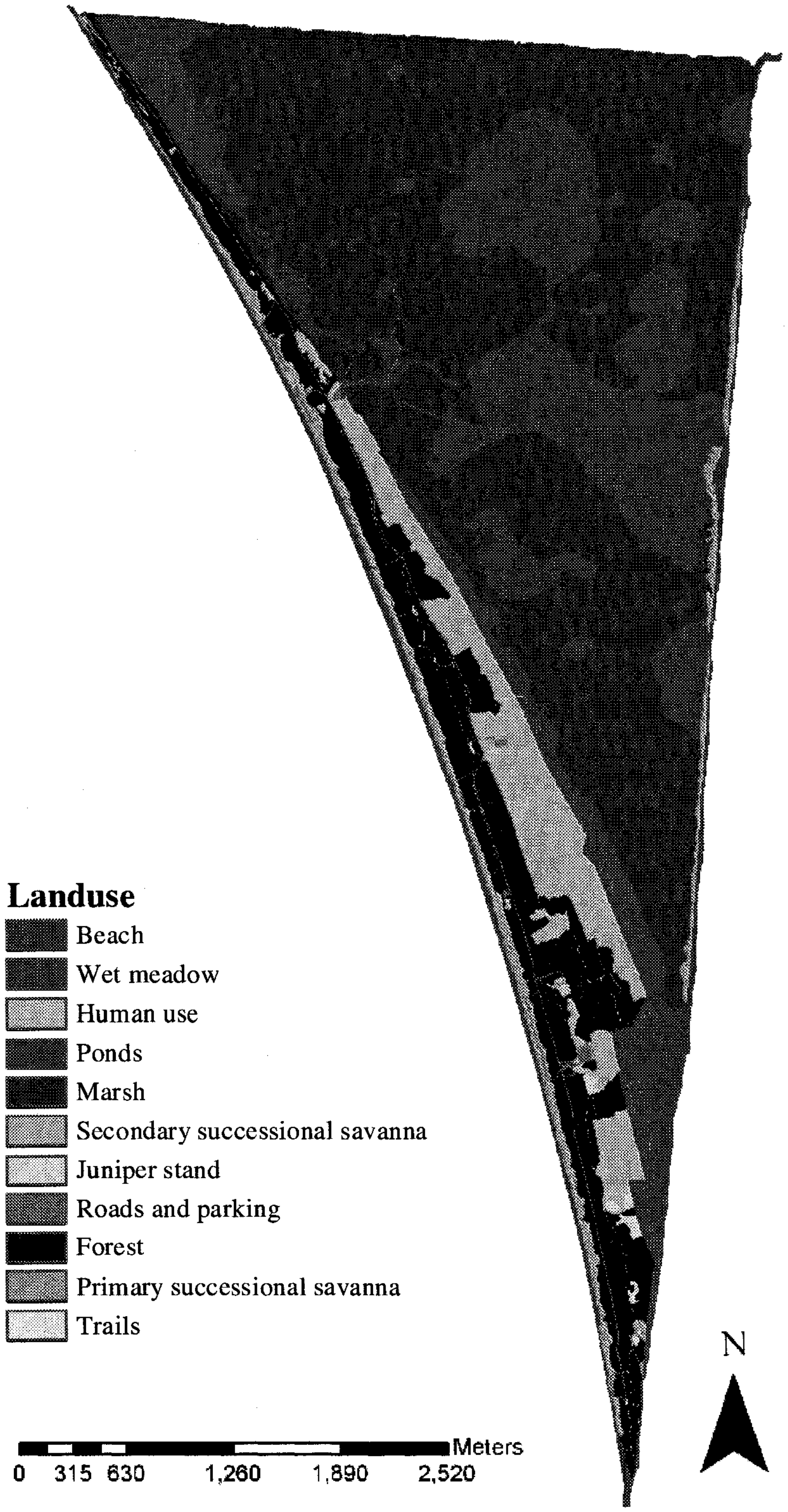


Figure 5.1. Landuse classification of Point Pelee National Park, Ontario, supplied by park staff.

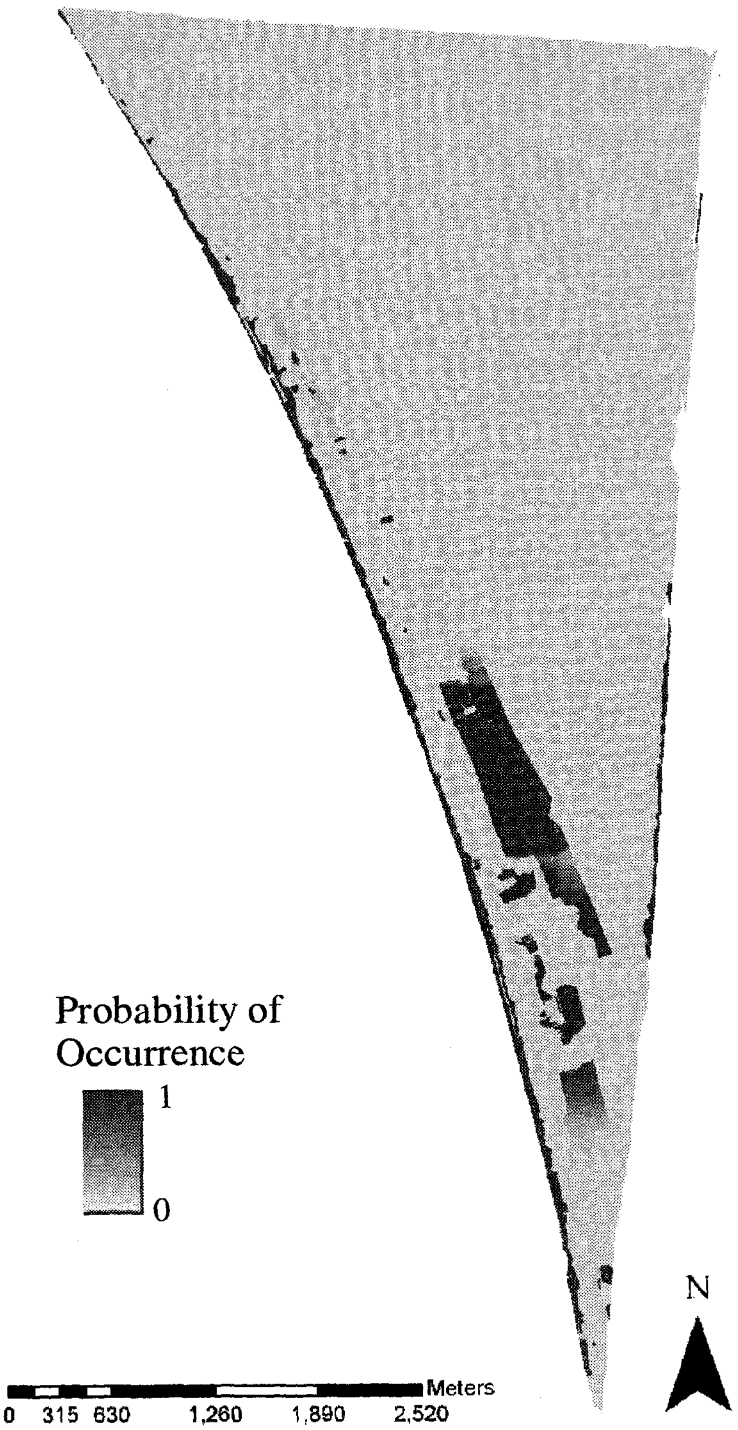


Figure 5.2. Potential niche map for *Opuntia humifusa* at Point Pelee National Park, Ontario. The predicted potential niche is based upon a naïve Bayesian classifier using occurrence information and relating it to environmental attributes.

Table 5.1. Logistic regression results describing the species-environment relationship at the buffer distance that best distinguishes occurrences from pseudo-absences. Landuse attributes includes both the point values as well as the results based on the proportion of the buffer zone that is each landuse category. With respect to DEM and the aerial photo, the means and associated variance for the buffer zones were regressed against the occurrence information from an intensive field survey for *Opuntia humifusa* individuals in July-August 2004 (VanDerWal et al. 2005a).

		Buffer distance (m)	Nagelkerke's R^2		Sig.	
			Mean	SE	mean	SE
Landuse	beach	50	0.272	0.0286	<0.001	<0.0001
	forest	300	0.398	0.0006	0.001	<0.0001
	human use	500	0.412	0.0044	0.041	0.0014
	juniper stand	10	0.013	<0.001	0.046	<0.0001
	marsh	300	0.416	0.0034	0.001	<0.0001
	ponds	700	0.389	0.0028	0.040	0.0002
	primary successional savanna	10	0.428	<0.001	<0.001	<0.0001
	road and parking	700	0.506	0.0058	0.047	0.0014
	secondary successional savanna	500	0.369	0.0021	0.022	0.0002
	trails	300	0.272	0.0004	0.006	0.0001
	wet meadow	150	0.091	0.0007	0.008	<0.0001
	beach	50	0.272	0.0286	<0.001	<0.0001
DEM	mean	0	0.111	0.0001	<0.001	<0.0001
	SD	50	0.253	0.0000	<0.001	<0.0001
Aerial photo	mean	200	0.184	0.0002	0.002	<0.0001
	SD	40	0.245	0.0000	<0.001	<0.0001

Table 5.2. The ruleset derived by training a naïve Bayesian algorithm using the independent environmental attributes. Only landuse attributes were used; these included both the point values as well as the proportion of the buffer zone that is beach, human use, ponds and roads / parking at the appropriate buffer size listed in Table 5.1. The point values are conditional probabilities while proportionate landuse attributes report the mean and SD (in brackets) used to calculate the condition probabilities.

		Probability of	
		occurrence	absence
a priori probability		0.15	0.85
Point values	beach	0	0
	forest	0	0.163
	human use	0.007	0
	juniper stand	0	0.006
	marsh	0	0.502
	ponds	0	0.198
	primary successional savanna	0.568	0.001
	road and parking	0	0.001
	secondary successional savanna	0.425	0.037
	trails	0	0.016
	wet meadow	0	0.078
Proportion of the buffer zone	beach	11.55 (11.374)	0.88 (3.084)
	human use	1.65 (0.843)	0.26 (0.502)
	ponds	1.51 (2.338)	21.27 (18.470)
	roads / parking	2.15 (0.723)	0.66 (0.772)

Table 5.3. Contingency tables describing the accuracy of the modelling using the training, testing and overall (training and testing) datasets. For actual and expected values, 1 represents an occurrence and 0 an absence.

Training dataset

		Expected		
		1	0	% correct
Actual	1	170	4	97.7%
	0	18	139	88.5%
	Overall %			93.1%

Testing dataset

		Expected		
		1	0	% correct
Actual	1	165	6	96.5%
	0	16	173	91.5%
	Overall %			94.0%

Total dataset

		Expected		
Actual		1	0	% correct
	1	335	10	97.1%
	0	34	312	90.2%
	Overall %			93.6%

Chapter 6 – Applying a species distribution model testing the "neutrality" of rare species within southwestern Ontario, Canada.

Summary

The concept of the niche exists as a part of the underlying paradigm in most species distribution models. In general, quantifed environmental correlates of species occurrences have been taken to characterize a species' niche. If each species is assumed to have its own unique niche, modeling the distribution of groups of ecologically similar species is ecologically unjustifiable. The "null" alternative to the niche concept presumes that all species are functionally equivalent (e.g., have the same niche and the same demographic rates, etc.). Given this assumption, distribution models can be developed for groups of species such that deviations from such a null model could be used to describe the potential niche differentiation within the group of species.

Species distribution models were created for each of four broad groups of rare biota (birds, insects, plants and reptiles). Subsequently, the rare plants were further subdivided: first into trees, shrubs, herbs, grasses, and sedges; then (to examine aspects of biological rarity) the rare plants were grouped based upon their designated species rarity (subnational rank = S1, S2 and S3). The rare herbs were also grouped in this way, based upon rarity; finally, rare plants were grouped into aquatic and terrestrial categories. The models pertaining to each of these groups were created using a Bayesian analystic approach (as described in Chapter 5). Inputs included occurrence / pseudo-absence information, and environmental attributes (slope, aspect, elevation, distance-to-road,

building density, land classification for agriculture, soil texture and landcover). The species-environment relationship was examined at a series of buffer distances (or scales), ranging from 30m to *c.* 10 km. Despite large species-specific differences in life history, resource utilization, etc., in the broad distinctions of biotic groupings, distribution models performed relatively well for birds, insects, reptiles, sedges, aquatic and terrestrial plants (< 27% false negatives). Dominating all of the models was the immediate landcover type and the proportion of the surrounding landscape that was cropland. The probability of occurrence decreased with increasing cropland in the area surrounding, and the probability of an absence at a location that was cropland was approximately two times more likely than an occurrence.

These results provide support for the neutral assumption. Implications of such results are discussed in an applied conservation perspective, since such an approach could be used in potential prioritization of areas for conservation / restoration of groups of rare biota rather than for single species. For example, results suggest that by conserving a selected 21% of the regional landscape, one would be conserving the potential niche of both 77% of rare aquatic plants and 74.5% of rare terrestrial plants of SW Ontario. While these results supported the neutral assumption, it was not completely supported by any biotic group. Deviations from the null model potentially describe the niche differentiation within a group (e.g., aquatic vs. terrestrial, open field vs. closed forest). Finer scale studies, both in terms of biotic groups and study area, may further clarify the relative importance of niche differentiation and its null alternative in determining the distribution of species.

Introduction

Species distribution models (SDM) often attempt to determine the non-random associations between environmental attributes of known occurrences and those of absences (Pearce and Boyce 2005). Kearney and Porter (2004) suggested that the concept of the niche provides an important and useful starting point for understanding the distribution of species. Indeed, many SDMs assume an underlying niche concept (Guisan and Thuiller 2005). The unique attributes of locations where a species occurs are often inferred as characteristics of a species' niche and, thus a common approach used to determine the extent of a species niche (a species' geographic distribution) is to characterize, quantitatively, a suite of environmental conditions for known occurrences of the species (Peterson 2001, Kearney and Porter 2004).

Surveys of species often only report occurrences of species; rarely are absences recorded and thus the methods for modelling species distributions are limited (Pearce and Boyce 2005). One way to model presence-only data is to use pseudo-absences (locations not sampled in which species are assumed to be absent). Although pseudo-absences are assumed to be true absences (locations where the species does not occur), they may represent presence locations (occurrence at that location has not been reported) (Graham et al. 2004). In some models, pseudo-absences are chosen at random to represent the background set of environmental attributes of the study area (Pearce and Boyce 2005, Soberón and Peterson 2005). Statistical models of the probability of occurrence can be derived for the species-environment relationship using occurrence / pseudo-absence; and then these models can be interpolated to other parts of the landscape, potentially under differing environmental conditions (see e.g., Lindenmayer et al. 1991, Sykes et al. 1996, Peterson et al. 1999, Peterson 2001, Kearney and Moussalli 2003, Kearney et al. 2003).

Guissan and Thuiller (2005) suggest that one challenge for SDMs is to examine applications with respect to larger groups of taxa or guilds of species. Limiting such research may be the underlying niche assumption of many of SDMs. A problem with assuming an underlying niche concept is that no ecologically justifiable hypothesis emerges for modeling groups of species, since each species is assumed to have its own "niche". Neutral theories such as MacArthur and Wilson's (1967) theory of island biogeography and Hubbell's (2001) neutral theory of macroecology may offer a justifiable assumption to model distributions of groups of species.

MacArthur and Wilson (1967) assumed island species richness was a function of the mainland richness, the size of the island and distance from the mainland; species niche requirements and life history characteristics were ignored. More recently, Hubbell (2001) proposed that "ecologically similar" species can be assumed functionally equivalent. Ecologically similar species are those in which the individuals may potentially compete with one another, not accounting for trophic interactions (Bell 2001).

Hubbell's neutral theory assumes that individuals, independent of species, may differ in many ways (size, shape, color, etc.), but they do no differ demographically with respect to vital rates of birth, death, dispersal and speciation, or in the way they interact with their environment (no habitat specificity – all have the same niche requirements) (Hubbell

2001, 2005, Bell 2001, Ricklefs 2003). While all species violate this assumption to some degree, the question posed by neutral models is, how good is this approximation? This counter-intuitive assumption has proven controversial (see discussions in e.g., Zhang and Lin 1997, Yu et al. 1998, Hubbell 2001, Enquist et al. 2002, Chase and Leibold 2003, Hubbell and Lake 2003, Ricklefs 2003, Chave 2004, Poulin 2004, Hubbell 2005), however, community assemblages and species distributions have been accurately modelled based on this assumption (see review in Chave 2004, and see e.g., Dolman and Blackburn 2004, Tilman 2004, Bell 2005, He 2005, Rangel and Diniz-Filho 2005).

In this chapter, species distribution models were created for various groups of rare species such that the efficacy of the ecological equivalence assumption could be tested, specifically in the way species interact with their environment. Further, spatial predictions of groups of species were examined with respect to determining if some groups of species can be more accurately modelled than others. Finally, the implications in an applied conservation perspective are discussed, as this approach could be used for potential prioritization of areas of potential niche for conservation / restoration efforts for groups of rare species rather than for single species.

Methods

Environmental attributes and species occurrence data

Species occurrences, in the form of a location representing the centroid of a population, were provided by the Natural Heritage Information Center (NHIC; http://nhic.mnr.gov.on.ca/) for south-western Ontario. NHIC is a branch of the Ontario

Ministry of Natural Resources that compiles, maintains and provides information on species in Ontario that are threatened, endangered or of special concern (further described here as rare species). The rare species are prioritized using subnational ranks (S-rank), where by S1 species are extremely rare (with usually 5 or fewer occurrences in the province); S2 are very rare (with usually between 5 and 20 occurrences); S3 are rare to uncommon (with between 20 and 100 occurrences); S4 are common and apparently secure (with usually with more than 100 occurrences); and S5 are very common and demonstrably secure in Ontario (Master 1991). NHIC tracks rare species ranked S1-S3.

Environmental information was provided through Ontario Geospatial Data Exchange (OGDE; http://www.lio.mnr.gov.on.ca/ogdedescription.cfm)(digital elevation model [DEM], landcover, road and building locations) and Ontario Ministry of Food, Agriculture and Rural Affairs (soils; http://www.omafra.gov.on.ca/). In total, seven environmental attributes were used for the SDM: slope, aspect, elevation, distance-to-road, building density, land classification for agriculture, soil texture and landcover. Slope and aspect were calculated from the DEM. Distance to a road was the straight line distance from any location within the study area to the nearest road. Building density was the density of buildings per square kilometre.

The landcover classifications included water, freshwater marsh, deciduous swamp, coniferous swamp, open fen, dense deciduous forest, dense coniferous forest, coniferous plantation, deciduous mixed forest, coniferous mixed forest, sparse deciduous forest, bedrock outcrops (this included clearings for mining activity, aggregate quarries and bedrock outcrops), developed land (this included clearings for human settlement and economic activity, as well as major transportation routes), pasture (and abandoned fields), cropland and unclassified. The land classification for agriculture included seven classes: Class 0 are organic soils and were not classified for agriculture; Class 1 soils have no significant limitation in use for crops; Class 2 soils have moderate limitation that restrict the types of crops grown or require moderate conservation practices; Class 3 soils have moderately severe limitations that restrict the range of crops or require special conservation practices; Class 4 soils have severe limitations that restrict the range of crops or require special conservation practices, or both; Class 5 soils have very severe limitations that restrict their capability to producing perennial forage crops, and improvement practices are feasible; and Class 7 soils have no capability for arable culture or permanent pasture. Soil texture classes included water, bedrock, gravel, sand, sandy loam, silt loam, silt clay loam, silty clay, clay, mixed fluvial deposits and organic classifications.

All environmental data was rasterized or resampled to 30m resolution.

Species distribution model

Models of rare species occurrences were created for several groups of species. Broad taxonomic differences were first examined by grouping occurrences into birds (274 occurrences of some 75 species), insects (232 occurrences of some 146 species), plants (2904 occurrences of some 79 species) and reptiles (200 occurrences of some 20 species). While the occurrences were within the SW Ontario study area, not all species (e.g., insect species) were found within the study area. The rare plants were further subdivided: first into trees, shrubs, herbs, grasses and sedges; then to examine aspects of rarity, the rare plants were grouped by S-rank (S1, S2 and S3; n = 422, 1095 and 1356 occurrences, respectively) and the rare herbs were grouped by S-rank (S1, S2 and S3; n = 234, 740 and 564 occurrences, respectively); and finally, rare plants were grouped into aquatic and terrestrial categories (n = 304 and 514 occurrences, respectively).

Models of the niche of each biotic group were created using a Bayesian analysis approach, as described in Chapter 5. Input into the models included occurrence / pseudo-absence information, the environmental attributes (slope, aspect, elevation, distance-to-road, building density, CLI, soil texture and landcover), and a series of buffer distances ranging from 30m to nearly 10km (30, 60, 90, 150, 300, 450, 600, 900, 1200, 2400, 4800, 7200, 9600m). Here, pseudo-absences were sites randomly drawn from within the geographic limits of SW Ontario such that the number of pseudo-absences equalled that of occurrences of the group of species being modelled.

Recognizing that species may perceive different environmental cues at different scale resolutions, and that the species-environment relationship is best observed at different scales has become increasingly important concept in much of landscape ecology (Turner et al. 2001, Holland et al. 2004, Guisan and Thuiller 2005). Thus the method used here was not limited to examining the species-environment relationship at a single scale. The environmental attributes at each occurrence / pseudo-absence location were recorded; however, rather than only using the environmental data of the specific raster cell in which the occurrence / pseudo-absence is located, environmental data were also recorded at each buffer distance. In other words, the environmental conditions were quantified at the specific occurrence / pseudo-absence location and for the surrounding landscape, at the specified buffered sizes. Surrounding landscapes were treated as square delineations where the size was defined by the buffer distance (e.g., a buffer size of 1m on a grid with a resolution of 1m would describe a 3x3 cell area in which the edges are defined as 1m from the focal cell in the cardinal directions).

Quantification of the environmental attributes varied depending upon the type of data used. For categorical datasets (e.g., landcover, soil type), the point value at the occurrence / pseudo-absence location and the proportion of the surrounding landscape at each buffer distance for each category was recorded. With respect to continuous data, the point value at the occurrence / pseudo-absence location and the mean and standard error of the values representing the surrounding landscape was recorded for each buffer distance.

To determine the scale (buffer distance) that was the best predictor of a species occurrence, a series of multiple single logistic regressions was carried out on the occurrence data with each environmental parameter. At each buffer distance and for each environmental parameter, 100 spatially independent subsets of the data were used. Spatial independence here refers to the criterion that areas within which the predictor variable was measured may not overlap; in other words, the distance between any two points in a subset must be at least two times the radius (or buffer size). The mean and SE of the regression statistics (log-likelihood and Nagelkerke's R^2) for each environmental attribute at each buffer distance were used such that the optimal buffer size for each environmental parameter was determined based on significant model fit.

Based on fit of the logistic regressions and independence of predictors (determined by Pearson's correlation), five to six environmental attributes were then used in a naïve Bayesian algorithm (Porwal et al. 2006) to model the species distribution for each biotic group. The naïve Bayesian algorithm was trained using 100 subsets of a training dataset (a randomly drawn 50% of the entire occurrence / pseudo-absence dataset). A composite ruleset (i.e., a list of the *a priori* and *a posteriori* probabilities) was derived, tested with the training, testing and overall occurrence / pseudo-absence datasets, and then applied to SW Ontario as a whole to map the potential niche of the particular biotic group.

All modelling was done using Visual Basic (VB.NET, Microsoft Corporation, Redmond, WA) and the developer libraries / functionality of ArcGIS 9.0 (ESRI, Redlands, CA).

Results

Models were first constructed for the broad taxonomic groups of rare species of birds, insects, plants and reptiles. The model inputs varied for each group, but common to all models were the environmental attributes of proportion of the surrounding landscape that was cropland and the immediate landcover type (Table 6.1). The proportion of the surrounding landscape that was cropland was the greatest single predictor of occurrences, with respect to the multiple single logistic regression results (Table 6.1), such that probability of occurrence decreased with increasing cropland in the area surrounding an occurrence. Differences in the use of this attribute occurred in regard to the buffer distance that was most significant (300 - 9600m). Furthermore, the immediate landcover types for cropland dominated the *a posteriori* probabilities of the Bayesian classifier, with the probability of an absence at a location that was cropland being approximately two times more likely than an occurrence. Given that a particular location is classified as cropland, the average probability of absence was 91.1% (± 2.20 SE), whereas the average probability of an occurrence was 45.8% (± 6.58 SE) (Appendix 6.1 - 6.4).

Table 6.2 describes the accuracy of the four biotically-based models. Despite the diversity of life histories within each group, overall modeling accuracy for the groups was on average 80.8% with the greatest accuracy recorded for the birds (86.4%), followed by insects (83.3%), reptiles (81.1%) and plants (72.3%). More importantly, with respect to the proportion of false negatives (i.e., predicted absence but actual occurrence), models performed relatively well for the birds (17.4% false negatives), reptiles (24.5%) and insects (27.2%). However, 42.4% of predicted values proved to be false negatives for the plants.

The models created to examine aspects of rarity and life-form within the rare plants all included, as with the broad taxonomic groups, the proportion of the surrounding landscape that was cropland and immediate landcover types (see Tables 6.3 - 6.5). For all these models, the proportion of the surrounding landscape that was cropland was again the greatest single predictor of occurrences with respect to the multiple single logistic regressions (Tables 6.3 - 6.5). Based on the *a posteriori* probabilities, the probability of occurrence decreased with increasing cropland, in the area surrounding an occurrence (Appendix 6.5 - 6.17). The buffer distance used to describe this differed depending on the model, ranging from 90 - 4800m (Tables 6.3 - 6.5). Thus for example, the proportion of the surrounding landscape that was cropland within a 300m buffer distance was used for further modeling of S2 plants and herbs, and rare terrestrial plants, whereas a buffer distance of 4800m was used for modeling S3 plants. Again as with the models for the major taxonomic groups, the *a posteriori* probability of a specific location being cropland dominated the Bayesian classifier with respect to the immediate landcover types. Given a location is classified as cropland, the average probability of absence was 88.7% (± 0.55 SE), whereas the average probability of an occurrence was 46.6% (± 1.96 SE) (Appendix 6.5 - 6.17).

With respect to grouping the plants into trees, shrubs, herbs, grasses and sedges, besides landcover and cropland, distance to a road was an important environmental attribute used in four of the five models (Table 6.3). In three of these models, the mean distance to a road at 7200m buffer distance was used. Overall model accuracy was greatest for the sedges (81.1%) and did not differ much for the remaining groups, with accuracy ranging from 72.1 to 74.4% (Table 6.6). Differences in model accuracy were noted with respect to the number of false negatives. Sedges had the lowest proportion of false negatives at 21.1%, with trees, shrubs and grasses ranging from 32.7 to 35% and the herbs with 40.4%.

Common in five of the six models that examined levels of rarity in the plants was the use of the proportion of the surrounding landscape with sandy soil, at buffer distances between 4800 to 9600m (Table 6.4). Overall model accuracy ranged from 71.7 to 74.4% with respect to plant rarity rankings (S1, S2 or S3) and ranged from 74.9 to 78.1% with respect to herb rarity rankings (S1, S2 or S3). Similarly with the proportion of false negatives, the models for herb rarity showed greater accuracy, having, on average, 33.4% (± 0.81 SE), compared with the models for plant rarity which had, on average, 37.7% (± 2.81 SE).

The models built based on the broad "habitat" classification of aquatic or terrestrial differed in model inputs with respect to the buffer distance used for deciduous swamp, dense deciduous forest and cropland (Table 6.5). Included in the aquatic plant model was class 7 of the land classification for agriculture and included in the terrestrial plant model was the proportion of the surrounding landscape that was bedrock outcrop. Model accuracy for these models was relatively high at 79.4% (± 0.45 SE; Table 6.8), similar to that of the models for birds, insects, reptiles and herbs (Tables 6.2, 6.6). The proportion of false negatives was 23% for the aquatic plant model and 25.5% for the terrestrial plant model. The SDM was applied to the SW Ontario study area and the potential niche of the rare aquatic and terrestrial plants was mapped and is shown in Figs. 6.1 and 6.2. The overlapping potential niche of the two biotic groups is shown in Fig. 6.3.

Discussion

Examining the environmental correlates of distributions of groups or guilds of species is not uncommon in the literature (see e.g., Steger 1987, Marone 1991, Lovett-Doust and Kuntz 2001, Lovett-Doust et al. 2003, Fernández-Juricic 2004, Segurado and Araújo 2004). While such studies describe the correlative relationship, often it is not used as a predictive model of the potential niche for the group of species. Species distribution models have been further limited to modeling distributions of single species, not groups of species (Guisan and Thuiller 2005). Here, the applicability of a Bayesian SDM methodology applied to groups of rare species in SW Ontario was demonstrated. Despite the species-specific differences in life history, resource utilization, etc. in the broad distinctions of species groupings, distribution models performed relatively well for birds, insects, reptiles, sedges, aquatic and terrestrial plants (less than 27% false negatives).

Rabinowitz (1981) suggested that rare species can be classified into rarity levels based on three factors: habitat specificity, local population abundances and range size. In the absence of such information about each and every species, subnational ranks specify species rarity based primarily on the number of populations (occurrences) (Master 1991). The distribution of the number of occurrences within a study area occupied by species follows a "hollow curve" (log-normal) distribution (Brown et al. 1996) – in other words, many rare species with few common ones. Mechanisms underpinning such patterns are believed to be due to attributes of the landscape (e.g., habitat availability), of the species (e.g., habitat generality, breadth of environmental tolerances and dispersal ability) or both (Gaston 1994, 1996). In contrast, Hubbell's (2001) neutral theory suggests that rare species are not rare due to some niche-based attributes of the species but rather due to chance (and see Bell 2001, Chave 2004).

The *per capita* equivalence assumption of Hubbell's neutral theory suggests that, independent of the species, each individual in the study area is identical with respect to basic demographic parameters, including probability of dying, reproducing, speciating and dispersing. In other words, for example, each individual has the same probability of dying independent of species or its biotic and abiotic environment. Examining these probabilities at the species level, species which are rare (defined by a small number of individuals), have a higher probability to remain rare simply because the overall probability of reproduction and dispersal is lower, and the probability of extinction is higher, than that of a species with a greater number of individuals (Bell 2001, Maurer and McGill 2004). If neutral theory's oversimplifying assumption is correct, all species, independent of whether it is rare or common, should have no habitat preferences (the same niche requirements).

While all species clearly violate this assumption to some degree, the question posed by neutral models is, how good is this approximation? It is the deviation from this neutral assumption that is most interesting. In the present study, the deviation, in terms of false negatives, ranged from 17.4 to 42.4% (Tables 6.2, 6.6 - 6.8). Hubbell (2005) suggested the assumption would hold best for sessile organisms, such as plants, and least for more mobile animals. Here, the opposite was observed for the four broad taxonomic categories (birds, insects, plants and reptiles). Rare birds showed fewest false negatives and plants showed the greatest number (Table 6.2). The relatively low proportion of false negatives for bird, insect and reptile SDMs were unexpected, since these included mobile species with varied life histories that included differences in broad habitat classifications such as aquatic and terrestrial species. While the poor accuracy with respect to plant model was unexpected (compared to other taxonomic groups), the rare plants were potentially too broad of a group and thus were further reclassified by taxonomic and rarity classifications.

In examining the finer taxonomic and rarity grouping of rare plants, little improvement in model accuracy with respect to the number of false-negatives was observed, except with respect to the sedges. Improvement in model accuracy (reduction in the number of false negatives) was observed when rare plants were reclassified into aquatic or terrestrial groups (Tables 6.2, 6.6 - 6.8).

The framework used here for modelling species distributions, in essence, describes aspects of a species' distribution in a Hutchinsonian n-dimensional niche space (Hutchinson 1957). Here, environmental attributes of the realized niche for groups of rare species were quantitatively characterized and then, given a set of environmental attributes for any location, the probability that the location was part of the species potential niche was estimated. Many different methods have been proposed to model species distributions: expert opinion models (e.g., Burgman et al. 2001, Kuhnert et al. 2005), bioclimatic envelopes (e.g., Nix 1986, Walther et al. 2004), algorithms such as logistic regression (see Keating and Cherry 2004), generalized linear models (e.g., Ferrier et al.

2002), classification and regression trees (e.g., Breiman et al. 1984, Bourg et al. 2005), genetic algorithms (e.g., Stockwell and Peters 1999, Peterson 2001) and Bayesian logic (e.g., Bayliss et al. 2005, Mac Nally 2005). They have all been used to determine the unique attributes of locations where a species occurs and all have particular benefits and limitations. Many of the statistical methods (e.g., generalized linear models) are restricted by parametric assumptions and linear interactions (Bourg et al. 2005), while bioclimatic envelopes and machine learning techniques (e.g., genetic algorithms) can overestimate a species niche (Peterson 2001).

While logistic regression has become increasingly popular for modelling species distributions, such models may be limited in their suitability (Keating and Cherry 2004). The SDM methodology employed here first uses logistic regression to examine the species-environment relationship at each of a set number of buffer distances. From this, the scale at which the species-environment relationship can be best observed is determined. Logistic regression was not used to examine the probability of occurrence (or of potential niche) with respect to the multiple environmental attributes as the logistic regression assumption of randomness may not be valid, and also due to the difficulty of integrating the results of multiple logistic regression analyses based on different subsets of the data. Many species occurrences are gathered in areas traversed by people and thus the distribution of occurrences may not be a random sample of all possible habitats. Further, using multiple logistic regression with spatially independent subsets of the data will yield a unique set of models (unique dependent variables, coefficients, etc.), one model for each of the spatially independent subsets used. Integration of the unique models into a single composite model would be difficult to justify. Thus, a naïve Bayesian algorithm was used such that the prediction of the probability of a location being part of a species niche could be made without the limitations of the logistic regression.

Naïve Bayesian (NB) classifiers have been used in a wide variety of ecological / environmental applications (e.g., species niche modelling, landcover and soil classifications, reconstructing ancestral genealogies). A major advantage of Bayesian classifiers is that the attribute variables may be binary or multistate categorical or continuous, but also quantitative or qualitative information (Ames 2002, Sadoddin et al. 2005). This is especially useful for resource management and species distribution models, in that both quantitative and qualitative (e.g., expert opinion, social views) can be incorporated into the classifier in its application as a discussion support system (Sadoddin et al. 2005).

While many SDMs have no formal means of representing uncertainty (Burgman et al. 2005), Bayesian classifiers are capable of representing and considering uncertainty (Saddodin et al. 2005). Bayesian classifiers provide a formal reasoning about partial beliefs under conditions of uncertainty; conditional probabilities between different knowledge components represent uncertainty (Varis and Kuikka 1999). In the present application, the model provides the probability that a specified location is within the potential niche of the biotic group of interest. By applying probability theory (in using a Bayesian classifier), uncertainty is represented by that probability.

Furthermore, while many SDMs are limited in that quantification of the species-environment relationship is done at the scale of the patch or some arbitrary buffer distance from where the species occurs (Holland et al. 2004), the SDM methodology used here was not. It is difficult to determine the appropriate scale at which the species-environment relationship can be best observed; different environmental attributes of species occurrences may be best observed at different scales (at different grain or extent) (Turner et al. 2001, Holland et al. 2004, Guisan and Thuiller 2005). The method used here examines the species-environment relationship at multiple scales (grains or buffer distances) to determine the "optimal" scale distinguishing a species occurrence from absence.

Furthermore, the methodology employed here uses multiple subsets of spatially independent data in both the logistic regression and naïve Bayesian training. This type of procedure ensures that the sites included in the subset is not affected by the first site chosen and sites at different buffer sizes are not nested (Holland et al. 2004). The use of spatially independent subsets also increases the power of the analysis and maximizes the use of the data (Holland et al. 2004).

Turner et al. (2001) warn that modellers should "know thy model". Indeed, limiting assumptions of the model may be present and errors in the data may propagate. Here, the proportion of false positives ranged from 6.3 to 22.2% (Tables 6.2, 6.6 - 6.8). If pseudo-absences, as assumed, represent the study area with respect to all environmental conditions, this would suggest that 6.3 to 22.2% of SW Ontario would be the potential niche of these rare biotic groups. For example, with the rare birds, 9.7% of SW Ontario is suitable habitat for 82.6% of the occurrences (Table 6.2). The model developed for the birds used landcover information. The landcover classification was noted to be accurate 80-95% of the time (depending on the landcover class; Ontario Geospatial Data Exchange metadata) and NHIC occurrence locations were, in general, accurate to within 100m. This may explain why 38% of occurrences were recorded in cropland (Appendix 6.1 - 6.17). The benefit of a Bayesian algorithm is that such errors can be accounted. With the immediate lancover type defined as cropland, although the location has a 38% chance of being part of the rare bird species' potential niche, it is more than twice as likely (some 92%) chance of not. Defining the cropland as part of the biota's potential niche would be mitigated by the other environmental attributes.

Kerr and Cihlar (2004) examined the patterns and causes of species endangerment in Canada. The authors noted that the greatest proportion of rare species could be found in areas with greatest amount of agriculture. Indeed, they noted that SW Ontario was a hotspot for rare species and agriculture. SW Ontario is the most densely populated, urbanized, and intensively farmed area in Canada (Allen et al. 1990). The remnant natural areas have become highly fragmented and dominated by agriculture (Freisen et al. 1999, Henson et al. 2005); there remains less than 11% forest cover (Riley & Mohr 1994) and less than 10% and 3% of presettlement wetlands and prairie cover, respectively (Henson et al. 2005).

While Kerr and Cihlar (2004) discussed landscape fragmentation and agricultural intensity as the greatest causes of species endangerment, this study was aimed at examining attributes of the species-environment relationship to describe where rare species may occur in such an agriculturally dominated landscape. Indeed, in all models, as the proportion of landscape that was cropland increased, the probability of that location was part of the biota's potential niche decreased (Appendix 6.1 - 6.17). From an applied conservation perspective, this suggests, perhaps unsurprisingly, that agriculture negatively impacts native species and that larger parcels of natural areas should be maintained to minimize the agricultural impact.

There are further potential implications from an applied conservation perspective, specifically with respect to conservation reserve design. Soulé and Simberloff (1986) suggest that conservation reserves are typically developed with one or more of three primary biological motivations. The first is preservation of large, intact ecosystems such that ecosystem functioning is preserved (e.g., watersheds for flood control); this is difficult to do since most of the earth's surface is heavily influenced by humans (Meffe et al. 1997). The second motivation for conservation reserves is to preserve biodiversity. The focus of this is to protect areas of high species diversity. The third motivation is to protect particular species or groups of species (e.g., endangered or highly visible species). The second and third of these motivations are not obviously independent of each other. Often areas of highest species richness include many rare species (Meffe et al. 1997; and see log-normal relative abundance curves, e.g., Brown et al. 1996, Hubbell 2001).

The use of SDMs, as applied here, not only supports Hubbell's functional equivalence assumption but also may help in defining areas with the greatest probability of containing rare species of wide biotic groups. Meffe et al. (1997) note that a species-by-species approach to conservation can only secure a miniscule fraction of overall biological diversity; it may be more appropriate to take a larger-scale approach. Georgiadis and Balmford (1992) discussed how a systematic approach whereby examining the non-random clustering of species occurrences, the identification of biodiversity "hotspots" could be used for conservation prioritization. They authors describe how, for example, 16% of South Africa's land area (if selected properly) could conserve 95% of the vascular plants, and how conservation of 14 of 90 studied locations in Thailand could conserve all hawkmoth species in that country. In the present context, one could suggest that by conserving a selected 17.6% of the landscape, one would be conserving the potential niche of 77% of the rare aquatic plants; and by conserving a selected 16.4% of the landscape, one would be conserving the potential niche of 74.5% of the rare terrestrial plants in SW Ontario. However these areas are not exclusive. Indeed much of the predicted potential niche of rare aquatic and terrestrial plants overlaps (Fig. 6.3), such that by conserving a selected 21% of the landscape, one would be conserving the potential niche of both 77% of rare aquatic plants and 74.5% of rare terrestrial plants of SW Ontario.

References

Allen, G. M., Eagles, P.F.J. & Price, S.D., 1990. Conserving Carolinian Canada. University of Waterloo Press, Waterloo, ON.

Ames, D., 2002. Bayesian decision networks for watershed management, Ph.D. Thesis, Civil and Environmental Engineering Department, Utah State University, Logan, UT.

Bell, G., 2001. Neutral macroecology. Science 293: 2413-2418.

Bell, G., 2005. The co-distribution of species in relation to the neutral theory of community ecology. Ecology 86: 757-770.

Bourg, N.A., McShea, W.J. & Gilla, D.E., 2005. Putting a cart before the search: successful habitat prediction for a rare forest herb. Ecology 86: 2793-2804.

Brown, J.H., Stevens, C. & Kaufman, D.M., 1996. The Geographic Range: Size, Shape, Boundaries, and Internal Structure. Annual Review of Ecology and Systematics 27: 597-623.

Burgman, M.A., Lindenmayer, D.B. & Elith, J., 2005. Managing landscapes for conservation under uncertainty. Ecology 86: 2007-2017.

Chase, J.M. & Leibold, M.A., 2003. Ecological Niches: Linking Classical and Contemporary Approaches. University of Chicago Press, Chicago, IL.

Chave, J., 2004. Neutral theory and community ecology. Ecology Letters 7: 241-253.

Dolman, A.M. & Blackburn, T.M., 2004. A comparison of random draw and locally neutral models for the avifauna of anEnglish woodland. BioMed Central Ecology 4: 8.

Enquist, B.J., Sanderson, J. & Weiser, M.D., 2002. Modeling macroscopic patterns in ecology. Science 295: 1835-1837.

Fernández-Juricic, E., 2004. Spatial and temporal analysis of the distribution of forest specialists in an urban-fragmented landscape (Madrid, Spain): implications for local and regional bird conservation. Landscape and Urban Planning 36: 17-32.

Friesen, L., Cadman, M.D. & MacKay, R.J., 1999. Nesting success of neotropical migrant songbirds in a highly fragmented landscape. Conservation Biology 13: 338-346.

Gaston, K.J., 1994. Rarity. Chapman & Hall, London, UK.

Gaston, K.J., 1996. Species-range-size distributions: patterns, mechanisms and implications. Trends in Ecology and Evolution 11: 197-201.

Georgiadis, N. & Balmford, A., 1992. The calculus of conserving biological diversity. Trends in Ecology and Evolution 7: 321-322.

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T., 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology and Evolution 19: 497-503.

Guisan, A. & Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8: 993-1009.

Guisan, A. & Zimmerman, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135: 147-186.

He, F., 2005. Deriving a neutral model of species abundance from fundamental mechanisms of population dynamics. Functional Ecology 19:187-193.

Henson, B.L., Brodribb, K.E. & Riley, J.L., 2005. Great Lakes conservation blueprint for terrestrial biodiversity. Nature Conservancy of Canada, Ontario Ministry of Natural Resources, Queens Printer, Ottawa, ON.

Holland, J.D. Bert, D.G. & Fahrig, L., 2004. Determining the spatial scale of species' response to habitat. Bioscience 54: 227-233.

Hubbell, S.P. & Lake, J.K., 2003. The neutral theory of biodiversity and biogeography, and beyond. pp. 45-63 In T.M. Blackburn & K.J. Gaston (eds). Macroecology: Concepts and Consequences. Blackwell Publishing, Oxford, UK.

Hubbell, S.P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton, NJ.

Hubbell, S.P., 2005. Neutral theory in community ecology and the hypothesis of functional equivalence. Functional Ecology 19, 166-172.

Hutchinson, G.E., 1957. Concluding remarks. Cold Spring Harbor Symposium on Quantitative Biology 22: 415-427.

Kearney, M. & Moussalli, A., 2003. Geographic parthenogenesis in the Australian arid zone. II. Climatic analyses of orthopteroid insects of the genus *Warramaba* and *Sipyloidea*. Evolutionary Ecology Research 5: 977-997.

Kearney, M. & Porter, W.P., 2004. Mapping the fundamental niche, physiology, climate and the distribution of a nocturnal lizard. Ecology 85: 3119-3131.

Kearney, M., Moussalli, A., Strasburg, J., Lindenmayer, D. & Moritz, C., 2003. Geographic parthenogenesis in the Australian arid zone. I. A climatic analysis of the *Heteronotia binoei* complex (Gekkonidae). Evolutionary Ecology Research 5: 953-976.

Keating, K.A. & Cherry, S., 2004. Use and interpretation of logistic regression in habitat-selection studies. Journal of Wildlife Management 68: 774-789.

Kerr, J.T. & Cihlar, J., 2004. Patterns and causes of species endangerment in Canada. Ecological Applications 14: 743-753.

Lindenmayer, D., Nix, H., McMahon, J.P., Hutchinson, G.E. & Tanton, M.T., 1991. The conservation of Leadbeater's possum, *Gymnobelideus leadbeateri* (McCoy): a case study of the use of bioclimatic modeling. Journal of Biogeography 18: 371-383.

Lovett-Doust, J. & Kuntz, K., 2001. Land ownership and other landscape-level effects on biodiversity in southern Ontario's Niagra Escarpment Biosphere Reserve, Canada. Landscape Ecology 16: 743-755.

Lovett-Doust, J., Biernacki, M., Page, R. Chan, M., Natgunarajah, R. & Timis, G., 2003. Effects of land ownership and landscape-level factors on rare-species richness in natural areas of southern Ontario, Canada. Landscape Ecology 18: 621-633.

MacArthur, R.H. & Wilson, E.O., 1967. The Theory of Island Biogeography. Princeton University Press, Princeton, N.J.

Marone, L., 1991. Habitat features affecting bird spatial distribution in the Monte Desert, Argentina. Ecologia Austal 1: 77-86.

Master, L.L., 1991. Assessing threats and setting priorities for conservation. Conservation Biology 5: 559-563.

Maurer, B.A. & McGill, B.J., 2004. Neutral and non-neutral macroecology. Basic and Applied Ecology 5: 413-422.

Meffe, G.K., Carroll, C.R. & contributors, 1997. Principles of Conservation Biology, 2nd edition. Sinauer Associates, Inc., Sunderland, MA.

Pearce, J.L. & Boyce, M.S., 2005. Modelling distribution and abundance with presence-only data. Journal of Applied Ecology, In Press.

Peterson, A.T., 2001. Predicting species' geographic distributions based on ecological niche modeling. Condor 103: 599-605.

Peterson, A.T., Soberón, J. & Sánchesz-Cordero, V., 1999. Conservatism of ecological niches in evolutionary time. Science 285: 1265-1267.

Poulin, R., 2004. Parasites and the neutral theory of bio-diversity. Ecography 27: 119-123.

Rabinowittz, D., 1981. Seven forms of rarity. pp 205-217 In H. Synge [ed]. John Wiley and Sons Ltd., New York, NY.

Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F., 2005. Neutral community dynamics, the mid-domain effect and spatial patterns in species richness. Ecology Letters 8: 783-790.

Ricklefs, R.E., 2003. A comment on Hubbell's zero-sum ecological drift model. Oikos 100: 185-192.

Riley, J.L. & Mohr, P., 1994. The natural heritage of southern Ontario's settled landscapes. Technical report TR-001. Science and Technology Transfer, Ontario Ministry of Natural Resources, Aurora, Ontario.

Sadoddin, A., Letcher, R.A., Jakeman, A.J. & Newham, L.T.H., 2005. A Bayesian decision network approach for assessing the ecological impacts of salinity management. Mathematics and Computers in Simulation 69: 162-176.

Seguardo, P. & Araújo, M.B., 2004. An evaluation of methods for modelling species distributions. Journal of Biogeography 31: 1555-1568.

Soberón, J. & Peterson, A.T., 2005. Interpretation of models of fundamental ecological niches and species' distributions. Biodiversity Informatics 2: 1-10.

Soulé, M.E. & Simberloff, D., 1986. What do genetics and ecology tell us about the design of nature reserves? Biological Conservation 35: 19-40.

Steger, R., 1987. Effects of refuges and recruitment on Gonodactylid Stomatopods, a guild of mobile prey. Ecology 35: 1520-1533.

Sykes, M.T., Prentice, I.C. & Cramer, W., 1996. A bioclimatic model for the potential distributions of north European tree species present and future climates. Journal of Biogeography 23: 203-233.

Tilman, D., 2004. Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assembly. Proceedings of the National Academy of Sciences 101: 10854-10861.

Turner, M.G., Gardner, R.H. & O'Neill, R.V., 2001. Landscape ecology: in theory and practice. Springer-Verlag New York, Inc. New York, NY.

Varis, O. & Kuikka, S., 1999. Learning Bayesian decision analysis by doing: lessons from environmental and natural resources management. Ecological Modelling 119: 177-195.

Yu, D.W., Terborgh, J.W. & Potts, M.D., 1998. Can high tree species richness be explained by Hubbell's null model? Ecology Letters 1: 193-199.

Zhang, D.-Y. & Lin, K., 1997. The effects of competitive asymmetry on the rate of competitive displacement: how robust is Hubbell's community drift model? Journal of Theoretical Biology 188: 361-367.

Figure 6.1. Potential niche map for rare aquatic plants in SW Ontario. The predicted potential niche is based upon a naïve Bayesian classifier using population occurrence information and relating it to environmental attributes. A probability of 1 represents a high likelihood that a specified location is part of the biota's potential niche.

Figure 6.2. Potential niche map for rare terrestrial plants in SW Ontario. The predicted potential niche is based upon a naïve Bayesian classifier using population occurrence information and relating it to environmental attributes. A probability of 1 represents a high likelihood of a specified location is part of the biota's potential niche.

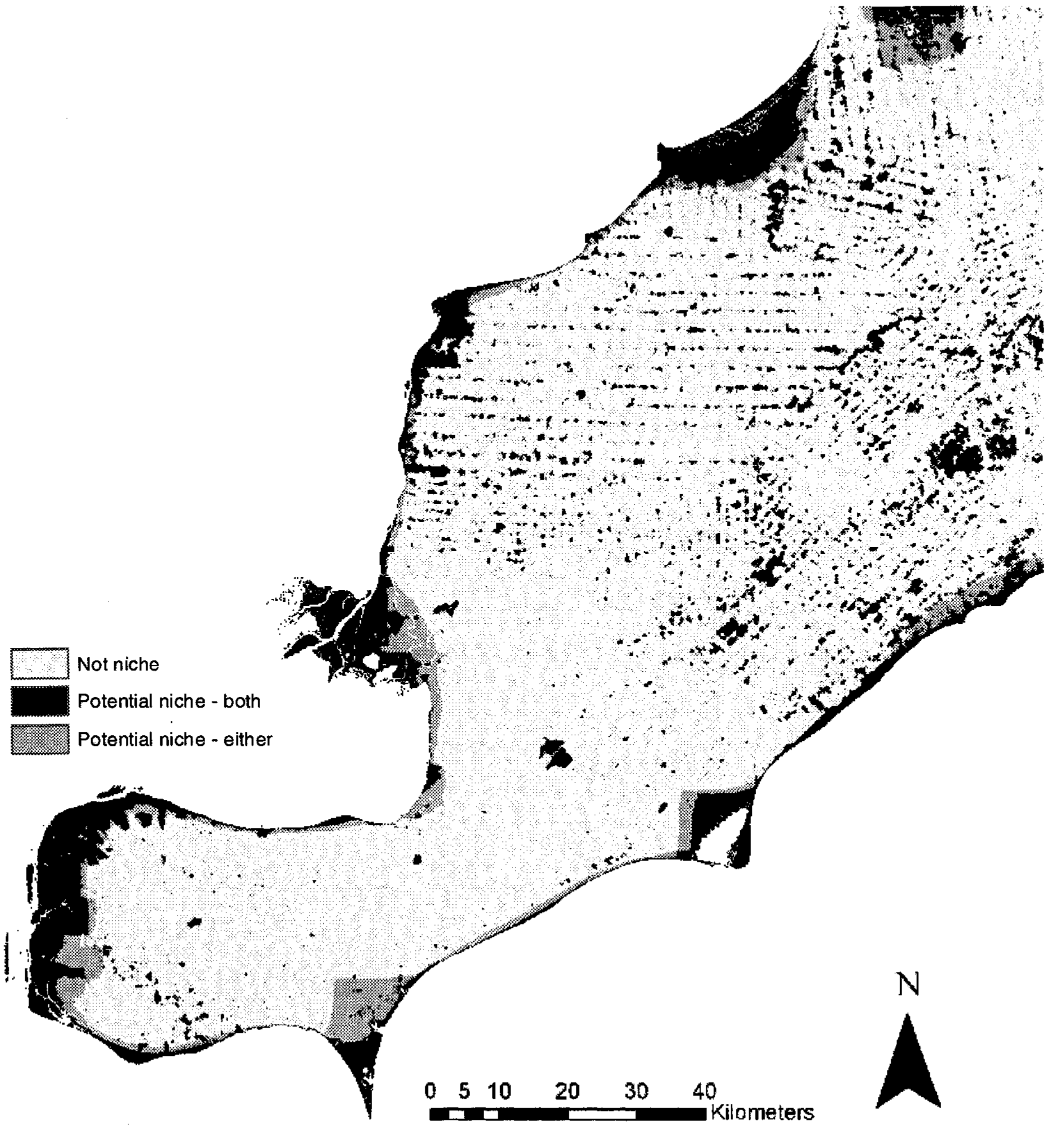


Figure 6.3. Potential niche map for both rare aquatic and terrestrial plants in SW Ontario. Lighter areas represent limited-to-no potential of the location as potential niche, and the darkest areas represent areas that have been defined as having high probability as potential niche for both aquatic and terrestrial plants. The light grey shading represents areas that are potential niche for either aquatic or terrestrial plants.

Table 6.1. Results of multiple single logistic regressions between occurrences and pseudo-absences of different groups of biota (birds, insects, plants and reptiles). Reported is the model fit (Nagelkerke's R^2) and buffer distance (scale) at which the environmental attribute had the greatest model fit. All regression results shown were significant at the 0.05 level. Bolded values represent environmental attributes included in subsequent models and empty values represent no significant regression results.

		Bird		Insect		Plant		Reptile	
		Scale (m)	R^2	Scale (m)	R^2	Scale (m)	R^2	Scale (m)	R^2
Aspect	mean	600	0.024	900	0.037	2400	0.017	4800	0.038
	SE			30	0.011	150	0.005		
Building density	mean	2400	0.019	4800	0.053	0	0.091	0	0.043
	SE			300	0.017	4800	0.038	300	0.023
Land classification for agriculture	Class 0	9600	0.305	7200	0.209	9600	0.14	4800	0.262
	Class 1	7200	0.082			4800	0.034	4800	0.04
	Class 2	9600	0.172	**600**	**0.183**	4800	0.054	9600	0.181
	Class 3	900	0.016	30	0.041				
	Class 4			2400	0.028				
	Class 5	1200	0.046	2400	0.032			30	0.037
	Class 7	450	0.254	9600	0.196	9600	0.116	**9600**	**0.235**
	Point values	0	0.301	**0**	**0.232**	**0**	**0.07**	**0**	**0.222**
Elevation	mean	0	0.053	30	0.073	0	0.084	0	0.099
	SE	2400	0.054	300	0.135	150	0.059	150	0.044
Landcover	Water	9600	0.331	**900**	**0.321**	9600	0.18	4800	0.312
	Freshwater marsh	450	0.177	7200	0.104	2400	0.011	150	0.089
	Deciduous swamp	**600**	**0.153**	900	0.135	**9600**	**0.128**	**900**	**0.204**

	Conifer swamp	7200	0.1	2400	0.03	2400	0.013	2400	0.053
	Open fen	900	0.012	30	0.011				
	Dense deciduous forest	150	0.224	300	0.163	2400	0.023	450	0.152
	Dense coniferous forest	4800	0.1	2400	0.106	2400	0.02	2400	0.066
	Coniferous plantation							900	0.027
	Deciduous mixed forest	450	0.068	900	0.116	2400	0.018	600	0.076
	Coniferous mixed forest	4800	0.056	600	0.092	2400	0.011	4800	0.054
	Sparse deciduous forest	150	0.037	600	0.047			300	0.039
	Bedrock outcrop	**4800**	**0.207**	7200	0.181	9600	0.098	7200	0.126
	Developed land	2400	0.042	4800	0.06	4800	0.044	1200	0.042
	Pasture	900	0.014	900	0.044	2400	0.041	2400	0.044
	Cropland	**300**	**0.616**	**4800**	**0.417**	**9600**	**0.229**	**600**	**0.445**
	Point values	**0**	**0.482**	**0**	**0.287**	**0**	**0.216**	**0**	**0.24**
Distance to a road	mean	9600	0.377	7200	0.344	9600	0.194	9600	0.368
	SE	9600	0.378	7200	0.336	9600	0.191	9600	0.342
Slope	mean	2400	0.052	300	0.147	150	0.063	150	0.04
	SE	1200	0.069	300	0.175	150	0.078	150	0.057
Soil Texture	Gravel	2400	0.085	**1200**	**0.199**	4800	0.052	4800	0.115
	Sand	7200	0.111	300	0.112	**7200**	**0.119**	7200	0.111
	Sandy loam	4800	0.105	2400	0.037	4800	0.034		
	Clay loam			300	0.013				
	Silty clay			4800	0.046				

Clay	2400	0.146	300	0.136			4800	0.054
Organics	4800	0.142	7200	0.14			9600	0.147
Point values	0	0.082	**0**	**0.23**	**0**	**0.093**	0	0.095

Table 6.2. Accuracy of models built for rare birds, insects, plants and reptiles. The percentage of pseudo-absences and occurrences predicted correctly are reported for each of a training, testing and overall (training + testing) datasets. The overall column represents the percentage of correct predictions independent of occurrence or absence.

	Datasets	Occurrence	Pseudo-absence	Overall
Birds	Training	84.4	92.9	88.7
	Testing	80.3	88.1	84.2
	Overall	82.6	90.3	86.4
Insects	Training	70.3	95.0	82.7
	Testing	76.0	92.8	84.4
	Overall	72.8	93.7	83.3
plants	Training	58.3	87.1	72.7
	Testing	56.9	86.7	71.8
	Overall	57.6	86.9	72.3
reptiles	Training	72.2	88.9	80.6
	Testing	81.1	85.1	83.1
	Overall	75.5	86.7	81.1

Table 6.3. Results of multiple single logistic regressions between occurrences and pseudo-absences of different groups of rare plants (trees, shrubs, herbs, grasses and sedges). Reported is the model fit (Nagelkerke's R^2) and buffer distance (scale) at which the environmental attribute had the greatest model fit. All regression results shown were significant at the 0.05 level. Bolded values represent environmental attributes included in subsequent models and empty values represent no significant regression results.

		Trees		Shrubs		Herbs		Grasses		Sedges	
		Scale (m)	R^2	Scale (m)	R^2	Scale (m)	R^2	Scale (m)	R^2	Scale (m)	R^2
Aspect	mean	600	0.014	**7200**	**0.089**	2400	0.013				
	SE							300	0.009	450	0.011
Building density	mean	150	0.013	7200	0.124	0	0.067	30	0.041	60	0.046
	SE	30	0.013	7200	0.115	60	0.039	30	0.041	30	0.046
Land classification for agriculture	Class 0	4800	0.074	2400	0.109	**4800**	**0.16**	7200	0.098	4800	0.154
	Class 1	7200	0.073	2400	0.036					7200	0.11
	Class 2	30	0.025			7200	0.071			4800	0.071
	Class 3	60	0.017	30	0.016						
	Class 5	2400	0.036			2400	0.023				
	Class 7	900	0.027			7200	0.098	7200	0.116	7200	0.082
	Point values	0	0.052			0	0.058	0	0.091	0	0.057
Elevation	mean	7200	0.065	0	0.056	0	0.091	150	0.007	0	0.085
	SE	300	0.069			300	0.05	300	0.037	300	0.091

Landcover	Water	4800	0.082	7200	0.104	4800	0.188	7200	0.134	4800	0.199
	Deciduous swamp	600	0.041	450	0.072	**9600**	**0.108**	7200	0.115	7200	0.102
	Conifer swamp	900	0.011			2400	0.014	2400	0.022	900	0.01
	Open fen									300	0.009
	Dense deciduous forest	150	0.283	60	0.209	2400	0.03	**450**	**0.137**	**90**	**0.251**
	Dense coniferous forest	600	0.03	90	0.016	2400	0.027	2400	0.082	450	0.034
	Deciduous mixed forest	300	0.044	900	0.05	2400	0.025	1200	0.041	900	0.026
	Coniferous mixed forest	450	0.023	1200	0.026	2400	0.018	1200	0.063	300	0.008
	Sparse deciduous forest	900	0.055	900	0.038			900	0.026	600	0.032
	Bedrock outcrop	900	0.074	1200	0.03	**7200**	**0.114**	**9600**	**0.145**	1200	0.096
	Developed land			**7200**	**0.119**	2400	0.019	2400	0.029	2400	0.037
	Pasture	2400	0.071	**900**	**0.096**	2400	0.046	2400	0.037	900	0.043
	Cropland	**90**	**0.305**	**60**	**0.26**	**4800**	**0.254**	**600**	**0.269**	**300**	**0.389**
	Point values	**0**	**0.226**	**0**	**0.155**	**0**	**0.182**	**0**	**0.11**	**0**	**0.261**
Distance to a road	mean	**7200**	**0.089**	**7200**	**0.124**	7200	0.189	**7200**	**0.177**	9600	0.231
	SE	4800	0.081	7200	0.114	4800	0.198	7200	0.169	**4800**	**0.222**
Slope	mean	150	0.083	300	0.019	300	0.054	600	0.043	150	0.099
	SE	**600**	**0.098**	300	0.04	300	0.067	600	0.054	**150**	**0.114**

Soil Texture	Gravel	2400	0.052	4800	0.063	4800	0.08			4800	0.092
	Sand	**7200**	**0.101**	4800	0.083	**7200**	**0.124**	7200	0.1	**4800**	**0.117**
	Sandy loam	4800	0.06			4800	0.035	7200	0.119	4800	0.043
	Silt loam	600	0.025	1200	0.018						
	Silty clay	2400	0.022							4800	0.032
	Clay	600	0.038			4800	0.037	7200	0.087		
	Organics					4800	0.051				
	Point values	**0**	**0.1**			0	0.106	**0**	**0.156**	0	0.074

Table 6.4. Results of multiple single logistic regressions between occurrences and pseudo-absences of different groups of rare plants based on rarity (subnational rarity ranks of S1, S2 and S3 for all rare plants and for rare herbs). Reported is the model fit (Nagelkerke's R^2) and buffer distance (scale) at which the environmental attribute had the greatest model fit. All regression results shown were significant at the 0.05 level. Bolded values represent environmental attributes included in subsequent models and empty values represent no significant regression results.

		Rare plants						Rare herbs					
		S1		S2		S3		S1		S2		S3	
		Scale (m)	R^2	Scale (m)	R^2	Scale (m)	R^2	Scale (m)	R^2	Scale (m)	R^2	Scale (m)	R^2
Aspect	mean			900	0.005	2400	0.014	1200	0.019	1200	0.006	4800	0.036
	SE					300	0.005					300	0.008
Building density	mean	0	0.105	0	0.056	0	0.041	0	0.136	0	0.074	0	0.033
	SE	30	0.075	60	0.036	90	0.029	30	0.089	60	0.047	60	0.023
Land classification for agriculture	Class 0	4800	0.103	4800	0.136	**4800**	**0.154**	**2400**	**0.223**	4800	0.149	**4800**	**0.175**
	Class 1	4800	0.048	4800	0.04	4800	0.043	7200	0.105				
	Class 2	4800	0.036	7200	0.066	4800	0.044	150	0.039	7200	0.081	4800	0.081
	Class 4			900	0.008					1200	0.006		
	Class 5			2400	0.026	2400	0.029	90	0.048	2400	0.017	2400	0.031
	Class 7	**9600**	**0.133**	7200	0.103	7200	0.074	4800	0.084	9600	0.134	4800	0.078
	Point values	0	0.049	0	0.06	0	0.053	0	0.044	0	0.057	0	0.068
Elevation	mean	0	0.047	0	0.109	0	0.051	90	0.083	0	0.136	0	0.048
	SE	2400	0.022	300	0.03	150	0.092	300	0.015	300	0.023	300	0.107
Landcover	Water	9600	0.146	4800	0.161	4800	0.181	**9600**	**0.21**	4800	0.181	4800	0.203

	Deciduous swamp	**9600**	**0.128**	9600	0.112	7200	0.074	2400	0.1	9600	0.126	**7200**	**0.113**
	Conifer swamp	900	0.008	2400	0.02	2400	0.014	900	0.015	900	0.01		
	Open fen	900	0.008							1200	0.005		
	Dense deciduous forest	**300**	**0.131**	**90**	0.186	2400	0.038	**150**	**0.171**	**150**	**0.165**	2400	0.057
	Dense coniferous forest	2400	0.038	450	0.052	2400	0.031	2400	0.049	450	0.042	2400	0.032
	Deciduous mixed forest	2400	0.018	2400	0.033	2400	0.027	900	0.06	1200	0.033	2400	0.031
	Coniferous mixed forest	2400	0.021	450	0.041	2400	0.021	2400	0.039	450	0.039	2400	0.022
	Sparse deciduous forest	300	0.023	300	0.035	2400	0.013	900	0.03	300	0.034		
	Bedrock outcrop	7200	0.079	**7200**	**0.122**	2400	0.089	2400	0.115	7200	0.119	7200	0.109
	Developed land	4800	0.044	2400	0.026	4800	0.038	4800	0.052	2400	0.025	2400	0.018
	Pasture	900	0.046	900	0.051	2400	0.054	900	0.067	900	0.054	2400	0.052
	Cropland	**600**	**0.229**	**300**	**0.293**	**4800**	**0.255**	**2400**	**0.309**	**300**	**0.276**	**2400**	**0.295**
	Point values	**0**	**0.135**	**0**	**0.157**	**0**	**0.247**	**0**	**0.138**	**0**	**0.143**	**0**	**0.271**
Distance to a road	mean	**9600**	**0.191**	**9600**	**0.188**	7200	0.191	9600	0.26	**9600**	**0.215**	7200	0.213
	SE	9600	0.161	7200	0.174	4800	0.199	9600	0.237	7200	0.2	4800	0.223
Slope	mean	2400	0.022	300	0.034	150	0.102	2400	0.03	300	0.026	300	0.116
	SE	2400	0.029	300	0.045	**150**	**0.12**	2400	0.054	300	0.035	**150**	**0.141**
Soil Texture	Gravel	4800	0.07	4800	0.082	**2400**	**0.094**	2400	0.122	4800	0.084	2400	0.091

Sand	4800	0.095	**7200**	**0.122**	**4800**	**0.112**	**9600**	**0.153**	**4800**	**0.138**	**4800**	**0.114**
Sandy loam	4800	0.045	7200	0.08	2400	0.045	600	0.049	4800	0.037	2400	0.033
Silt loam					2400	0.018	600	0.041			2400	0.014
Clay	4800	0.045	4800	0.045	2400	0.041	450	0.123	4800	0.034	2400	0.059
Organics	4800	0.049			4800	0.037			4800	0.044	2400	0.031
Point values	0	0.136	0	0.118	0	0.083	**0**	**0.19**	0	0.111	0	0.083

Table 6.5. Results of multiple single logistic regressions between occurrences and pseudo-absences of different groups of rare plants based on broad description of aquatic and terrestrial plants. Reported is the model fit (Nagelkerke's R^2) and buffer distance (scale) at which the environmental attribute had the greatest model fit. All regression results were shown significant at the 0.05 level. Bolded values represent environmental attributes included in subsequent models and empty values represent no significant regression results.

		Aquatic plants		Terrestrial plants	
		Scale (m)	R^2	Scale (m)	R^2
Aspect	mean	2400	0.0157		
Building density	mean	0	0.0523	90	0.0581
	SE	60	0.0452	30	0.0613
Land classification for agriculture	Class 0	4800	0.2045	7200	0.1245
	Class 1	7200	0.0947	7200	0.1273
	Class 2	7200	0.0779		
	Class 7	**9600**	**0.126**	9600	0.1242
	Point values	0	0.0735	0	0.0738
Elevation	mean	0	0.1082	0	0.1101
	SE	300	0.0531	300	0.0677
Landcover	Water	2400	0.2857	7200	0.1688
	Freshwater marsh	2400	0.0257		
	Deciduous swamp	**7200**	**0.1192**	**9600**	**0.1397**
	Conifer swamp	450	0.0079	2400	0.0265
	Open fen	2400	0.0134	450	0.0067
	Dense deciduous forest	**90**	**0.1233**	**300**	**0.2772**

	Dense coniferous forest	450	0.0259	450	0.052
	Deciduous mixed forest	150	0.0111	900	0.068
	Coniferous mixed forest			900	0.0543
	Sparse deciduous forest	900	0.0158	300	0.0545
	Bedrock outcrop	1200	0.0843	**7200**	**0.1303**
	Developed land	2400	0.0261	2400	0.0381
	Pasture	2400	0.0471	2400	0.0517
	Cropland	**450**	**0.3751**	**300**	**0.3644**
	Point values	**0**	**0.2228**	**0**	**0.2365**
Distance to a road	mean	**9600**	**0.3029**	**9600**	**0.2216**
	SE	4800	0.2998	7200	0.1955
Slope	mean	150	0.055	300	0.0742
	SE	150	0.0748	300	0.0879
Soil Texture	Gravel	4800	0.087		
	Sand	4800	0.0822	7200	0.1463
	Sandy loam	4800	0.0469		
	Clay	4800	0.0371		
	Organics	7200	0.0974		
	Point values	0	0.0738	0	0.1109

Table 6.6. Accuracy of models built for rare trees, shrubs, herbs, grasses and sedges. The percentage of pseudo-absences and occurrences predicted correctly are reported for each of a training, testing and overall (training + testing) datasets. The overall column represents the percentage of correct predictions independent of occurrence or absence.

	Datasets	Occurrence	Pseudo-absence	Overall
Trees	Training	66.2	79.1	72.6
	Testing	69.0	76.8	72.9
	Overall	67.3	77.8	72.6
Shrubs	Training	63.3	83.3	73.3
	Testing	67.2	82.7	74.9
	Overall	65.0	83.0	74.0
Herbs	Training	58.9	84.4	71.7
	Testing	60.1	84.6	72.4
	Overall	59.6	84.5	72.1
Grasses	Training	66.2	85.2	75.7
	Testing	65.3	81.4	73.4
	Overall	65.9	83.0	74.4
Sedges	Training	77.8	85.7	81.7
	Testing	80.0	81.5	80.8
	Overall	78.9	83.4	81.1

Table 6.7. Accuracy of models built for rarity groupings in plants and herbs separately. The rarity groupings are based on the subnational rarity rankings of S1, S2 and S3. The percentage of pseudo-absences and occurrences predicted correctly are reported for each of a training, testing and overall (training + testing) datasets. The overall column represents the percentage of correct predictions independent of occurrence or absence.

		Datasets	Occurrence	Pseudo-absence	Overall
All rare plants	S1	Training	62.5	84.9	73.7
		Testing	62.6	85.8	74.2
		Overall	62.6	85.4	74.0
	S2	Training	65.5	83.6	74.6
		Testing	66.8	81.5	74.2
		Overall	66.2	82.6	74.4
	S3	Training	58.2	83.6	70.9
		Testing	58.3	87.6	72.9
		Overall	58.3	85.2	71.7
Rare herbs	S1	Training	65.0	91.8	78.4
		Testing	67.9	86.7	77.3
		Overall	66.7	89.5	78.1
	S2	Training	65.1	85.5	75.3
		Testing	71.2	83.4	77.3
		Overall	67.7	84.3	76.0
	S3	Training	65.7	85.2	75.4
		Testing	65.2	83.5	74.4
		Overall	65.4	84.4	74.9

Table 6.8. Accuracy of models built for rare aquatic and terrestrial plants. The percentage of pseudo-absences and occurrences predicted correctly are reported for each of a training, testing and overall (training + testing) datasets. The overall column represents the percentage of correct predictions independent of occurrence or absence.

	Datasets	Occurrence	Pseudo-absence	Overall
Aquatic plants	Training	75.2	83.7	79.5
	Testing	80.4	81.5	81.0
	Overall	77.0	82.4	79.7
Terrestrial plants	Training	72.2	85.5	78.8
	Testing	76.8	81.6	79.2
	Overall	74.5	83.6	79.1

Chapter 7 - The niche and its alternatives: a continuum?

"And NUH is the letter I use to spell Nutches,
Who live in small caves, known as Niches, for Nutches.
These Nutches have troubles, the biggest of which is
The fact there are many more Nutches than Niches.
Each Nutch in a Nich knows that some other Nutch
Would like to move into his Nich very much.
So each Nutch in a Nich has to watch that small Nich
Or Nutches who haven't got Niches will snitch."
Dr. Seuss (1955)

Introduction

Hutchinson (1957) provided a formalized concept of the niche (an n-dimensional hypervolume) which has become a central organizing tenet of modern ecology (Leibold 1995, Pulliam 2000). While the notion of the niche is often emphasized as a concept from the domain of community ecology (see, e.g., Begon et al. 1990, Pianka 1994, Ricklefs 2001), it can and has been applied with respect to studies at most levels of ecological organisation (Liebold 1995). For example, identifying environmental conditions limiting an organism's performance or fitness is often the analytical path taken by physiologists (e.g., Anthony and Connolly 2004, Ochocinska and Taylor 2005, Welsh et al. 2005); population biologists examine the limiting factors that alter population dynamics (e.g., Halpern et al. 2005, Lõhmus and Remm 2005); biogeographers examine the environmental constraints limiting species distributions (e.g., Peterson et al. 1999, Peterson 2001); and ecosystem ecologists seek to determine how the functional traits of taxa alter ecosystem structure and process (e.g., Hunter and Simons 2004). Thus, moving towards larger scales, one can examine an organism's niche, a population's niche, a species' niche, and the niche of a taxon or group of species.

The study of how organisms are distributed in space and time has long interested ecologists and has inspired many modeling approaches that quantify the species-environment relationship (e.g., Rushton et al. 2004, Guisan and Thuiller 2005). Kearney and Porter (2004) suggested that the concept of the niche provides a useful starting point for understanding the distribution of species. Indeed, many models examining distributions of species assume an underlying niche concept (Guisan and Zimmermann 2000, Guisan and Thuiller 2005, Soberón and Peterson 2005). Traditional explanations for patterns of species distribution, abundance and coexistence have all argued that each species evolves adaptations for niche exploitation (Whitfield 2002).

In general, ecologists work from the premise that the niche is a central organizing tenet in ecology and that attributes of the niche structure biogeographic patterns of diversity, abundance and distribution (Gaston and Chown 2005). One major emergent alternative to the niche concept is the inference of stochasticity, and its application through null models. For example, rather than competitive interactions of species shaping the coexistence of species, "historical accidents of dispersal" have been suggested (Ostling 2005).

In this thesis I explored the concept of niche through the use of null models. While analytical and simulation models attempt to mimic reality, null models exclude the mechanism or factor of interest, offering a baseline for comparison (Harvey et al. 1983, Colwell and Winkler 1984, Gotelli and Graves 1996, Gotelli 2001). The deviation from null thus describes the influence of the factor or mechanism of interest (Colwell et al. 2004). Here I have tried to exclude niche-based attributes of species in order to determine the relative influence of such attributes in examining species distributions.

Continental to global patterns of species distributions

Broad scale patterns in species richness have long intrigued ecologists (e.g., Brown and Lomolino 1998, Gaston 2003). The study of latitudinal gradients in species richness is one of the oldest and most fundamental patterns, whereby richness increases from the earth's poles to the equator (Willig et al. 2003). Proposed mechanisms for broad-scale patterns in species richness have been abundant (see reviews in Rohde 1992, Rahbek 1995, Willig et al. 2003, Currie et al. 2004). Most of the proposed mechanisms require or at least imply that the biota respond to environmental gradients, and are influenced by different biotic and abiotic interactions.

Willig et al. (2003) reviewed more than 30 proposed mechanisms of broad-scale patterns in species richness and noted that many of them were too specific to be generally applicable; that they were tightly interlinked, insufficiently supported and/or circular. They further proposed six hypotheses that have the most support, potential and/or generality; for each the authors reviewed the evidence supporting and refuting the hypotheses. These included the geographic-area hypothesis, productivity hypothesis, ambient-energy hypothesis, Rapoport-rescue hypothesis, evolutionary-speed hypothesis and geometric-constraints hypothesis (also known as the mid-domain effect).

The first five of these six hypotheses postulate mechanisms whereby the species respond to particular biotic or abiotic environmental gradients. The geographic-area hypothesis proposes that since the tropics have greater area (due to the fact that equatorial latitudinal parallels are larger than those at temperate or polar latitudes, and that N and S tropics are adjacent while N and S temperate or polar regions are separated), they are assumed to have greater habitat heterogeneity facilitating specialization, adaptation and speciation - thus greater species richness. The productivity hypothesis assumes that the annual input of solar radiation describes the available energy which in turn influences the productivity and biomass of biota; greater productivity is assumed to be correlated with greater species richness. The ambient-energy hypothesis is essentially an "umbrella" hypothesis that includes other explanations (such as seasonality, harshness, and environmental stability / predictability); species richness is assumed to be lower in unstable, unpredictable and "harsh" environments. Rapoport's-rescue hypothesis combines aspects of other hypotheses in suggesting that in more stable environments (e.g., tropical zones, lower elevations), species ranges should be smaller (greater adaptation for the local environment), and thus a greater number of species creating greater habitat heterogeneity should be observed; overall species richness is further supplemented by dispersal of organisms from favorable areas to areas where the species would not normally persist. The evolutionary-speed hypothesis suggests that areas of increased energy (e.g., temperature) should have increased rates of adaptation and speciation due to shorter generation times (induced by greater available energy); shorter generation times, increased mutation and selection pressures are assumed to increase speciation and thus, for example tropical zones would be evolutionarily older than temperate or polar zones (Rohde 1992, Willig et al. 2003).

The final hypothesis which Willig et al (2003) proposed to show promise was the geometric-constraints hypothesis. This is best known as the mid-domain effect (MDE). The MDE, as discussed in Chapters 2-4, proposes that non-biological gradients in species richness occur as a consequence of random placement of species ranges within a bounded domain (Colwell and Hurtt 1994). Not only are predictions of MDE models powerful in elucidating patterns of species richness along gradients of latitude, longitude and altitude (reviewed in Zapata et al. 2003, Colwell et al. 2004; and see Chapters 3 and 4 of this thesis), but they also speak to the relative importance of the niche concept influencing such patterns.

In Chapter 3, geometric constraints were shown to have greatest predictive power on the distribution of North American tree species richness for species having ranges that are large relative to the domain (ranges < 30% of NA, Table 3.1). For species with large ranges, MDE alone often explained more of the variation in species richness than all of the environmental variables combined (Tables 3.1 and 3.2). For medium- and small-range species, geometric constraints clearly have little influence on patterns of species richness and, as expected, environmental parameters are much more important (Table 3.1). None of the MDE models alone was able to capture the empirical peak in tree species richness in the southeastern part of the US. When MDE was combined with environmental variables, particularly precipitation seasonality and slope, the percent of variation in species richness explained was often > 90%, and did capture this natural peak. Species richness is generally higher in regions where precipitation seasonality and slope are low.

Large-range tree species richness patterns generated by geometric constraints alone often have as much, if not more, predictive power than those generated by environmental gradients alone (see Table 3.1). Colwell et al. (2005; and see Colwell and Lees 2000) stated that the question is not *whether* geometry affects patterns of species richness, but what the magnitude of the contribution is. The results presented here show that the magnitude of the contribution can be quite substantial, especially for large ranged tree species. For medium- and small-range species, geometric constraints had little influence on patterns of species richness and environmental parameters were much more important.

The fact that precipitation seasonality and slope were consistently significant in predicting species richness, in both one-dimensional latitudinal, longitudinal and altitudinal models as well as two- and three-dimensional models, points to their considerable importance in structuring tree species richness. The inclusion of precipitation seasonality may support the ambient energy hypothesis. However alone, neither precipitation seasonality nor slope were able to capture the peak in species richness in the southeastern corner of the continent (see Appendix 3B).

The MDE tends to be weaker where ranges are small relative to the domain extent (Laurie and Silander 2002), and stronger where ranges are large relative to the extent of the domain (McCain 2003). Small-ranged species are less likely to experience 'hard' boundaries than large-range species, and thus the impact of boundaries on their richness patterns should be smaller (Jetz and Rahbek 2002), and more likely to reflect environmental and historical factors (Colwell and Lees 2000). Colwell and Lees (2000) further noted that large-ranged species are more likely to be affected by continental geometry and to occupy the centre of the bounded area. This was supported by the results of both Chapters 3 and 4.

In Chapter 4, the effects of MDE increased with increasing range size; MDE accounted for some 73%, 77% and 84% of the variation in amphibian (smallest average range size), mammal (mid-average range size) and avian (largest average range size) species richness gradients, respectively (see Table 4.1) across North, Central and South America.

Studies that have partitioned datasets into range size categories have found consistently stronger support for MDE among large-ranged species than among the smaller-range species in the dataset (Hawkins and Diniz-Filho 2002, Jetz and Rahbek 2002, Vetaas and Grytnes 2002, Mora and Robertson 2005, Murphy et al. 2006).

As an additional variable together with climatic variables, MDE added little to the explanatory power of the analysis. On average, MDE explained an additional 1.1%, 0.5% and 6.9% of the variation in amphibian, avian and mammal species richness, respectively. For all groups, precipitation seasonality and isothermality were the most commonly used predictors of species richness. It could be hypothesized that the patterns of species richness examined here may provide further support to the ambient energy hypothesis (environmental stability or predictability hypotheses) (see e.g., Thiery 1982, Begon et al. 1990, Willig et al. 2003). These climatic correlates of species richness could not alone explain the observed patterns. In addition, these hypotheses are generally accepted as unsupported (Rohde 1992, Willig et al. 2003) and testing these potential hypotheses is beyond the scope of this thesis.

Although in Chapters 3 and 4, MDE and environmental correlates provided nearly equal explanatory value, the principle of parsimony suggests that random placement of ranges within the limits of a domain should be used. Thus, this correlative approach suggests, with respect to the predictive power of MDE, species are randomly placed with the domain limits, or the richness pattern appears random but is a consequence of some larger set of interacting factors. Most likely, richness patterns are the product of variation in relative strength of stochastic and ecological (niche-based) processes. The relative strength is partially determined by the range size of the species and vice versa, the range size itself may be partially determined by the ecological processes.

MDE effectively ignores spatial environmental gradients when placing species within the bounds of a domain; it assumes that "environmental conditions vary but that species' responses to environmental conditions would be sufficiently individualistic that, in the aggregate, no part of the domain would be more hospitable to species than any other part" (Connolly 2005). The question being asked by the MDE paradigm seems to be not "what constrains the distribution of a species" but rather, to what degree would stochastic processes be sufficient to explain actual species richness patterns? The second section of this thesis does examine some of the factors potentially constraining the distribution of a species.

Local to regional scale mapping of the niche

Understanding that a species' range is not homogeneously and uniformly suitable throughout, finer-scale habitat suitability mapping, quantitative habitat models and predictive distribution maps (of a species' niche) are all potentially important tools to guide management and restoration as approaches to conservation of rare species (Guisan and Zimmerman 2000, Johnson et al. 2004).

Spatial randomness is a null hypothesis used in many tests to detect spatial patterns (e.g., point clustering, spatial autocorrelation), including clustering of species occurrences (e.g., Peterson et al. 1999, Peterson 2001). Often, such a null hypothesis is not really relevant for complex ecological systems (Sokal et al. 1998, Fortin and Jacquez 2000), so rejection of the null hypothesis may represent little scientific value. For example, when testing spatial patterns of fish occurrences, it would be unrealistic and unhelpful to use a null distribution pattern that included random locations within the terrestrial environment. The appropriate null model is that which captures the notion of a plausible system state, as in Chapter 5.

While the species distribution model in Chapter 5 was created to present a plausible system state (a null model of the spatial clustering of the species), it also describes the niche of a species with respect to attributes of its environment. Deviation from such a null model would describe potential biotic / abiotic interactions that further influence the niche of a species. The null species distribution model created for *Opuntia humifusa* at Point Pelee National Park, SW Ontario demonstrated the methodology. The derived model accurately predicted species occurrences 9.36% of the time, using the landuse category at the location of interest and the proportions of the surrounding landscape with respect to beach (at 50m radius), human use (at 500m), ponds (at 700m) and roads/parking (at 700m). Deviations from this null model (2.9%) may be attributed to biotic interactions. Indeed, six of the 10 individuals, although believed to be natural, are located in an area that was previously an orchard.

Hutchinson (1957) formalization of the concept of the niche has become a central organizing tenet of modern ecology (Leibold 1995, Pulliam 2000). The key difference between Hutchinson's concept of niche and that of earlier definitions (i.e., Grinnell's [1917, 1924, 1928] definition as the habitat in which an organism resides, or Elton's [1927] definition as the ecological role an organism fills within a community) was that Hutchinson (1957) used the niche to represent the environmental requirements of a species rather than a place in the environment that has the potential to support a species (Schoener 1989, Colwell 1992) - thus Hutchinson (1957) emphasised attributes of the species, not the environment (Pulliam 2000). With the emphasis on attributes of the species, each species is thought to have a unique niche (Gravel et al. 2006, Tuomisto 2006).

There is ongoing debate over the relative importance of niche-assembly vs. dispersal-assembly theories of species coexistence (see Potts et al. 2004). Niche-assembly theories posit that biological interactions and environmental heterogeneity are the processes underlying species coexistence and community structure (Tilman 1982, Lieberman et al. 1985, Hubbell and Foster 1986, Denslow 1987, Kohyama 1994, Terborgh et al. 1996, Clark et al. 1998). It is believed that species can only coexist when they differ from each other in the resources they use most efficiently, or in their adaptation to the local environmental conditions (Ostling 2005). This theory makes the assumption that coexisting species must have different niches.

In contrast, chance, history, and dispersal explain species coexistence in dispersal-assembly theories (Hubbell 1997, 2001, Bell 2001). These theories suggest that "historical accidents of dispersal," rather than competitive interactions of species shape the coexistence of species (Ostling 2005). Rather than being quickly out-competed, it is suggested that species that are less efficient at using a resource evolve to be as efficient as their competitors. Dispersal to the same habitable region is the main criterion for coexistence (Ostling 2005); thus ecologically similar species need not have unique niches.

If indeed ecologically similar species need not have a unique niche, they can be assumed equivalent with respect to their interactions with their environment and therefore show no habitat preference (Hubbell 2001, 2005). In chapter 6, I test the validity of this assumption by creating null species distribution models for groups of rare species in SW Ontario.

Intuitively, all species violate this assumption to some degree, the question posed by Hubbell's neutral models is, how good is this approximation? It is the deviation from this neutral assumption that is most interesting. Here the deviation, in terms of false negatives, ranged from 17.4 to 42.4% (accuracy ranged from 82.6 to 57.6%; Table 6.2, 6.6-6.8). Hubbell (2005) suggested the assumption should hold best for sessile organisms, such as plants, and least for more mobile animals. In Chapter 6, the opposite was observed upon examining the four broad taxonomic categories (birds, insects, plants and reptiles). The rare birds showed the fewest false negatives and plants showed the greatest number of false negatives (Table 6.2). The relatively low proportion of false negatives for bird, insect and reptile species distribution models was unexpected since these included mobile species having varied life histories that included differences in broad habitat classifications such as aquatic and terrestrial species. While the poor accuracy with respect to the plant model was unexpected (compared to other taxonomic groups), the rare plants were potentially too broad of a group and thus were further reclassified by taxonomic and rarity classifications. Little improvement with respect to the number of false-negatives was observed, except with respect to the rare sedges in reclassifying rare plants into finer taxonomic and rarity based groups. Improvement in model accuracy (reduction in the number of false negatives) was observed when rare plants were reclassified into aquatic or terrestrial groups (Tables 6.2, 6.6 - 6.8).

Despite the species-specific differences in life history, resource utilization, etc. (niche differences) in the broad distinctions of species groupings, distribution models performed relatively well for birds, insects, reptiles, sedges, aquatic and terrestrial plants (less than 27% false negatives). Results for these groups of species lend support to Hubbell's neutral assumption.

Reconciling stocasticity and the niche

Models and their underlying assumptions are in the end only abstractions of reality. As such, even null models should therefore not be taken to be free from bias; all models make assumptions and simplifications of reality and these assumptions must be considered when interpreting the model results and utility (Gotelli 2001, Colwell et al. 2004). In this thesis, two detailed and quite different null models were presented, each with assumptions that countered conventional ideas of the niche (the niche of a species influences richness patterns at the continental scale and each species has a unique niche at a more local scale). While both null models were supported, neither was supported in its entirety.

The mid-domain models explained more of the variation in species richness with larger ranged species, and the regional, null species distribution models performed best for birds, insects, reptiles, sedges, and aquatic and terrestrial plants. It seems most likely that real species distributions are the product of variation in relative strength of stochastic and ecological processes (Bokma et al. 2001, Colwell et al. 2005, Rangel and Diniz-Filho 2005). Gravel et al. (2006) reviewed the literature and proposed ways of reconciling the niche and neutral concepts (or null alternatives to the niche concept). The authors suggest that niche and neutral models form the ends, the extremes of a continuum of events, from deterministic to stochastic processes.

Gravel et al. (2006) noted that while some studies support either niche or neutral models, most studies reported results that were intermediate between the two. Indeed, the present thesis similarly reports results intermediate (often toward the null) between the extremes of the continuum. Thus, in supporting such a "continuum hypothesis", this thesis has started by characterizing the attributes of groups of species (e.g., large-ranged species are influenced by the MDE more than small-ranged species) that will determine where on the continuum, from niche to null alternatives, a particular species or group of species will lie.

References

Anthony, K.R.N. & Connolly, S.R., 2004. Environmental limits to growth: physiological niche boundaries of corals along turbidity-light gradients. Oecologia 141: 373-384.

Begon, M., Harper, J.L. & Townsend, C. R., 1990. Ecology: individuals, populations and communities. Blackwell Scientific, Boston, MA.

Bell, G., 2001. Neutral macroecology. Science 293: 2413-2418.

Bokma, F., Bokma, J. & Mönkkönen, M., 2001. Random processes and geographic species richness patterns: why so few species in the north? Ecography 24: 43-49.

Brown, J.H. & Lomolino, M.V., 1998. Biogeography. Sinauer Associates, Sunderland, MA.

Clark, D.B., Clark, D.A. & Read, J.M., 1998. Edaphic variation and the mesoscale distribution of tree species in a neotropical rain forest. Journal of Ecology 86: 101-112.

Colwell, R.K., 1992. Niche: a bifurcation in the conceptual lineage of the term. Pp 241-248 in E. Fox-Keller and E. A. Lloyd, eds. Keywords in evolutionary biology. Harvard University Press, Cambridge, MA.

Colwell, R.K. & Hurtt, G.C., 1994. Nonbiological gradients in species richness and a spurious Rapoport effect. The American Naturalist 144: 570-595.

Colwell, R.K. & Lees, D.C., 2000. The mid-domain effect: geometric constraints on the geography of species richness. Trends in Ecology and Evolution 15, 70-76.

Colwell, R.K. & Winkler, D.W., 1984. A null model for null models in biogeography. Pp 344-359 in D. R. Strong, Jr., D. Simberloff, L. G. Abele, and A. B Thistle, eds. Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton, N.J.

Colwell, R.K., Rahbek, C. & Gotelli, N.J., 2004. The mid-domain effect and species richness patterns: What have we learned so far? The American Naturalist 163: E1-E23.

Colwell, R.K., Rahbek, C. & Gotelli, N.J., 2005. The mid-domain effect: there's a baby in the bathwater. The American Naturalist 166: E149-E154.

Connolly, S.R., 2005. Process-based models of species distributions and the mid-domain effect. The American Naturalist 166: 1-11.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J., Hawkins, B.A., Kaufman, D.M., Kerr, J.T., Oberdorff, T., O'Brien, E. & Turner, J.R.G., 2004.

Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. Ecology Letters 7: 1121-1134.

Denslow, J.S., 1987. Tropical rainforest gaps and tree species diversity. Annual Review of Ecology and Systematics 18: 431-451

Dr. Seuss, 1955. On Beyond Zebra. Random House, Inc. New York, NY.

Elton, C., 1927. Animal Ecology. Sidgwick and Jackson, London.

Fortin, M.J. & Jacquez, G.M., 2000. Randomization tests and spatially autocorrelated data. Bulletin of the Ecological Society of America 81: 201-205.

Gaston, K.J., 2003. The Structure and Dynamics of Geographic Ranges. Oxford University Press, Oxford, UK.

Gaston, K.J. & Chown, S.L., 2005. Neutrality and the niche. Functional Ecology 19: 1-6.

Gotelli, N.J., 2001. Research frontiers in null model analysis. Global Ecology & Biogeography 10: 337-343.

Gotelli, N.J. & Graves, G.R., 1996. Null Models in Ecology. Smithsonian Institution Press, Washington, DC.

Gravel, D., Canham, C.D., Beaudet, M. & Messier, C., 2006. Reconciling niche and neutrality: the continuum hypothesis. Ecology Letters 9: 399-409.

Grinnell, J., 1917. The niche-relationships of the California Thrasher. Auk 34: 427-433.

Grinnell, J., 1924. Geography and evolution. Ecology 5: 225-229.

Grinnell, J., 1928. A distributional summation of the ornithology of Lower California. University of California Publications in Zoology 32: 1-30.

Guisan, A. & Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8: 993-1009.

Guisan, A. & Zimmerman, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135: 147-186.

Halpern, B.S., Gaines, S.D. & Warner, R.R., 2005. Habitat size, recruitment, and longevity as factors limiting population size in stage-structured species. The American Naturalist 165: 82-94.

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

Hawkins, B.A. & Diniz-Filho, J.A.F., 2002. The mid-domain effect cannot explain the diversity gradient of Nearctic birds. Global Ecology & Biogeography 11: 419-426.

Hubbell, S.P., 1997. A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. Coral Reefs 16: S9-S21.

Hubbell, S.P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton, NJ.

Hubbell, S.P., 2005. Neutral theory in community ecology and the hypothesis of functional equivalence. Functional Ecology 19: 166-172.

Hubbell, S.P. & Foster, R.B., 1986. Biology, chance and history and the structure of tropical rain forest tree communities. pp 314-329 In: J. Diamond & T.J. Case (eds). Community ecology. Harper and Row, New York, NY.

Hunter, R.D. & Simons, K.A., 2004. Dreissenids in Lake St. Clair in 2001: evidence for population regulation. Journal of Great Lakes Research 30: 528-537.

Hutchinson, G.E., 1957. Concluding remarks. Cold Spring Harbor Symposium on Quantitative Biology 22: 415-427.

Jetz, W. & Rahbek, C., 2002. Geographic range size and determinants of avian species richness. Science 297: 1548-1551.

Johnson, C.J., Seip, D.R. & Boyce, M.S., 2004. A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales. Journal of Applied Ecology 41: 238-251.

Kearney, M. & Porter, W.P., 2004. Mapping the fundamental niche, physiology, climate and the distribution of a nocturnal lizard. Ecology 85: 3119-3131.

Kohyama T., 1994. Size-structure-based models of forest dynamics to interpret population-level and community-level mechanisms. Journal of Plant Research 107: 107-116.

Laurie, H. & Silander Jr., J.A., 2002. Geometric constraints and spatial patterns of species richness: critique of range-based null models. Diversity and Disturbance 8: 351-364.

Leibold, M.A., 1995. The niche concept revisited: mechanistic models and community context. Ecology 76: 1371-1382.

Lieberman, M., Lieberman, D., Hartshorn, G.S. & Peralta, R., 1985. Smallscale altitudinal variation in low-land wet tropical forest vegetation. Journal of Ecology 73: 505-516.

Lõhmus, A. & Remm, J., 2005. Nest quality limits the number of hole-nesting passerines in their natural cavity-rich habitat. Acta Oecologica 27: 125-128.

McCain, C.M., 2003. North American desert rodents: a test of the mid-domain effect in species richness. Journal of Mammalology 84: 967-980.

Mora, C. & Robertson, D.R., 2005. Causes of latitudinal gradients in species richness: a test with fishes of the tropical Eastern Pacific. Ecology 86: 1771-1782.

Murphy, H.T., VanDerWal, J., Khalatkhar, N. & Lovett-Doust, J., 2006. Geometric and environmental correlates of tree species richness in North America. Ecography. In Review.

Ochocinska, D. & Taylor, J.R.E., 2005. Living at the physiological limits: field and maximum metabolic rates of the common shrew (Sorex araneus). Physiological and Biochemical Zoology 78: 808-818.

Ostling, A., 2005. Ecology - neutral theory tested by birds. Nature 436: 635-636.

Peterson, A.T., 2001. Predicting species' geographic distributions based on ecological niche modeling. Condor 103: 599-605.

Peterson, A.T., Soberón, J. & Sánchesz-Cordero, V., 1999. Conservatism of ecological niches in evolutionary time. Science 285:1265-1267.

Pianka, E.R., 1994. Comparative ecology of *Varanus* in the Great Victoria desert. Australian Journal of Ecology 19, 395-408.

Potts, M.D., Davies, S.J., Bossert, W.H., Tan, S. & Supardi, M.N.N., 2004. Habitat heterogenity and niche structure of trees in two tropical rain forests. Oceologia 139: 446-453.

Pulliam, H.R., 2000. On the relationship between niche and distribution. Ecology Letters 3: 349-361.

Rahbek, C., 1995. The elevational gradient of species richness - a uniform pattern. Ecography 18: 200-205

Rangel, T.F.L.V.B. & Diniz-Filho, J.A.F., 2005. An evolutionary tolerance model explaining spatial patterns in species richness under environmental gradients and geometric constraints. Ecography 28: 253-263.

Ricklefs, R.E., 2001. The Economy of Nature, 5th ed. WH Freeman, New York, NY.

Rohde, K., 1992. Latitudinal gradients in species diversity: the search for the primary cause. Oikos 65: 514-527.

Rushton, S.P., Ormerod, S.J. & Kerby, G., 2004. New paradigms for modeling species distributions? Journal of Applied Ecology 41: 193-200.

Schoener, T.W., 1989. The ecological niche. pg 79-114 in J. M. Cherrett, ed. Ecological concepts: the contribution of ecology to an understanding of the natural world. Blackwell Scientific, Oxford, UK.

Soberón, J. & Peterson, A.T., 2005. Interpretation of models of fundamental ecological niches and species' distributions. Biodiversity Informatics 2: 1-10.

Sokal, R.R., Oden, N.L. & Thomson, B.A., 1998. Local spatial autocorrelation in a biological model. Geographical Analysis 30: 331-354.

Terborgh, J., Foster, R.B. & Nunez, P.V., 1996. Tropical tree communities: A test of the non-equilibrium hypothesis. Ecology 77: 561-567.

Thiery, R.G., 1982. Environmental instability and community diversity. Biological Reviews 57: 671-710.

Tilman, D., 1982. Resource competition and community structure. Princeton University Press, Princeton, N.J.

Tuomisto, H., 2006. Edaphic niche differentiation among *Polybotrya* ferns in western Amazonia: implications for coexistence and speciation. Ecography. In Press.

Vetaas, O.R. & Grytnes, J.A., 2002. Distribution of vascular plant species richness and endemic richness along the Himalayan elevation gradient in Nepal. Global Ecology and Biogeography 11: 291-301.

Welsh, H.H. Jr., Hodgson, G.R. & Lind, A.J., 2005. Ecography of the herpetofauna of a northern California watershed: linking species patterns to landscape processes. Ecography 28: 521-536.

Whitfield, J., 2002. Nuetrality versus the niche. Nature 417: 480-481.

Willig, M.R., Kaufman, D.M. & Stevens, R.D., 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. Annual Review of Ecology, Evolution and Systematics 34: 273-309.

Zapata, F.A., Gaston, K.J. & Chown, S.L., 2003. Mid-domain models of species richness gradients: assumptions, methods and evidence. Journal of Animal Ecology 72: 677-690.

Appendix 3A. List of the 431 tree species from the USDA forest database: species with * have been excluded from the analyses.

Scientific name	**Common name**
Abies amabilis	Pacific silver fir
Abies balsamea	Balsam fir
Abies concolor	White fir
Abies grandis	Grand fir
Abies lasiocarpa	Subalpine fir
Abies magnifica	California red fir
Abies procera	Noble fir
Acacia greggii	Catclaw acacia
Acer barbatum	Florida maple
Acer circinatum	Vine maple
Acer glabrum	Rocky Mountain maple
Acer grandidentatum	Bigtooth maple
Acer leucoderme	Chalk maple
Acer macrophyllum	Bigleaf maple
Acer negundo	Boxelder
Acer nigrum	Black maple
Acer pensylvanicum	Striped maple
Acer rubrum	Red maple
Acer saccharinum	Silver maple
Acer saccharum	Sugar maple
Acer spicatum	Mountain maple
Aesculus glabra	Ohio buckeye
Aesculus octandra	Yellow buckeye
Aesculus parviflora	Bottlebrush buckeye
Aesculus pavia	Red buckeye
Aesculus sylvatica	Painted buckeye
Alnus oblongifolia	Arizona alder
Alnus rhombifolia	White alder
Alnus rubra	Red alder
Alnus serrulata	Hazel alder
Alnus sinuata	Sitka alder
Alnus tenuifolia	Thinleaf alder
Amelanchier alnifolia	Western serviceberry
Amelanchier arborea	Downy serviceberry
Amelanchier interior	Inland serviceberry
Amelanchier sanguinea	Roundleaf serviceberry
Amelanchier utahensis	Utah serviceberry
Aralia spinosa	Devils-walkingstick
Arbutus arizonica	Arizona madrone
Arbutus menziesii	Pacific madrone
Artemisia tridentata	Big sagebrush
Asimina triloba	Pawpaw
Betula alleghaniensis	Yellow birch

Betula lenta	Sweet birch
Betula nigra	River birch
Betula occidentalis	Water birch
Betula papyrifera	Paper birch
Bumelia lanuginosa	Gum bumelia
Bumelia lycioides	Buckthorn bumelia
Bumelia tenax	Tough bumelia
Carpinus caroliniana	American hornbeam
Carya aquatica	Water hickory
Carya cordiformis	Bitternut hickory
Carya glabra	Pignut hickory
Carya illinoensis	Pecan
Carya laciniosa	Shellbark hickory
Carya ovata	Shagbark hickory
Carya pallida	Sand hickory
Carya tomentosa	Mockernut hickory
Castanea alnifolia	Florida chinkapin
Castanopsis chrysophylla	Golden chinkapin
Castanea dentata	American chestnut
Castanea ozarkensis	Ozark chinkapin
Castanea pumila	Allegheny chinkapin
Catalpa bignonioides	Southern catalpa
Ceanothus thyrsiflorus	Blueblossom
Celtis laevigata	Sugarberry
Celtis occidentalis	Hackberry
Celtis reticulata	Netleaf hackberry
Celtis tenuifolia	Georgia hackberry
Cephalanthus occidentalis	Common buttonbush
Cercocarpus betuloides	Birchleaf cercocarpus
Cercocarpus breviflorus	Hairy cercocarpus
Cercis canadensis	Eastern redbud
Cercocarpus ledifolius	Curlleaf cercocarpus
Cercidium microphyllum	Yellow paloverde
Cercis occidentalis	California redbud
*Cercocarpus traskiae**	Catalina cercocarpus
Cereus giganteus	Saguaro
Chamaecyparis lawsoniana	Port-Orford-cedar
Chamaecyparis nootkatensis	Alaska-cedar
Chamaecyparis thyoides	Atlantic white-cedar
Chionanthus virginicus	Fringetree
Cladrastis kentukea	Yellowwood
Clethra acuminata	Cinnamon clethra
Cliftonia monophylla	Buckwheat-tree
Cornus alternifolia	Alternate-leaf dogwood
Cornus drummondii	Roughleaf dogwood
Cornus florida	Flowering dogwood
Cornus glabrata	Brown dogwood
Cornus nuttallii	Pacific dogwood

Cornus occidentalis	Western dogwood
Cornus stolonifera	Red-osier dogwood
Cornus stricta	Stiffcornel dogwood
Corylus cornuta	Beaked hazel
Cotinus obovatus	American smoketree
Cowania mexicana	Cliffrose
Crataegus columbiana	Columbia hawthorn
Crataegus douglasii	Black hawthorn
Cupressus arizonica	Arizona cypress
Cupressus bakeri	Modoc cypress
Cyrilla racemiflora	Swamp cyrilla
Diospyros virginiana	Common persimmon
Elliottia racemosa	Elliottia
Erythrina flabelliformis	Southwestern coralbean
Euonymus atropurpureus	Eastern wahoo
Euonymus occidentalis	Western wahoo
Fagus grandifolia	American beech
Forestiera acuminata	Swamp-privet
Forestiera angustifolia	Texas forestiera
Forestiera phillyreoides	Desert-olive forestiera
Forestiera segregata	Florida-privet
*Franklinia alatamaha**	Franklinia
Fraxinus americana	White ash
Fraxinus anomala	Singleleaf ash
Fraxinus berlandieriana	Berlandier ash
Fraxinus caroliniana	Carolina ash
Fraxinus cuspidata	Fragrant ash
Fraxinus dipetala	Two-petal ash
Fraxinus gooddingii	Goodding ash
Fraxinus greggii	Gregg ash
Fraxinus latifolia	Oregon ash
Fraxinus nigra	Black ash
Fraxinus papillosa	Chihuahua ash
Fraxinus pennsylvanica	Green ash
Fraxinus profunda	Pumpkin ash
Fraxinus quadrangulata	Blue ash
Fraxinus texensis	Texas ash
Fraxinus velutina	Velvet ash
Garrya elliptica	Wavyleaf silktassel
Gleditsia aquatica	Waterlocust
Gleditsia triacanthos	Honeylocust
Gordonia lasianthus	Loblolly-bay
Halesia carolina	Carolina silverbell
Halesia diptera	Two-wing silverbell
Halesia parviflora	Little silverbell
Hamamelis virginiana	Witch-hazel
Ilex ambigua	Carolina holly
Ilex amelanchier	Sarvis holly

Ilex cassine	Dahoon
Ilex coriacea	Large gallberry
Ilex decidua	Possumhaw
Ilex longipes	Georgia holly
Ilex montana	Mountain winterberry
Ilex myrtifolia	Myrtle dahoon
Ilex opaca	American holly
Ilex verticillata	Common winterberry
Ilex vomitoria	Yaupon
Illicium floridanum	Florida anise-tree
Juglans cinerea	Butternut
Juglans major	Arizona walnut
Juglans nigra	Black walnut
Juniperus ashei	Ashe juniper
Juniperus californica	California juniper
Juniperus communis	Common juniper
Juniperus deppeana	Alligator juniper
Juniperus flaccida	Drooping juniper
Juniperus horizontalis	Creeping juniper
Juniperus monosperma	One-seed juniper
Juniperus occidentalis	Western juniper
Juniperus osteosperma	Utah juniper
Juniperus pinchotii	Pinchot juniper
Juniperus scopulorum	Rocky Mountain juniper
Juniperus silicicola	Southern redcedar
Juniperus virginiana	Eastern redcedar
Kalmia latifolia	Mountain-laurel
Larix laricina	Tamarack
Larix lyallii	Subalpine larch
Larix occidentalis	Western larch
Leitneria floridana	Corkwood
Libocedrus decurrens	Incense-cedar
Liquidambar styraciflua	Sweetgum
Liriodendron tulipifera	Yellow-poplar
Lithocarpus densiflorus	Tanoak
*Lyonothamnus floribundus**	Lyontree
Magnolia acuminata	Cucumbertree
Magnolia fraseri	Fraser magnolia
Magnolia macrophylla	Bigleaf magnolia
Magnolia pyramidata	Pyramid magnolia
Magnolia tripetala	Umbrella magnolia
Magnolia virginiana	Sweetbay
Malus angustifolia	Southern crab apple
Malus coronaria	Sweet crab apple
Malus diversifolia	Oregon crab apple
Malus ioensis	Prairie crab apple
Morus microphylla	Texas mulberry
Morus rubra	Red mulberry

Myrica californica	Pacific bayberry
Myrica cerifera	Southern bayberry
Myrica heterophylla	Evergreen bayberry
Myrica inodora	Odorless bayberry
Nyssa aquatica	Water tupelo
Nyssa ogeche	Ogeechee tupelo
Nyssa sylvatica	Black tupelo, blackgum
Osmanthus americanus	Devilwood
Ostrya knowltonii	Knowlton hophornbeam
Ostrya virginiana	Eastern hophornbeam
Oxydendrum arboreum	Sourwood
Persea borbonia	Redbay
Picea breweriana	Brewer spruce
Picea chihuahuana	Chihuahua spruce
Picea engelmannii	Engelmann spruce
Picea glauca	White spruce
Picea mariana	Black spruce
Picea pungens	Blue spruce
Picea rubens	Red spruce
*Picea sitchensis**	Sitka spruce
Pinckneya pubens	Pinckneya
Pinus albicaulis	Whitebark pine
Pinus aristata	Bristlecone pine
Pinus attenuata	Knobcone pine
Pinus ayacahuite	Mexican white pine
Pinus balfouriana	Foxtail pine
Pinus banksiana	Jack pine
*Pinus caribaea**	Caribbean pine
Pinus cembroides	Mexican pinyon
Pinus clausa	Sand pine
Pinus contorta	Lodgepole pine
Pinus cooperi	Cooper pine
Pinus coulteri	Coulter pine
*Pinus cubensis**	Cuban pine
Pinus culminicola	Potosi pinyon
Pinus douglasiana	Douglas pine
Pinus durangensis	Durango pine
Pinus echinata	Shortleaf pine
Pinus edulis	Pinyon
Pinus elliottii	Slash pine
Pinus engelmannii	Apache pine
Pinus flexilis	Limber pine
Pinus glabra	Spruce pine
Pinus greggii	Gregg pine
Pinus hartwegii	Hartweg pine
Pinus jeffreyi	Jeffrey pine
Pinus lambertiana	Sugar pine
Pinus lawsonii	Lawson pine

Pinus leiophylla	Chihuahua pine
Pinus longaeva	Intermountain bristlecone pine
Pinus lumholtzii	Lumholtz pine
Pinus maximartinezii	Martínez pinyon
Pinus michoacana	Michoacán pine
Pinus monophylla	Singleleaf pinyon
Pinus monticola	Western white pine
Pinus montezumae	Montezuma pine
Pinus muricata	Bishop pine
Pinus nelsonii	Nelson pinyon
*Pinus occidentalis**	West Indian pine
Pinus oocarpa	Nicaraguan pitch pine; ocote pine
Pinus palustris	Longleaf pine
Pinus patula	Mexican weeping pine
Pinus pinceana	Pince pinyon
Pinus ponderosa	Ponderosa pine
Pinus pringlei	Pringle pine
Pinus pseudostrobus	False Weymouth pine
Pinus pungens	Table-Mountain pine
Pinus quadrifolia	Parry pinyon
Pinus radiata	Monterey pine
Pinus resinosa	Red pine
Pinus rigida	Pitch pine
Pinus sabiniana	Digger pine
Pinus serotina	Pond pine
Pinus strobus	Eastern white pine
Pinus strobiformis	Southwestern white pine
Pinus taeda	Loblolly pine
Pinus teocote	Twisted -leaves pine; Aztec pine
*Pinus torreyana**	Torrey pine
*Pinus tropicalis**	Tropical pine
Pinus virginiana	Virginia pine
Pinus washoensis	Washoe pine
Planera aquatica	Planertree
Platanus occidentalis	American sycamore
Platanus wrightii	Arizona sycamore
Populus angustifolia	Narrowleaf cottonwood
Populus balsamifera	Balsam poplar
Populus deltoides	Eastern cottonwood
Populus fremontii	Fremont cottonwood
Populus grandidentata	Bigtooth aspen
Populus heterophylla	Swamp cottonwood
Populus tremuloides	Quaking aspen
Populus trichocarpa	Black cottonwood
Prosopis juliflora	Mesquite
Prosopis pubescens	Screwbean mesquite
Prunus alleghaniensis	Allegheny plum

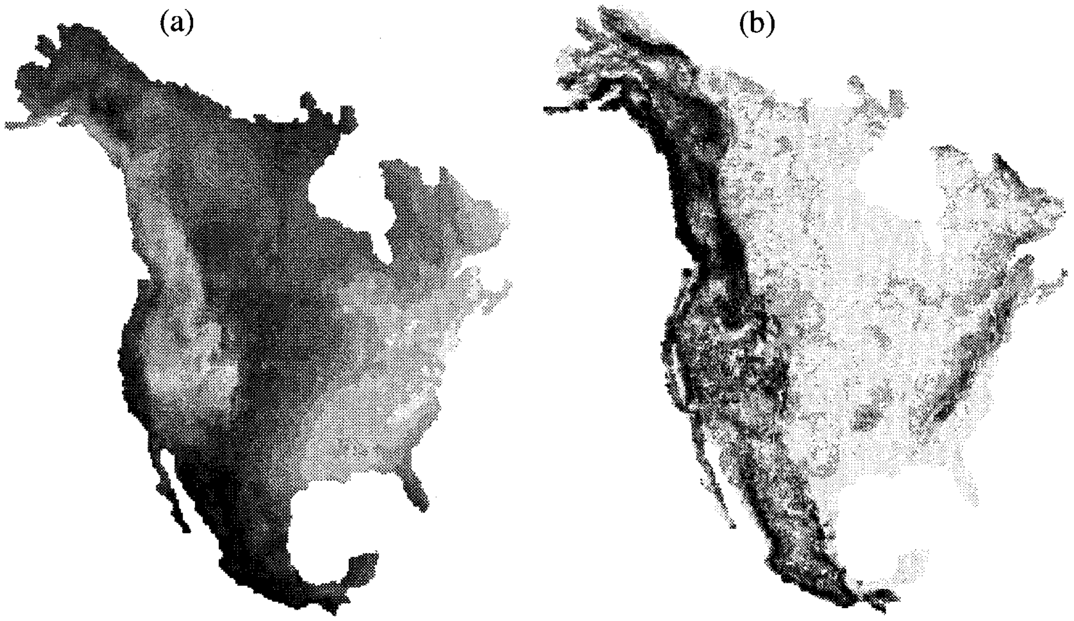
Prunus americana	American plum
Prunus angustifolia	Chickasaw plum
Prunus caroliniana	Carolina laurelcherry
Prunus emarginata	Bitter cherry
Prunus fremontii	Desert apricot
Prunus hortulana	Hortulan plum
Prunus ilicifolia	Hollyleaf cherry
*Prunus lyonii**	Catalina cherry
Prunus mexicana	Mexican plum
Prunus munsoniana	Wildgoose plum
Prunus myrtifolia	West Indies cherry
Prunus nigra	Canada plum
Prunus pensylvanica	Pin cherry
Prunus serotina	Black cherry
Prunus subcordata	Klamath plum
Prunus umbellata	Flatwoods plum
Prunus virginiana	Common chokecherry
Pseudotsuga menziesii	Douglas-fir
Ptelea trifoliata	Common hoptree
Quercus agrifolia	California live oak
Quercus ajoensis	Ajo oak
Quercus alba	White oak
Quercus arizonica	Arizona white oak
Quercus arkansana	Arkansas oak
Quercus bicolor	Swamp white oak
Quercus chapmanii	Chapman oak
Quercus chrysolepis	Canyon live oak
Quercus coccinea	Scarlet oak
Quercus douglasii	Blue oak
Quercus dunnii	Dunn oak
Quercus durandii	Durand oak
Quercus ellipsoidalis	Northern pin oak
Quercus emoryi	Emory oak
Quercus engelmannii	Engelmann oak
Quercus falcata	Southern red oak
Quercus gambelii	Gambel oak
Quercus garryana	Oregon white oak
Quercus georgiana	Georgia oak
Quercus glaucoides	Lacey oak
Quercus graciliformis	Chisos oak
Quercus gravesii	Graves oak incl. Q. tardifolia (Muller) lateleaf oak
Quercus grisea	gray oak
Quercus havardii	Havard oak
Quercus hypoleucoides	Silverleaf oak
Quercus ilicifolia	Bear oak
Quercus imbricaria	Shingle oak
Quercus incana	Bluejack oak
Quercus kelloggii	California black oak

Quercus laevis	Turkey oak
Quercus laurifolia	Laurel oak
Quercus lobata	California white oak
Quercus lyrata	Overcup oak
*Quercus macdonaldii**	McDonald oak
Quercus macrocarpa	Bur oak
Quercus marilandica	Blackjack oak
Quercus michauxii	Swamp chestnut oak
Quercus mohriana	Mohrs oak
Quercus muehlenbergii	Chinkapin oak
Quercus myrtifolia	Myrtle oak
Quercus nigra	Water oak
Quercus nuttallii	Nuttall oak
Quercus oblongifolia	Mexican blue oak
Quercus oglethorpensis	Oglethorpe oak
Quercus palustris	Pin oak
Quercus phellos	Willow oak
Quercus prinus	Chestnut oak
Quercus pungens	Sandpaper oak
Quercus rubra	Northern red oak
Quercus rugosa	Netleaf oak
Quercus shumardii	Shumard oak
Quercus stellata	Post oak
*Quercus tomentella**	Island live oak
Quercus toumeyi	Toumey oak
Quercus turbinella	Shrub live oak
Quercus velutina	Black oak
Quercus virginiana	Live oak
Quercus wislizeni	Interior live oak
Rhamnus betulaefolia	Birchleaf buckthorn
Rhamnus californica	California buckthorn
Rhamnus caroliniana	Carolina buckthorn
Rhamnus purshiana	Cascara buckthorn
Rhododendron catawbiense	Catawba rhododendron
Rhododendron macrophyllum	Pacific rhododendron
Rhus copallina	Shining sumac
Rhus glabra	Smooth sumac
Rhus typhina	Staghorn sumac
Robinia neomexicana	New Mexican locust
Robinia pseudoacacia	Black locust
Sabal palmetto	Cabbage palmetto
Salix amygdaloides	Peachleaf willow
Salix bebbiana	Bebb willow
Salix caroliniana	Coastal Plain willow
Salix exigua	Coyote willow
Salix floridana	Florida willow
Salix geyeriana	Geyer willow
Salix hindsiana	Hinds willow

Salix hookeriana	Hooker willow
Salix lasiolepis	Arroyo willow
Salix lucida	Shining willow
Salix mackenzieana	Mackenzie willow
Salix nigra	Black willow
Salix scouleriana	Scouler willow
Salix sericea	Silky willow
Salix sessilifolia	Northwest willow
Salix sitchensis	Sitka willow
Sambucus callicarpa	Pacific red elder
Sambucus canadensis	American elder
Sambucus glauca	Blueberry elder
Sambucus melanocarpa	Blackbead elder
Sapindus saponaria	Wingleaf soapberry
Sassafras albidum	Sassafras
Sequoia sempervirens	Redwood
Shepherdia argentea	Silver buffaloberry
Sorbus americana	American mountain-ash
Sorbus decora	Showy mountain-ash
Sorbus scopulina	Greene mountain-ash
Sorbus sitchensis	Sitka mountain-ash
Staphylea trifolia	American bladdernut
Stewartia malacodendron	Virginia stewartia
Stewartia ovata	Mountain stewartia
Styrax americana	American snowbell
Styrax grandifolia	Bigleaf snowbell
Symplocos tinctoria	Common sweetleaf
Taxodium distichum	Baldcypress
Taxus brevifolia	Pacific yew
Thuja occidentalis	Northern white-cedar
Thuja plicata	Western redcedar
Tilia americana	American basswood
Torreya taxifolia	Florida torreya
Toxicodendron vernix	Poison-sumac
Tsuga canadensis	Eastern hemlock
Tsuga caroliniana	Carolina hemlock
Tsuga heterophylla	Western hemlock
Tsuga mertensiana	Mountain hemlock
Ulmus alata	Winged elm
Ulmus americana	American elm
Ulmus rubra	Slippery elm
Ulmus serotina	September elm
Umbellularia californica	California-laurel
Vaccinium arboreum	Tree sparkleberry
Vauquelinia californica	Torrey vauquelinia
*Vauquelinia pauciflora**	Fewflower vauquelinia
Viburnum nudum	Possumhaw viburnum
Viburnum obovatum	Walter viburnum

Viburnum prunifolium	Blackhaw
Viburnum rufidulum	Rusty blackhaw
Yucca brevifolia	Joshua-tree
Zanthoxylum americanum	Common prickly-ash
Zanthoxylum clava-herculis	Hercules-club

Appendix 3B. Map of precipitation seasonality and slope



Appendix 3B.1. Patterns of variation in (a) precipitation seasonality (lighter shading indicates lower seasonality) and (b) slope (lighter shading indicates less slope) across North America

Appendix 4A. A list of the 19 interpolated global climate layers WorldClim Version 1.3 (Hijmans et al. 2005), the variable name used to represent each and the number of times they were used in the spatial regression models used to examine patterns of species richness for each of three groups of species (amphibians, birds and mammals).

	Variable ID	all	Amphibian	Bird	Mammal
Annual Mean Temperature	Bio01	2	0	1	1
Mean Diurnal Range (Mean of monthly (max temp - min temp))	Bio02	7	4	1	2
Isothermality (Bio02 / Bio07) (* 100)	Bio03	11	5	5	1
Temperature Seasonality (standard deviation)	Bio04	7	2	4	1
Max Temperature of Warmest Month	Bio05	7	2	2	3
Min Temperature of Coldest Month	Bio06	3	0	3	0
Temperature Annual Range (Bio05-Bio06)	Bio07	4	2	1	1
Mean Temperature of Wettest Quarter	Bio08	6	0	2	4
Mean Temperature of Driest Quarter	Bio09	3	0	1	2
Mean Temperature of Warmest Quarter	Bio10	8	2	2	4
Mean Temperature of Coldest Quarter	Bio11	3	0	2	1
Annual Precipitation	Bio12	8	4	3	1
Precipitation of Wettest Month	Bio13	4	3	0	1
Precipitation of Driest Month	Bio14	2	1	1	0
Precipitation Seasonality (Coefficient of Variation)	Bio15	13	4	4	5

Precipitation of Wettest Quarter	Bio16	11	4	2	5
Precipitation of Driest Quarter	Bio17	10	2	3	5
Precipitation of Warmest Quarter	Bio18	9	4	3	2
Precipitation of Coldest Quarter	Bio19	5	3	1	1

Appendix 4B. A list of the climatic variables of the 19 interpolated global climate layers WorldClim Version 1.3 (Hijmans et al. 2005) used in the spatial regression models examining patterns of species richness for each of three groups of species (amphibians, birds and mammals). See Appendix 4A for a description of variables.

	Longitude	Latitude	Altitude	Long x Lat	Long x Alt	Lat x Alt	Long x Lat x Alt
Amphibian	Bio02	Bio02	Bio03	Bio02	Bio12	Bio03	Bio02
	Bio15	Bio03	Bio04	Bio03	Bio14	Bio04	Bio03
	Bio16	Bio05	Bio05	Bio12	Bio15	Bio07	Bio12
	Bio17	Bio10	Bio07	Bio13	Bio17	Bio12	Bio13
	Bio18	Bio15	Bio10	Bio16	Bio18	Bio13	Bio16
	Bio19	Bio19	Bio15	Bio18	Bio19	Bio16	Bio18
Bird	Bio03	Bio02	Bio04	Bio03	Bio03	Bio01	Bio03
	Bio08	Bio03	Bio05	Bio04	Bio12	Bio04	Bio04
	Bio10	Bio05	Bio07	Bio06	Bio14	Bio06	Bio06
	Bio12	Bio15	Bio08	Bio16	Bio15	Bio09	Bio11
	Bio15	Bio17	Bio10	Bio18	Bio17	Bio11	Bio12
	Bio17		Bio15	Bio19	Bio18	Bio16	Bio18
Mammal	Bio03	Bio02	Bio04	Bio02	Bio12	Bio01	Bio05
	Bio08	Bio10	Bio05	Bio05	Bio15	Bio08	Bio08
	Bio15	Bio15	Bio07	Bio10	Bio16	Bio09	Bio09
	Bio16	Bio16	Bio08	Bio15	Bio17	Bio11	Bio10
	Bio17	Bio17	Bio10	Bio16	Bio18	Bio13	Bio16
	Bio19		Bio15	Bio17		Bio18	Bio17

Appendix 6A. The *a priori* and *a posteriori* probabilities derived from a naïve Bayesian classifier, used for modeling probability of occurrence of groups of rare species. First the *a priori* probability is reported, followed by the *a posteriori* probabilities. For point values (PV), the *a posteriori* probability (conditional probability) is reported, whereas for all others, the mean and SD for the environmental attribute is reported such that the probability can be calculated.

Appendix 6A.1. The *a priori* and *a posteriori* probabilities derived for rare bird species.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.6122	.	0.3878	.
Landcover	Bedrock outcrop	4800	0.0085	0.0592	0.2995	0.6420
		PV	0.0000	.	0.0130	.
	Cropland	300	92.0114	14.1795	37.5868	31.3582
		PV	0.9504	.	0.2844	.
	Deciduous mixed forest	PV	0.0021	.	0.0000	.
	Deciduous swamp	600	0.0700	0.2813	1.5887	5.3411
		PV	0.0000	.	0.0189	.
	Dense deciduous forest	150	2.9074	8.5594	35.8857	36.6207
		PV	0.0191	.	0.4499	.
	Developed land	PV	0.0000	.	0.0499	.
	Freshwater marsh	PV	0.0000	.	0.0546	.
	Pasture	PV	0.0167	.	0.0140	.
	Sparse deciduous forest	PV	0.0000	.	0.0073	.
	Water	PV	0.0118	.	0.1079	.

Appendix 6A.2. The *a priori* and *a posteriori* probabilities derived for rare insect species.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.5984	.	0.4016	.
Land classification for agriculture	Class 1	PV	0.0659	.	0.1249	.
	Class 2	600	80.4892	29.4326	51.6618	41.5331
		PV	0.8398	.	0.4566	.
	Class 3	PV	0.0701	.	0.2280	.
	Class 4	PV	0.0000	.	0.0055	.
	Class 5	PV	0.0143	.	0.0179	.
	Class 7	PV	0.0098	.	0.1671	.
Landcover	Bedrock outcrop	PV	0.0000	.	0.0199	.
	Coniferous mixed forest	PV	0.0000	.	0.0024	.
	Cropland	4800	89.2228	7.9175	63.5910	25.3000
		PV	0.9222	.	0.5778	.
	Deciduous mixed forest	PV	0.0000	.	0.0000	.
	Deciduous swamp	PV	0.0000	.	0.0598	.
	Dense coniferous forest	PV	0.0000	.	0.0089	.
	Dense deciduous forest	PV	0.0452	.	0.1789	.
	Developed land	PV	0.0182	.	0.0506	.
	Freshwater marsh	PV	0.0000	.	0.0107	.
	Open fen	PV	0.0000	.	0.0062	.
	Pasture	PV	0.0000	.	0.0067	.
	Sparse deciduous forest	PV	0.0143	.	0.0052	.
	Water	900	0.2646	1.0242	12.7828	21.1178
		PV	0.0000	.	0.0730	.
Soil texture	Clay	PV	0.4116	.	0.1837	.
	Clay loam	PV	0.0524	.	0.0000	.
	Fluvial deposits	PV	0.0143	.	0.0000	.
	Gravel	1200	0.8256	3.5348	11.3370	21.0516

	PV	0.0000	.	0.1967	.
Organics	PV	0.0000	.	0.0473	.
Sand	PV	0.0416	.	0.0702	.
Sandy loam	PV	0.2029	.	0.2267	.
Silt	PV	0.0000	.	0.0199	.
Silt loam	PV	0.1064	.	0.0504	.
Silty clay	PV	0.1707	.	0.1654	.
Water	PV	0.0000	.	0.0397	.

Appendix 6A.3. The *a priori* and *a posteriori* probabilities derived for rare plant species.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.6154	.	0.3846	.
Land classification for agriculture	Class 0	PV	0.0122	.	0.0016	.
	Class 1	PV	0.1668	.	0.1138	.
	Class 2	PV	0.6639	.	0.5523	.
	Class 3	PV	0.1013	.	0.1342	.
	Class 4	PV	0.0022	.	0.0042	.
	Class 5	PV	0.0165	.	0.0319	.
	Class 7	PV	0.0371	.	0.1621	.
Landcover	Bedrock outcrop	PV	0.0000	.	0.0174	.
	Conifer swamp	PV	0.0000	.	0.0006	.
	Coniferous mixed forest	PV	0.0000	.	0.0000	.
	Coniferous plantation	PV	0.0007	.	0.0011	.
	Cropland	9600	80.3891	14.6131	63.9057	23.3579
		PV	0.8482	.	0.4290	.
	Deciduous mixed forest	PV	0.0023	.	0.0047	.
	Deciduous swamp	9600	0.1172	0.1440	0.2448	0.2840
		PV	0.0000	.	0.0253	.
	Dense coniferous forest	PV	0.0018	.	0.0090	.
	Dense deciduous forest	PV	0.0808	.	0.3080	.
	Developed land	PV	0.0081	.	0.0327	.
	Freshwater marsh	PV	0.0048	.	0.0116	.
	Open fen	PV	0.0000	.	0.0037	.
	Pasture	PV	0.0122	.	0.0321	.
	Sparse deciduous forest	PV	0.0113	.	0.0120	.
	Unclassified	PV	0.0000	.	0.0016	.
	Water	PV	0.0299	.	0.1114	.
Soil texture	Bedrock	PV	0.0004	.	0.0016	.

Clay	PV	0.3743	.	0.1775	.
Clay loam	PV	.	.	0.0672	.
Fluvial deposits	PV	0.0014	.	0.0080	.
Gravel	PV	0.0254	.	0.0618	.
Organics	PV	0.0104	.	0.0487	.
Sand	7200	1.7787	3.8205	5.7691	9.0028
	PV	0.0106	.	0.0953	.
Sandy loam	PV	0.2000	.	0.2679	.
Silt	PV	0.0059	.	0.0258	.
Silt loam	PV	0.0932	.	0.0316	.
Silty clay	PV	0.1437	.	0.1485	.
Water	PV	0.0201	.	0.0661	.

Appendix 6A.4. The *a priori* and *a posteriori* probabilities derived for rare reptile species.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.5000	.	0.5000	.
Land classification for agriculture	Class 0	PV	0.0000	.	0.0365	.
	Class 1	PV	0.1061	.	0.0737	.
	Class 2	PV	0.8484	.	0.5665	.
	Class 3	PV	0.0337	.	0.0977	.
	Class 5	PV	0.0039	.	0.0309	.
	Class 7	9600	1.8189	4.1613	8.5433	10.3056
		PV	0.0079	.	0.1946	.
Landcover	Coniferous mixed forest	PV	0.0000	.	0.0000	.
	Coniferous plantation	PV	0.0000	.	0.0038	.
	Cropland	600	89.6395	12.5130	49.4691	33.3813
		PV	0.9249	.	0.5399	.
	Deciduous mixed forest	PV	0.0000	.	0.0161	.
	Deciduous swamp	900	0.0350	0.1793	1.2786	2.6477
		PV	0.0000	.	0.0349	.
	Dense deciduous forest	PV	0.0493	.	0.1306	.
	Developed land	PV	0.0124	.	0.0642	.
	Freshwater marsh	PV	0.0000	.	0.0453	.
	Open fen	PV	0.0000	.	0.0091	.
	Pasture	PV	0.0000	.	0.0365	.
	Sparse deciduous forest	PV	0.0134	.	0.0204	.
	Water	4800	2.3270	5.2189	25.3645	28.0846
		PV	0.0000	.	0.0992	.

Appendix 6A.5. The *a priori* and *a posteriori* probabilities derived for rare tree species.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.5922	.	0.4078	.
Landcover	Bedrock outcrop	PV	0.0000	.	0.0021	.
	Conifer swamp	PV	0.0000	.	0.0015	.
	Cropland	90	87.6856	26.4828	48.7803	35.7744
		PV	0.9083	.	0.4737	.
	Deciduous mixed forest	PV	0.0000	.	0.0181	.
	Deciduous swamp	PV	0.0000	.	0.0050	.
	Dense coniferous forest	PV	0.0000	.	0.0077	.
	Dense deciduous forest	PV	0.0645	.	0.3933	.
	Developed land	PV	0.0058	.	0.0040	.
	Freshwater marsh	PV	0.0070	.	0.0133	.
	Pasture	PV	0.0018	.	0.0514	.
	Sparse deciduous forest	PV	0.0000	.	0.0069	.
	Water	PV	0.0126	.	0.0231	.
Distance-to-roads	Mean	7200	338.7711	323.4827	645.8037	719.7808
Slope	SE	600	0.6177	0.8180	1.0470	1.1800
Soil texture	Clay	PV	0.3648	.	0.2271	.
	Clay loam	PV	0.1050	.	0.0562	.
	Fluvial deposits	PV	0.0051	.	0.0171	.
	Gravel	PV	0.0232	.	0.0399	.
	Organics	PV	0.0178	.	0.0275	.
	Sand	7200	1.8485	3.7604	5.7101	8.4532
		PV	0.0354	.	0.1114	.
	Sandy loam	PV	0.1950	.	0.2933	.
	Silt	PV	0.0028	.	0.0168	.
	Silt loam	PV	0.0730	.	0.0237	.
	Silty clay	PV	0.1688	.	0.1518	.
	Water	PV	0.0091	.	0.0351	.

Appendix 6A.6. The *a priori* and *a posteriori* probabilities derived for rare shrub species.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.6364	.	0.3636	.
Aspect	Mean	7200	196.2710	16.2678	190.7824	16.0159
Landcover	Cropland	60	85.0116	29.6315	49.5345	42.8580
		PV	0.8989	.	0.4907	.
	Deciduous mixed forest	PV	0.0000	.	0.0040	.
	Deciduous swamp	PV	0.0000	.	0.0000	.
	Dense deciduous forest	PV	0.0475	.	0.3476	.
	Developed land	7200	1.1490	3.5280	3.7022	7.9092
		PV	0.0036	.	0.0241	.
	Freshwater marsh	PV	0.0267	.	0.0149	.
	Pasture	900	1.3262	1.6113	3.4750	4.0078
		PV	0.0100	.	0.0476	.
	Sparse deciduous forest	PV	0.0000	.	0.0067	.
	Water	PV	0.0133	.	0.0643	.
Distance-to-roads	Mean	7200	360.8836	349.8311	699.6482	781.5955

Appendix 6A.7. The *a priori* and *a posteriori* probabilities derived for rare herb species.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.5556	.	0.4444	.
Land classification for agriculture	Class 0	4800	2.0202	6.2190	12.5717	19.2857
Landcover	Bedrock outcrop	7200	0.0414	0.1317	0.1842	0.3870
		PV	0.0000	.	0.0119	.
	Conifer swamp	PV	0.0000	.	0.0009	.
	Coniferous mixed forest	PV	.	.	0.0037	.
	Coniferous plantation	PV	0.0000	.	0.0016	.
	Cropland	4800	84.1299	14.3109	62.4430	27.4496
		PV	0.8854	.	0.4760	.
	Deciduous mixed forest	PV	0.0000	.	0.0069	.
	Deciduous swamp	9600	0.1235	0.1647	0.2496	0.2822
		PV	0.0020	.	0.0221	.
	Dense coniferous forest	PV	0.0000	.	0.0066	.
	Dense deciduous forest	PV	0.0679	.	0.2929	.
	Developed land	PV	0.0160	.	0.0321	.
	Freshwater marsh	PV	0.0029	.	0.0139	.
	Open fen	PV	0.0000	.	0.0069	.
	Pasture	PV	0.0086	.	0.0177	.
	Sparse deciduous forest	PV	0.0008	.	0.0112	.
	Unclassified	PV	0.0000	.	0.0019	.
	Water	PV	0.0163	.	0.0937	.
Soil texture	Sand	7200	1.8428	3.8970	5.9743	9.1378

Appendix 6A.8. The *a priori* and *a posteriori* probabilities derived for rare grass species.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.5735	.	0.4265	.
Landcover	Bedrock outcrop	9600	0.0419	0.1167	0.1667	0.2947
		PV	0.0000	.	0.0270	.
	Conifer swamp	PV	0.0000	.	0.0037	.
	Coniferous mixed forest	PV	0.0000	.	0.0059	.
	Cropland	600	88.2884	15.9600	55.7214	33.4230
		PV	0.8851	.	0.5234	.
	Deciduous swamp	PV	0.0035	.	0.0347	.
	Dense coniferous forest	PV	0.0000	.	0.0079	.
	Dense deciduous forest	450	6.1050	9.4218	17.3534	20.8114
		PV	0.0611	.	0.1899	.
	Developed land	PV	0.0155	.	0.0641	.
	Freshwater marsh	PV	0.0016	.	0.0124	.
	Pasture	PV	0.0233	.	0.0203	.
	Sparse deciduous forest	PV	0.0000	.	0.0254	.
	Unclassified	PV	0.0000	.	0.0000	.
	Water	PV	0.0099	.	0.0854	.
Distance-to-roads	Mean	7200	364.4151	345.2124	783.7989	831.0342
Soil texture	Clay	PV	0.4654	.	0.0949	.
	Clay loam	PV	0.0720	.	0.0922	.
	Fluvial deposits	PV	0.0012	.	0.0019	.
	Gravel	PV	0.0073	.	0.0655	.
	Organics	PV	0.0082	.	0.0729	.
	Sand	PV	0.0284	.	0.1221	.
	Sandy loam	PV	0.1656	.	0.3552	.
	Silt	PV	0.0027	.	0.0284	.
	Silt loam	PV	0.0947	.	0.0311	.
	Silty clay	PV	0.1339	.	0.0976	.
	Water	PV	0.0206	.	0.0382	.

Appendix 6A.9. The *a priori* and *a posteriori* probabilities derived for rare sedge species.

		Scale (m)	Absence Mean or conditional probability	Absence SD	Presence Mean or conditional probability	Presence SD
a priori			0.6359	.	0.3641	.
Landcover	Bedrock outcrop	PV	0.0000	.	0.0167	.
	Conifer swamp	PV	0.0000	.	0.0000	.
	Coniferous plantation	PV	0.0000	.	0.0000	.
	Cropland	300	87.9985	19.8070	44.8118	29.1154
		PV	0.9069	.	0.3597	.
	Deciduous mixed forest	PV	0.0000	.	0.0042	.
	Deciduous swamp	PV	0.0018	.	0.0074	.
	Dense coniferous forest	PV	0.0000	.	0.0029	.
	Dense deciduous forest	90	5.0048	14.6777	40.3087	38.3197
		PV	0.0493	.	0.4050	.
	Developed land	PV	0.0151	.	0.0212	.
	Freshwater marsh	PV	0.0067	.	0.0202	.
	Open fen	PV	0.0000	.	0.0028	.
	Pasture	PV	0.0060	.	0.0187	.
	Sparse deciduous forest	PV	0.0000	.	0.0035	.
	Water	PV	0.0142	.	0.1377	.
Distance-to-roads	SE	4800	253.7167	247.4097	629.5783	625.6881
Slope	SE	150	0.4777	0.8165	1.1125	1.4931
Soil texture	Sand	4800	1.8795	4.6875	6.9734	11.5486

Appendix 6A.10. The *a priori* and *a posteriori* probabilities derived for rare plant species with a subnational rarity rank of S1.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.6615	.	0.3385	.
Land classification for agriculture	Class 7	9600	2.7036	5.4252	8.7154	14.2093
Landcover	Bedrock outcrop	PV	0.0000	.	0.0078	.
	Coniferous mixed forest	PV	0.0000	.	0.0088	.
	Cropland	600	85.7912	18.8269	58.5522	33.4095
		PV	0.8529	.	0.5435	.
	Deciduous mixed forest	PV	0.0000	.	0.0091	.
	Deciduous swamp	9600	0.1280	0.1650	0.2729	0.2966
		PV	0.0101	.	0.0131	.
	Dense coniferous forest	PV	0.0000	.	0.0103	.
	Dense deciduous forest	300	8.7581	14.9993	20.5369	25.5282
		PV	0.0831	.	0.2223	.
	Developed land	PV	0.0192	.	0.0360	.
	Freshwater marsh	PV	0.0016	.	0.0160	.
	Pasture	PV	0.0113	.	0.0196	.
	Sparse deciduous forest	PV	0.0000	.	0.0134	.
	Unclassified	PV	0.0000	.	0.0098	.
	Water	PV	0.0218	.	0.0904	.
Distance-to-roads	Mean	9600	470.7072	500.3790	1262.1028	1384.3600

Appendix 6A.11. The *a priori* and *a posteriori* probabilities derived for rare plant species with a subnational rarity rank of S2.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.5570	.	0.4430	.
Landcover	Bedrock outcrop	7200	0.0437	0.1332	0.1747	0.3767
		PV	0.0000	.	0.0193	.
	Conifer swamp	PV	0.0000	.	0.0015	.
	Coniferous mixed forest	PV	0.0000	.	0.0042	.
	Coniferous plantation	PV	0.0000	.	0.0000	.
	Cropland	300	87.3552	20.6816	53.4868	35.1467
		PV	0.8657	.	0.5106	.
	Deciduous mixed forest	PV	0.0000	.	0.0097	.
	Deciduous swamp	PV	0.0000	.	0.0151	.
	Dense coniferous forest	PV	0.0000	.	0.0157	.
	Dense deciduous forest	90	8.6228	21.7043	27.2443	35.0037
		PV	0.0855	.	0.2656	.
	Developed land	PV	0.0179	.	0.0320	.
	Freshwater marsh	PV	0.0065	.	0.0099	.
	Open fen	PV	0.0000	.	0.0070	.
	Pasture	PV	0.0153	.	0.0174	.
	Sparse deciduous forest	PV	0.0021	.	0.0083	.
	Water	PV	0.0069	.	0.0836	.
Distance-to-roads	Mean	9600	456.2549	469.6473	1172.2800	1271.7911
Soil texture	Sand	7200	1.9122	3.9985	5.9050	9.3068

Appendix 6A.12. The *a priori* and *a posteriori* probabilities derived for rare plant species with a subnational rarity rank of S3.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.6708	.	0.3292	.
Land classification for agriculture	Class 0	4800	1.9347	6.6725	11.6316	17.6547
Landcover	Bedrock outcrop	PV	0.0000	.	0.0201	.
	Conifer swamp	PV	0.0000	.	0.0020	.
	Coniferous mixed forest	PV	0.0000	.	0.0000	.
	Coniferous plantation	PV	0.0000	.	0.0000	.
	Cropland	4800	84.6975	13.7837	64.9958	26.1526
		PV	0.8665	.	0.3978	.
	Deciduous mixed forest	PV	0.0000	.	0.0037	.
	Deciduous swamp	PV	0.0014	.	0.0200	.
	Dense coniferous forest	PV	0.0000	.	0.0106	.
	Dense deciduous forest	PV	0.0755	.	0.3716	.
	Developed land	PV	0.0149	.	0.0192	.
	Freshwater marsh	PV	0.0080	.	0.0055	.
	Open fen	PV	0.0000	.	0.0000	.
	Pasture	PV	0.0154	.	0.0332	.
	Sparse deciduous forest	PV	0.0000	.	0.0052	.
	Unclassified	PV	0.0000	.	0.0000	.
	Water	PV	0.0183	.	0.1113	.
Slope	SE	150	0.3922	0.6657	1.1437	1.3914
Soil texture	Gravel	2400	1.7051	4.6235	5.9561	10.8769
	Sand	4800	1.6986	4.4717	6.3802	10.8478

Appendix 6A.13. The *a priori* and *a posteriori* probabilities derived for rare herb species with a subnational rarity rank of S1.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.6346	.	0.3654	.
Land classification for agriculture	Class 0	2400	0.3111	1.3183	5.9548	10.7674
Landcover	Bedrock outcrop	PV	0.0000	.	0.0000	.
	Cropland	2400	84.9392	17.3905	59.2296	28.2845
		PV	0.8917	.	0.5500	.
	Deciduous mixed forest	PV	0.0000	.	0.0166	.
	Deciduous swamp	PV	0.0000	.	0.0097	.
	Dense coniferous forest	PV	0.0000	.	0.0156	.
	Dense deciduous forest	150	9.4423	19.1697	18.2705	30.4934
		PV	0.0639	.	0.1762	.
	Developed land	PV	0.0000	.	0.0691	.
	Freshwater marsh	PV	0.0094	.	0.0000	.
	Pasture	PV	0.0000	.	0.0170	.
	Sparse deciduous forest	PV	0.0000	.	0.0028	.
	Water	9600	10.5148	16.4815	20.8607	20.8478
		PV	0.0349	.	0.1428	.
Soil texture	Clay	PV	0.3970	.	0.1432	.
	Clay loam	PV	0.0815	.	0.1074	.
	Fluvial deposits	PV	0.0082	.	0.0000	.
	Gravel	PV	0.0492	.	0.1312	.
	Organics	PV	0.0364	.	0.0554	.
	Sand	9600	2.4184	4.3339	5.9506	7.0532
		PV	0.0375	.	0.1174	.
	Sandy loam	PV	0.1940	.	0.2390	.

Silt loam	PV	0.0457	.	0.0000	.
Silty clay	PV	0.1400	.	0.1547	.
Water	PV	0.0104	.	0.0518	.

Appendix 6A.14. The *a priori* and *a posteriori* probabilities derived for rare herb species with a subnational rarity rank of S2.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.6125	.	0.3875	.
Elevation	Mean	0	210.2204	32.0899	194.0221	24.2603
Landcover	Bedrock outcrop	PV	0.0000	.	0.0103	.
	Coniferous mixed forest	PV	0.0000	.	0.0078	.
	Cropland	300	87.4828	19.9662	54.5959	34.7476
		PV	0.8801	.	0.5270	.
	Deciduous mixed forest	PV	0.0000	.	0.0064	.
	Deciduous swamp	PV	0.0006	.	0.0182	.
	Dense coniferous forest	PV	0.0000	.	0.0183	.
	Dense deciduous forest	150	8.1507	18.0178	24.0576	30.3476
		PV	0.0773	.	0.2218	.
	Developed land	PV	0.0138	.	0.0493	.
	Freshwater marsh	PV	0.0059	.	0.0131	.
	Open fen	PV	0.0000	.	0.0153	.
	Pasture	PV	0.0145	.	0.0224	.
	Sparse deciduous forest	PV	0.0006	.	0.0132	.
	Water	PV	0.0072	.	0.0769	.
Distance-to-roads	Mean	9600	474.8483	547.0297	1155.9813	1153.7045
Soil texture	Sand	4800	1.7397	4.4140	7.8233	13.5172

Appendix 6A.15. The *a priori* and *a posteriori* probabilities derived for rare herb species with a subnational rarity rank of S3.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.6182	.	0.3818	.
Land classification for agriculture	Class 0	4800	1.9161	6.1377	11.4157	17.8684
Landcover	Bedrock outcrop	PV	0.0000	.	0.0238	.
	Conifer swamp	PV	0.0000	.	0.0000	.
	Coniferous plantation	PV	0.0000	.	0.0018	.
	Cropland	2400	86.5641	13.5479	61.6414	28.0311
		PV	0.9078	.	0.3448	.
	Deciduous mixed forest	PV	0.0000	.	0.0085	.
	Deciduous swamp	7200	0.1345	0.2105	0.3479	0.4714
		PV	0.0028	.	0.0228	.
	Dense coniferous forest	PV	0.0000	.	0.0199	.
	Dense deciduous forest	PV	0.0534	.	0.4196	.
	Developed land	PV	0.0123	.	0.0282	.
	Freshwater marsh	PV	0.0080	.	0.0070	.
	Open fen	PV	0.0000	.	0.0000	.
	Pasture	PV	0.0063	.	0.0163	.
	Sparse deciduous forest	PV	0.0014	.	0.0074	.
	Unclassified	PV	0.0000	.	0.0064	.
	Water	PV	0.0080	.	0.0936	.
Slope	SE	150	0.3437	0.6179	1.1892	1.4879
Soil texture	Sand	4800	2.2789	5.4937	7.7037	13.2102

Appendix 6A.16. The *a priori* and *a posteriori* probabilities derived for rare aquatic plant species.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.6029	.	0.3971	.
Land classification for agriculture	Class 7	9600	2.3565	4.8231	6.2670	9.6625
Landcover	Bedrock outcrop	PV	0.0000	.	0.0051	.
	Conifer swamp	PV	0.0000	.	0.0057	.
	Cropland	450	88.9912	16.2757	51.1343	30.3393
		PV	0.9145	.	0.4727	.
	Deciduous mixed forest	PV	0.0000	.	0.0059	.
	Deciduous swamp	7200	0.1138	0.1837	0.3266	0.4498
		PV	0.0009	.	0.0227	.
	Dense coniferous forest	PV	0.0000	.	0.0261	.
	Dense deciduous forest	90	6.7540	18.7510	26.0399	35.1898
		PV	0.0508	.	0.2509	.
	Developed land	PV	0.0098	.	0.0296	.
	Freshwater marsh	PV	0.0021	.	0.0164	.
	Open fen	PV	0.0000	.	0.0042	.
	Pasture	PV	0.0115	.	0.0155	.
	Sparse deciduous forest	PV	0.0009	.	0.0034	.
	Water	PV	0.0095	.	0.1416	.
Distance-to-roads	Mean	9600	423.8681	422.8946	1278.3169	1129.0468

Appendix 6A.17. The *a priori* and *a posteriori* probabilities derived for rare terrestrial plant species.

			Absence		Presence	
		Scale (m)	Mean or conditional probability	SD	Mean or conditional probability	SD
a priori			0.6087	.	0.3913	.
Landcover	Bedrock outcrop	7200	0.0409	0.1287	0.2106	0.4105
		PV	0.0000	.	0.0155	.
	Coniferous mixed forest	PV	0.0000	.	0.0000	.
	Coniferous plantation	PV	0.0000	.	0.0000	.
	Cropland	300	87.5913	20.1405	47.4179	34.8286
		PV	0.8645	.	0.3862	.
	Deciduous mixed forest	PV	0.0000	.	0.0080	.
	Deciduous swamp	9600	0.1267	0.1636	0.2721	0.2955
		PV	0.0000	.	0.0145	.
	Dense coniferous forest	PV	0.0000	.	0.0189	.
	Dense deciduous forest	300	6.7646	12.2675	29.0499	27.7431
		PV	0.0776	.	0.3959	.
	Developed land	PV	0.0169	.	0.0553	.
	Freshwater marsh	PV	0.0072	.	0.0071	.
	Open fen	PV	0.0000	.	0.0000	.
	Pasture	PV	0.0079	.	0.0099	.
	Sparse deciduous forest	PV	0.0009	.	0.0089	.
	Unclassified	PV	0.0000	.	0.0000	.
	Water	PV	0.0249	.	0.0797	.
Distance-to-roads	Mean	9600	454.9936	480.9138	1098.7791	1086.8535

VITA AUCTORIS

NAME: Jeremy VanDerWal

PLACE OF BIRTH: Sarnia, Ontario

DATE OF BIRTH: January 6, 1975

EDUCATION: University of Windsor, Windsor, ON

1994-1999 B.Sc. (Honors Biological Sciences)

University of Western Ontario, London, ON

1999-2001 M.Sc. (Zoology)

University of Windsor, Windsor, ON

2001-2006 Ph.D (Biological Sciences)