ABSTRACT

Title of dissertation: CORRELATES OF TERRESTRIAL VERTEBRATE SPECIES RICHNESS: AN EVALUATION OF ENVIRONMENTAL HYPOTHESES OVER THE WESTERN CONTINENTAL USA

Daniel Andrew Slayback, Doctor of Philosophy, 2006

Directed By: Professor Stephen Prince
Department of Geography

An explanation for the unequal distribution of life forms across the Earth's surface has been a persistent and problematic question in modern ecology ever since these patterns were first noted, over 100 years ago. Most empirical research supports one of three environmental hypotheses to explain these patterns: environmental energy (ambient environmental energy or ecosystem productivity); climatic variability; or habitat heterogeneity. This research examines these hypotheses using better datasets than those commonly considered, and using a consistent methodology that addresses often neglected statistical and analytic details.

The environmental datasets used in this study are derived from time series of satellite and ground station data, including the Daymet climate data, and net primary productivity data from the GLOPEM model. Species richness is derived from the individually modeled vertebrate distributions provided by the individual state Gap

Analysis Projects for the western US states of California, Oregon, Washington, Idaho, Montana, Wyoming, Utah, and Colorado, which define the spatial extent of this study.

The study methodology relies upon the summary of results from many model variants for each hypothesis. These variants are constructed by creating regression models at each of four different spatial scales (8, 16, 32, and 64 km grid cells), for each class of vertebrates (amphibians, birds, mammals, reptiles, and all), and over each of the eight states considered. Preliminary studies found that ordinary least squares would be a sufficient model form, although conditional autoregressive models were extensively considered. Other preliminary work examined issues of spatial autocorrelation and variable selection.

The results indicate that the energy/productivity hypothesis consistently outperforms all other hypotheses in explaining species richness, across almost all spatial scales, geographic regions, and vertebrate classes. The performance of the climatic variability and habitat heterogeneity hypotheses varies for particular states or vertebrate classes. Vertebrate data quality was important; results for Colorado and Washington were frequently unusual, suggesting an incompatibility between their modeled vertebrate distributions and those of other states. Models of reptile richness also often showed substantially different characteristics than those for other vertebrates. Overall the results provide additional support to the energy/productivity hypothesis, from a more comprehensive methodological basis.

CORRELATES OF TERRESTRIAL VERTEBRATE SPECIES RICHNESS: AN EVALUATION OF ENVIRONMENTAL HYPOTHESES OVER THE WESTERN CONTINENTAL USA

By

Daniel Andrew Slayback

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:

- Professor Stephen Prince, Chair
- Professor Ruth DeFries
- Professor William Fagan
- Dr. Scott Goetz
- Professor Axel Kleidon

Acknowledgements

In addition to the important support provided by my advisory committee, I would also like to acknowledge the particular assistance of the following individuals in completing this piece of research:

Professor Mary Christman, who was formerly on my advisory committee before leaving the University of Maryland, for initial critical guidance on many statistical issues.

Dr Jim Tucker, of NASA's Goddard Space Flight Center, for continued encouragement to pursue a doctorate, and for flexibility in working arrangements.

I would also like to acknowledge the generous financial support provided by my employer, Science Systems & Applications, Inc, who have also allowed liberal scheduling flexibility.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Category	Abbrev	Description
Mathematical / Statistical	AIC	Akaike's information criterion
	CAR	Conditional auto-regressive model
	GLIM	Generalized linear model
	HP	Hierarchical partitioning
	LR	Likelihood ratio
	OLS	Ordinary least squares
	PC	Principal components
	PV	Predictor variable
	VIF	Variance inflation factor
Datasets	AET	Actual evapotranspiration
	DEM	Digital elevation model
	GLOPEM	Global production efficiency model
	NPP	Net primary productivity
	PET	Potential evapotranspiration
Detrended Variables	HVVAR	Detrended HVVA
	NPPSDR	Detrended NPPSD
	TXSSR	Detrended TXSS
	PFSR	Detrended PFS
	PTSR	Detrended PTS
Productivity variables	NPPAV	Average NPP
	NPPSD	SD of mean annual NPP
	NPPSUMAV	Averaged annual sums of NPP
Heterogeneity variables	GLCV	GAP landcover variety, within cell
	GLCF	GAP landcover variety, focal area surrounding cell
	ELD	Elevation difference within cell

Category	Abbrev	Description
Temperature Variables	TAA	Average daily air temp
	TAS	SD of average daily air temp
	TXAS	Mean max July temp
	TXSS	SD of max July temp
	TXAW	Mean max of Jan temp
	TXSW	SD of max Jan temp
	TXD	Annual temp range (=TXAS – TXAW)
	TVNA	Annual mean day-to-day mean temp variation
	TVNS	SD of day-to-day mean temperature variation
	TDA	Annual mean growing degree days
	TDS	SD of growing degree days
	TRA	Annual max – annual min daily temp range
	TRS	Summer max – summer min daily temp range
	TRW	Winter max – winter min daily temp range
	TFA	Annual mean number of days with frost
	TFS	SD of annual number of days with frost
Precipitation variables	PTA	Mean annual precip
	PTS	SD of annual precip
	PEA	Average precip event size
	PFA	Mean precip frequency
	PFS	SD of precip frequency
Humidity variables	HVA	Average daily water vapor pressure (WVP)
	HVS	SD of daily WVP
	HVVA	Mean day-to-day variability of WVP
Richness	MAR	Mammals, amphibians, and reptiles (all non-bird animals)

Chapter 1: Introduction

1.1. Question of Interest

The variation in species diversity between geographic regions is a well-known but poorly understood phenomenon. The latitudinal gradient in species richness, with numbers of species generally declining from the equator to the poles, is the primary and most striking global-scale pattern in species richness. In other words, it is the most fundamental pattern concerning the distribution of terrestrial life on the surface of the Earth. Because of this, the source of this pattern is an extremely interesting, and often hotly debated, research topic in modern ecology. Given the recent and developing climatological changes and the accelerating alteration of the planet's surface, the interest in this pattern is growing, as its relevance to modern-day life. By many accounts, the biosphere is now entering its sixth major extinction episode (Leakey and Lewin 1995). An inventory of life, including the density of species, or species richness, and its distribution is thus of major concern, and subject to increasing research effort.

Zoomed in from the global scale, at regional or local scales, other patterns often obscure the global latitudinal gradient. However, neither the latitudinal nor more local patterns have been satisfactorily explained, despite a good deal of both theoretical and empirical work. On the contrary, there is a longstanding and often contentious debate about the factors responsible for generating these diversity patterns. One of the primary disagreements is between those promoting historical explanations for diversity patterns, and those promoting environmental explanations. In the context of understanding what is happening to the biosphere, this is of great importance, because if these patterns are largely driven by historical events, and are not particularly responding to environment, there will be implications for both ecological theory as well as conservation policy.

Unfortunately, ecological research has not yet come to a consensus on either the drivers or even, more simply, the correlates of biodiversity. As pointed out by Currie et al. (1999), the number of hypotheses explaining diversity has been rising dramatically, from the six reviewed by Pianka in 1966, to the 120 listed by Palmer in 1994. Although there appears to be much more effort spent on generating new hypotheses than on testing existing ones, a fair amount of research has been conducted to explore some of these hypotheses. Much of this research suffers from limitations or inadequacies of the data and methods, or is of only limited generality. Few studies, for example, explicitly account for the spatial features of the analysis, or use anything other than general climate atlas data and generalized species range maps as inputs. The research presented here attempts to shed some light on this topic by simultaneously examining several longstanding environmental hypotheses, and to do so with improved data sources and methods.

1.2. Objective

The main objective of this research is to evaluate the empirical support for several different hypotheses that propose mechanisms to account for species richness using environmental factors. This will be done by examining the relationships between species richness and environmental variables, while considering several important biological as well as statistical issues, such as spatial scale, variable selection, spatial autocorrelation, and differing relationships for different vertebrate groups. This research is *not* attempting to establish any causal link between richness and factors hypothesized to influence it. Given the natural experiment method required for the present study, where manipulative experiments are simply not possible, causality is extremely difficult, if not impossible, to prove. Some recent work has attempted to address causal mechanisms using structural equation modeling (Williams, Seabloom et al. 2005), but the present study aims to reexamine the correlational relationships using significantly improved data and more appropriate analytical tools, as a preface to further work.

1.3. Hypotheses

This goal of this research is to evaluate the three predominant hypotheses concerning the environmental controls on species richness; this research does not propose one particular hypothesized mechanism to be the most important or best supported. Rather, the hypothesis of this research is simply that through the use of improved datasets and statistical methods, one of the proposed hypotheses will emerge as better supported, empirically. The null hypothesis is that the improved datasets and methods do not significantly improve the ability to discriminate between the hypotheses considered. Such a result could indicate either that none of the tested hypotheses adequately explain richness, or that the methods and datasets utilized are not appropriate to answer the question.

The three hypotheses being tested are: energy/productivity, climatic variability, and habitat heterogeneity. Within this text, they will be denoted by the following abbreviations: H1 = energy/productivity (with a further distinction, described below, of H1a = productivity, and H1b = environmental energy); H2 = climatic variability; and H3 = habitat heterogeneity. These three hypotheses will be briefly reviewed here for introductory purposes, but a more thorough summary is provided in the next chapter (section 2.1.3.3).

The energy/productivity hypothesis suggests that the amount of available energy controls the amount of growth, and that higher growth leads to higher reproductive rates, whereby more speciation is likely to occur. An important related effect is that along with the higher growth rates allowed by the increased energy, larger populations are less susceptible to extinction through stochastic variations in population numbers. The energy / productivity hypothesis can be further separated into two versions, one supporting productivity (H1a), and the other supporting climatic energy (H1b) as the main driver. For the productivity hypothesis (H1a), the suggested mechanism is a trophic one: increased primary productivity leads to increased plant biomass, which represents increased energy resources, and so leads via the growth mechanism to higher speciation. For the climatic energy hypothesis (H1b), the mechanism is more direct, suggesting that the ambient climatic energy (temperature, actual evapotranspiration, etc.), helps regulate speciation by controlling the rate of growth; at higher temperatures, metabolism increases, and higher metabolism leads to higher growth rates and the possibility for increased speciation.

The climatic variability hypothesis (H2), also sometimes referred to as the “niche assembly theory”, suggests that more stable and equable environments allow organisms to specialize more intensively, and thus allow more niche space for differently specialized organisms to fill (Connell and Orias 1964; Dynesius and Jansson 2000). If climate is more variable, then a given species may require more niche space to meet its survival needs, as it will be forced to be adaptable to differing environmental conditions.

An alternate formulation suggests the opposite: that in some cases, regular variability can partition the environment into a larger number of ecological niches. The key may lie with the definition of variability.

The habitat heterogeneity hypothesis (H3) simply suggests that the more physically diverse an environment is, the more niche space should be available, and thus its capacity for higher species richness will be increased. Biotic diversity, and the related effects of competition and predation, may also contribute to an effect of heterogeneity, but these are generally considered to be a second-order effects on richness. They are also difficult to quantify across large regions.

1.4. Unique contribution of this research

The research presented here attempts to make several contributions to the topic of biodiversity. First and primarily, this research tests several existing hypothesized correlates of species richness, simultaneously. Most existing research tests only one hypothesized influence on species richness at a time. Partly, this is expected, since the dataset considered may only be relevant to one hypothesis. However, the specific methods used to analyze the data may easily have as much importance as the datasets themselves in finding support for a hypothesis. Using the same methodological framework for all factors, as is done here, may help illuminate the extent to which the data, versus the methods, are responsible for the hypothesized relationships.

Further, most existing research on environmental correlates of richness typically uses generalized environmental datasets, such as synoptic climate data.

Although useful, such data are limited in the environmental parameters that it can measure. This research utilizes time series of satellite and ground station observations, which allows construction of measures of climate variability. Similarly, most published research relies upon generalized range maps for species distributions. Here, species distributions are based on detailed models which have been individually parameterized for each species.

Finally, because species richness and the underlying species distributions are intrinsically spatial patterns and, as natural phenomena, are inherently spatially autocorrelated, the spatial aspects of statistical analysis are carefully considered. Most published research considers only one spatial scale, which leaves open the question as to how well the data would support the hypothesis at alternate spatial scales. Furthermore, most research does not consider the effects of the spatial autocorrelation of the data. Here, these issues are examined.

In summary, this research examines an important ecological question by using much improved data to evaluate existing hypotheses of vertebrate diversity with a single comprehensive methodological framework, and with explicit consideration of several important statistical issues.

Chapter 2: Background

2.1. Biodiversity and Species Richness: theoretical considerations

2.1.1. Diversity definitions

The general term “diversity” or “biodiversity” may refer to any of several characterizations, depending on the physical and ecological scales of interest (Purvis and Hector 2000). At the micro scale, genetic or character diversity might be of interest (Bisby 1995); these measures can be useful for studies of a single species, such as its population viability or evolutionary history. At the larger scales, diversity of ecosystems, sometimes characterized by landcover, can be important for studying interactions of ecosystems with global climate and biogeochemical cycles. Between the micro and macro scales is the diversity of types of organisms. At its simplest interpretation, this is taxonomic diversity. Taxonomic diversity can be used to quantify the diversity of higher taxonomic groupings, such as family or genus, but it is more typically used to quantify species diversity. This is most commonly referred to as species richness (SR), and less frequently as species density. In either case, species (or family, or genera) richness is defined simply as the count of species in a given area.

For this research, only species richness will be considered, for two important reasons. First, species richness is a simple concept, and generally simple to measure, requiring few additional definitions or constructs. Second, most existing literature on the topic focuses on species richness. Even so, measures of diversity with more of an ecological basis might be more appropriate or relevant. For example, functional diversity (reflected in trophic or niche structure (Bisby 1995)), or body size diversity (Badgley and Fox 2000), are sometimes claimed to be more closely linked to environmental factors than a simple count of species.

Several indices are frequently used to define species richness. Alpha diversity, or within-area diversity, refers to the simple count of species (or other taxonomic unit) in the area of interest. Beta diversity refers to the difference in species composition between two nearby but different areas or habitats. Gamma diversity generally refers to the overall species richness of a large region, such as a landscape or a country (Bisby 1995; Brown and Lomolino 1998). Because they depend on the specific area of concern, the definitions for beta and gamma diversity require careful specification for any given application. Several alternate diversity index definitions have also been proposed, such as the Shannon-Wiener index (Noss 1990). However, to avoid the many issues inherent in selecting and computing some optimal index of diversity, and to maintain a more straightforward correspondence to existing literature, this research only considers total species richness, which is equivalent to alpha diversity at local scales and gamma diversity at coarser, more regional scales.

2.1.2. Diversity components

The species richness of a particular location arises from the balance of the rates of speciation, extinction, immigration, and removal. Speciation occurs when the genetic differences in one population of an organism are sufficient to define a new species. This is generally thought to occur after isolation of sub-populations by one of two mechanisms: either a dispersal event, in which a population of the original species migrates across a barrier, or a vicariance event, in which a population of the original species becomes geographically isolated by the formation of a barrier, such as a mountain range or body of water (Brown and Lomolino 1998). Extinction also occurs by one of two general processes: stochastic extinction (or background extinction), in which random fluctuations in populations can lead to chance extermination of a species; and catastrophic extinction, in which some external environmental event precipitates high rates of extinction in particular taxa or regions (Bisby 1995). Immigration depends on the suitability of the habitat, its distance from the source population, and the dispersal ability of the organisms. Removal (e.g., local extinction) can occur via the same factors as extinction—random population fluctuations or as a result of catastrophic environmental changes—or by biotic exclusion, in which biotic interactions either force a species to migrate out of the region or cause its local extermination. The balance of these four rates will then determine the species richness of a particular region.

An additional important factor when considering the balance of speciation, extinction, immigration, and removal is the issue of ecological equilibrium, over century to millennial time-scales. The essential question is simply whether ecological communities are, at present, relatively stable, or whether they are still undergoing non-equilibrium changes following specific historical events, such as the Pleistocene glaciations. Note that in this context, "stable" communities may still be subject to significant shorter-term (decadal to century) changes due to local community dynamics; these changes may, or may not, be independent of any longer term trends. If communities are considered stable (in this longer time-scale context), then some suggest that contemporary environmental conditions should be most relevant for setting the balance (Kerr and Currie 1999). If ecosystems are not at equilibrium, but rather still recovering from some past event, then the details of this historical event – its duration and location – may have an important influence on the balance (Rohde 1998), although the current environment may also affect the non-equilibrium rates (Bisby 1995). The equilibrium / non-equilibrium state of modern ecosystems is a current topic of debate, with some arguing for present-day equilibrium (Rosenzweig and Sandlin 1997), and others insisting on the opposite (Rohde 1998).

Deconstructing diversity into these fundamental components helps to clarify the mechanisms at work that the major diversity theories attempt to explain. For example, the theories emphasizing the importance of current environment on diversity are also suggesting that diversity is dominated by immigration and removal rates. Conversely, theories proposing historical controls on diversity place more emphasis on speciation and extinction rates. However, diversity may be more than simply the sum of these components; it may be an emergent ecological property that influences the balance of these components and mechanisms. Kleidon & Mooney (2000) describe a theoretical model that predicts global plant diversity based on climatic inputs and simple prototype plant models. Thus, their model appears to capture an emergent diversity property, and without having to model the actual complex interactions of real plants. That their world of simple prototype plants reflects, to a surprisingly good approximation, the diversity found in nature indicates that diversity may, indeed, be an emergent ecological property in its own right. For plants at a coarse, global scale, the implication is that climate, and basic plant biology, are largely responsible for these patterns of diversity.

Others have pointed out that diversity may in fact just be the sum of different parts; Blackburn & Gaston (1996) suggest that there are two primary drivers for diversity: a ‘raw material’ driver, that essentially effects high populations, and a partitioning driver, that effects the generation of species from that material pool. The two work together synergistically to support a given level of diversity. If this is true, as they suggest, then a single predictive model or hypothesis will never be fully satisfactory.

2.1.3. Influences on diversity

Many hypotheses have been proposed to explain patterns in species richness, and for some time; Pianka (1966) reviewed the major hypotheses forty years ago, and more recently, entire books have been published on the subject (Ricklefs and Schluter 1993; Huston 1994; Heywood and Watson 1995; Rosenzweig 1995). The major hypotheses can be broken into three categories: (1) biotic hypotheses, with competition and predation as the major explanatory factors; (2) historical hypotheses, with habitat age and history as the major factors; and (3) environmental hypotheses, with area, habitat heterogeneity, energy, and climatic variability as the major factors. Additionally, the human impact on diversity must be considered.

2.1.3.1. Biotic hypotheses

Biotic factors arise from interactions between organisms in an ecosystem, and between those organisms and their environment. Competition and predation are two such influences that may affect the distribution of individual species, and therefore might affect diversity generally. However, as Currie et al. (1999) note, a clear quantification of predation and competition has never been done, and for good reason—it is not an easy problem. Thus, competition and predation make weak hypotheses. Begon et al. (1990) suggest that biotic factors are of only secondary importance because they characterize internal biotic interactions within an ecosystem, not external to it. Bisby (1995) notes that many of these interactions are poorly understood because they involve complex feedback mechanisms.

2.1.3.2. Historical hypotheses

Historical hypotheses contend that the age of a habitat is an important factor determining its diversity. These hypotheses suggest that with increasing age, a habitat will experience more speciation, leaving it enriched in species compared to younger habitats. Defining ‘age’ then becomes an issue, particularly if there are no major local extinction events, such as glaciation, to set a clear starting date. Rohde (1992) also suggests that higher temperatures, in stable habitats, further increase the speciation rate. Thus, tropical areas have high diversity due to their higher temperatures and the lessened impact of Pleistocene glaciations at tropical latitudes. An important component of these hypotheses is a non-equilibrium view of current ecosystems; they are considered still recovering from the Pleistocene extinctions, and therefore have plenty of empty niche space. Rohde (1998) argues that “the glass is 90% empty”, and habitats continue to accrue species, across the globe. However, others argue that habitats at present are essentially filled with species (Rosenzweig and Sandlin 1997), and therefore history should have relatively little effect. As far back as 1966, Pianka (1966) noted that such “ecological time” theories had been largely discounted due to evidence that most terrestrial habitats are presently saturated. These hypotheses have also been criticized as being weak and post-hoc, making no testable predictions, and with little empirical support (Currie, Francis et al. 1999; Kerr and Currie 1999). However, they may have some merit for explaining some of the observed differences between areas subject to the recent glaciations; Hawkins & Porter (2003) recently found a small but detectable effect of glaciation history on diversity patterns in North America.

2.1.3.3. Environmental hypotheses

In contrast to the difficult-to-test biotic and historical hypotheses, hypotheses linking species richness to various environmental characteristics are frequently supported by empirical evidence. Essentially, these hypotheses propose that some environmental factor, or combination of factors, exerts a control on the partitioning of the environment, both spatially and temporally, into niches, which can then be occupied by any immigrant or evolved species. The principal proposed factors include area, habitat heterogeneity, climatic variability, energy, and other physical environmental attributes. A common mechanism by which all of these factors might affect diversity is via population dynamics; if an environmental characteristic supports higher populations of a species, the risk of extinction for that species, due to both catastrophic environmental changes and stochastic population variations, should be reduced. Also, over sufficiently long periods, increased populations will carry an increased diversity of genetic material, potentially allowing greater rates of speciation – if appropriate isolation events also occur. To avoid repetition, these general effects are omitted from the specific descriptions below.

2.1.3.3.1. Area

Area is one of the most obvious influences on species richness; a larger region is, by chance alone, simply more likely to contain a greater variety of species. Partly, the effect is due to the increased number of habitats and environments, and therefore accompanying species, that larger areas are likely to encompass. Also, larger areas will more likely contain dispersal barriers, and thereby stimulate allopatric speciation within isolated populations. But the number of species also increases when the physical environment is relatively unchanging. MacArthur & Wilson (1967) took advantage of this in their historic theory of island biogeography. They related extinction rates to island size, via the mechanism of population stability. Outside of islands, however, defining the relevant area of interest poses a major difficulty for practical application. Also, using area as a predictive variable complicates any evaluation of the impact of scale on the emergent relationships (Currie, Francis et al. 1999; Whittaker, Willis et al. 2001). Nevertheless, some insist on the independent importance of area (Rosenzweig 1995; Rosenzweig and Sandlin 1997), but others find its effect swamped, at global and continental scales, by other factors (Chown and Gaston 2000; Hawkins and Porter 2001).

2.1.3.3.2. Habitat heterogeneity

Habitat heterogeneity may affect species richness by influencing the niche space available to immigrant or evolving species. At extremely low heterogeneity, for example, one species may be able to utilize all available resources simply by increasing its population, leaving little ecological space for any species less adapted to that habitat. Quinn & Harrison (1988) observed that natural or manmade habitat fragmentation increases species diversity over what would be expected by area alone. Turner et al. (1987) found a correlation between butterfly and moth diversity and habitat diversity in Britain. However, in Amazonian experiments, fragmentation has been seen to have a negative effect on natural species diversity and ecosystem functioning (Laurance and Bierregaard 1997). The key factor here may be the difference between the long-term evolutionary effects of natural or stable heterogeneity, and the short-term effects of introduced fragmentation.

Physical variation due to complex topography or the structure of vegetation also falls into this category. Processes that cause disturbance, such as periodic extreme weather, may increase habitat heterogeneity, and therefore lead to increased diversity. Note that this may have the opposite effect on diversity than would be expected due to climatic variability (below). Human induced physical variation, such as landcover conversion, may also affect diversity, but the time-scale is critically important; any introduced variation is likely to effect an immediate loss of diversity, as preexisting species unsuited to the new conditions exit the area. Over a longer time frame, however, the new physical configuration will attract a different complement of species, with a different total richness. Thus, for example, habitat heterogeneity in Britain, which is human induced but longstanding, may affect diversity in a different manner than heterogeneity recently introduced, by clearing perhaps, to a forest ecosystem. Scale is also an important factor here—heterogeneity at one scale may appear quite homogenous at another.

There is, however, some debate about the underlying mechanisms here. The question arises as to whether species richness might actually be driving habitat heterogeneity, or, more commonly, the density of niches (Rosenzweig and Abramsky 1993), rather than what has been assumed here—that the availability of more niches will promote more richness. The argument is simply that if a larger number of species are present, and successfully coexisting, they will by necessity subdivide the available ecological resources into smaller niches.

2.1.3.3.3. Energy

Energy is frequently hypothesized to influence diversity patterns. Wright (1983) proposed energy simply as an adjunct to area, in his extension of MacArthur & Wilson's (1967) theory of island biogeography. Most hypotheses generally suggest that the available environmental energy in an ecosystem affects the density and range of available resources (Hutchinson 1959; Connell and Orias 1964; Begon, Harper et al. 1990; Currie 1991), both of which impact species numbers; an increased density of resources allows higher populations, while an increased range of resources provide new types of resources and therefore new niche space for different species. Within the energy hypothesis, two different mechanisms have been proposed: productivity and ambient energy (Hawkins, Field et al. 2003; Turner 2004).

Primary productivity (NPP) is usually proposed to affect richness through trophic cascades, in which the availability of more food (plant material) at the base of the food chain leads to increased diversity of consumers up the trophic chain, to herbivores and carnivores (Wright, Currie et al. 1993). Productivity itself is the product of the interaction of environmental energy with the biosphere, but can more simply be viewed in this context as a measure of the resources available to herbivorous animals. More resources lead to higher populations, and, as discussed, higher populations lead to increased diversity through a variety of mechanisms. Much empirical research supports a link between diversity and productivity (Waide, Willig et al. 1999), although the form of the relationship appears to vary and theoretical support needs further development (Rosenzweig and Abramsky 1993). The relationship between productivity and richness varies between positively correlated, negatively correlated, and a humped pattern which shows richness peaking at moderate levels of productivity (Rosenzweig 1995). The general consensus appears to be that an overall humped pattern is that expected, and observations of either monotonically increasing or decreasing relationships are merely capturing one part of the humped relationship.

However, the effect of productivity on diversity may not be one-way; diversity may also affect the productivity of an ecosystem. There are two main mechanisms for this feedback: (1) the sampling effect, in which higher diversity simply increases the chances of a particularly productive species being present; and (2) the complementarity effect, in which more species are able to more effectively utilize all available resources (Waide, Willig et al. 1999). Experimentally, Tilman (1996) has verified that increased productivity correlates with increased species diversity, at least in grasslands. Nevertheless, the causality of the link continues to be debated (Tilman 1999).

Other forms of energy, associated with productivity, often appear in diversity hypotheses, such as actual evapotranspiration (AET), temperature, precipitation, and solar radiation. For the most part, these variables, or combinations of them, correlate well to productivity, and so are likely capturing the same relationship. NPP, for example, is often well correlated with AET, from which it can be estimated, and AET, in turn, is calculated from temperature and precipitation.

While the productivity hypothesis relies upon the realized interaction of environmental energy with the biosphere (generating primary productivity), to affect species richness, the second proposed energy mechanism is far more direct. This hypothesis proposes that ambient environmental energy affects diversity through its impact on physiological requirements and thermoregulatory needs; it has been proposed that speciation rates would increase because of the increased metabolic rates at higher temperatures (Rohde 1992; Allen, Brown et al. 2002), and that higher temperatures allow more of a creature's metabolized energy to go towards reproduction, instead of maintaining body warmth (Currie 1991; Hawkins, Field et al. 2003). For modeling purposes, ambient energy is often quantified as some measure of mean temperature, such heating degree days (the number of days per year that the temperature exceeds some threshold, often 0° C), which focuses on temperatures most relevant to physiological processes. Potential evapotranspiration (PET) is also often used, as it reflects the amount of energy available.

2.1.3.3.4. Climatic variability

Along with the direct energy component of climate just discussed, climatic variability is also proposed to influence species richness. Climatic stability is hypothesized to allow the evolution of more finely specialized species, since climatic flexibility is not required in a stable climate (Pianka 1966). The more specialized species occupy smaller resource niches, allowing more species to populate the environment. Stability need not imply monotony, however; climates with stable temporal variation (e.g. seasons) may allow increased niche density over the temporal domain (Rosenzweig 1995). For example, while one species is dormant, waiting for its optimal seasonal conditions, another may be in its active phase, thereby reducing the amount of interspecies competition for resources at any given time. Seasonality provides a natural timing mechanism for such temporal differentiation. Stability may also allow small populations to persist, with less danger of extinction due to environmental extremes. At the other extreme, explicit climatic harshness, or unpredictability, is thought to limit diversity by requiring the evolution of special mechanisms to 'weather the storm', so to speak. Species that cannot survive the extremes are quickly driven extinct, at least locally (Richerson and Lum 1980). The evolution of adaptation to climatic extremes may require the use of more resources and relatively large ecological niches – since the constancy of any one resource may be unreliable – and thereby lead to reduced diversity.

2.1.3.3.5. Other physical environmental characteristics

Aside from those addressed above, other physical environmental characteristics are sometimes hypothesized to affect diversity. Nutrient concentrations, pH, salinity, and other soil characteristics, for example, may affect the number or size of species niches, and therefore species diversity, in a particular environment (Bisby 1995). Geomorphology, such as slope, aspect, and elevation, can have a strong influence on the local effect of other variables, such as precipitation or temperature, and so are often included in analyses, although a direct causal link may not be specified.

2.1.3.4. Human impact

By modifying natural landscapes, whether for agriculture, settlement, or transportation, humans have, across large areas, significantly altered the natural balance of species. There are two primary modes of human impact: direct landscape modification, and the introduction of exotic or invasive species. Direct landscape modification, such as conversion from forest to farm, alters the biophysical environment. The local mix of species is then affected based on both the new environmental conditions, and the communities of species the new conditions attract. The introduction of exotics directly affects the biotic environment, and the impact can be significant. As the biotic interactions ensuing from either type of impact can be complex, the impact on species richness is difficult to predict. Furthermore, a significant amount of time may be required for a new community, and diversity level, to stabilize.

2.1.3.5. Other issues

Some associated issues in understanding diversity are related to the geometry of species ranges. The biogeographic principle called Rapoport's rule (Stevens 1989) suggests a positive correlation between the size of species geographic ranges and increasing latitude, and thus richness gradients could simply be a manifestation of the smaller ranges of tropical species. However, Rapoport's rule currently appears to be discredited, with many authors finding no empirical support (Colwell and Hurtt 1994; Hughes, Cawsey et al. 1996; Rahbek 1997; Kerr 1999), although Taylor and Gaines (1999) find some salvageable utility for the idea. Other recent work has indicated species richness is influenced by the geometry of species ranges, at least for taxa with relatively large geographic ranges (Colwell and Lees 2000); the idea is simply that more range overlap will occur towards the middle of geographic range constraints, thereby maximizing diversity. Narrow-range species will contribute little to such patterns.

2.1.4. Scale

Scale is an important consideration when examining questions of diversity, for both purely spatial and for ecological reasons. Spatially, scale affects analyses via its impact on sampling and on resolution. Stoms (1994) found that species richness maps changed in inconsistent ways as sampling unit size increases, particularly for sampling units smaller than 10,000 hectares (square cells 10 km on a side) or larger than 100,000 ha (~32 km on a side). Scheiner et al. (2000) argue that the scale of observation can critically affect the form of species-productivity relationships, and, by extension, any species-environmental variable relationship. They also argue that the scale of observation may interact with the study extent and have a significant effect; for example, richness may decline as productivity increases within landscapes inside a larger region, while across these landscapes (treating each as a unit), richness may increase with productivity. Thus, to capture the underlying spatial relationships, the possible effects of scale must be considered.

Ecologically, scale affects the strength of relationships between species richness and environmental variables. As with species richness, environmental variables may not scale in an expected manner. Furthermore, their contribution to a diversity relationship may vary nonlinearly with scale (Whittaker, Willis et al. 2001). Not surprisingly, variables appear most important over the scale at which they show the most variation, leading to regional variations in modeled processes(Huston 1999). Currie (1991) notes that local-scale studies often correlate richness with local habitat diversity, while regional-scale studies find correlations with regional energy measures. Thus, climatic measures are considered important over regional and larger scales, setting the broader outlines of diversity, while topographic and habitat heterogeneity effects are proposed to help define the variation locally. Whittaker et al. (2001) make a useful contribution by isolating different biogeographic response variables: “ecological” aspects of diversity, such as species richness, and “biogeographic” aspects of diversity, such as endemism and taxonomic distinctiveness. They suggest that, to first order, contemporary ecological processes and their controls, such as climate, determine ecological diversity, while evolutionary and historical processes largely determine biogeographic diversity.

Unfortunately, much of the existing literature that examines questions of species richness fails to adequately address scaling issues. Partly, this follows from the methodologies used; frequently, data are compiled or aggregated from essentially point sources with no explicit scale, or from a variety of sources with different scales. Since the vast majority of ecological research is conducted at very localized scales (e.g., less than one hectare (Kareiva and Anderson 1988)), this is not surprising. It also follows from a lack of guidance in ecological theory, which has generally failed to adequately address scale (Whittaker, Willis et al. 2001). The empirical and experimental literature that discusses scale in this context does generally find an important connection. For example, Gross et al. (2000) found that with herbaceous plant communities, the focus and extent of analysis did affect the form of the diversity – productivity relationship, changing it from unimodal, to negative linear, to no correlation. In his extensive reviews of existing studies, Rosenzweig (1993; 1995) suggests that the form of the diversity-productivity relationship changes with scale, and that this effect helps explain the disparate (increasing, unimodal, decreasing) results found in particular studies; these various patterns may simply reflect one section of an overall unimodal pattern.

2.1.5. Spatial autocorrelation

There has been an increasing recognition of the importance of spatial autocorrelation in biogeographic studies over recent years (Carroll and Pearson 2000; Lichstein, Simons et al. 2002; Foody 2004; Tognelli and Kelt 2004). The crux of the matter is simply that spatially autocorrelated variables will affect regression analyses by distorting the coefficients and significance of the included variables, similar to the effects of collinear variables. The effect is most pronounced for ordinary linear regression models, since these are based on the assumption of independent and normally distributed errors, an assumption often violated when spatially autocorrelated variables are included. And since many environmental variables are also correlated to one another – independent of any spatial autocorrelation – the effects of collinearity and spatial autocorrelation can become compounded. Further, the effect can vary with the magnitude of the spatial autocorrelation (Lennon 2000). The overall result is that variables are retained in models when they should not be, and that their inclusion might then preclude the inclusion of variables with less autocorrelation but more significance for the process under study.

2.1.6. Summary: theoretical considerations

The list of possible influences on diversity is long, and the above summary provides only a cursory review. Unfortunately, as Currie et al. (1999) note, the list continues to get longer, with little progress being made in testing and discarding weak hypotheses. Specifically, the biotic and historical hypotheses, as currently formulated, provide very little in terms of testable hypotheses. The proponents of historical influences have argued that this is not a relevant criticism (Latham and Ricklefs 1993; McGlone 1996; Ricklefs, Latham et al. 1999). Although this provides a distracting argument, it does nothing for improving our ability to predict species richness based on available data. It appears that at most a ‘glaciated/not-glaciated’ dummy variable has been included in some analyses, to test for the existence of an effect (Currie and Paquin 1987; Adams and Woodward 1989). Additionally, it is not unreasonable to assume that some contemporary environmental characteristics continue to reflect historical events (McGlone 1996), so the two hypotheses may not be so far removed. Perhaps more importantly, the issues of scale and autocorrelation, which have only been dealt with occasionally or in a cursory manner until recently, can have an important effect, and should be considered explicitly to gain a deeper understanding of these relationships.

2.2. Existing research

There is an increasing body of empirical work examining species richness, using a variety of methods, and for a variety of regions and spatial scales. The most common approach is to generate a multiple regression model to test a specific dataset, of either richness or environmental variables, or both, and, based upon the variables that are found significantly different from random effect, come to conclusions about specific hypotheses. A significant drawback of this approach is that unless the predictor variables are uncorrelated, it is difficult to draw solid conclusions about hypotheses that rely upon collinear variables for validation. Nevertheless, this is a fairly widespread approach, and the literature is filled with many examples, often with divergent conclusions.

For relevance to the present work, it is informative to review existing studies that either examine similar relationships (vertebrate diversity over large regions), or utilize similar methods. Several such studies will be summarized in this section, along with a few others that provide useful or interesting results. Table 1 provides some details of the research reviewed in this section.

Table 1: Selected studies of relevance to present work. HabHet = some type of habitat heterogeneity measure. var=variability. LS=low support for listed measure. NS=no support.

Study	Region	Species	Hypotheses tested	Results: important variables
Adams & Woodward 1989	North America, E Asia, Europe	Trees	History (glaciation), climate, climatic var, productivity, topography	NPP
Badgley & Fox 2000	North America	Mammals	(none specifically, but tested many environmental vars)	Annual temp min & max, AET, relief, elevation
Boone & Krohn 2000	Maine	Vertebrates	Geomorphology, climate, woody plant distributions	Climate
Currie & Paquin 1987, Currie 1991	US & Canada	Vertebrates, trees	Energy, climate, climatic var, habitat heterogeneity	PET (verts), AET (trees), solar radiation, temp
Fraser 1998	Wyoming	Vertebrates	Energy, productivity, HabHet	Heterogeneity, landcover metrics. NS: NPP, energy
H-Acevedo & Currie 2003	North America	Birds	Productivity, climatic stability, seasonality, HabHet	Mean & var of climate (temp & precip), NDVI. NS: HabHet, soil fertility
Hurlbert & Haskell 2003	North America	Birds	Energy, seasonality, HabHet	NDVI: fine scale HabHet: coarse
Jetz & Rahbek 2002	Sub-Saharan Africa	Birds	HabHet, NPP, Elev var, max temp, range geometry	NPP, HabHet. LS: geometric constraints.
Kerr & Packer 1997	North America	Vertebrates	energy, habitat heterogeneity	Topographic var, PET var NS: mean PET (S of Canada)
O'Brien 2000	Southern Africa	Woody plants	(developed hypothesis)	Energy, water, topography
O'Connor et al. 1996	Continental US	Birds	Climate (temp, precip, seasonality), HabHet, landuse	July T, long, precip., landuse, HabHet
Rahbek & Graves 2001	South America	Birds	Climate (temp & precip means), area, HabHet, elev range, latitude × elev range	Lat × elev range, precip, HabHet, cloud cover, elev range. LS: mean, max temp.
Ruggiero & Kitzberger 2004	South America	Mammals	Energy, climatic variability, HabHet	Productivity (AET), elev var NS: energy variability
Simpson 1964	North America	Mammals	(none tested; descriptive)	Topography within continental US
Tognelli & Kelt 2004	South America	Mammals	Energy, HabHet, climate.	Productivity LS: HabHet, climate..

2.2.1. Continental scale

Continental scale studies are very useful because they can address the generality of the hypotheses without being restricted to some geographic subset, and while sampling a much fuller range of environmental conditions than available in most smaller regions. The more recent availability of satellite datasets and computer aided compilation of species datasets has spurred an increase in the number of such studies appearing in the literature. However, significant continental-scale studies are not only a recent invention.

More than forty years ago, Simpson (1964) published a classic study of North American mammal species richness. Using quadrats ~240 km on a side (150 miles), he observed a significant variation in richness patterns that roughly coincided with the US – Mexico and US – Canada borders. Both north of the Canadian border and south of the Mexican border he found a fairly regular correlation of richness to latitude. Within the US, this pattern breaks down due to topographic complexity, which allows variations in richness due to both geographic isolation and to the higher niche density in topographically intricate regions. Although his analysis was purely descriptive and did not contain any quantitative measures of correlation, his observations are still valid and provide a convenient continental overview.

More recently, studies conducted by David Currie (Currie and Paquin 1987; Currie 1991) appear to have reinitiated interest in continental richness patterns and their relationship to environmental variables. The 1987 paper focused on North American tree richness, while the 1991 paper expanded the analysis to terrestrial vertebrates. Using simple univariate regression and an exponential model, they explained 80-93% of vertebrate class richness with potential evapotranspiration (PET) (92% for all vertebrates) and 76% of tree richness with actual evapotranspiration (AET). They also used multiple regression to investigate relationships, but the issues of scale and spatial autocorrelation were not considered. A latitude and longitude based quadrat system was used, allowing area to vary; equivalent square cells would have sides from 134 km to 237 km. Even so, area effects were tested for and found of no or low significance.

However, Kerr & Packer (1997) analyzed the same dataset as Currie (1991), and found the energy-richness relationship to only hold at lower levels of energy (potential evapotranspiration (PET) < 1000 mm/yr), corresponding, roughly, to areas north of the US-Canadian border, and echoing the observations of Simpson (1964). They found areas with higher PET to be much more closely related to topographic variation and the spatial variability in PET, and not to PET itself. These spatial and topographic variables explained 77% of the species richness variation for the high PET regions.

Following on the Currie & Paquin study of 1987, Adams & Woodward (1989) examined relationships between tree species richness and environment in North America, Europe, eastern Asia, and New Zealand, and found similar climatic variables as the most important factors in all regions. The study included a computed measure of net primary productivity, and this had the highest single correlation to richness in all regions, explaining from 84% to 90% of richness. They also found a potential role for glacial history in explaining residual differences in Asia.

Much recent work focuses on either birds or mammals, presumably due to difficulties in assembling datasets for both simultaneously. Since there are far fewer amphibian and reptile species, omitting them may have a relatively minor impact on the ability to assess hypotheses concerning overall vertebrate species richness—since the hypotheses do not discriminate between vertebrate class—but, amphibians and reptiles may represent significant amounts of richness in areas not well represented by birds or mammals. Thus, these studies are not optimal, but they are nevertheless frequently interesting and informative in their scope and methods.

O'Connor et al. (1996) examined the environmental correlates of bird richness across the coterminous US, and found they could predict 47.5% of richness using a regression tree algorithm. The most important variables were average July temperature, longitude, and average precipitation. Although they included a variety of landcover type and heterogeneity metrics, none of these explained more than 3.3% of the richness. Clearly, there is a significant difference between these results and those of Currie (1991), who was able to predict 81% of bird richness given only PET. The difference might be a result of any number of factors, such as the different datasets employed, the scale, or the analysis methods. Notably, O'Connor et al. did not include a measure of either potential or actual evapotranspiration—common proxies for productivity. They do include several climate variables, however, that might be expected to provide similar information: January and July mean temperatures, and annual precipitation. Also, Currie used a map of bird richness published in 1969, while this study used the 1981-1990 Breeding Bird Survey (BBS) data. Presumably, the BBS data provide a better, and more current, indication of actual bird ranges, but it might suffer from sampling effects. Furthermore, the very different scales used may have had an impact on the results; this study was conducted on a regular, equal-area grid (~27 km cells), while Currie's cell sizes ranged from 134 to 237 km.

Two more recent studies have also examined the relationship between the seasonality of bird richness and climatic seasonality, along with the more regular correlates, such as habitat heterogeneity and general climate. H-Acevedo and Currie (2003) modeled summer and winter richness separately, and found that the relationships between bird richness and climate and NDVI (normalized difference vegetation index, a satellite measured index strongly correlated to productivity) changed between seasons. They also found little relationship to their measures of habitat heterogeneity, including range of elevation, but did find relationships to climatic stability; richness was negatively related to temperature variation, indicating the favorability of equable temperatures, and positively correlated to precipitation variation. Hurlbert and Haskell (2003) found a significant relationship between the seasonality of NDVI and energy and that of bird richness. They also used spatial regression models, and conducted the analysis at four different spatial scales. NDVI's effect was found to decrease with increasing spatial scale (larger grid cells), while the effect of elevation range increased with spatial scale.

Badgley & Fox (2000) conducted a detailed analysis of the relationship between mammal species richness and environment across North America, and were able to relate 88% of mammal species richness to five environmental variables. Notably, this set of variables includes actual evapotranspiration (AET). They often found PET to be important, but not for tropical (< 30°) or high (> 60°) latitudes, nor for the continent as a whole. Their analysis was conducted on an equal area grid of 240 km and using multiple regression for the species richness analyses. Other interesting results include the variability of the factors important for predicting richness in different regions of North

America, and in doing so for different categories of mammals, sorted by size or trophic category. Unfortunately, although they do describe the spatial autocorrelation of the variables, this information is not incorporated into the predictive analyses in any way. They also do not vary the scale.

Several studies have looked at the richness of birds or mammals in South America, with sometimes contradictory results. Rahbek & Graves (2001) looked at bird richness across South America using multiple regression, and did so at ten different spatial scales, one of the few studies to explicitly address the issue of scale in such a way. They found that topography, precipitation, and cloud cover were the most important variables overall. They also evaluated separate climate, ecosystem diversity, and topography-latitude (an elevation difference variable multiplied by latitude) models, and found climate and topography-latitude models to perform best. Energy measures were only minimally useful in the climate models; precipitation and cloud cover were more important. As scale varied, precipitation was found important at finer scales, and cloud cover and elevation variation at coarser. Stated more descriptively, humid conditions with high topographic variation, such as at the base of the Andes, were found to support the highest richnesses.

Two recent studies examine mammals richness in South America. Ruggiero & Kitzberger (2004) tested the energy (both productivity and ambient energy), environmental stability, and habitat heterogeneity hypotheses, and found the best support for productivity and elevation variability. Notably, they also found little general support for environmental variability, although it did have an effect on the richness of narrow-ranging species. Tognelli & Kelt (2004) also looked at mammal richness in South

America, and evaluated the same three basic hypotheses, and found similar results, with productivity as the most important predictor. They found little support for contemporary climate or ambient energy, and indicate these are often found to be important due to spatial autocorrelation issues. Habitat heterogeneity was found of intermediate importance. This latter study was one of the few to fully utilize spatial models at the continental scale; they evaluated both ordinary least squares (OLS) and spatial autoregressive models. Both model forms led to similar results, but with differences in emphasis. The spatial models showed a lower importance for highly spatially correlated predictor variables, as would be expected.

In Africa, Jetz & Rahbek (2002) examined bird richness, utilizing both OLS and spatial regression models – in this case, conditional autoregressive models (CAR). Although they included variables from the several hypotheses, they did not explicitly test each hypothesis against the others; a combined model was used. They found that productivity and habitat heterogeneity were among the most important predictors, and that geometric constraints, one of the more unique variables they tested, were of only secondary importance. They also examined their results by range size, and found that productivity's effect decreases with decreasing range size, while topographic heterogeneity's effect increases with decreasing range size. This difference in effects for narrow-ranging versus wide-ranging species was fairly unique to this study, and provides useful insight.

A few other studies have addresses richness questions at continental scales, but for even more limited taxonomic subsets. For example, Lyons & Willig (1999) examined latitudinal gradients of richness across both North and South America, with an explicit focus on issues of scale, but only for bats and marsupials. They found the latitudinal gradient to be rather scale-invariant at their grid cell sizes, from 32 to 150 km. Kerr & Currie (1999) found climate to best relate to the richness of tiger beetles and several groups of freshwater fishes in North America. Using phylogenetic information, they also tested the historical hypotheses that propose richness is related to evolutionary advancement, but found little support.

2.2.2. Other studies

A number of other studies at regional scales are also of relevance, because of their methodology or results. Boone & Krohn (2000) examined environmental correlations to vertebrate richness in Maine, finding climatic variation (and not geomorphology, or woody plant richness) to best explain richness, using both multiple regression and regression tree techniques. They explained 78% of richness with multiple regression techniques, and 93% with regression tree methods. Although they only examined these relationships at one scale (18 km), they did attempt to account for spatial autocorrelation via a method of partitioning the variance into spatially correlated and uncorrelated components. Unfortunately, they used principal components to summarize the information in the climate variables, and thus do not report the relative importance of the different variables. However, they did include PET in the set of climate variables.

Fraser (1998) examined the relationship between species richness and both environmental energy and spatial heterogeneity, across Wyoming. Using species data from the Wyoming GAP project and conducting the analysis at a range of grid resolutions, from 10 km to 120 km, he found that environmental heterogeneity, defined from landcover measures and the variance of climatic variables, always explained species richness much better than energy. When there was any significant correlation, his energy variables were only weakly correlated to species richness, with a maximum R^2 of 0.15. He included temperature, precipitation, insolation, and AET in the set of energy variables.

O'Brien (1998; 2000) developed a nonlinear model to predict woody plant diversity in southern Africa, based on water-energy dynamics but also including topography. Her model, developed from theoretical considerations, includes water, energy (minimum mean monthly PET), and a measure of topography as predictor variables. After fitting to the data, it explains 86% of the variation in species richness.

As the importance of spatial autocorrelation has become increasingly apparent, a number of recent studies have begun addressing this issue, using a variety of methods. The simplest approach is to simply require a higher significance threshold for inclusion of variables in linear regression model (e.g., see H-Acevedo & Currie (2003) for an example). Typically, for models without these concerns, a p-value of 0.05 may be used to identify significant variables. This threshold can be decreased, or the effective number of degrees of freedom can be reduced, for a similar effect. Another approach is to use a series of preliminary regressions and analyses to attempt to isolate important variables (Jetz and Rahbek 2002), but the effectiveness of such an approach is questionable; it appears to address the collinearity issue more so than that of spatial autocorrelation. The partitioning of the variance of fitted models into spatial and non-spatial components is an approach advocated by Legendre & Legendre (1998) to identify how much of the explanatory power of a model is based purely on spatial autocorrelation, and how much is based on the independent effects of variables. More recently, spatial regression models (Haining 1990; Cressie 1993) have increasingly been used to explicitly account for spatial autocorrelation (Jetz and Rahbek 2002; Lichstein, Simons et al. 2002; Hurlbert and Haskell 2003; Tognelli and Kelt 2004). These methods may have the best theoretical basis, but are often difficult to implement. An alternative approach is to use locally varying regression models to explicitly model the spatial effects (Fotheringham, Brunsdon et al. 2002; Foody 2004), but the utility of this approach for examining global relationships is unclear.

2.2.3. State of existing research

As the above summary indicates, there has been a fair amount of work done examining the environmental correlates of diversity, with a wide variety of methods, and a varied set of conclusions. In more recent years, there appears to be an increasing awareness of scale and spatial autocorrelation, and increasing attempts to adequately address these issues (Lennon 2000). This awareness is relatively new, however, and there are therefore a wide variety of ways in which researchers have attempted to tackle the issues, with no clear-cut solution for all situations. The related issue of collinearity is sometimes discussed, but is often not adequately addressed, considering how pervasive this issue is with the types of correlated environmental variables used in most of these studies. Some researchers attempt to circumvent this through carefully considered stepwise regressions (e.g., see Richerson & Lum (1980) for an example), while others use preliminary screening of variables (Jetz and Rahbek 2002), but most simply do not address the issue.

The potentially important issues of scale are also not frequently addressed. Partly this may be the result of the general lack of recognition of the importance of scale, but it may also reflect the technical difficulties in approaching a multi-scale analysis. Such an analysis would almost certainly require both spatially comprehensive datasets (e.g., not just a sparse collection of site data) at a relatively fine scale and the use of specialized software (as available in a GIS) to appropriately resample, or rescale, the data. Most of the above studies did use spatially comprehensive datasets, but many of these datasets were derived from rather general species range maps. And although the scale of analyses may have been appropriate for the scale of those maps, it would be difficult at best to extrapolate species ranges—and therefore richness—at the finer scales necessary for a multi-scale analysis from such general, continent-wide range maps. In other words, a multi-scale analysis requires data at a relatively fine grain and over a large region.

Fortunately, there does seem to be a growing recognition that different hypotheses need to be simultaneously tested and their performance compared, to really begin answering the fundamental questions about the validity of the various hypotheses. Unfortunately, the most commonly implemented approach is that of the single model; a single model is built from the variables that would support the different hypotheses of interest, and those variables found significant in the model are then cited to validate a particular hypothesis. Although seemingly straightforward, this method can easily lead to ambiguous results if, as they almost always are, variables are collinear or spatially autocorrelated. Alternate methods for selecting the best supported hypotheses based on information theory are under development (Burnham and Anderson 2002), but may be too immature at this stage for reliable use.

Chapter 3: Data

3.1. Species Richness: GAP datasets

3.1.1. Theoretical considerations

There are several practical and theoretical problems with the development of datasets of species richness. Species richness is computed by counting the number of species believed to occur in a specified area. That statement encompasses the two most problematic issues involved in defining species richness: the issue of what constitutes presence of a species, and the issue of scale. Given that the presence of each individual of a given species cannot be observed, presence must be predicted, on the basis of information regarding habits, biological and environmental requirements, and preferences – information that may be poorly known for many species. Furthermore, there is the issue of how much presence is required to register a particular species as "present." Is a single individual sufficient and necessary, or might more of a population, or an intermittent population, be sufficient? For example, wide-ranging species may migrate, leaving an area with or without any individuals of the species at particular times. For birds, presence is usually defined as the areas the birds inhabit while breeding.

This question also raises the related issue of the theoretical versus the realized niche; a theoretical niche is one in which the given species should be able to exist (environmental and biotic conditions are suitable), while a realized niche is one with an actual occupant. A theoretical niche may not be realized for any number of reasons, including stochastic population variations, manmade predation or disturbance, or theoretical considerations; some biologists suggest that most niches are presently filled, while others insist that most are not.

The issue of scale also impacts measures of species distribution, and therefore richness. As scale changes, so does the relevance of the above issues of prediction vs observation, and actual vs realized niches. At very local scales, given a suitable ecological niche, the presence or absence of a species may become very dependent on both local heterogeneity and stochastic population variations. At coarser scales, these factors may be largely irrelevant. The result is simply that species distributions, and therefore richness, have differences in meaning at different scales.

Compiling species richness across any substantial continental-sized region requires the use of predicted distributions. Although the observational record for some taxa, such as trees or birds, and for some regions, such as the UK, might be relatively complete, this record is simply not complete for most taxonomic groups across any large, continental-sized region. Furthermore, use of the observational data would require a tremendous effort to collate the vast amounts of data scattered throughout museum collections and research institutions. Range maps reflect a combination of observational data and species ecology, preferably synthesized by an expert. However, such range maps presumably are developed at a wide range of scales, with a wide variety of methodologies, and with mostly unknown accuracies, limiting their usefulness for the type of multi-scale analysis undertaken here.

3.1.2. GAP datasets

Over the past decade or so, a series of state-level analyses have been underway to predict the distribution of terrestrial vertebrate species at a relatively fine scale. These projects, termed Gap Analysis Projects, or simply GAP, were initially conceived in order to highlight regions in need (e.g., “gaps”) of conservation protection, due to the local assemblage of species and current landuse and protection status. One component of these projects is to develop modeled distributions for all terrestrial vertebrate species, not just those of particular conservation interest. The projects have been coordinated at a national level by the national Gap Analysis Program (http://gapanalysis.nbii.gov) in order to standardize methods and products, but are implemented at the state level by universities or government agencies. At the present time, most states have completed the vertebrate distribution mapping.

The standard GAP analysis methodologies are described by Scott (1993) and Csuti & Crist (2000). Additionally, upon completion, each state project publishes a final report describing the details of their specific methodology. In general, there are two main stages to the methodology: construction of a preliminary habitat suitability map, and a more detailed modeling within the suitable habitats on this map, based on ancillary, species-specific data.

The habitat suitability maps are generated from potential species range limits, a database of wildlife-habitat relationships (WHR), and a landcover classification. Potential species ranges are determined from existing literature and observation records, with expert review, and might be developed from preexisting, traditional range maps. The wildlife-habitat relationship databases are developed in consultation with experts to indicate categories of use (e.g., absent, rarely occurs, occasional occurrence, common, abundant) for each species within each of the state’s habitat types. These habitat types are then associated with the set of habitat types mapped by the project’s landcover classification, which is generally derived from Landsat TM imagery at 30 meters. Using these inputs, a preliminary map of habitat suitability is generated.

Next, predicted distributions are modeled based on this habitat suitability map and information on any important associations of each species to ancillary data layers, such as a required proximity to water, for example. These layers may include variables such as elevation, wetland type, and hydrology, and are determined from the literature. The final model then combines these associations for each species with the habitat suitability map to generate the output map of predicted species distribution. Usually there is some effort to validate some portion of these distribution maps, although the extent and depth of these validations varies greatly from project to project.

Most GAP projects provide the final species distributions at either a 90-meter or 1-km resolution, although some provide distributions at the much coarser resolution of the EMAP hexagon scheme (about 18 km). The metadata indicate at what scale the data were produced (despite the nominal resolution of the output format), and thus at what scale the data are suitable for analyses. For example, most state projects specifically state the data are produced at a 1:100,000 scale and are not suitable for analyses at their nominal resolution, which is usually 90 or 100 meter. They often suggest a minimum resolution for analyses of approximately 1000 ha, corresponding to a square grid cell of roughly 3 km on a side.

The advantages of using GAP distribution data are several. Primarily, these data provide fine-resolution distribution information for all vertebrates across a large swath of the US. They improve on existing species distribution maps by incorporating knowledge of the landcover and specific information on the habitat and other environmental requirements of each species. The widespread use of independent expert review at various stages in the process helps to ensure a reasonable representation of reality, as do the validation exercises, where and when conducted.

Nonetheless, the GAP data are less than ideal for several reasons. Due to the differing procedures, which are developed at the state level, species distributions may appear discontinuous across state boundaries. In some cases, this translates into a corresponding discontinuous species richness across state boundaries; however, since richness is the sum of many individual species distributions, many, but not all, of the differences at the species level tend to cancel each other out when summed to richness.

There may also be some concern that the environmental datasets used by the GAP projects in modeling species distributions could include some of the same data used for the present research. In such a case, there could be potential issues with circularity. However, this concern is not valid for two reasons. First, the specific methodologies for each state used in this analysis have been examined, and none use, as ancillary data, any of the predictor variables specifically used in this research. Some do use elevation ranges, but for the present research, only the difference in elevations within a grid cell are used (section 3.4.2). Furthermore, as the results show, this variable was not found to be a strong predictor of species richness. More common is the use of a required distance to water; however, the present research includes no similar specific measure of water presence. Landcover, however, is used by both the GAP projects to predict vertebrate distributions, and by this research to develop predictor variables for the richness models. In the GAP vertebrate distribution modeling, landcover is used to link the known habitat affinities of each species to mapped landcover types, and thus provides the first cut of the vertebrate distribution. The distribution is then modified by other known affinities (with additional ancillary layers, as mentioned above), and by known distribution limits (e.g., if a species is known to not occur in a specific county, then its predicted distribution will be removed from that county). This research also uses the GAP landcover product, but not directly as a predictor of richness; rather, it is used only to generate a measure of spatial heterogeneity (section 3.4.1). These measures of spatial heterogeneity are not particularly linked to the landcover types; it is, rather, simply a measure of the diversity of landcover types. Thus, although the GAP vertebrate distributions depend on specific landcovers, they do not depend on their variety, and so should not pose a significant issue of circularity.

Second, species richness may be considered to be an emergent ecological property, not a directly observable quantity, and so is fundamentally different from species ranges. Although it is computed as the spatial sum of species range maps, richness may well be expected to have a different relationship to environmental variables than that of any given species. Mathematical sums can often be separated into their parts rather trivially, but the same may not hold true ecologically.

3.2. Climate: DAYMET datasets

Climate datasets were acquired from the DAYMET project, which provides a variety of daily and summary modeled climate variables across the coterminous US at 1 km spatial resolution (Thornton, Running et al. 1997; Thornton, Hasenauer et al. 2000).

Notably, these data are produced entirely from ground-based sources. The project website (http://www.daymet.org) provides a useful brief summary of the data products:

> Daymet is a model that generates daily surfaces of temperature, precipitation, humidity, and radiation over large regions of complex terrain. Daymet was developed at the University of Montana, Numerical Terradynamic Simulation Group (NTSG), to fulfill the need for fine resolution, daily meteorological and climatological data necessary for plant growth model inputs
>
> Using a digital elevation model and daily observations of minimum and maximum temperatures and precipitation from ground-based meteorological stations, an 18 year daily data set (1980 – 1997) of temperature, precipitation, humidity and radiation has been produced as a continuous surface at a 1 km resolution. A wide range of summary and point daily data over the conterminous United States are now available.

For this research, several summary datasets were chosen as potentially useful for characterizing either general climate, or climatic variability. Summary datasets are those that provide a general (often an average) representation of some climate variable over the 18 year period of the source data—the summary measures are not time or date specific. Table 2 lists all the DAYMET summary variables acquired for initial consideration, from each of three categories: temperature, precipitation, and humidity. These layers are provided in a Lambert azimuthal equal area projection from the DAYMET project.

Table 2: DAYMET variables (SD=standard deviation)

Category	Sub-category	Variable	Description
Temperature	Daily air temperature	TAA	Average daily air temperature
		TAS	SD of average daily air temperature
	Annual extremes	TXAS	Mean maximum July temperature
		TXSS	SD of maximum July temperature
		TXAW	Mean maximum of January temperature
		TXSW	SD of maximum January temperature
	Day-to-day mean variation	TVNA	Annual mean day-to-day mean temperature variation
		TVNS	SD of day-to-day mean temperature variation
	Growing degree days	TDA	Annual mean growing degree days (sum of positive Celsius temperatures)
		TDS	SD of growing degree days
	Daily range	TRA	Annual maximum – annual minimum daily temperature range
		TRS	Summer maximum – summer minimum daily temperature range
		TRW	Winter maximum – winter minimum daily temperature range
	Frost	TFA	Annual mean number of days with frost
		TFS	SD of annual number of days with frost
Precipitation	Annual	PTA	Mean annual precipitation
		PTS	SD of annual precipitation
	Event size	PEA	Average precipitation event size
	Frequency	PFA	Mean precipitation frequency
		PFS	SD of precipitation frequency
Humidity	Daily	HVA	Average daily water vapor pressure
		HVS	SD of daily water vapor pressure
	Day-to-day variation	HVVA	Mean day-to-day variability of water vapor pressure

3.3. Productivity: GLOPEM NPP

Due to the difficulty of computing actual measures of primary productivity (NPP), most existing research on this topic uses typically uses proxy measures such as satellite-measured NDVI (normalized difference vegetation index), or potential or actual evapotranspiration (PET and AET) , which are generally derived from climate atlas data.

For this research, satellite derived productivity data were acquired in an attempt to improve models with a more accurate measure of productivity.

GLOPEM (Goetz, Prince et al. 1999; Cao, Prince et al. 2004), a global, satellite-derived dataset of net primary productivity (NPP), was acquired from Stephen Prince at the University of Maryland. This dataset is produced at an 8 km grid cell resolution, and at a 10 day time-step. Data for the period 1982-1999 were used for this research. The GLOPEM dataset was selected because, at the beginning of this research, it was the only available measure of NPP produced globally, at regular time intervals, and available for several years. Having values of NPP every 10 days for an 18 year period is important because it allows measures of temporal variance to be computed.

The acquired GLOPEM dataset consists of NPP values for every 10 day period for every 8 km grid cell. From these raw data, the following summary variables were computed: the 18-year mean of the mean annual NPP (NPPAV); the standard deviation of the 18 yearly means (NPPSD); the 18-year mean of the mean annual sum of NPP (NPPSUMAV); the standard deviation of the 18 yearly NPP sums (NPPSUMSD); the 18-year mean of the annual range of NPP (NPPRNGAV); and the standard deviation of the 18 yearly NPP ranges (NPPRNGSD). That is, for each year in the 1982-1999 period for which GLOPEM data are available, an annual average, an annual sum, and an annual range (maximum – minimum) were calculated. The means and standard deviations of each of these produced the set of variables initially considered in the analysis.

3.4. Habitat Heterogeneity

Habitat heterogeneity was quantified with measures of landcover complexity, and with a measure of topographic complexity.

3.4.1. Landcover

Two simple measures of landscape complexity were defined using a landcover map. The landcover map used was that produced by the state GAP projects. These individual state maps were chosen over global or continental maps because they appeared to be more accurate, particularly in their delineation of farmland. Also, as the landcover map was an important part of each state GAP project (since it would then be used to map wildlife habitats) and so was likely carefully developed, it seemed probable that these products would provide more accurate landcover delineations than maps developed at continental or global scales. The GAP landcover maps also provided sufficient spatial resolution to compute the wanted landcover metrics. For the eight states of this study, they were provided at either 90 or 100 m resolution, although one state – Washington – provided landcover (and all its products) in vector format, but with approximately the same working resolution.

For some of the preliminary analyses, the UMD 1 km global landcover classification (Hansen, DeFries et al. 2000) was also used. This classification provides 13 different landcover classes.

3.4.2. Elevation difference

A measure of topographic complexity was computed from the difference between maximum and minimum elevation, per grid-cell. For these computations, the

Hydro1k (USGS EROS Data Center) 1 km resolution digital elevation model (DEM) was acquired.

Chapter 4: Methodology

4.1. Overview of analytical approach

The aim of this research is to determine which of several competing hypotheses is best supported by the best available data. Thus, the problem here is somewhat unlike more traditional research driven by a single hypothesis, which is proposed and tested. Rather, several existing hypotheses need to be evaluated, and that with most support identified; the hypothesis of this study is that one of these existing hypotheses will emerge as better supported by the data than the alternative hypotheses. As the research was initiated as a response to the diversity of opinions on this question, no particular hypothesis has been suggested to be, or treated as, likely or preferred; to do so might bias the results.

A trivial evaluation of the working hypothesis might simply evaluate a single model for each hypothesis, compare the results, and based upon best model fit or best predictive ability, pronounce one hypothesis to be best. However, if such a simple approach could work, there would not likely be the current contentious debate on this topic. In reality, there are many factors unrelated to the validity of the various hypotheses that may nevertheless affect the outcome of such a test, including: spatial scale, type of model used, geographic area examined, vertebrate group considered, and choice of variables to include in the models. To try to avoid the specific results that might derive from such a particular choice of factors, this research attempts to come to more universal conclusions by considering the weight of evidence from many permutations on evaluating the hypothesis models. Relying upon the weight of evidence also avoids the otherwise difficult issue of defining a single statistical test to validate one hypothesis over the others. Although some work has been done in this type of model selection (Burnham and Anderson 2002), the work is new and not fully developed; thus it was decided using more traditional statistical approaches over many model permutations would provide more reliable and defendable conclusions.

To address the fundamental question – identification of the best supported hypothesis – three evaluative measures were used. First, individual hypothesis models (models testing a specific hypothesis) were constructed and their performance evaluated using standard criteria, such as AIC (Akaike's Information Criterion (Akaike 1973)) and R^2, that evaluate overall model performance. Second, the coefficients for the variables included in these models were examined, for consistencies in sign, magnitude, and significance, across all models. Finally, to directly compare the performance of the different hypotheses, likelihood ratio (LR) comparisons were conducted against a set of combined models.

Within that overarching analytic framework, there were several important considerations, such as the selection of variables to use in each hypothesis, and the various model permutations. These and other decisions about the modeling framework are described below in section 4.2. For some modeling decisions, preliminary and more detailed work was conducted in limited geographic areas – either California, or over a few states – to evaluate the best way to proceed with the full, eight state analysis. These elements are described in section 4.3 on preliminary work. The details of bringing the disparate data sets together into a uniform map grid so that data points could be extracted for the analyses is reviewed in section 4.4 on data preparation. Section 4.5 describes the procedures used for running the models, model diagnostics are discussed in section 4.6, and the particulars of the three measures of hypothesis evaluation mentioned above are described in section 4.7. Finally, this chapter finishes up with a list of the software used for this work in section 4.8.

4.2. Modeling framework

The modeling framework defines what models will be constructed to test the hypotheses under study. There are several parts to this: initial variable selection for each hypothesis; the use of a multi-dimensional analytic structure; the choice of model form; the number of predictor variables allowed per model; the selection of variables to actually include in the models; and the procedures used to run the models.

4.2.1. Initial variable selections

Three main hypotheses are under examination here: energy/productivity (H1), climatic variability (H2), and habitat heterogeneity (H3). For each, a set of candidate variables was chosen as possible predictor variables, based upon both the hypothesized mechanism, and the available datasets. This initial selection of variables for each will be discussed in some detail below. Before the final models were run, a detailed variable selection procedure was conducted to determine which of these variables should actually be included in the models; this is discussed below in section 4.2.5. For abbreviation purposes in the text, figures, and tables, each variable is given an abbreviation (such as, for example, NPPAV for average NPP). For reference, the List of Abbreviations on page vii has a complete list of all variable abbreviations.

4.2.1.1. Energy / productivity

The energy/productivity hypothesis has been broken out into 3 separate hypotheses: H1a: productivity, H1b: energy, and H1: productivity + energy. For H1a (productivity), three variables were considered: mean annual NPP (NPPAV), mean annual sum of NPP (NPPSUMAV), and the mean annual range of NPP (NPPRNGAV). For H1b (climatic energy), two main variables were considered: growing degree days (TDA), and mean daily temperature (TAA). For H1 (combined, energy and productivity), all the variables from H1a and H1b were considered.

4.2.1.2. Climatic variability

The climatic variability hypothesis posits a dependence of species richness on variability (or lack thereof) in climatic conditions. The DAYMET dataset provides many relevant measures of climatic variability, including measures of temperature, precipitation and humidity; table 2 (page 43) lists all DAYMET variables considered in this study. For the climatic variability hypothesis, the candidate temperature variables were: the variability of daily air temperature measured as the standard deviation (SD) of average daily air temperature (TAS), as well as the mean and SD of the day-to-day mean temperature variations (TVNA, TVNS); annual extremes of temperature measured as the mean and SD of July and January temperatures (TXAS, TXSS, TXAW, TXSW); and measures of the variability of freezing, via the SD of the number of growing degree days per year (TDS), and the mean and SD of the number of days per year with frost (TFA, TFS). For precipitation, the candidate variables were: mean and SD of annual precipitation (PTA, PTS); average precipitation event size (PEA); and the mean and SD of precipitation frequency (PFA, PFS). The candidate variables for humidity were: the mean and SD of average daily water vapor pressure (HVA, HVS); and the mean day-to-day variability of water vapor pressure (HVVA).

4.2.1.3. Habitat heterogeneity

The habitat heterogeneity hypothesis suggests a dependence of richness on the complexity of habitats. Although “complex habitat” will have a varying meaning depending on the type of organism considered, for this research it was simply interpreted as spatial complexity, either horizontally or vertically. For horizontal spatial complexity, two measures were chosen: the within-pixel variability (GLCV), and the variability within the pixel’s focal area (GLCF) – which here is defined simply as the surrounding eight pixels. For vertical complexity, the difference in elevation between the highest and lowest points in the pixel was used (ELD).

4.2.2. Multi-dimensional analysis

“Multi-dimensional” analysis refers to the generation and use of hypothesis model variants along each of several different dimensions: four spatial scales, six groups of vertebrates, and nine geographic areas. For each hypothesis, then, there will be 216 (4 x 6 x 9) models generated to cover all these dimensions. The advantages of such a multi-dimensional analysis are threefold. First, by not selecting a single choice for each dimension (e.g., a particular spatial scale or one particular geographic region), the results are not constrained, or biased, by essentially arbitrary methodological decisions. Second, by including variants across each dimension, the limitations of the results can be more finely explored, in terms of where they hold best, and where they may fail. Finally, the multi-dimensional analysis multiplies the number of model results generated, and this allows conclusions to be drawn from the weight of evidence of many different models, rather than from a single or a few models that may be biased by certain choices, or subject to possibly spurious or non-characteristic statistical effects.

Note that there are other dimensions to the analysis, mentioned elsewhere, but which are not specifically included in what are termed “dimensions” here. For example: the three different groups of models based upon the number of included predictor variables (all, two-variable, and one-variable; section 4.2.4); the use of different model forms (ordinary least squares and spatial regression models) in some cases (section 4.2.3); and other more limited situations that were only tested in limited geographic contexts (California or Pacific northwest states) (section 4.3). In some cases these could also be termed additional “dimensions”, but generally these additional dimensions were not applied universally, or for explanatory and methodology reasons they are dealt with in separate sections. What is described here as “dimensions” are model variants that were universally applied for all analyses.

4.2.2.1. Spatial scale

Analyses were conducted at several spatial scales in order to test the effect of scale, as well as to eliminate the requirement to choose one particular scale for analysis. As published work on the topic has been conducted at a very wide range of scales, there is no preferred choice. For this research, scale was also constrained by the resolution of some of the environmental data. Specifically, the GLOPEM NPP data is only available at 8-km grid cell resolution, so this set the lower-limit (finest spatial scale) of the analysis to cells of 8-km on a side. Three multiples of this finest scale were also chosen: cells 16, 32, and 64-km on a side, giving an overall 8-fold increase in cell size, measured by grid cell spacing, from finest to coarsest scale. The equivalent areas per cell are 64, 256, 1024, and 4096 km^2, giving a 64-fold increase in area between cells at the finest and coarsest resolutions. Using spatial resolutions coarser than 64-km is not practical due to an unacceptable drop in number of observations that then occurs, especially for smaller states. Thus, for analysis, the dimension of scale was broken out into four groups, with grid-cell dimensions of: 8, 16, 32, and 64 km.

4.2.2.2. Vertebrate class

Although inclusion of aquatic vertebrates and invertebrates would fit the overall goal of this research, the current analysis is confined to terrestrial vertebrates for purely practical reasons. Primarily, neither all aquatic vertebrates nor, certainly, all invertebrates have had their distributions mapped in any detail; the data that would be required for this analysis simply do not exist. Further, aquatic vertebrates and many invertebrates inhabit landscapes (rivers, lakes, oceans), that would require a substantially different analytic methodology than terrestrial vertebrates for many reasons. Invertebrates, and aquatic vertebrates, tend to be fairly small animals, with some invertebrates even being microscopic, so inclusion of them would require explicit attention to animal size, and a corresponding explicit effort to include size in the analysis. Although useful (see Badgley & Fox (2000), who examined size among North American mammals), gathering useful and representative size information for all terrestrial vertebrates, let alone larger invertebrates or aquatic vertebrates, would be difficult if not impossible. Thus, this study only examines animals in the four classes of terrestrial vertebrates: amphibians, birds, mammals, and reptiles.

Because the hypothesized influences on richness are not dependent on vertebrate taxonomic class, but rather apply to all living things, all four vertebrate classes are combined into one group, of all terrestrial vertebrates (referred to simply as ‘all vertebrates’ henceforth). However, as different groups of animals may well be expected to respond differently to the hypothesized influences on richness, it is also informative to look at each group individually, and attempt to draw out or highlight peculiarities in the generality of the hypotheses. Since the group of all vertebrates is heavily dominated birds (approximately 75% of species in these data), and because birds are able to move about their environment much more easily than non-volant vertebrates, it is also interesting to examine how the hypotheses fare for the non-volant vertebrates (mammals, amphibians, and reptiles). Thus, the dimension of vertebrate class was broken out into six different groupings of vertebrates: amphibians, birds, mammals, reptiles, all vertebrates (referred to simply as ‘all’ in tables and graphs), and non-volant vertebrates (mammals, amphibians, reptiles: referred to simply as ‘MAR’ in tables and graphs).

4.2.2.3. Geographic area

The question being examined here is suited to data at continental scales, because the environmental variables under consideration as hypothesized influences on richness show variation at continental scales. Thus, the ideal geographic size would be continental. Unfortunately, data availability issues have required a focus on a smaller geographic area. Partly, this resulted because the GAP projects producing the state-level species distributions are only now being completed for most states. When this research was begun, the western US provided the largest contiguous block of available GAP vertebrate data, and so was chosen as the primary study area.

Within this western US study area, it would be preferred if the analysis were conducted over the entire area simultaneously, and not divided by state. Unfortunately, due to differences in GAP methodology in each state, the vertebrate distributions, and resulting richness layers, show discontinuities across state borders (this will be discussed more thoroughly in the results section). Whether the discontinuities are fundamental enough to bias an analysis conducted over the entire study region can – and will be, in the Discussion (chapter 6:) – speculated upon, but is otherwise beyond the scope of the present work. To diagnose how these differences have arisen would require a thorough evaluation of each included state GAP project. Thus, instead, a dimension of geographic area was used, and simply broken into 9 groupings: each of eight individual states (California, Oregon, Washington, Idaho, Montana, Utah, Wyoming, and Colorado), and all eight combined. An obvious missing piece to this group is Nevada; although data for Nevada have been available, correspondence with the Nevada GAP project led to its elimination from analyses, due to both quality concerns and practical issues of the dataset's format. Nevada GAP is currently being redone by a consortium of southwestern states, and could be included in any future update to this work. Some of the preliminary work (section 4.3) was conducted on the four Pacific northwest states of California, Oregon, Washington, and Idaho – these will be referred to collectively as 'NW' in graphs, tables, and text.

4.2.3. Regression model forms

In determining the model form to use, the first decision was whether to use a regression model, or some other type of model, such as a classification, or a decision or regression tree. For this work, regression models were chosen, partly because almost all existing work on this topic uses regression modeling, and partly because the tools for parametric regressions are well developed in both the statistical literature and in statistical software packages.

With a regression model decided upon, the next decision is what form of model to use – whether linear, or non-linear, or some other option such as ridge or support vector regression. With little theoretical support for a non-linear model, the logical fit, and the choice supported by most existing work, is the ordinary least squares linear regression model (OLS). However, some other related model forms were also investigated and used for some of the analysis: general linear models (GLIM), which can help linearize non-linear relationships and address non-Gaussian effects; and spatial autoregressive models, which can help properly account for the spatial autocorrelation of the data.

4.2.3.1. Ordinary Least Squares

For the bulk of this research, simple ordinary least squares (OLS) (Fox 1997) models were constructed, for several reasons. Primarily, they are simple, widely-understood, and fast to compute. Also, as the vast majority of previous work on this topic uses OLS models, using them in this work allows the results to be more easily compared to other studies. Because of their ubiquity and utility, OLS model forms also have a wide variety of diagnostic tools available, both theoretically (proven and expounded upon in the literature) and practically (actually pre-programmed and available in software packages). Thus, OLS was a very attractive option, and was made the standard model form for this work. However, other types of models were also considered, and were used for some of the regressions.

4.2.3.2. General Linear Models

General linear models (GLIMs) are a class of models similar to OLS, but they possess some advanced features, such as allowing for different residual error structures, or for non-normal input data (Fox 1997). Due to these features, GLIMS were investigated for some of the California work, and are discussed more extensively in the section on preliminary analyses (4.3)

4.2.3.3. Conditional Auto-Regressive Models

Another model type was used extensively, particularly for the California work: conditional autoregressive (CAR) models (Haining 1990; Cressie 1993). CAR models are spatial regression models, which explicitly account for the spatial autocorrelation of the modeled process. When data in an OLS model are spatially autocorrelated, as is the case with most geographic data, there is a risk of artificially deflated standard errors for the coefficients of the autocorrelated variables, and thus inflated t-scores for testing the significance of those variables against the null hypothesis (models without those variables). The result is that variables may be retained in models – particularly if variable selection techniques, such as step-wise regression, are being used – due solely to their spatial correlation with other included variables, both response and predictor, and not due to any intrinsic explanatory value. To avoid such padding of models with unimportant variables, CAR models were used to ensure that all variables retained in the OLS models were indeed important.

The basic equation for a CAR model, following Kaluzny et al (1998), is simply: $Z_i = \mu_i + \delta$, where Z is the random process being modeled, at site *i*; μ_i is the mean response at site *i* (essentially the OLS component); and δ is the spatial component, which is normally distributed with zero mean and a covariance matrix described by $\Sigma = (I - \rho N)^{-1} D\sigma^2$, where I is the identity matrix; ρ and σ are scalars estimated by the regression, with ρ indicating the magnitude of the spatial component; N is a matrix describing the spatial arrangement of neighbors to location i; and D is a weights matrix accounting for the variance of the marginal distribution. Thus to implement a CAR model, a neighborhood relationship must be specified to define N. This includes both the distance out to which values are expected to be correlated, and the strength of the spatial dependence, which is essentially a distance weighting. For example, the neighborhood might be defined as pixels within 100 km, with a uniform spatial dependence, in which case all pixels within 100 km have an equal influence on the pixel of interest. More typically, the dependence might be $1/d$ or $1/d^2$, where d is the separation distance, indicating decreasing dependence on the pixels further out.

Specifying the appropriate neighborhood function requires knowledge of the spatial dependence of the data. A simple method for describing this is to construct a spatial variogram (also known as a semi-variogram or correlogram) (Ripley 1981), which provides a graph of the correlation of all pairs of points according to their separation in space. This is done by identifying all points separated by a series of increasing lag distances, such as 0-10 km, 10-20 km, 20-30 km, 30-40 km, and so forth, as appropriate. Then, the sum of the squared differences between all pairs of points within each distance class are computed, averaged, and plotted as a function of that distance class. This sum of squared differences is denoted by (Kaluzny, Vega et al. 1998):

$$\gamma(h) = \frac{1}{|N(h)|} \sum_{N(h)} (z_i - z_j)^2$$

A typical variogram shows decreasing correlation out to some distance, termed the range, after which the variance stabilizes.

However, these standard variograms are constructed from Pearson r^2 correlations, and r^2 is only a simple measure of correlation, not a measure of spatial correlation. Thus, some (Legendre and Legendre 1998; Lichstein, Simons et al. 2002) have suggested using a variogram constructed from an appropriate index of spatial randomness, such as Moran's I (Cliff and Ord 1981), in place of the standard r^2 variogram. Moran's I is a statistic that indicates whether two points in space are positively correlated, negatively correlated, or not significantly correlated. Moran's I returns a scalar value (from -1 to 1), but there are no default thresholds to indicate at which value the correlation, whether positive or negative, is deemed significant. Thus, to ensure that the returned Moran's I values were indicative of spatial autocorrelation, a randomization was conducted in which the value computed from the data was compared to 100 permutations with spatially randomized data (Lichstein, Simons et al. 2002).

The spatial dependence in the OLS residuals is the spatial dependence that the OLS model is unable to account for, and thus the range of this dependence defines the spatial neighborhood that the CAR model should target. To determine the appropriate spatial neighborhood distance, both regular r^2 variograms and Moran's I variograms were constructed for the residuals of OLS models, and the appropriate range extracted from the variogram. Both types of variograms were used in order to investigate the necessity of using the much more computationally intensive Moran's I variogram, but only for the preliminary models in California (4.3), due to the practical issues of computing the Moran's I variograms over much larger areas.

To determine the appropriate spatial dependence function, CAR models were run with three different dependence functions, and the results evaluated to see which produced the best output (Lichstein, Simons et al. 2002). The three evaluated functions were: *1* (uniform), *1/d*, and $1/d^2$, where *d* is the separation distance. Results from CAR models using each dependence function were evaluated for residual spatial autocorrelation using, again, both regular variograms and Moran's I variograms. The dependence function found to be most generally useful was then selected.

4.2.4. Number of predictor variables

A crucial feature of the methodology was the introduction of essentially another 'dimension' of the analysis, related to the number of predictor variables in each model. Initially, all models were constructed using all predictor variables (PVs) – but only after the variable selection step (details in the next section) had reduced the number of PVs to a reasonable subset. However, a preliminary examination of the results revealed that the models with the largest number of PVs were also the models performing best. It was not clear if this correlation was purely due to the number of variables, so to ensure that was not the case, two additional sets of models were developed for all hypotheses: one with just one predictor variable, to match the hypothesis models which would only ever have one PV given the available datasets (H1a and H1b); and one with just two PVs, to match the hypothesis models that would only ever have two PVs (H1, the combination of H1a and H1b). H3, the heterogeneity hypothesis, has three potential contributory PVs, but in most cases one of these drops out (as it is fairly well correlated to one of the others), and so H3 most often only contains two significant PVs. The end result of adding this "dimension" became key to deriving useful and sensible conclusions from this work.

4.2.5. Variable selection methods

Variable selection is the process in which an initial set of variables that might potentially be used in a given model is reduced to a more manageable and meaningful subset. The naïve way to proceed might be to simply include all possible predictor variables (PVs) in a given model. However, this can quickly lead to serious problems that may compromise the interpretability of the results. Most importantly, if PVs are correlated to one another to any substantial degree, the resulting multicollinearity can lead to incorrect conclusions about the significance of any one variable in the regression, based upon standard parametric significance tests. Also, the contributions of other variables can be masked by the inclusion of a collinear variable. The basic problem, then, is to remove extraneous collinear variables, while retaining the optimal set of variables capable of explaining the response variable.

For this research, the issue quickly becomes more complicated, because, as it turns out, the relationships between the environmental variables and species richness does tend to vary significantly with the vertebrate class considered. Thus, for example, if all vertebrates are being considered, a given variable might be removed as not significant, or because it is substantially collinear with other retained variables. However, this variable might also have the strongest relationship to the class of reptiles, so removing it could significantly affect the subsequent performance of the reptile models. Since "all vertebrates" is largely dominated by birds, one must specifically consider all of the other vertebrate groups used as response variables, to ensure any such dependencies are not overlooked.

One tack would be to determine a different set of PVs for each response variable. The disadvantage here is that models (and their parent hypotheses) cannot be effectively compared across the different response variables (vertebrate groupings). It was decided against such an approach, because one of the important goals of this research is to describe the cases and conditions under which the different hypotheses are best supported.

Given that objective, it was then necessary to find a way to screen all the relevant variables for each different response variable, and extract optimal subsets of PVs to use in the modeling. Note that this issue is most serious for the selection of variables for the H2 (climatic variability) hypothesis, because there were many potential PVs available. For the H1 and H3 hypotheses, the choice of appropriate variables was more limited.

To be clear, the objective of variable selection in the context of this research is to find an optimal subset of predictor variables for each hypothesis: H1a, H1b, H1, H2, and H3. Each of these sets of PVs will then be used as the starting point for constructing the models that test each hypothesis. Several methods were investigated for finding an optimal subset of PVs, including: pairwise r^2 comparisons; all-subsets regression; hierarchical partitioning; and principal components analysis.

4.2.5.1. Pairwise r^2 correlations

A first cut at generating an optimal subset of predictor variables is to examine the pairwise Pearson r^2 correlations between all variables. This will immediately indicate which variables are highly correlated to one another. It will not reliably indicate, however, correlations dependent on other variables being present – e.g., interaction effects. Since such effects can be important, and are one of the main complicating factors for multiple regressions, pairwise correlations are not sufficient to define an optimal set, but they are useful for identifying and removing strongly collinear variables.

4.2.5.2. All-subsets regression

All-subsets regression (Miller 2002) is a method for determining which of a set of predictor variables are most frequently correlated to the response variable, when included in multivariate regressions. For a set of k PVs, the method computes a model for each possible combination of those k PVs, for all cases from one-variable models up to the case with all k variables. For k variables, the total number of models computed is equal to $2^k - 1$. The output of interest is not the details of each model, but simply the summary counts of how often each PV was found to be statistically significant, in all $2^k - 1$ models. One can then quickly ascertain which PVs are consistently found to contribute to regressions against the given response variable, and which PVs contribute very infrequently. One practical issue with all subsets regression is that the computation time increases exponentially with the number of PVs; 10 PVs would require ~1000 regressions to be computed, while 20 would require ~1,000,000. Thus, practically, the variables to include may need to be preselected, by, for example, the pairwise r^2 comparisons previously mentioned.

For this research, all-subsets regressions were computed for the H2 (climatic variability) PVs, against each of the four vertebrate classes (birds, mammals, amphibians, and reptiles), at each of the four spatial scales, for the combined NW dataset

(California, Oregon, Washington, and Idaho). The results were then examined, and the PVs found to most frequently contribute to the models, across all vertebrate classes and spatial scales, were identified.

4.2.5.3. Hierarchical partitioning

Hierarchical partitioning (HP) (Chevan and Sutherland 1991) is another method that attempts to identify the most consistently important predictor variables in multiple regression. The method is similar to all-subsets regression in that all combinations of PVs are evaluated. However, whereas all-subsets regression simply indicates which PVs were significant in the computed regressions, hierarchical partitioning evaluates and summarizes the contribution to model fit of each PV, both independently and jointly, from the superset of all possible models (MacNally 2002). HP was designed not to find the best model, but to find the most likely causal variables, by identifying consistent patterns of explanatory power. To be able to select from the HP results a subset of variables to use in modeling, a randomization approach was also implemented. This approach compares the output of HP for each PV with the output generated after randomizing the data matrix, and thereby allows one to test the significance of the actual values against the distribution of those obtained from the randomized data. For this study, 100 randomizations were run for each case. However, because the method is computationally expensive, all variables of interest could not be run simultaneously. To overcome this, candidate variables were divided into groups and run in batches, with each group being run with all other groups. The PVs were divided into three groups A, B, and C, (each containing 4-5 PVs), and then HP was run three times, on the combined groups A+B, A+C, and B+C. The results from these runs were then manually examined to identify consistently important variables across all combinations.

4.2.5.4. Principal Components

Due to limitations from the results of the preceding methods, principal components (Jackson 2003) were used as the basis for a simpler and more common-sense approach to generating an optimal subset of predictor variables. The idea is simply to manually examine the variance in the candidate set of PVs, and select a set that are both fairly well correlated to the response variable, and that are mostly orthogonal to one another. In other words, a subset of PVs are selected that best represent the full data space of all the PVs, but only in terms of the components of that data space that show some correlation to the response variables. Principal components are used simply to summarize all the variance into just a few relevant variables. Although there was a chance that the relevant variance (e.g., that related to the response variables) would be spread across many or all of the PCs, which would render this method of little use, that, as the Results chapter will show, did not turn out to be the case. A disadvantage of this method, compared to all-subsets and hierarchical partitioning, is that significant interaction effects between PVs may be overlooked, since the PCs are constructed with no reference to the response variables. However, such interaction effects are not necessarily excluded, and may be represented in the PCs, but it is not possible to specifically ensure they are included.

To implement this screening method, principal components (PC) were generated from the full candidate set of PVs. These PCs were then correlated to the response variables, to see which PCs exhibited any significant relationship; only those PCs would be further examined. A significant relationship was defined as one in which the simple pairwise r^2 was higher than 0.20. A low threshold was chosen here because data spaces (PCs) with even minor correlations to the response variable might still be relevant, particularly in the realm of higher order interaction effects. This was examined for each of the four vertebrate classes, and at each of the four spatial scales, for the combined NW data. Then, for each pair of PCs that did have a significant correlation to at least one of the vertebrate richness response variables, biplots (Gower and Hand 1996) were generated. Biplots of principal components show how the original variables are distributed in the two-dimensional space of two plotted PCs. Based on these plots, then, PVs were selected with the following criteria: (1) good coverage of the data space by the selected set of PVs; (2) good magnitude (indicating that a PV is contributing significantly to that data space); and (3) as orthogonal as possible to the other selected PVs. This exercise was repeated for the four spatial scales, and the results compared across the four scales.

4.2.6. Other preparations

Aside from the above described framework, two additional concerns were normalizing and detrending the input datasets. Typically in OLS regression, neither predictor nor response variables need to be normally distributed, although the regression residuals do need to be normal. Thus, normalizing the input variables was not initially a concern. However, it later became clear that having normalized input variables would aid in the interpretation of the model coefficients; a higher coefficient in a model with normalized input variables simply means that that variable has a higher influence on the response variable. If the predictor variables are not normalized, one cannot make that assumption. The assumption can only be strictly valid, however, if the predictor variables are normally distributed, such that 'normalizing' them has the intended effect (returning a variable with zero mean and a standard deviation of one). Normalizing variables, whether they are truly normally distributed or not, does not affect other regression outputs, such as R^2, AIC, or model or predictor significance; it only affects the output coefficients. For this research, a parallel set of models was constructed with normalized predictor and response variables, so the differences in effect of individual predictor variables could be easily compared by simply examining coefficient magnitudes.

Another concern with the predictor variables for the H2 (climatic variability) hypothesis was that some of those variables might, along with the variability they quantify, also contain a significant trend signal. For variables that are derived as standard deviations, this is certainly not unexpected, as variance is often correlated (perhaps inversely) to the trend. For example, in precipitation datasets, the areas with low precipitation – deserts – often have highly variable precipitation, whereas some tropical areas often have high precipitation all year round – with low variance. In the context of this analysis, if a variable that is meant to represent variability is also correlated to the trend, these two aspects could become confounded, and lead to a less informative evaluation of the hypothesis. To address this issue, the predictor variables for the H2 hypothesis that quantify some type of variability were tested for a significant correlation to their parent trend variable. When this correlation was deemed substantial (r^2 above 0.5), they were detrended by simply removing the linear association between the two variables with an OLS regression.

4.3. Preliminary investigations in a restricted geographic context

An important feature of the methodology was the use of a limited geographic region to test out certain analytic methods and choices. Primarily, this was done using California, as it is a large and geographically diverse state; it was felt that lessons learned there should generally apply elsewhere. Furthermore, by conducting some of the analyses in a single state, the issues about integrating the GAP datasets across state lines would not be complicating factors. In particular, California was used to investigate: the use of generalized linear models (GLIMs); the use of spatial regression models (CARs); the inclusion of a landcover variable; and extensive model diagnostics. Unlike the other dimensions mentioned previously, it was not felt that replicating these analytic alternatives in detail across the full dataset would be particularly useful, while it would be considerably time consuming. The results did lead to the use of some of these methods – spatial models and model diagnostics – in a more limited fashion for the full analyses across the entire eight-state region.

Generalized linear models (GLIMs) (Nelder and Wedderburn 1972) can help address several common problems in linear regression modeling. Ordinary least squares (OLS) models are one form of a GLIM, but GLIMs add three important features to a standard OLS model: the ability to specify a structure for non-Normal errors; a use of a linear predictor, which is constructed from the PVs and can include interactions and transformations; and the specification of a link function that allows a specified transformation of predicted values, in order to match the response variable (Crawley 2002). The main value in this research is to be able to specify different error structures, as often the errors were not entirely normally distributed, and, due to the count nature of the data, may have been expected to be Poisson distributed. Thus, the primary focus of using GLIMs was to see if using Poisson error structures would improve the output models. Different link functions were also tested, but, lacking strong theoretical support for anything other than a linear relationship, this was not a priority, and not extensively evaluated.

Conditional autoregressive (CAR) spatial models were used extensively in California to evaluate their utility for this research. As described above in section 4.2.3.3, CAR models were primarily used to determine which variables to retain in OLS models because of their significance in the corresponding CAR model. Any variables identified as non-significant in the CAR model were then removed from the OLS model.

Another purpose of the California models was to evaluate the use of a landcover variable in the models. The motivation for this was to examine how models behaved if different relationships were allowed in areas that are natural, versus those regions that are significantly impacted by society (agriculture, orchards, urban areas). To do this, a landcover classification (Hansen, DeFries et al. 2000)was simplified to two classes (natural, and human-influenced), and included as a categorical factor in the hypothesis models.

Finally, the California models were also subject to intensive review of several model diagnostics, including: normality of residuals; non-constant error variance; variance inflation (a result of collinear predictor variables); outlier identification; and nonlinearity. These are discussed in more detail in the below section on model diagnostics.

4.4. Data preparation

All data, both predictor variables and response variables, were imported into or converted to ArcInfo (ESRI 1982-2005) grids (raster data format) from their native format. Within the ArcInfo environment, additional data compilation operations as required (described below) were conducted, and the resulting grids projected into a common output grid and projection. An Albers equal area conical projection was chosen for this, with a latitudinal origin of 45° N , a longitudinal origin of 100° W, and standard parallels of 29° 30’ and 45° 30’ – a commonly used projection for the continental US. A grid corner for all grids (1 km, 8 km, 16 km, 32 km, and 64 km) was chosen at coordinate 12970 E, 56200 N. After all data were properly compiled at all four spatial resolutions, the cell values, along with grid location, were output from the ArcInfo environment, for subsequent import to statistical packages. The following sections describe the particular manipulations required to import and prepare each dataset.

4.4.1. Vertebrate data

The species richness layers were generated by summing grids of vertebrate distributions. For example, if a species is predicted to occur in a given area, a binary distribution map is created by setting the grid values for the predicted area to 1, and the values elsewhere to zero. These binary distribution maps can then be summed in a GIS system, providing a single output grid whose cell values are the species richness at that location. To arrive at this point, many preliminary steps were required, generally including: import raw GAP vertebrate distributions to ArcInfo workspace; correct species identification nomenclature and remove unwanted species; project individual species distribution grids to the analysis’ standard projection and grid; aggregate distributions to a 1 km resolution grid; aggregate distributions to 8, 16, 32, and 64 km resolution grids; sum species of interest to species richness.

The vertebrate distributions from each state GAP project were not provided in a uniform format, so each state's data required individual attention. For example, most states provide the vertebrate distributions as raster layers, of varying spatial resolution and different projections, but Washington state provided them as vector polygon coverages.

The naming conventions used by different states also varied and needed to be harmonized, as different states included varying combinations of scientific name, common name, TNC ELCODE (a 10 digit taxonomic coding system developed by The Nature Conservancy), or state-specific species codes. The primary issue was species nomenclature, with occasional discrepancies between different states on scientific names, common names, or the ELCODEs. Due to these occasional discrepancies, each state's list of species was cross checked against those of neighboring states. Because the different GAP projects did not universally include or exclude exotic or introduced species, all such species were removed from further analysis at this stage.

Aggregation of the imported distribution grids to species richness was done in a multi-step process. The first step was to produce a 1 km resolution template layer for each species, in the analysis' final output projection and grid. Then, each state's vertebrate grids were inserted into the 1 km grid template, through a combination of projection and aggregation. Vertebrate distributions were provided in either 90 m or 100 m resolution grids, or as vector coverages (Utah and Washington). Cells in the 1 km grid were filled, indicating predicted presence, if 10% or more of that cell's area was predicted to have the species present. Initially, a 50% threshold was used, but this was found to exacerbate differences, between states, of the resulting richness layers. Because each state's distribution grids were in different projections and sometimes at different spatial resolutions, this step of aggregating to the 1 km base layer was customized for each state.

Once these 1 km base distribution layers were generated, the aggregation to coarser scales could then be done uniformly. As with the aggregation to 1 km, the upwards aggregation was done using a 10% rule as well. That is, for example, if 10% of an output 32 km grid cell is covered by predicted habitat in the 1 km grid, then that 32 km grid cell will be defined as predicted habitat. Each spatial scale was aggregated in this manner, directly from the 1 km base layers, resulting in a distribution layers for each species, at each spatial scale. These layers were then summed to species richness.

4.4.2. NPP

GLOPEM NPP data, at 8 km resolution and 10-day intervals from 1982 through 1999, was first projected to the output Albers projection and grid. Each NPP grid was then averaged to generate 16, 32, and 64 km resolution versions. Then, productivity variables were calculated, using custom IDL (RSI 2005) code, at each spatial resolution. The calculated productivity variables were: 18-year average NPP (NPPAV); 18-year NPP standard deviation (SD) (NPPSD); average of yearly summed NPP (NPPSUMAV); SD of yearly summed NPP (NPPSUMSD); average yearly range of NPP (NPPRNGAV); and SD of yearly NPP range (NPPRNGSD).

4.4.3. Climate data

DAYMET climate variables (table 2)(Thornton, Running et al. 1997; Thornton, Hasenauer et al. 2000), which are provided at 1 km resolution, were imported to the ArcInfo workspace and projected to the output Albers projection and grid. Then, variables were generated at 8, 16, 32, and 64 km resolutions by simply averaging the 1 km grids.

Aside from the raw DAYMET variables, an additional variable was constructed to quantify yearly temperature ranges, via the difference between summer and winter maximum temperatures: TXD = TXAS – TXAW.

4.4.4. Landcover

Landcover classifications were used in three ways in this study: to generate the heterogeneity measures used in the analysis; to test inclusion of a landcover variable, in the exploratory California models; and to eliminate areas modified by humans from further analysis for the full analysis.

The California landcover models use the UMD 1-km landcover product (Hansen, DeFries et al. 2000). This product was projected and regridded to the output analysis resolutions (8, 16, 32, and 64 km) with a plurality rule – the output pixel was assigned the landcover common to the highest number of constituent 1 km pixels. The classification was then regrouped to three broad landcover classes: natural, human modified, and arid/desert (with arid/desert assumed to be largely natural). For some of the California analyses, the ‘arid/desert’ class was combined into the ‘natural’ class, to produce a binary classification of simply human-modified and natural classes.

For the heterogeneity variables used for the H3 hypotheses, and for the purpose of screening out areas modified by humans, the landcover classifications of the individual GAP projects were used. These were preferred to the UMD landcover simply because it was assumed they would more accurately represent each state's landcover, as they were generated at a state level with local expertise. As with the vertebrate distributions, importing these individual classifications required an approach customized for each state. Once imported to ArcInfo workspaces, the state landcover grids were projected to the output projection, aggregated, via the plurality rule, to 1, 2, 4, 8, 16, 32, and 64 km resolution grids, and inserted into a single landcover grid covering the entire analysis region. The grids at 8, 16, 32, and 64 km resolution were also carried along and exported with the data, to allow removal of pixels of unwanted landcover (human modified) in the statistical modeling environment.

The GAP landcover grids formed the basis for constructing two derived spatial heterogeneity measures: the within-pixel landcover variety (GLCV), and the surrounding-pixel landcover variety (GLCF). GLCV was created by counting the number of different landcovers represented in the constituent pixels from the version of the landcover at 1/8 the desired resolution. Thus, for the 8 km GLCV, the cell values were set equal to the number of different landcovers in the 1 km grid cells that comprise the 8 km cell. This method was felt to be more resistant to scaling issues than the alternative of aggregating upwards sequentially, from 1 to 8 km, from 8 to 16 km, and so on. However, due to the different number of classes in each state's landcover classification, the GLCV values exhibit inconsistencies across state borders. To overcome this, the raw GLCV values for each state were normalized by dividing by the maximum raw GLCV value for that state.

For GLCF, each output cell was assigned the number of landcover types present in the immediately adjacent 8 grid cells, from the landcover at that output resolution. Thus, for the 16 km GLCF, the cell values were set to the number of different landcovers in the 8 adjacent grid cells of the16 km landcover grid. Because this number could only vary from 1 to 8, and did not appear markedly discontinuous across state borders, it was not further normalized or adjusted.

4.4.5. Elevation

The elevation difference (ELD) variable was constructed from the Hydro1k (USGS EROS Data Center) 1 km resolution digital elevation model (DEM). For each output resolution layer, ELD cell values were defined as the difference between the maximum and minimum elevations present in the constituent 1 km cells.

4.5. Model execution

OLS and GLIM modeling was conducted in the R statistical environment (R Development Core Team 2005). Initial models were constructed using a backward elimination approach (Crawley 2002); all candidate variables (after variable selection had reduced this to the final set) were included in each model, and then the variables with no significance to the model were removed, one at a time, starting with the least significant. After each variable was removed, the model was recomputed and variable significance reevaluated. Parsimony was an important consideration in deciding when to stop removing variables, with a preference for removing those with only marginal significance. Given that the spatial autocorrelation of the datasets tends to increase the significance of included variables, significance levels 99% or better were required to allow a variable to remain in a model. VIFs (variance inflation factors) were also used to determine which variables were contributing most to high levels of collinearity, and thus should be removed. For the preliminary modeling, this was a very important step, but for the final, eight-state, modeling, the variable selection methods had already removed collinear variables, and so reference to VIF values was not useful.

CAR models were run in the Splus (Insightful Corporation 1988-2002) statistical environment, which, for the present purposes, is practically identical to the R environment. However, Splus has an additional module for spatial statistics, S+SpatialStats (Kaluzny, Vega et al. 1998), which is one of the few available statistical packages capable of running CAR models. CAR models were run on final OLS models, after backward elimination, to determine if any variables included by the OLS model were only being included due to spatial autocorrelation. Variables that were not found significant in the CAR model were identified, and then removed from the corresponding OLS model. This occurred in less than 10% of all models. Due to the computational intensity of running CAR models, they could not practically be used for all regressions in this study, particularly with iterative procedures such as backward elimination, so they were only used as described, to identify variables to remove from the OLS models.

Before models could be run, in either statistical environment, the appropriate set of data points to analyze was selected. Primarily, this required the removal of pixels with human-modified landcover, or that were primarily water, both of which were determined from the individual state GAP landcover products (section 4.4.4).

4.6. Model diagnostics

An important feature of any analysis is the use of various standard model diagnostics to evaluate the fit and performance of the regression models; one cannot simply assume they are performing as expected, and that the resulting parameter coefficients, predicted values, and R^2 are meaningful. The most fundamental model diagnostics include the standard errors (and significance) of the parameter coefficients, the residual standard error of the model, an F-test evaluating the null hypothesis (that the model explains significantly more variance than no model), and a measure of model fit, such as R^2. These diagnostics provide a useful quick evaluation of model performance, and thus were computed for all models analyzed. However, other diagnostics are necessary to determine if fundamental modeling assumptions have been violated, and to assure that the preliminary diagnostics are indeed representative – e.g., that the specified model is indeed providing an appropriate fit to the data. The additional diagnostics evaluated here include measures to identify outliers, collinearity (variance inflation), non-normal errors, non-constant error variance, non-linearity (Fox 1997), and the spatial dependence of the model residuals. For this study, given the large number of models generated (on the order of 5000, not including the preliminary CA and NW modeling), all the above listed model diagnostics were evaluated only for the California investigations. Diagnostics that were deemed useful there were then applied to the full analysis.

Outliers were identified by several means: plots of Cook's D statistic; quantile comparison plots of the studentized residuals; and by examining other diagnostic plots for outlying points. Cook's D statistic is a measure of the difference between regression coefficients when a given observation is removed or added (Fox 2002); thus, there is a value for each observation, and these can be plotted to identify outliers. Another method is examination of studentized residuals. Studentized residuals are generated by removing each observation in turn, and calculating the resulting residuals (Fox 1997). Overlaying confidence intervals on quantile comparison plots of the studentized residuals allows identification of outliers. Finally, a number of additional plots, as described in the following sections, also highlight data outliers, although that is not their primary purpose.

Collinearity (also referred to as multicollinearity) of predictor variables can affect the standard errors of the regression coefficients, and thereby detrimentally impact the interpretation of the model. Individual variables may be included or excluded from the model due more to their correlation with other predictor variables than due to their relationship to the response variable. One sign of collinearity is instability of regression coefficients when a variable is added or deleted. Variance inflation factors (VIF) (Fox 1997) present one solution to this problem, by quantifying the level of collinearity for each PV that is adversely affecting the model. General guidelines suggest VIFs be no larger than 2 – 4. This diagnostic was used extensively to ensure that the variable selection routines mentioned in section 4.2.5 were adequately reducing multicollinearity in all test models.

For ordinary least squares regressions, a key assumption is that the regression errors are normally distributed. If they are not, this can indicate that either the linear regression model is inadequate, or that a non-normal error distribution should be assumed (as is possible with GLIM models). Two plots are used here to evaluate the normality of regression residuals: a quantile comparison plot of the studentized residuals against the *t* distribution, and a non-parametric density plot. Both help identify conditions that may warrant a transformation of the response. In such cases, the Box-Cox transformation (Box and Cox 1964; Venables and Ripley 1999) provides an automated way to select the most appropriate transformation for the response variable. For models showing problems with non-normal residuals, the Box-Cox transformation was applied, and the new model evaluated for an improvement in distribution of error residuals.

Another key assumption of OLS is that the variance of the errors is constant. If this assumption is violated, transformations of predictor variables may be necessary, or other measures, such as weighted least squares, considered. Error variance was of some concern in this work because the count nature of the response variables can indicate the potential for Poisson errors. Thus, error variance was evaluated with several tests. First, studentized residuals were plotted against the model's predicted (response) values. Such plots should show no spread in residuals, but as they can be difficult to definitely assess, a nonconstant error variance score test was also computed (Fox 2002). This test follows a χ^2 distribution, and so provides a simple way to assess nonconstant error variance. A final test of error variance was implemented with Tukey's spread-level plot (Fox 2002). This plots the absolute value of the studentized residuals against the log of the fitted values, and also suggests an appropriate power transformation of the response to stabilize the variance.

Assessing model non-linearity is another important diagnostic. The OLS regression model assumes a linear relationship between the response and each predictor variable. If any of these relationships are substantially non-linear, the integrity and relevance of the model coefficients can be questioned. Non-linearity was evaluated with the use of partial-residual plots (Fox 2002), which plot, for each predictor variable, that variables' contribution to the predicted response plus the residuals, against that predictor variable. If the resulting plot is substantially non-linear, this suggests the relationship between that predictor and the response is also non-linear, and may require transformation.

A final and important diagnostic is the evaluation of the spatial autocorrelation of the regression residuals. Autocorrelated residuals indicate whether, in the OLS case, spatial autocorrelation may be adversely affecting the model outputs and a spatial model should be considered, or, in the CAR case, whether the CAR model parameters have effectively accounted for the spatial autocorrelation. To evaluate, r^2 and Moran's I variograms were constructed from the residuals of both OLS and CAR models, and evaluated for spatial dependence. Section 4.2.3.3 discusses variograms in more detail.

4.7. Hypothesis evaluation

A critical feature of the analysis are the methods used to evaluate and compare the hypotheses, and thereby come to well supported conclusions. Two strategies were employed: (1) generate models to test the individual hypotheses, and manually (e.g., without a statistical test) compare their performance across hypotheses, and (2) generate 'superset' aggregate models in addition to the individual hypothesis models, and use likelihood ratio (LR) tests to statistically compare performance between hypotheses.

The LR method has the advantage that a statistical test is used to determine which hypotheses are better supported; however, it has the disadvantage that interaction effects in the superset model may affect the output. Thus, both methods were used to provide a more robust set of conclusions.

4.7.1. Individual model comparisons

For the individual hypothesis model comparisons, both model performance and model outputs were evaluated and compared. Model performance can be quantified in several ways, including R^2, residual error, and newer methods such as Akaike's Information Criterion (AIC), which provides a measure of performance that penalizes for lack of parsimony. The penalty helps counterbalance the apparent increase in model performance that accrues simply due to a larger number of predictor variables.

For this research, both R^2 and AIC were used to evaluate overall model performance. R^2 was chosen because it provides a direct measure of a model's predictive performance, and this measure can be used to compare any two multivariate models. However, R^2 values are not used to compute statistical tests of significance, and they do not include a penalty for lack of parsimony. To address those issues, AIC was also used to evaluate model performance. AIC values can be directly compared between models, but only between models with the same basic structure; they must have the same response variable and the same observations, although the predictor variables can differ. Within those constraints, model AIC values are compared directly, and the model with the lower AIC value is the better model. To be able to use AIC to compare all models of interest (e.g., models with different response variables, at different scales, and in different geographic regions), models with the lowest scoring AIC – within the allowed set of comparable models (same response variable, same geographic area and thus observations, and same scale) – were tallied. These tallies then indicate which hypotheses are emerging from the AIC comparisons as the best supported. However, they do not provide any indication of the margin by which the best hypotheses were selected over the alternatives.

Output model coefficients were also examined for consistency. If a hypothesis is well supported by the data, then the coefficients of its predictor variables should show some consistency in sign and magnitude across the various dimensions of analysis. Consistency may also help separate hypotheses that are capturing the fundamental driving processes – which should be the same everywhere – versus hypotheses that are merely fitting to the data, independent of underlying drivers. To evaluate this, the sign and magnitude of the coefficients for each variable for all models included in the final set of analyses were examined.

However, two issues arise here: the effects of atypical and of non-normal response (vertebrate richness) data. If data for a particular state or states is out of character when compared to data from other states, the coefficients output by the OLS model will likely be significantly different, in either or both of sign and magnitude, since, essentially, a different dataset is being fit. This study did not attempt to directly remedy any such issues with the richness data, but the results are interpreted with this consideration. Normality of the data is also a concern here, when coefficients are being examined, because of the large range of values of the response (richness) data; bird richness is an order of magnitude higher than amphibian richness – the maximum richness for a 32 km cell is 295 for birds, and 17 for amphibians. Thus, the magnitude of coefficients for models of non-normalized data will be strongly affected simply by the difference in magnitude of the different response variables. To address this, all input datasets were normalized by subtracting the mean and dividing by the root-mean-square, and a parallel set of models run with this normalized dataset. However, many of the variables used in this analysis, and especially including the species richness, are not normally distributed. Thus, the applied normalization does not produce genuinely normally distributed variables, although they will be much more similar (closer to 'normal') than had they not been normalized. Their residual non-normality may have an impact on the examination of coefficients.

4.7.2. Aggregate model comparisons

In addition to the individual model comparisons, likelihood ratio testing was used to test the effect of removing the subsets of variables, for each individual hypothesis, from a superset model. In this case, the superset model is simply a model including all predictor variables from all the hypotheses of interest; it consists of: NPPAV and TDA for H1; TRA, TVNA, TFS, and PTSR for H2; and GLCV, GLCF, and ELD for H3. The likelihood is computed for this superset model, and then ratioed against the likelihood for the superset model after removing the group of variables for a given hypothesis. Thus, to test the H2 hypothesis, for example, the likelihood of the superset model is divided by the likelihood of the model containing all variables not in H2, e.g., the H1 variables (NPPAV, TDA), and the H3 variables (GLCV, GLCF, ELD). The generalized likelihood ratio test statistic is given by $-2 \log_e(L_0/L_1)$ (where L_0 is the likelihood of the subset model and L_1 is the likelihood of the superset model), and has a χ^2 distribution (Fox 1997). The significance levels computed from the ratio can then be used to directly determine which of the tested hypotheses is best supported by the data. The test cannot be used to compare any model that is not a strict sub-model of the superset model. Thus, for different spatial scales, geographic regions, and response variables, likelihood ratio tests were also computed, and the p-values from the significance test were summed over the different dimensions of the analysis to identify trends in hypotheses support.

4.8. Software requirements

Several software packages were used for this research. ESRI's ArcInfo Workstation (ESRI 1982-2005), versions 8.3 and 9.1, was used to compile, georeference, and sample for analysis, all datasets. Data exported from ArcInfo was then imported to the statistical analysis packages S-Plus version 6.1.2 (Insightful Corporation 1988-2002), and R version 2.2 (R Development Core Team 2005), in which all modeling was conducted. The Spatial Stats module version 1.5 (Kaluzny, Vega et al. 1998) of S-Plus was used for the CAR models.

Chapter 5: Results

5.1. Data preparation

Figure 1 shows the compiled vertebrate richness at 1 km, per vertebrate class, with the resulting richness from both the 50% and 10% rules. The 10 and 50% rules require that if the given percentage of the species of interest's source distribution covers the output grid cell, then that cell is counted as present. As is clear, the 50% rule leads to more discontinuities across state borders, particularly for birds and mammals, in Montana and Wyoming. The 10% rule minimizes these issues, and thus richness from the 10% aggregation rule was used for all analyses. The figure also indicates the starkly higher richness for birds in Colorado and somewhat depauperate amphibian and reptile richness for Wyoming. Except for reptiles, California richness also seems low, compared to Oregon, while Washington's reptile richness seems low. Clearly, there are significant differences between some pairs of states in levels of richness. Due to this, the spatially combined models – where there is no indication of state – were not relied upon for conclusions, and individual models were run for each state.

The richness data at each output resolution (8, 16, 32, and 64 km) was generated from the 10% richness maps at 1 km, again using a 10% aggregation rule. Appendix I contains maps of these along with the predictor variables for all hypotheses, at 16 km. Maps at 8, 16, and 64 km are similar, and so are not presented.

Appendix II (figures 41 and 42) contains plots of the distributions of each variable. These are over-plotted with a fitted Gaussian distribution, to give some indication of the normality, or lack thereof, of each variable.

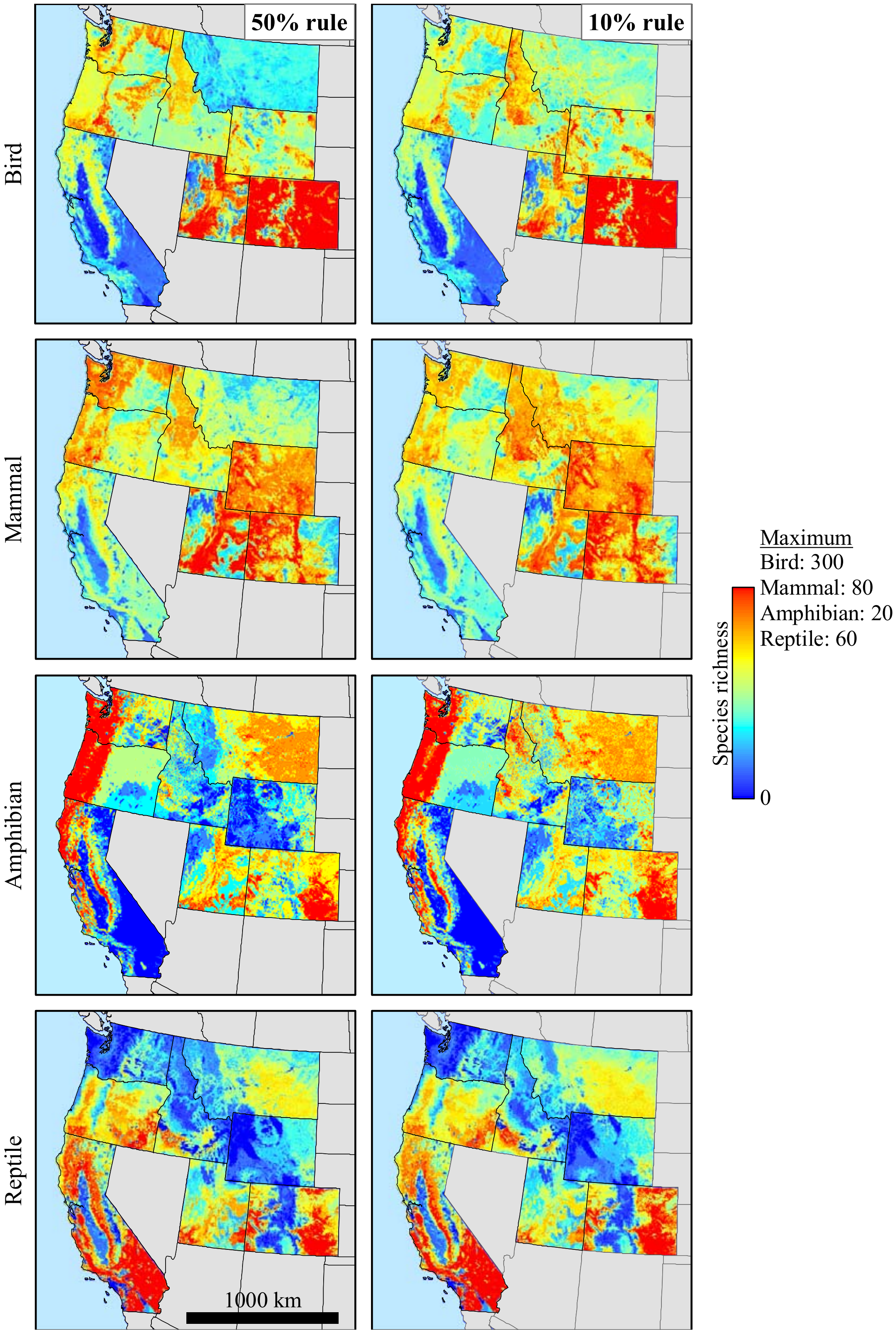


Figure 1: Species richness compiled from GAP vertebrate distributions. 1 km resolution. Left column shows maps resulting from 50% rule, and right from 10% rule. Note generally improved cross-border continuity with 10% rule; only 10% rule richness was used in this study.

5.2. Preliminary models: CA & NW

A great deal of preliminary work was conducted over a limited geographic area, to find the most effective methodological choices to use on the full, eight state analysis. Some of these choices were explored only for the state of California, while others were explored for a combination of four Pacific northwest states: California, Oregon, Washington, and Idaho. These will be referred to as the CA and NW studies, respectively. These preliminary studies were important for determining the variables to include in analysis, the types of models and model diagnostics to use, and to evaluate the utility of including a landcover variable in the analysis.

A note on terminology, particularly for figures and tables: “all” is used in two contexts, in reference to both geographic areas, and to vertebrate groups. In the former, “all” refers to the combination of either four or eight states; four in the NW context (CA, OR, WA, ID), and eight in the full analysis context (CA, OR, WA, ID, MT, UT, WY, CO). In either case, the data for the four or eight states has been combined into a single layer, which is then analyzed as a single unit. “All” in reference to vertebrate groups always refers to the total vertebrate richness; that is, the sum of amphibian, bird, mammal and reptile richness, analyzed as a single entity.

5.2.1. Detrended variables

The candidate H2 variables that are standard deviations of some measure were tested for the level of correlation to that measure for each NW state and all four states combined, and for all four spatial scales. The resulting r^2 values are listed in table 3.

Table 3: Correlation of variability and reference variables. r^2 values between variables measuring variability (VV) and their reference variable (Ref), for the four NW states, and all four combined (All), at each spatial scale. (Values higher than 0.50 are in bold.)

Variable		8-km					16-km				
VV	Ref	CA	ID	OR	WA	All	CA	ID	OR	WA	All
HVS	HVA	0.39	0.09	0.01	0.28	0.06	0.42	0.10	0.00	0.27	0.23
PFS	PFA	**0.90**	0.07	0.12	**0.54**	**0.52**	**0.91**	0.08	0.11	**0.56**	**0.57**
PTS	PTA	**0.95**	**0.79**	**0.90**	**0.93**	**0.72**	**0.95**	**0.79**	**0.90**	**0.94**	**0.76**
TDS	TDA	0.14	0.01	0.42	0.01	0.00	0.14	0.03	0.39	0.00	0.00
TFS	TFA	0.28	0.32	0.00	0.01	0.09	0.28	0.31	0.00	0.01	0.05
TVNS	TVNA	**0.61**	0.42	**0.55**	**0.76**	**0.78**	**0.60**	0.38	**0.55**	**0.75**	**0.81**
TXSS	TXAS	0.24	**0.54**	0.05	0.03	0.37	0.32	**0.54**	0.01	0.09	0.24
TXSW	TXAW	0.22	0.14	0.33	0.26	0.21	0.22	0.15	0.32	0.17	0.24
NPPSD	NPPAV	**0.84**	**0.89**	**0.79**	**0.73**	**0.58**	**0.86**	**0.90**	**0.81**	**0.73**	**0.56**
NPPSUMSD	NPPSUMAV	0.38	0.24	0.00	0.09	0.31	0.40	0.23	0.03	0.11	0.18
HVVA	HVA	**0.80**	0.27	**0.60**	0.06	**0.56**	**0.83**	0.27	**0.61**	0.06	**0.64**

Variable		32-km					64-km				
VV	Ref	CA	ID	OR	WA	All	CA	ID	OR	WA	All
HVS	HVA	0.45	0.11	0.01	0.23	0.24	0.40	0.13	0.00	0.39	0.24
PFS	PFA	**0.92**	0.08	0.10	**0.53**	**0.58**	**0.92**	0.07	0.17	**0.68**	**0.61**
PTS	PTA	**0.96**	**0.79**	**0.89**	**0.96**	**0.76**	**0.96**	**0.79**	**0.93**	**0.96**	**0.76**
TDS	TDA	0.14	0.05	0.34	0.01	0.00	0.13	0.07	0.34	0.00	0.00
TFS	TFA	0.30	0.30	0.00	0.03	0.04	0.27	0.26	0.03	0.01	0.03
TVNS	TVNA	**0.59**	0.32	**0.58**	**0.71**	**0.82**	**0.55**	0.31	**0.59**	**0.81**	**0.83**
TXSS	TXAS	0.42	**0.57**	0.00	0.16	0.25	0.33	**0.57**	0.14	0.01	0.26
TXSW	TXAW	0.22	0.17	0.37	0.04	0.24	0.18	0.15	**0.52**	0.28	0.25
NPPSD	NPPAV	**0.87**	**0.88**	**0.81**	**0.67**	**0.51**	**0.82**	**0.85**	**0.81**	0.43	0.45
NPPSUMSD	NPPSUMAV	0.42	0.21	0.09	0.18	0.14	0.40	0.23	0.22	0.13	0.11
HVVA	HVA	**0.86**	0.26	**0.62**	0.09	**0.65**	**0.84**	0.27	**0.72**	0.09	**0.66**

As the table indicates, PFS, PTS, TVNS, NPPSD, and HVVA all show significant and usually consistent correlation to their respective reference variables, across all spatial scales and for most states. After some consideration, it was decided not to detrend TVNS, because TVNA does not represent a mean value, as do the other reference variables here. Rather, it represents day-to-day variability, and thus it was not clear what the interpretation would be of a TVNS variable after removing TVNA. Despite a generally low correlation to its reference, TXSS was also detrended, because of the high correlation for Idaho, simply to avoid possible issues. The variables were detrended by simply removing the linear relationship derived from an OLS model relating the variability variable to its reference variable. In this document, detrended variables are indicated by appending an “R” to the variable name. Thus, the new variables generated, and subsequently used in the analysis, are: PFSR, PTSR, TXSSR, NPPSDR, and HVVAR.

5.2.2. Variable selection

The initial set of variables that were considered for evaluating each hypothesis are listed in table 4 (along with an indication of which were eventually selected). Several different methods were used to reduce the number of candidate variables to include in analyses, including: pairwise comparisons; all subsets regression; hierarchical partitioning; and principal components analysis.

Table 4: Candidate and final predictor variables, per hypothesis. “Selected” indicates whether a variable was selected for the final analysis.

Hypothesis	Variable	Selected?
H1a	**NPPAV**	**Yes**
	NPPSDR	No
	NPPSUMAV	No
H1b	**TDA**	**Yes**
	TAA	No
H3	**GLCV**	**Yes**
	GLCF	**Yes**
	ELD	**Yes**
H2	PTA	No
	PTSR	**Yes**
	PEA	No
	PFA	No
	PFS	No
	HVVAR	No
H2	TAS	No
	TXAS	No
	TXSSR	No
	TXAW	No
	TXSW	No
	TXD	No
	TVNA	**Yes**
	TVNS	No
	TDS	No
	TRA	**Yes**
	TRS	No
	TRW	No
	TFA	No
	TFS	**Yes**
	HVA	No
	HVS	No

5.2.2.1. Pairwise comparisons

The first step to evaluate variable selection was to examine the simple pairwise r^2 correlations between all candidate variables for each hypothesis. If any pair of variables show a high correlation, then one of the pair could be dropped from further consideration, or examined more closely with the other variable selection methods. Tables 5, 6, and 7 show the pairwise correlations for candidate variables for each hypothesis (H1, H2, and H3), for the 16 km NW data. Correlations were also examined at 8, 32, and 64 km scales, and were generally found to be similar, so are not detailed here.

Table 5: Pairwise correlations for candidate H1 variables (16 km).

	NPPAV	NPPSD	NPPSUMAV	TDA
NPPAV	1	**0.75**	**0.99**	0.08
NPPSD		1	**0.72**	-0.36
NPPSUMAV			1	0.12
TDA				1

Table 6: Pairwise correlations for candidate H2 variables (16 km).

	TRA	TRS	TRW	TVNA	TVNS	TXD	TXSSR	TXSW	TFS	PTSR	PFSR	HVVAR	NPPSDR
TRA	1	**0.61**	**0.88**	-0.12	-0.27	0.18	-0.27	0.01	-0.41	0.10	-0.06	0.44	-0.29
TRS		1	0.23	0.18	0.08	0.49	0.20	0.07	-0.06	0.17	0.31	0.28	0.09
TRW			1	-0.08	-0.23	0.02	-0.45	0.06	-0.45	-0.02	-0.24	0.35	-0.26
TVNA				1	**0.90**	**0.63**	**0.54**	**0.78**	0.01	-0.30	0.05	0.28	**0.75**
TVNS					1	**0.50**	**0.56**	**0.66**	0.19	-0.29	0.04	0.16	**0.70**
TXD						1	0.44	**0.64**	-0.28	-0.29	0	0.39	**0.56**
TXSSR							1	0.45	0.07	0.02	0.17	0.32	0.43
TXSW								1	-0.25	-0.23	-0.13	0.40	0.55
TFS									1	0.12	0.22	-0.34	0.05
PTSR										1	0.47	0.24	-0.38
PFSR											1	0.13	-0.01
HVVAR												1	0.01
NPPSDR													1

Table 7: Pairwise correlations for candidate H3 variables (16 km).

	GLCV	GLCF	ELD
GLCV	1	**0.86**	0.34
GLCF		1	0.29
ELD			1

Based on these correlations, both NPPSD and NPPSUMAV were dropped from the H1 models. For H2, it is clear that TRA, TRS, and TRW are somewhat mutually redundant, and probably one or two of those variables could be removed. Also, TVNA and TVNS are highly correlated, and there are high correlations among TVNS, TXD, TXSSR, and TXSW, and NPPSDR. The choice of best variables to remove here is less obvious, although clearly some need to be dropped. For H3, GLCF and GLCV are highly correlated, indicating one of the variables could be removed from the analysis. For both the sets of H2 and H3 variables, because it is difficult to decide, simply based on pairwise correlations, which variables to remove, all variables were also examined with the other variable selection methods.

5.2.2.2. All subsets

All subsets regression was conducted for the CA preliminary studies, using the LEAPS (Lumley and Miller 2005) package in R (R Development Core Team 2005). The regressions were run to return the best three models of each possible model size, where all possible models were ranked by their adjusted r^2 performance. Model size varies from one to the total number of variables included. The LEAPS routine returns a list of which variables were included in the resulting top three models, for each model size. Additional code was written to summarize this output, by summing, over the returned information on the top three models of each size, the number of times each variable appears. Since the information returned from large models – models with most of the included variables – is not very informative, an upper limit was set to filter out large models. Table 8 shows the output for 16-km CA datasets, and indicates that the most important variables appear to be PEA, PFS, TFA, PTA, PFA, TXAS, HVS, and TVNS, although there is sometimes considerable variation between vertebrate classes. Equally important, certain variables are shown to be of little relevance, such as TDS, TVNA, TXSS, TXSW, TRW, and TRS. Results were similar for other spatial scales, although they vary somewhat for different states. These results allow one to easily remove several variables, but it is still difficult to select a small subset of variables. To help further, alternative methods were also investigated, such as hierarchical partitioning.

Table 8: All subsets regression results. Numbers are number of times a variable was included in the top 3 models (ranked by adjusted r^2) for each size of model. These results are from CA 16 km dataset.

	Bird	Mammals	Amphibians	Reptiles	Total
TDA	1	5		2	8
TDS					
PEA	11	8	1	9	29
PFA	9	2	1	1	13
PFS	4		7	13	24
PTA	8	4	2		14
HVA	1	1	5		7
HVS	3		1	8	12
TFA	1	9	2	3	15
TFS	1		8		9
TRA			5	2	7
TRS		1		2	3
TRW			1		1
TVNA	1	1			2
TVNS	5	7			12
TXAS		6	6	1	13
TXAW		1	4	4	9
TXSS					0
TXSW			2		2

5.2.2.3. Hierarchical partitioning

Hierarchical partitioning returns both the independent and joint contributions of each predictor variable to model fit. It does this over all combinations of variables, and so is similar, in terms of the models generated, to all-subsets regression. To ensure that the returned values indicate a significant contribution, it is preferred to compare the returned values against those from a randomized version of the dataset, and then test for a significant difference. The R package HIER.PART (Walsh and MacNally 2005) was used to compute the partitioning. Due to limitations of both the software and the time required to compute partitionings for many more than 7 or 8 variables, the candidate set of variables were divided into three groups, and then all six combinations of all pairs of groups were run. This ensured that each variable was included with each other variable, at least once. After the three combinations of groups were run, the variables with high independent contributions were identified, and rerun together in another partitioning. This attempted to both further distill the most important variables, as well as to ensure that any high values from the first set of runs were maintained when all the best performing variables were present. Due to processing time requirements, this method was only run for CA, but was run for each individual vertebrate class.

For birds, table 9 shows the output, with the eight best performers highlighted in bold. Those eight variables were then run in the final partitioning. Table 10 shows the final results for all four vertebrate classes. In general, it is easier to find variables to definitely drop using HP results than to find a definitive best set of variables to include. Based on the results presented in table 10, TXSSR, TVNS, TRA, PTSR, PFSR, and HVVAR should all be dropped. This leaves the eight variables TXD, TXSWR, TFSR, TVNA, TRS, TRW, HVSR, and NPPSDR, as ones to potentially include.

Table 9: Hierarchical partitioning results for birds at 16 km in CA. Values are z-scores of the independent contributions against a randomized dataset. Group runs are initial runs including all candidate variables. Groups: A=TXD, TXSSR, TXSWR, TFSR, TVNA; B=TVNS, TRA, TRS, TRW, PTSR; C=PFSR, HVSR, HVVAR, NPPSDR. Final run only includes best performers (in bold) from three group runs.

Variable	Group				Final
	AB	AC	BC	Average	
TXD	**43**	**34**		**38**	**21**
TXSSR	6	9		8	
TXSWR	**24**	**13**		**19**	**6**
TFSR	**13**	**31**		**22**	**11**
TVNA	10	15		13	
TVNS	6		3	5	
TRA	**28**		**17**	**23**	**9**
TRS	**43**		**37**	**40**	**22**
TRW	**54**		**31**	**43**	**21**
PTSR	3		5	4	
PFSR		2	3	3	
HVSR		**127**	**86**	**107**	**42**
HVVAR		7	6	7	
NPPSDR		**36**	**13**	**25**	**13**

Table 10: Final hierarchical partitioning results for CA 16 km. Numbers are z-scores from second stage partitioning; blanks indicate that variable performed too poorly in first stage partitioning to continue to second. Average includes a 0 value for blanks.

Variable	Birds	Mammals	Amphibians	Reptiles	Average
TXD	21	22	29		18
TXSSR					
TXSWR	6		23	13	11
TFSR	11		33		11
TVNA		22		30	13
TVNS		6			2
TRA	9	7		7	6
TRS	22	10	13	19	16
TRW	21	7	16	19	16
PTSR					
PFSR					
HVSR	42	13	33	24	28
HVVAR				9	2
NPPSDR	13	48		16	19

Although HP appears to be a powerful tool able to help isolate important variables from a larger superset, the results from the CA study led to their abandonment in this research. There were several reasons for this, including the processing limitations, as described, in the number of variables that could be run; the inability to run all variables simultaneously undermined the ability to determine if all independent, and especially joint, contributions had been fully evaluated. The large number of samples in the datasets under study here also aggravated the processing time issue. Running the method for each state in the analysis would be excessively time consuming. Further, the results would vary somewhat if a different goodness of fit measure was used. The above work used a R^2 goodness-of-fit measure. Using log-likelihood provided similar results, but different enough that decisions about including or excluding variables could differ slightly. Because of these issues, a third method was devised, using principal components.

5.2.2.4. Principal components

Both the all-subsets and hierarchical partitioning methods attempt to identify the most consistently important variables from all combinations of predictor variables in OLS regressions against the response variables. A fundamental issue that ensues in the present analysis is that the rankings of the variables is substantially (but not entirely) dependent on the specific response variable of interest. Thus results for birds, mammals, amphibians, and reptiles must all be considered, and the best overall variables selected. This methodology, along with some of the computational issues encountered, was found less than ideal for the purposes of this research, so another approach, using instead principal components (PC), was developed.

The PC approach was run for the combined data of the four NW states, at each spatial scale, for the 13 following variables: TXD, TXSSR, TXSW, TFS, TVNA, TVNS, TRA, TRS, TRW, PTSR, PFSR, HVVAR, and NPPSDR. Principal components were computed using the R STATS (R Development Core Team 2005) package, which uses a singular value decomposition of the data matrix to compute principal components. Of the resulting principal components, the first three accounted for 70%, and the first seven 91%, of the variance. These principal components were then correlated to the vertebrate response variables. Table 11 shows the resulting R^2 correlations for the first seven PCs. The remaining PCs only had negligible R^2 (less than 0.05) with the response variables, so are not included in the table. The table shows that the first two PCs contain almost all of the variance that can be correlated to vertebrate richness, although there was some correlation for Amphibians with the third PC. However, for all other spatial resolutions, there were no correlations greater than 0.10 for any other than the first and second PC.

Table 11: R^2 correlation of principal components to vertebrate richness. For NW states, 8 km resolution. Values greater than 0.10 in bold.

R^2	Birds	Mammals	Amphibians	Reptiles
PC1	**0.13**	**0.14**	0.05	**0.15**
PC2	**0.11**	**0.14**	**0.27**	**0.26**
PC3	0.1	0.07	**0.12**	0.03
PC4				0.04
PC5	0.06	0.1	0.02	0.04
PC6			0.03	
PC7	0.03	0.02		

With the variance of interest almost entirely isolated in the first two principal components, a biplot (Gower and Hand 1996) was then generated with those two PCs as the two axes. The plot, as shown in figure 2, displays the PC loadings as vectors on the plot, and the data points as dots. Note that the plot has no relation to the response variables, it simply displays graphically the distribution of the data and of the PC loadings, within the space of the first two principal components. If there were substantial correlations between the vertebrate richness and PCs other than the first two, a multidimensional plot with all relevant PCs would be required. Fortunately, that is not necessary here, as it would quickly become very difficult to visually interpret. Given the plot, it is only then necessary to select a set of variables, as displayed by the loading vectors, that best represent the data space. The criteria for doing so are simply that the selected vectors be as orthogonal to one another as possible (to reduce multicollinearity), and that they have sufficient magnitude (length of the displayed vector) to adequately encompass the range of variation in the data space. Based on these criteria, TFS, TRA, PTSR, and TVNA were chosen for the 8 km case. Note that other sets of variables are

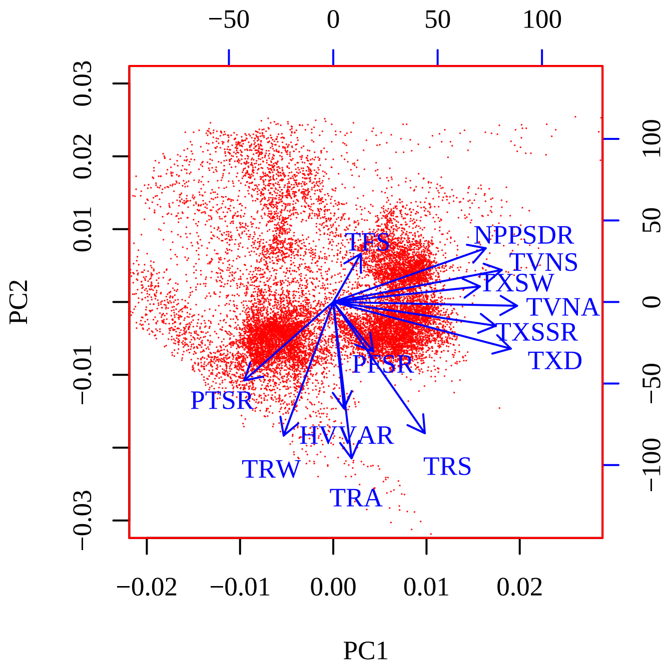


Figure 2: Biplot of PC loadings. Shows PC loadings (blue arrows and labels) within data space of first two principal components (red dots), for 8 km NW dataset.

possible, and would likely perform equally well, although alternate sets were not specifically evaluated. For the 16, 32, and 64 km biplots, the same set of four variables were chosen as good representatives of the data space. The same set of variables might not have been selected for those other spatial scales, if their biplots were examined solely on their own. However, as figure 2 shows, there are often several relatively equal choices; TXD might have been equally well selected instead of TVNA.

Based on these results from the principal components based selection methodology, the four variables TFS, TRA, PTSR, and TVNA were chosen as the four to use for the H2 hypothesis, for the final, eight-state analysis. Note that these four variables do not have high correlations among themselves (table 6). However, they also do not appear as likely choices from either the all-subsets or the hierarchical partitioning methods, which is somewhat troubling. Part of the reason for this may be that a different set of variables was used for the PC method, due to the inclusion of detrended variables. The initial work looking at all-subsets and HP for California did not include the exact same set of variables. Thus, interactions among variables may have contributed to different results here.

Despite these differences with the other methods, the PC results were selected because they produced a consistent set of relevant variables: the same four variables were easily selected from the biplots at the four spatial scales. The results for the other methods were more varied. The approach of the PC method was also considered superior because it circumvented the dependence on the response variable almost entirely; correlations to the response variables only identify the useful PCs, they do not affect the biplots or the selection of variables.

5.2.2.5. Selection for all, two-variable, and one-variable models

The added ‘dimension’ that generates three versions of each hypothesis model, with either all, two, or one predictor variables, was not initially considered in this research. However, as the results were being examined, it became obvious that the number of predictor variables was influencing model performance, particularly for the H2 hypothesis. Since adding any variable to a model can only improve model performance, it is likely to do so if the variable has any correlation to the response. Figure 3 shows the increase in R^2 with number of variables. Previous modeling iterations included five H2 variables—unlike the final results, which use only four—and thus the figure has a maximum of five predictor variables. Because the number of possible variables to include in the H1 and H3 models is already limited by the number available within these datasets, it should not be surprising that H2 models often outperform them, when several H2 variables are included in models. To prevent this from biasing the results, the dimension specifying number of predictor variables was added to the analysis. Note that the results in figure 3 were derived from only the four NW states; for the final eight-state results presented in section 5.3, H1 appears superior to H2 even in the all-variables case, because the four additional states all show a strong preference for H1 over H2.

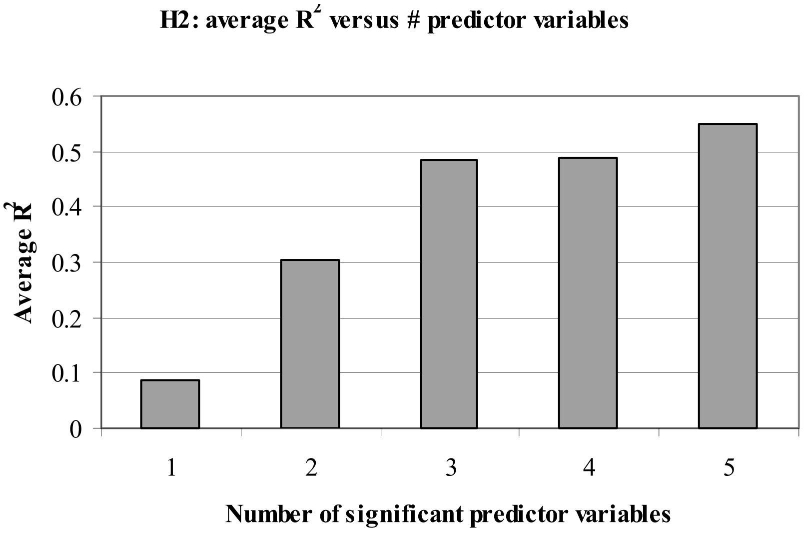


Figure 3: R^2 by number of significant predictor variables. OLS H2 models only, for NW at all spatial scales.

A final round of variable selection was then necessary to choose from the four selected H2 variables a subset of two for the two-variable models, and a single variable for the one-variable models. In this instance, all-subsets regression was used simply because it produced consistent sets of variables across all spatial scales and for all vertebrate classes; if it had not, the other methods would have been used to try to find consistent results. For this case, all subsets was run for all spatial scales, for each vertebrate class (and the combined non-volant, MAR class), and over the combined NW dataset. Figure 4 shows the results summed over all spatial scales. From this, TRA would be selected as the best individual variable, for the one-variable model, because it has the highest total value, indicating the most often appearance in all models, and because it has a good distribution between the different vertebrate classes. For the two-variable models, TVNA was also selected. Note, however, that for amphibians, it is not the best choice (TFS would be), but it otherwise seems to perform fairly well. PTSR is clearly the worst performing of the four variables, although for amphibians, it would be a better choice than the second ranked TVNA.

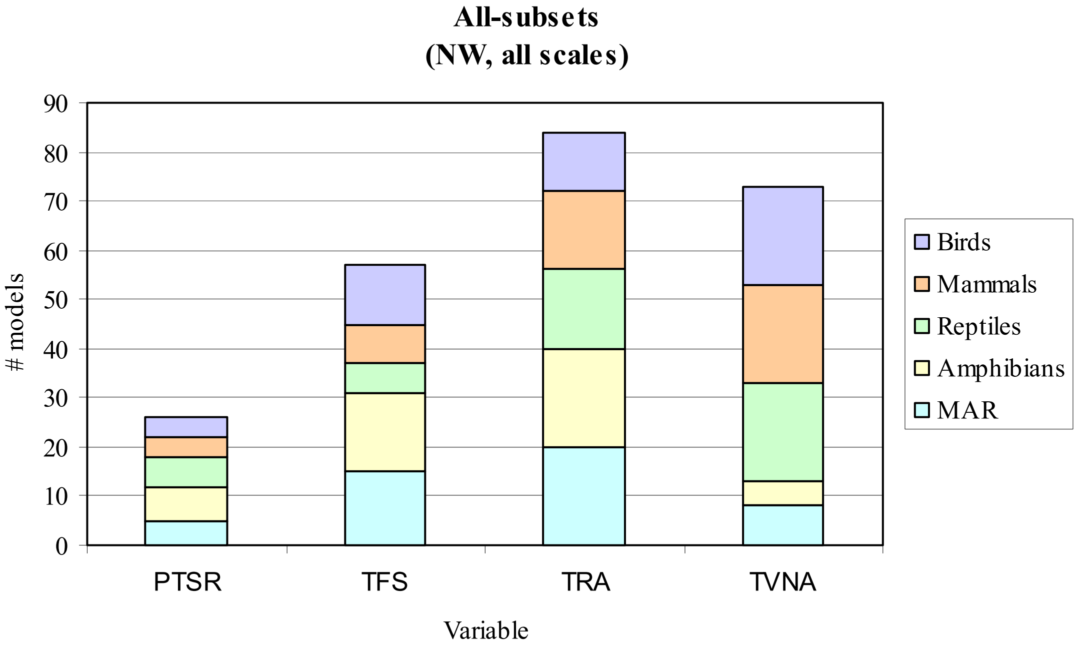


Figure 4: All-subsets regression results for selecting one and two-variable H2 model variables. Values indicate the number of times a given variable appears in all regression models, as one of the top three included variables, summed over all spatial scales, and run over the combined NW data.

5.2.2.6. Summary: variable selection

Based on the results from the above analyses, table 12 shows the variables used for the final analysis of the full eight-state dataset (section 5.3). Table 13 shows the correlation coefficients, among this final set of predictor variables, for the 16 km dataset. Note that among hypotheses, no correlations are greater than 0.34, except for that between GLCV and GLCF, which at 0.86 is substantial. GLCV and GLCF were both retained for the final analysis because it was unclear which to select. However, final two-variable models were not allowed to retain both GLCV and GLCF; one, based on high VIF or low significance, was always removed.

Table 12: Final set of variables selected for 1, 2, and all variable hypothesis tests.

# Variables	Hypothesis	Variables
1	H1a	NPPAV
	H1b	TDA
	H2	TRA
	H3	GLCV
2	H1	NPPAV, TDA
	H2	TRA, TVNA
	H3	GLCV, ELD
All	H1	NPPAV, TDA
	H2	TRA, TVNA, PTSR, TFS
	H3	GLCV, GLCF, ELD

Table 13: Correlations (r) among final set of predictor variables. 16 km dataset, NW.

	H1		**H2**				**H3**		
	NPPAV	**TDA**	**TRA**	**TVNA**	**PTSR**	**TFS**	**GLCV**	**GLCF**	**ELD**
NPPAV	1	0.08	-0.21	-0.21	0.27	0.18	0.38	0.39	0.09
TDA		1	0.32	-0.62	0.28	-0.40	0.01	-0.03	-0.30
TRA			1	-0.12	0.10	-0.41	-0.34	-0.38	-0.30
TVNA				1	-0.30	0.01	-0.27	-0.24	-0.04
PTSR					1	0.12	0.49	0.42	0.13
TFS						1	0.38	0.39	0.49
GLCV							1	0.86	0.34
GLCF								1	0.29
ELD									1

5.2.3. OLS model diagnostics

The quality of the California OLS models was extensively reviewed with several diagnostics, to better understand the limitations of using OLS with these data, which have known issues of spatial correlation and lack of normality. The diagnostics were used to check for: outliers, collinearity, normally distributed errors, constant error variance, and model non-linearity. To do so, several diagnostic plots and tests were generated for each model, printed to hardcopy, and reviewed manually for a large number of California models.

Figure 5 displays several diagnostic plots for the model testing the H2 hypothesis with all variables (TRA, TVNA, TFS, PTSR), for bird richness, in CA, at 32 km resolution.

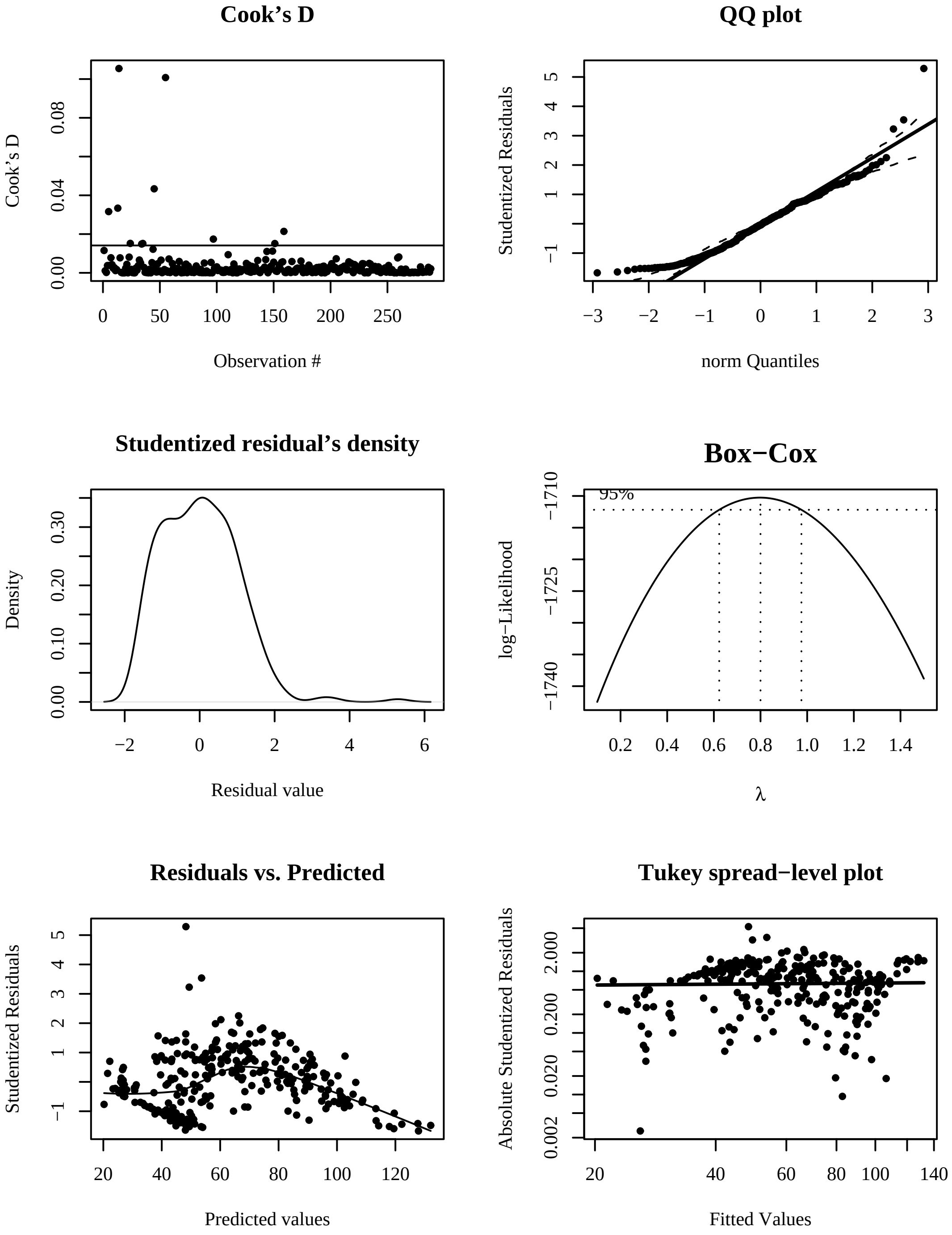


Figure 5: OLS diagnostic plots, for the H2 model fit to birds in CA at 32 km resolution.

The horizontal line in the Cook's D plot indicates the suggested threshold for significant values of D (Fox 2002); points above the line can be considered outliers. Although not displayed here, the points identified as outliers by Cook's D were also plotted geographically, to determine if their spatial layout. Typically, some of the outliers are randomly distributed across the state, and others are clustered. Often, in CA, clusters would occur near the central peaks of the Sierra Nevada, perhaps indicating a deficiency of the model, or of the datasets, in such terrain.

The QQ plot and the density plot display the statistical distribution of the model's studentized residuals. The QQ plot displays the distribution of actual residuals versus those for a normal distribution. Over-plotted on the graph is a 95% confidence interval computed by bootstrap; values outside the envelope are significantly non-normal. The QQ plot displayed in figure 5 is typical in this respect, as almost all residuals from all computed models had tails falling outside the normal distribution. The density plot shows the distribution of the studentized residuals after convolution with a kernel density estimation function (Fox 2002); this essentially provides a more detailed histogram of the model residuals, and is useful in conjunction with the QQ plot to diagnose non-normal residuals. In this case, although showing slight skew and not perfectly normal, there is no obvious issue with the normality of the residuals. The Box-Cox plot suggests a possible power transformation of the response variable in order to generate more normally distributed residuals. In this case, the 95% confidence interval (dashed lines on the plot) indicates that a power transformation of from 0.6 to 1.0 may help normalize the residuals. Since a power transformation of 1.0 would be the identity transform (no effect), in this case the Box-Cox plot does not definitively suggest a transformation is necessary. For the CA models, transforms were suggested in some cases, but there was no consistency in the types of models (e.g., a specific response variable, spatial scale, or predictor variables). In a few cases, the suggested transformations were applied, but this rarely appeared to have more than a minor effect. Thus, no transformations were applied to the response variable to attempt to rectify non-normal errors.

GLIM models, which allow specification of an alternate distribution for the errors, were investigated for their utility here in producing a model with properly distributed residuals. Theoretical considerations suggest that count variables follow a Poisson distribution. Species richness is a count variable, so using a Poisson error distribution in a GLIM model was investigated, but was not found to be very useful. Partly, the problem may be that the residuals for these models, although often not strictly normally distributed, are no better characterized by a Poisson distribution either, despite the theoretical support. Also, the distribution of species richness itself does not appear to particularly follow a Poisson distribution, and this may be because although it is a count variable, its relatively high magnitude distances it significantly from more typically Poisson distributed variables, which have relatively low total counts. Since the GLIM models with Poisson error distributions were not found to be very useful generally, it was decided to forgo their further evaluation.

Constant error variance was evaluated with plots of studentized residuals against the models' predicted richness. The 'Residuals vs Fitted' graph in figure 5 shows the basic plot, and is overlaid with a non-parametric smoothed line. Ideally, the variance will show no dependence on the predicted value, indicating constant error variance.

More typically, the variance increases with the value; however, in this case, the opposite may be occurring to some extent. Evaluating these plots becomes more difficult when the predicted values are not uniformly distributed, as a higher density of points in a certain range of values will inevitably display more variance. Thus, one expects a peak in the variance at the peak of the distribution of the predicted values, and this may be what is occurring in figure 5. To aid in evaluating nonconstant error variance, the Tukey spread-level plot is also graphed. This displays the log of the absolute value of the studentized residuals against the log of the predicted values, overplotted with a linear fit line. If the fit line has substantial slope, this indicates a nonconstant variance. The Tukey spread-level plot suggests a power transformation of the response to remove the slope; in this case, the suggested transformation is to the 0.95 power. This falls within the range suggested by the Box-Cox plot, but, being close to 1, does not suggest that the transformation will be very useful.

Model non-linearity was evaluated using partial-residual, or component + residual, plots. Figure 6 shows the plots for the full H2 model for bird richness in CA at 32 km resolution. Component + residual plots show the partial residuals for a given predictor plotted against that predictor variable (Fox 2002). Partial residuals are simply the sum of the residuals and the fitted component for the predictor variable of interest. The dotted line indicates the least squares fit for that predictor, and the solid line is simply a local regression smoother. A significant slope on these lines indicates there may be a non-linear relationship between that predictor variable and the response. The implementation of these plots in the CAR package (Fox 2002) in R allows the user to then investigate the effect of different transformations on the predictor variables. In this

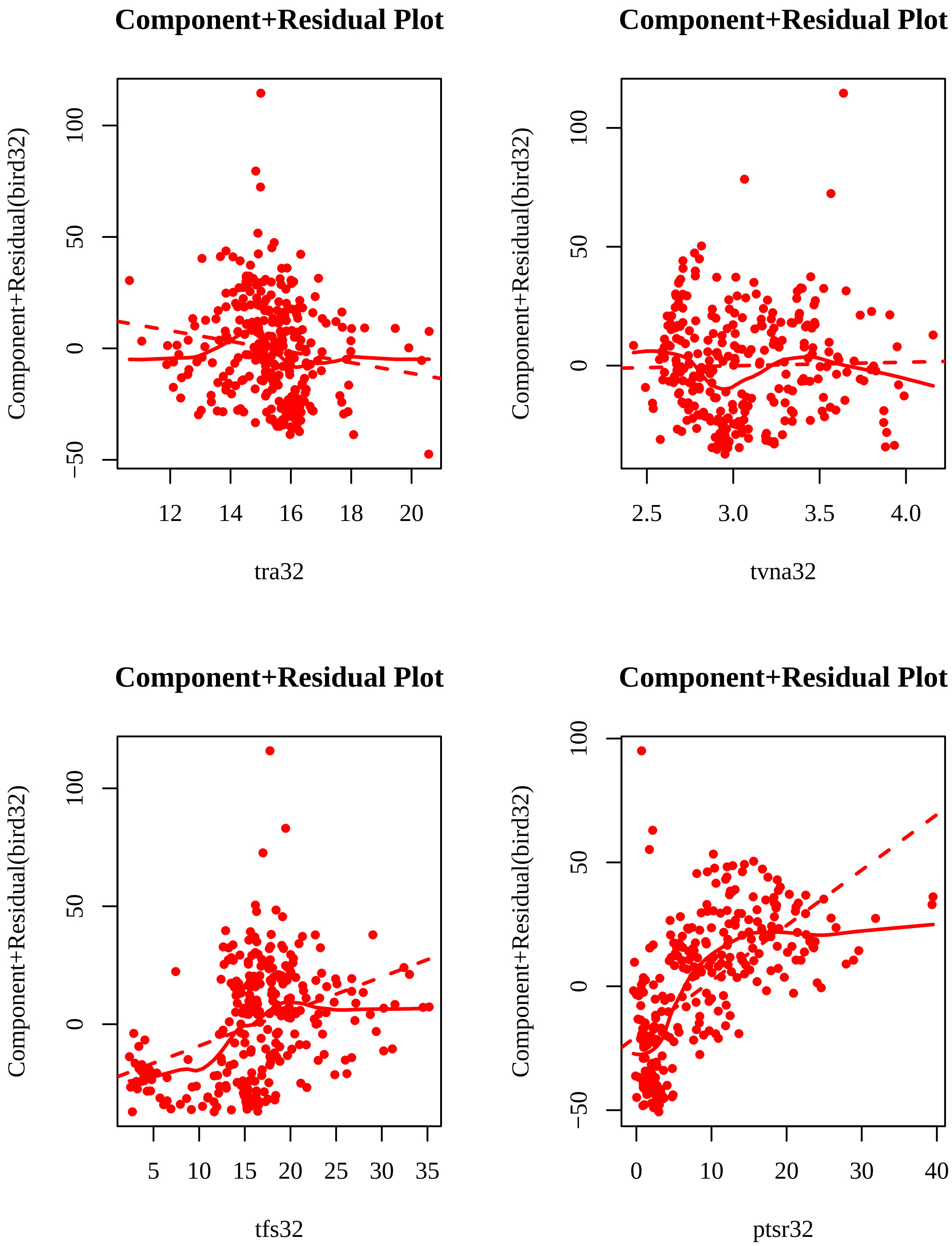


Figure 6: Component + residual plots, for the H2 model fit to birds in CA at 32 km resolution. PTSR and TFS exhibit some non-linearity. Dashed line is linear fit, and solid is local regression smoother.

case (figure 6), there appears to be possible non-linearity between bird richness and PTSR, and also possibly with TFS. This may explain why both of those variables were found less important than TRA and TVNA in the final variable selection process (section

5.2.2.5). However, as with the potential transformations of the response variable suggested by Box-Cox, it was found that transforming predictor variables was not very useful. Most importantly, different transformations were indicated for the different models and hypotheses, and there was little consistency in what transformations were needed for specific variables. Further, even when applied, the transformations did not appear to have a significant effect on the overall model, although they do indeed improve the component + residual plots.

The final diagnostic is a measure of the collinearity of the predictor variables, and this is provided by the variance-inflation factor (VIF). For the example model here, the VIFs are: 1.01 for TRA; 1.28 for TVNA; 1.82 for TFS; and 1.51 for PTSR. These values are all acceptable; ideally, VIFs should be less than 2 – 4, as the square root of the VIF indicates the inflation of the width of the confidence intervals for that predictor (Fox 2002). The VIF values for this model are not, however, simply coincidentally low. Most of the preliminary models constructed had significant problems with variance inflation, with values as high as 30, and frequently in the 4 – 10 range. Reducing the VIF of the remaining set of variables in each model was one of the key requirements pursued during model development with backward elimination.

The diagnostic measures reviewed here were used extensively for the preliminary California models. Although many of the diagnostics proved informative, relatively few regularly provided information that could actually be used to improve the analysis in a consistent manner. For example, as discussed, the various transformations of response and predictor variables suggested to correct non-normal errors, nonconstant error variance, or non-linear relationships between predictor and response variables, were not consistent, and so a given transformation could not be universally applied to any particular predictor or response variable. Certainly for a specific model, the suggested transformations might provide useful improvements. But lacking universal applicability, it was decided to forgo any such transformations for the full analysis. The use of variance inflation factors was, however, a very important tool, and was used throughout the analysis.

5.2.4. CAR models

5.2.4.1. Specification

Specification of the neighborhood distance for CAR models was done by first examining variograms of the OLS model residuals. The variograms indicate a range out to which there is still apparent autocorrelation in the data. Both standard r^2-based variograms and Moran’s I-based variograms were examined, to determine the necessity for the much more computationally complex Moran’s I version. Figure 7 shows the r^2 variograms for H1a, H1b, H2, and H3 models of bird richness in CA at 16 km resolution, and figure 8 shows the Moran’s I variogram for just the H2 model. Moran’s I values range from -1 for negative spatial correlation to +1 for positive spatial correlation, with 0 indicating no significant correlation. Both figures indicate model residuals retain significant autocorrelation out to approximately 125 – 150 km. Similar variograms, both r^2 and Moran’s I, were examined for many different models in California, and clearly indicated that a standard r^2 variogram provided information sufficiently equivalent to that provided by the Moran’s I variograms, in terms of estimating an autocorrelation distance. Thus, only r^2 variograms were used for determination of the appropriate spatial neighborhood.

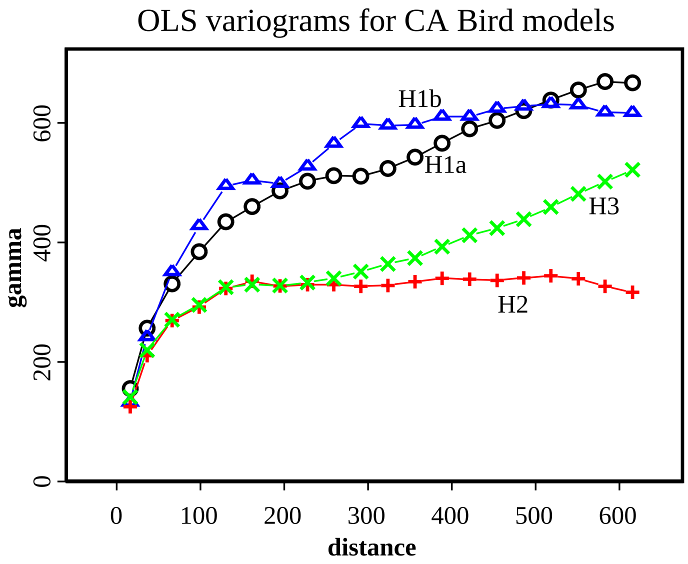


Figure 7: r^2 variograms for residuals from CA bird models at 16 km. Distance in kilometers. Symbols indicate different hypotheses (H1a = ○, H1b = Δ, H2 = +, H3 = ×)

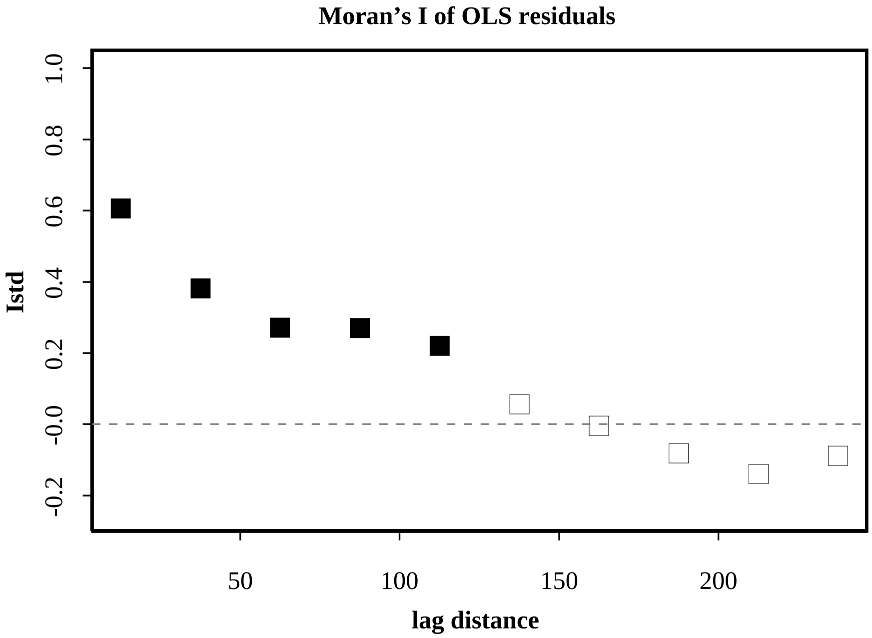


Figure 8: Moran's I variogram, of H2 bird model residuals, in CA at 16 km resolution. Filled squares indicate I is significant, based on 100 randomized permutations; empty squares indicate values are not signficantly different from zero. Lag distance in kilometers.

The r^2 variograms of OLS models clearly indicated an autocorrelation distance dependent on the spatial scale. It did not appear consistently dependent on either the specific state or the vertebrate richness response variable. Based on these variograms, the neighborhood distance for CAR models was chosen to be 40 km for 8 km scale, 160 km for both 16 and 32 km scale, and 256 km for the 64 km scale. These specific numbers were chosen to be multiples of the spatial resolution, so that a given number of pixels could be identified as the spatial neighborhood at each spatial scale: for the 8 km scale, this is a 5 pixel neighborhood; for 16 km, 10 pixels; for 32 km, 5 pixels; and for 64 km, 4 pixels.

Specification of the spatial dependence function was done by trial and error; the three obvious potential candidates of *1* (uniform), $1/d$, and $1/d^2$, where d is the separation between points, were all evaluated for the NW models, and the simplest function that consistently produced good results selected. A "good result" was indicated if the resulting regressions usefully reduced the spatial autocorrelation of the model residuals, as measured by a variogram of residuals, and did so as effectively as the other choices. Typically, there was not a substantial difference between performance in this regard, although the $1/d$ and $1/d^2$ dependences did produce models with slightly less autocorrelation. However, the $1/d^2$ also generated computation problems in some cases, preventing the iterative CAR modeling procedure from converging. Because of these issues, $1/d$ was chosen as the standard spatial dependence to use for all CAR models.

5.2.4.2. Diagnostics

Two diagnostics were used to assess CAR model performance: variograms of model residuals, and likelihood ratio tests of the estimated spatial correlation parameter, ρ (see section 4.2.3.3). Based on the close equivalence of the traditional r^2 based variograms and Moran's I variograms, only r^2 variograms were used extensively as a diagnostic. In all cases, the CAR models significantly reduced the spatial autocorrelation of model residuals. Figure 9 shows representative graphs of the variograms of OLS and CAR model residuals. When the residuals are mapped, OLS model residuals clearly mimic features of the input datasets, while the CAR residuals show very little recognizable spatial pattern.

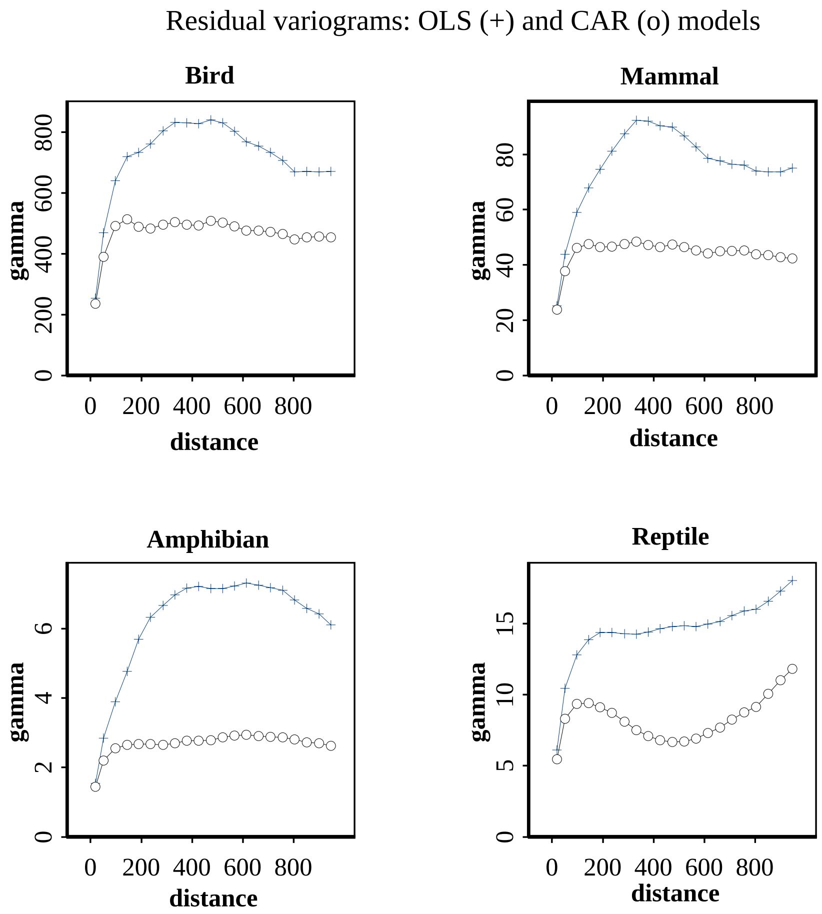


Figure 9: Variograms of OLS and CAR model residuals, for H2 NW models at 16 km resolution. + indicate OLS residuals, and ○ indicate CAR residuals. Distance in km.

A likelihood ratio test was also performed on the spatial autocorrelation parameter, ρ, which is estimated by the CAR model. A non-significant value of ρ would indicate no significant spatial autocorrelation was being modeled by the CAR model. In all cases, except those few cases in which the CAR model did not converge, ρ was found to be significant.

5.2.4.3. Utility for variable removal

CAR models were investigated for their ability to account for the spatial autocorrelation of the input datasets. Ignoring the autocorrelation effectively increases the number of observations, since each observation is not independent, which leads to increased statistical significance of the estimated model coefficients. In some cases, this can lead to the retention of a variable that would otherwise be removed, as non-significant. This does not significantly affect the model R^2, but can affect the estimated values and significance of model coefficients for variables with any degree of collinearity.

For NW models, CAR models were run on the final OLS models, and did not, over all models, identify many cases with non-significant variables; such redundant variables were found in approximately 10% of H2 models, significantly fewer H3 models, and no H1a or H1b models. Based on these results, it was decided that checking with CAR models was not necessary, particularly in light of the large number of models being computed, and the fact that removing the identified variables would affect only model coefficients, and not significantly model R^2 or overall performance.

5.2.5. Landcover

In the preliminary California models, it became apparent that there was a strong association between general landcover types and the individual observations, so the effect of landcover on the models was investigated. Residuals of many of the models showed suggestive clusters of points. Upon closer examination, these clusters corresponded to broad landcover types. The 1 km UMD landcover (Hansen, DeFries et al. 2000) was aggregated from 13 classes to 3 classes: a modified class, consisting of the ‘urban’ and ‘cropland’ classes; an arid class, consisting of the ‘bare’, and ‘open shrubland’ classes; and a ‘natural’ (non-arid) class, consisting of all other classes. Figure 10 shows the distribution of the aggregated landcover classes in California.

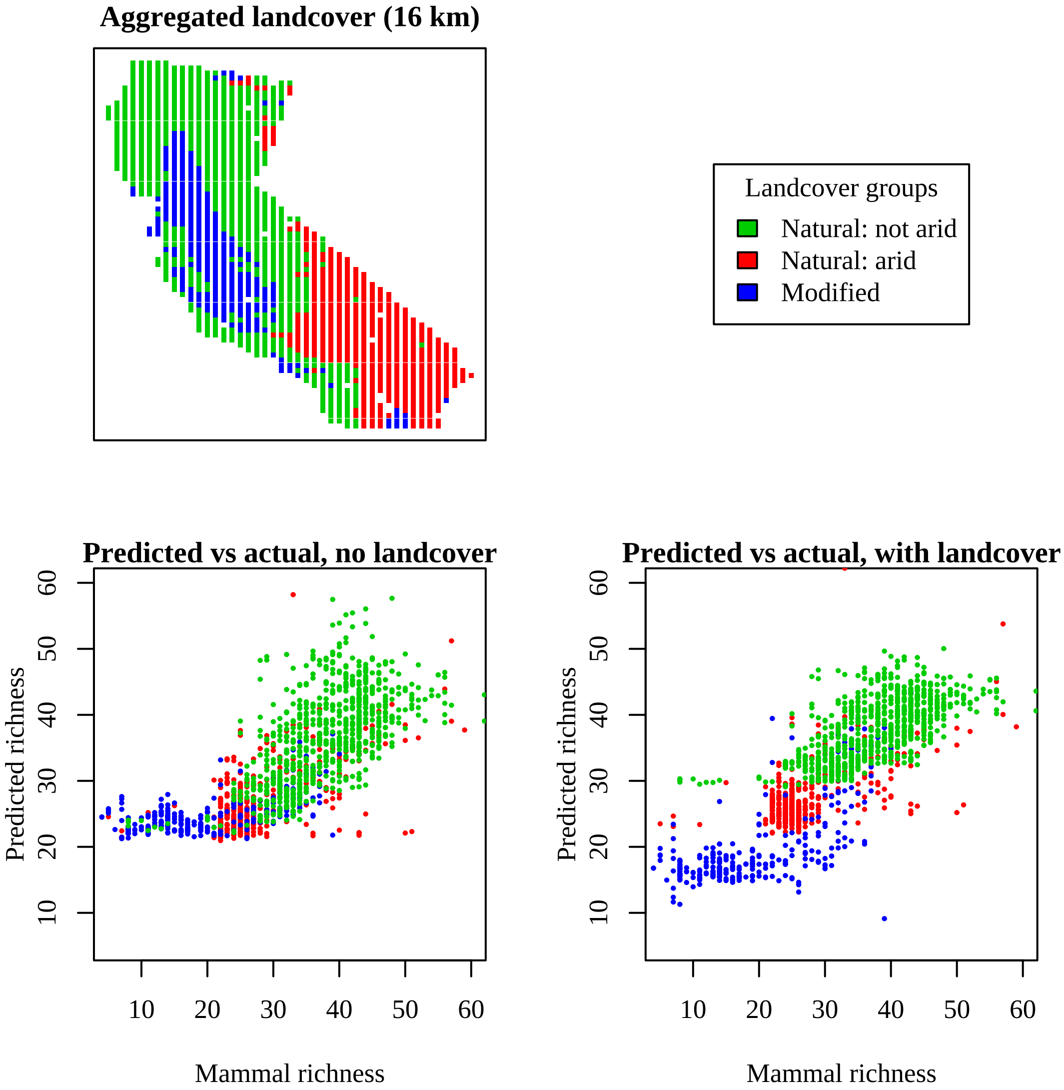


Figure 10: Landcover dependency. Upper panel shows the three landcover classes aggregated from the UMD 1 km global product. Lower graphs show the predicted versus actual richnesses, from a climate model for mammals over California at 16 km, with points colored according to their aggregated landcover class. The “with landcover” model included the three-level aggregated landcover class as a factor in the model; the “no landcover” model did not.

The lower panels of figure 10 show the predicted richness values against the actual, for a climate model of mammal richness in California, at 16 km scale. One model (left pane) contains only the climate variables, while the other contains a the three-level landcover variable as a factor. As the figure clearly shows, the ‘modified’ landcover (agriculture and urban), creates a non-linear feature in the model performance; the range of actual mammal richness in the “modified” landcovers is not well predicted by the model, particularly at the lowest levels of richness. This non-linearity persists in other diagnostic plots, and calls into question the suitability of the model. Fortunately, the observations causing the problem are almost entirely located in the modified landcover, as the colors in the figure indicate. A closer investigation of the location of the modified landcover points, in the model without the landcover variable, showed that the points with actual mammal richness less than 20 were located almost entirely within the central regions of either the Central valley, the Imperial valley, or the Los Angeles urban area.

Adding the three-level landcover variable as a factor in the model allows different offsets and slopes to be fit for each landcover, correcting, as the right pane shows, almost entirely the non-linearity. The resulting model is better behaved and its R^2 increased by 0.1 from 0.56 to 0.67. Tellingly, but perhaps unsurprising, the coefficients for two of the included three climate variables have the opposite sign for the ‘modified’ landcover as they do for both natural landcovers; clearly the process operating within the ‘modified’ landcovers is substantially different than that operating elsewhere.

Although the relationships within these landcovers might be quantifiable, it is difficult to know what such relationships represent, since the landscapes have been so fundamentally altered by humans. The environmental datasets represent actual physical conditions, which in these highly modified areas are quite opposite the conditions that would naturally prevail. Similarly, the GAP data may or may not provide an accurate representation of species richness in these modified landscapes. The richness derived from the GAP datasets may be strongly influenced by small remaining natural areas within larger agricultural mosaics. Perhaps most importantly, it is unknown how richness is expected to respond to environment in such heavily modified environments. The answer may be interesting, but as it does not directly support the research goals of this study – to evaluate hypothesized *natural* mechanisms driving species richness – it was decided to simply remove areas with highly modified landcovers from the analysis entirely. For the case highlighted in figure 10, this will allow a robust model to be generated without the need for inclusion of a landcover variable. Since the goal is to evaluate general hypotheses, which do not suggest limits based on landcover, this was decided more appropriate than the alternative of retaining landcover variables in order to maintain sufficient linearity of the models.

5.3. Full results: Western US

The results for the full, eight-state study, conducted with the methodological choices derived from the preliminary models (5.2), are presented below. First, the primary question of how these results support the different hypotheses is reviewed, followed by a look at model performance over the various analysis dimensions.

5.3.1. Hypothesis support

How well the data support the different hypotheses is evaluated by examining four evaluative diagnostics: AIC, R^2, coefficient values, and likelihood ratio comparisons of coefficient significance.

5.3.1.1. AIC

AIC values are directly comparable among models that only differ in the predictor variables; other model parameters, including the number of observations and the response variable, must be kept constant. Here, then, AIC can only be compared between models of a given response variable (a specific vertebrate group), at a given spatial scale, and over a given geographic region. Within those constraints, AIC values can then be compared among models with different sets of predictor variables, allowing a direct evaluation of hypothesis performance. To effectively evaluate overall hypothesis performance with AIC, the best performing hypotheses (lowest AIC) from all such groups (constant scale, vertebrate group, and geographic coverage) were tallied, and these tallies were summed over all dimensions of the analysis (varying scale, vertebrate group, and state). Figure 11 shows the results summed over geographic region, and figure 12 shows the sums over vertebrate class. Both figures show the three cases of all-variables, two-variables, and one-variable models.

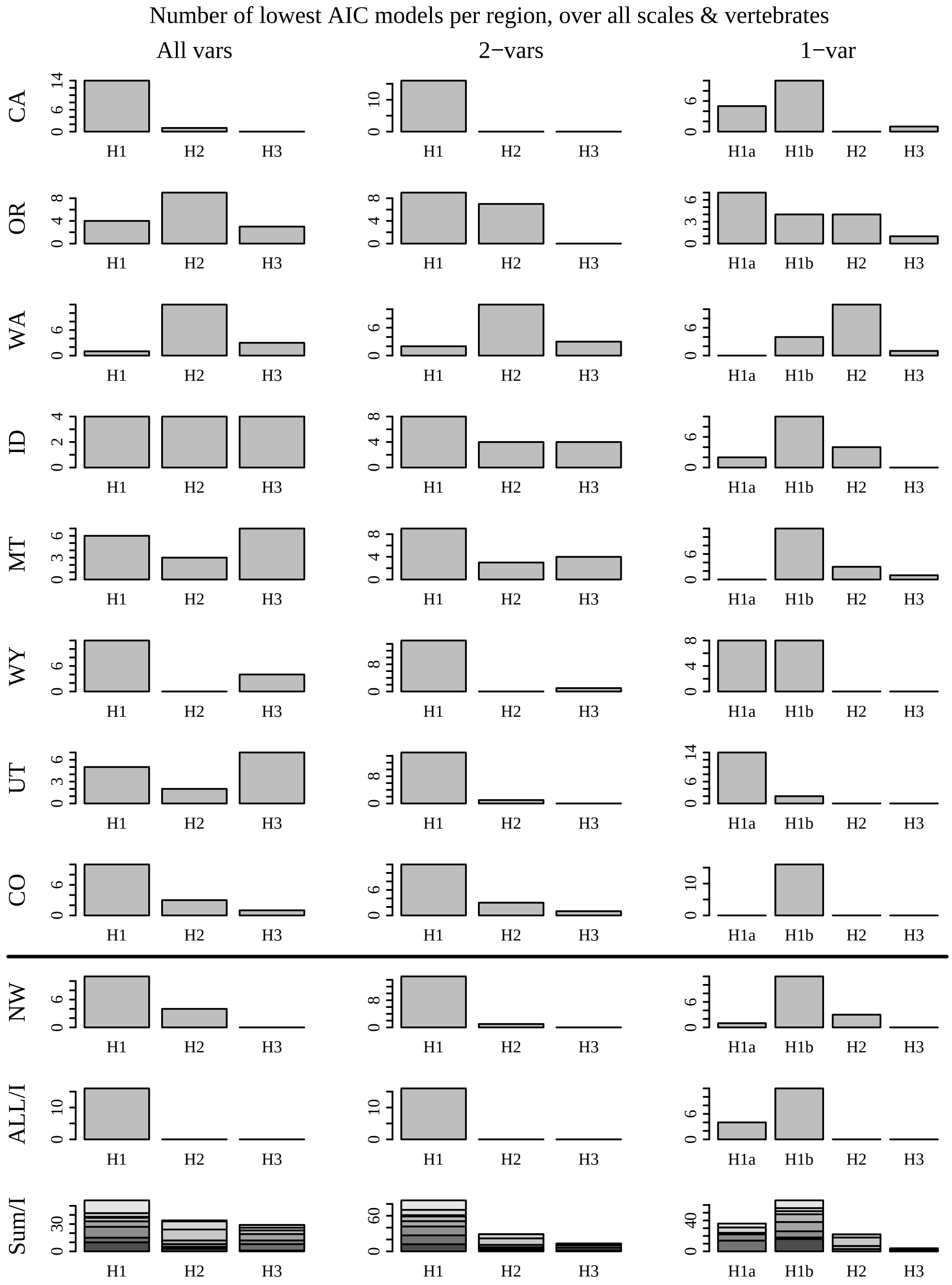


Figure 11: AIC summed over geographic region. The columns present the all-variable, two-variable, and one-variable cases. Above the horizontal line, the rows show results for each individual state, while below are summary results. NW indicates results over the four NW states. ALL/I indicates results for the four individual vertebrate groups over the geographic combination of all states. Sum/I simply stacks the sums from the eight individual states (all rows above horizontal line). Note that the results depicted here are for all species, and are therefore heavily dominated by birds (except for ALL/I, which equally weights the four classes).

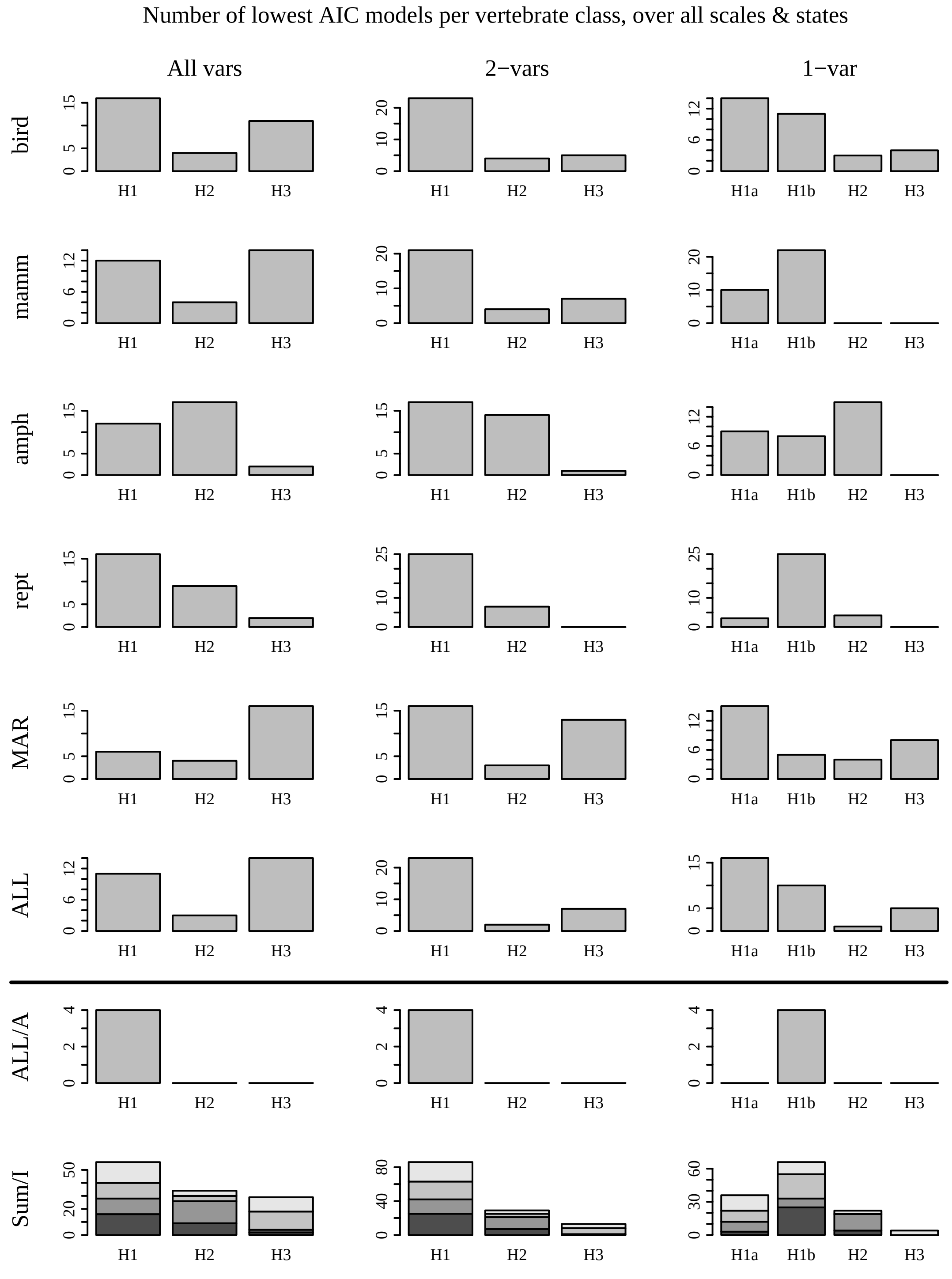


Figure 12: AIC summed over vertebrate group. The columns present the all-variable, two-variable, and one-variable cases. Above the horizontal line, the rows show results for each group of vertebrates, including the aggregations MAR and ALL. Below the line are summary results. ALL/A shows results for the ALL vertebrates group over the geographic combination of all states. Sum/I simply stacks the sums from the four individual states (first four rows). (The rows for individual vertebrate classes and MAR are summed only over all individual states, excluding the 'all' aggregated area).

These plots clearly indicate that, overall, the H1 energy/productivity hypotheses appear to be best supported by the data, for either models with all included variables, with two variables, or with a single variable. In most cases, the H2 models perform second best, and the H3 perform the worst, although there are significant differences between geographic regions and between vertebrate groups.

Between geographic regions (figure 11), the H1 hypothesis performs best for all three modeling cases (all, two-variable, and one-variable), for most regions. For the states of Oregon, Washington, Montana, and Utah, H2 (OR, WA) or H3 (MT, UT) perform better in the all-variables case. For the two-variables case, only in Washington does a hypothesis other than H1 perform best—in that case, H2, again, is better. For the one-variable case, H1 is always superior. Note that over large areas (all states (ALL/I), and the NW states), H1's advantage over H2 and H3 is even more pronounced. This suggests that H1's advantages in explanatory power increase with increasing geographic area. Within H1, the one-variable models indicate that climatic energy (H1b) is generally a better predictor than productivity (H1a), although this varies somewhat by region; H1a performs best in Oregon and Utah.

The results summed by vertebrate grouping (figure 12) again support H1 overall, but also show differences. Of the single class groups (amphibians, birds, mammals, and reptiles), only birds and reptiles show H1 to consistently outperform the other hypotheses for all three modeling cases (all variables, two-variables, and one-variable). For mammals, H3 outperforms H1 by a small margin, but only in the all-variable case. For amphibians, H2 performs best except in the two-variable case, in which H1 performs marginally better. H2 only performs reasonably well for amphibians, and for birds and mammals is the weakest performer. Although it rarely outperforms H1, H3 performs better than H2 for birds, mammals, MAR, and all vertebrates; conversely, it is always the worst performer for amphibians and reptiles. The sum of the individual class groups (the row “Sum/I”) shows H1 to clearly perform best, followed by H2 and H3, respectively. For the aggregation of non-volant vertebrates (MAR=mammals, amphibians, and reptiles), H3 actually performs best for the all variable case, and it performs well for the two-variable and one-variable cases – significantly better than H2, although not better than H1. Rather surprisingly, in the one-variable case, H3 performs reasonably well – second best after H1a – while it was never the best performer for each of its constituent groups (birds, mammals, reptiles, or amphibians), examined singly. For the all-vertebrates case, the results are, not surprisingly, similar to those for birds, although not identical. Examining the differences between H1a and H1b, H1a does marginally better for birds and amphibians, while H1b performs significantly better for mammals and reptiles. Interestingly, the MAR grouping is dominated by H1a, even though H1a outperforms H1b just for amphibians, and only nominally so. Also, for all vertebrates combined, tallied from each individual state (ALL/I), H1a performs best (dominated by birds), while the performance over all states combined into one region (e.g. one model versus eight: the ALL/A row in the figure) indicates H1b to be a significantly stronger performer.

When the results are presented for the group of all vertebrates, as in figure 11, they are, of course, heavily dominated by birds, since birds can constitute up to 70% of the total vertebrate richness of an area. Mammals also have some influence in “all” richness, and may typically account for 20% of overall richness; amphibians and reptiles are certainly not represented in these summary plots. Thus, more details about the support of different vertebrate groups for the hypotheses can be derived from examining the sums over geographic region (figure 11), but for each vertebrate separately. This essentially splits out the results of figure 12 by region. Doing so highlights a few anomalies; notable differences for the two-variable case will be mentioned here. Mammals in Colorado, it turns out, support the H2 hypothesis much better than H1. Birds, despite making up the bulk of the "all" category, also show some differences, with the Washington data supporting H2, and the Montana data supporting H3, instead of the H1 seen for all vertebrates. For amphibians, H2 is better supported for OR, WA, ID, and MT, and for reptiles, H2 is best supported for WA. The MAR aggregation of mammals, amphibians, and reptiles shows even more differences, with CA, WA, and ID data supporting H3, and OR supporting H2. Thus although the overall results support H1, there is a minor amount of variation around that overall median, and with increasing support apparent for H3 relative to H2.

5.3.1.2. R^2

Comparing model performance by R^2 allows a direct comparison across all of the modeled dimensions in terms of model predictive ability; the AIC comparisons are convenient for allowing a statistical test to determine the best model, but they do not indicate how much better a given model performs, in terms of its predictive ability. Figure 13 shows the mean R^2 averaged over all scales and all states. With a few exceptions, the same overall pattern, in terms of best performing hypothesis, is observed here as was seen for the AIC results (figure 12), with H1 performing best, followed by H2 and H3. For birds in the one-variable case, H1b provides better predictive ability

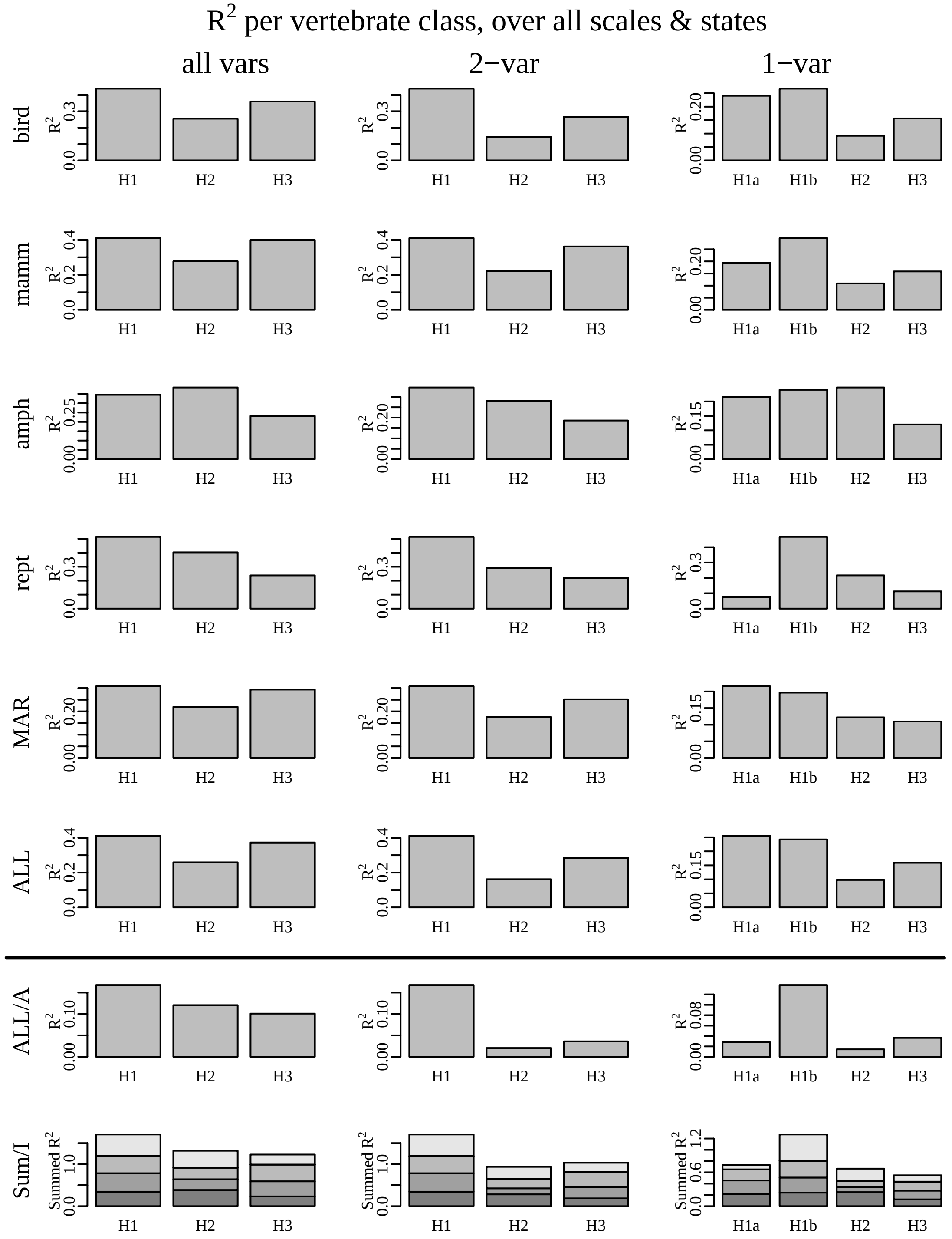


Figure 13: R^2 results, averaged over all states and spatial scales. See figure 12 caption for full explanation of rows and columns.

(R^2) even though models with H1a were found to be the 'best' model more frequently, in terms of having the lowest AIC value. This could potentially result from H1a more consistently producing a good model, while H1b might be less consistent, but produce a higher R^2 when it is significant. For the MAR group of vertebrates, the R^2 results consistently show H1 to be the best hypothesis, although by a small margin in the all-variables and two-variables cases, for which H3 is a very close second. For AIC, H3 was by far the best supported hypothesis for the all-variable case.

Figure 14 shows the R^2 results per geographic region, averaged over all four single vertebrate groups (and therefore dominated by birds) and over all spatial scales. Again, these results are similar to those for AIC (figure 11). H2 is consistently shown to provide the best predictive models for Washington state, while all other one and two-variable cases support H1 as the best hypothesis, except for the Oregon one-variable case. In that case, H2 provides best predictive ability, although it is very similar to that of H1a. The variation in hypothesis support by region, for each individual vertebrate class (as opposed to the results for all vertebrates, as just reviewed), mirrors that discussed for the AIC results. A significant difference, however, is clear in the increased power of the H3 hypothesis, relative to H2. Although either is second-best (after H1) for roughly half of the individual state or individual vertebrate class groupings (those rows above the horizontal line in the figures), in the R^2 evaluation, H3 outperforms H2 for the two-variable case, and performs nearly as well as H2 in the all-variable and one-variable cases. For AIC, H2's advantage was more decisive. The important point is simply that there is no clear worst performer among the three hypotheses, only a clear best performer.

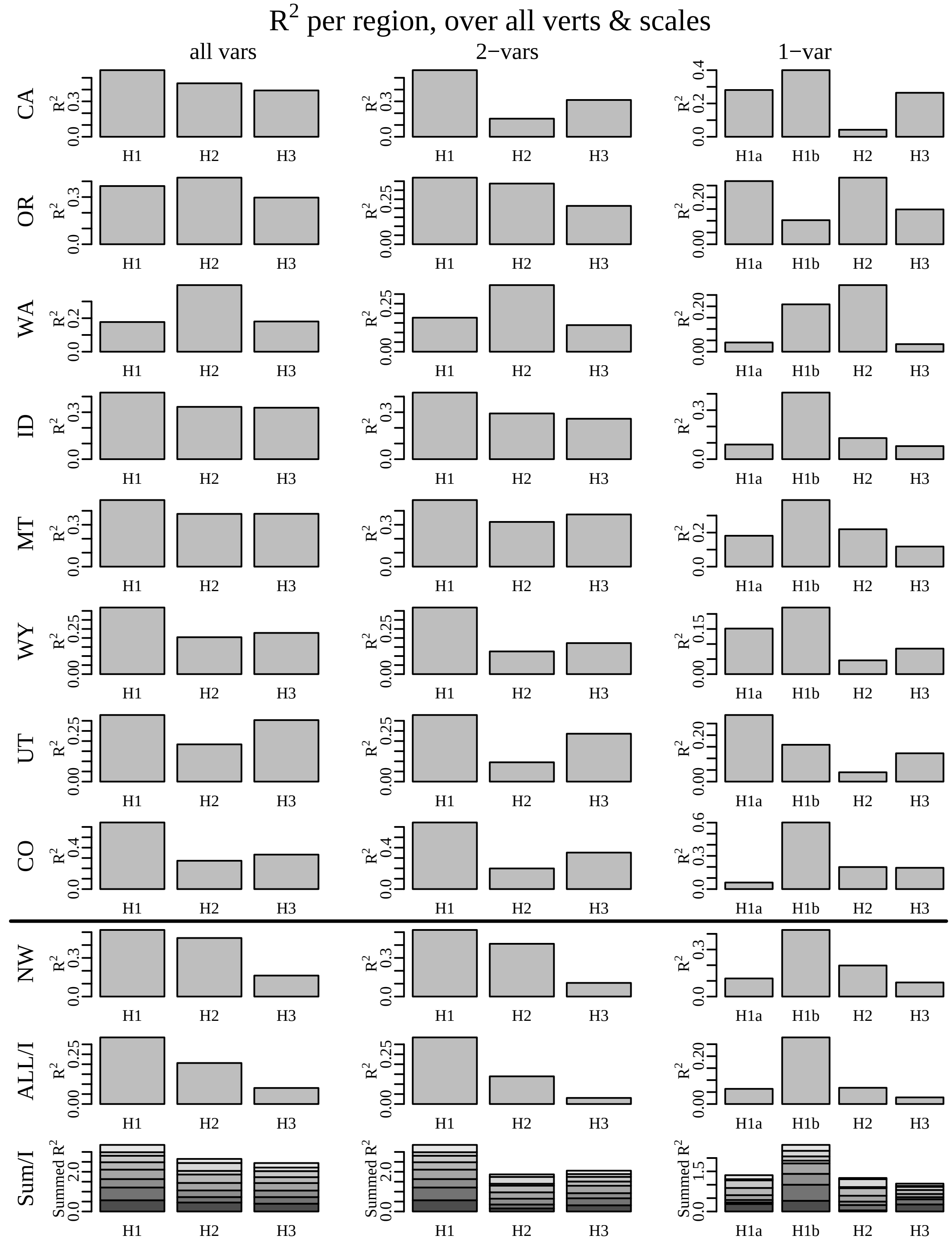


Figure 14: R^2 results, averaged over all vertebrate classes and spatial scales. See figure 11 caption for full explanation of rows and columns.

Overall, the R^2 results have less power for discriminating between hypotheses, since they are averaging R^2 for all models where the model had overall significance. However, they are useful for indicating that some hypotheses perform nearly as well as the best performing hypothesis, and thereby show that the overall differences among hypotheses, in terms of actual predictive performance, are not as different as the AIC results might imply.

5.3.1.3. Coefficients

The coefficients returned from the models provide additional insight into which hypotheses appear better supported by the data. Consistency of the coefficient values can help identify consistently performing hypotheses, and the signs of the coefficients can help indicate whether the variable's effect is as expected, given the proposed mechanism. To be able to assess coefficients more effectively across models of the different vertebrate classes—since they have widely varying richness levels—the results presented here are from models generated with normalized datasets. Using normalized datasets does not affect the R^2 or AIC values, but does allow one to compare the relative strength of each predictor variable against one another, and to compare their effect on different vertebrate groups without bias.

To assess coefficient consistency, four graphs were constructed for each variable, summarizing the coefficient values over the various dimensions, and these graphs were visually evaluated. Figure 15 shows the overall summary plot for NPPAV, the variable used in the H1a hypothesis. In this case, the coefficient is almost always positive, except for the full H1 hypothesis for reptiles. Other than that, the values are very similar between the H1 and the H1a models within each vertebrate group. A second plot breaks this out by scale; as figure 16 shows, there is some variance of the coefficients between spatial scales, with the value tending to increase with coarsening scale. A third graph shows the variation of the coefficient by geographic region (figure 17). Here more variation is apparent, although the values are fairly stable for most of the states, except Colorado, Washington, and to a lesser extent California, which exhibit substantial variation of the coefficient between different vertebrate groups. As expected from the previous figures, the coefficient for reptile models are anomalously negative, for several, but not all, regions.

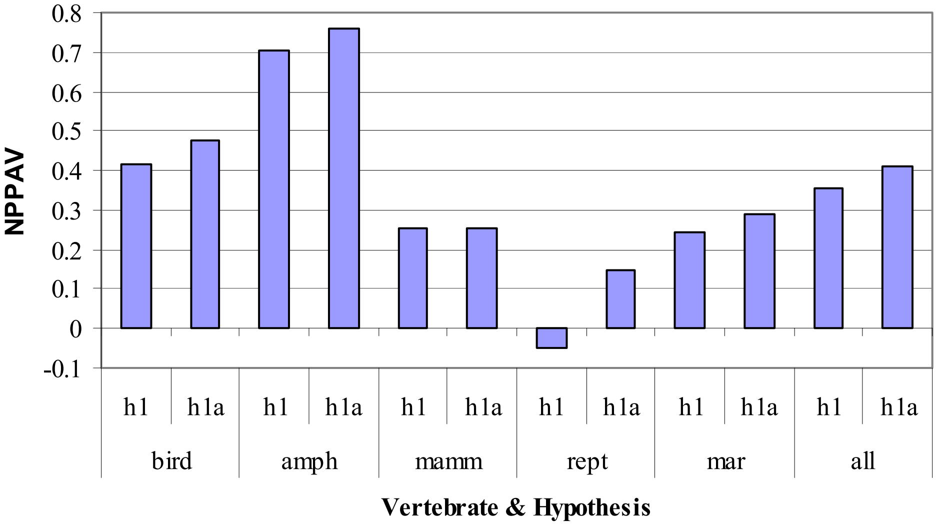


Figure 15: NPPAV coefficient (slope) summarized over vertebrate grouping.

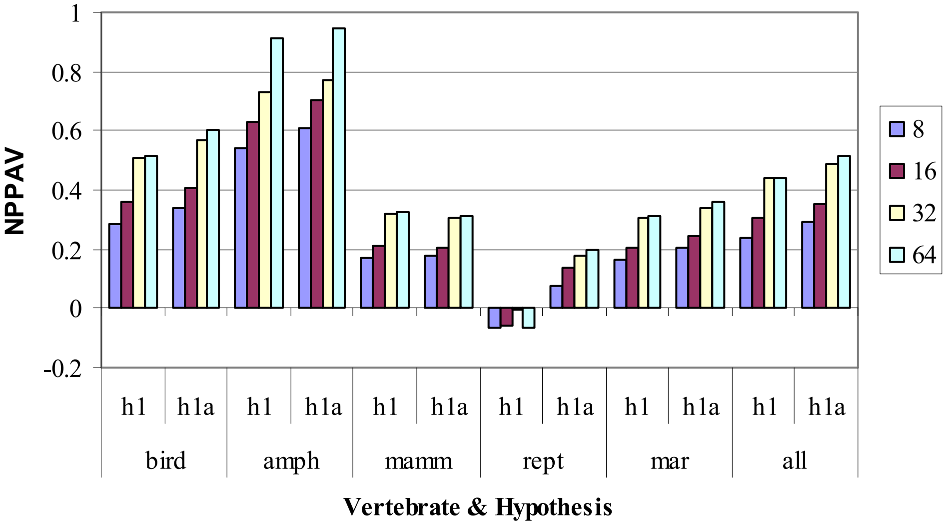


Figure 16: NPPAV slope summarized over vertebrate group and spatial scale.

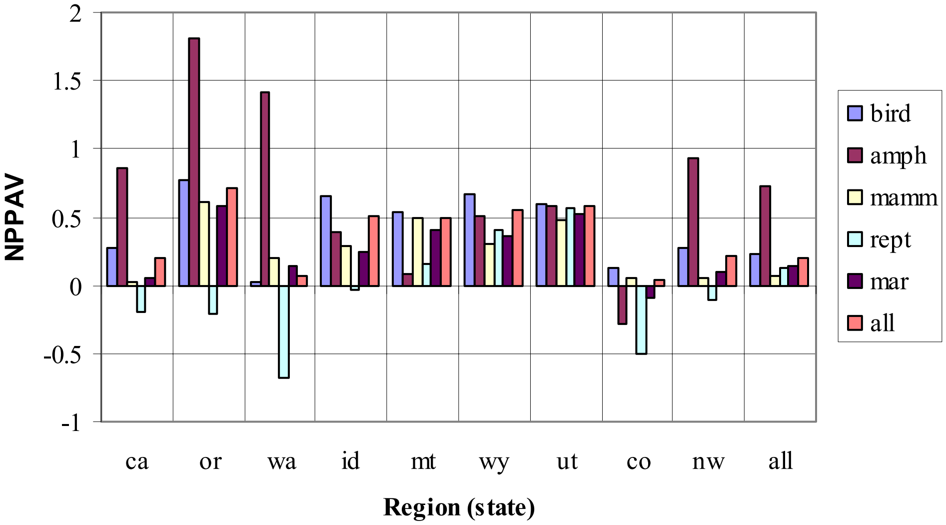


Figure 17: NPPAV slope summarized over region (state, or group of states).

Figure 18 shows the final of the four graphs produced per variable, which as a adaptation of the previous graph displays the variation of the coefficient within vertebrate groups, by region. Here the coefficients for the reptile models are seen to vary significantly between different states. Coefficients for the other vertebrate models also show substantial variance, but less than those for reptiles, and generally maintain the same sign.

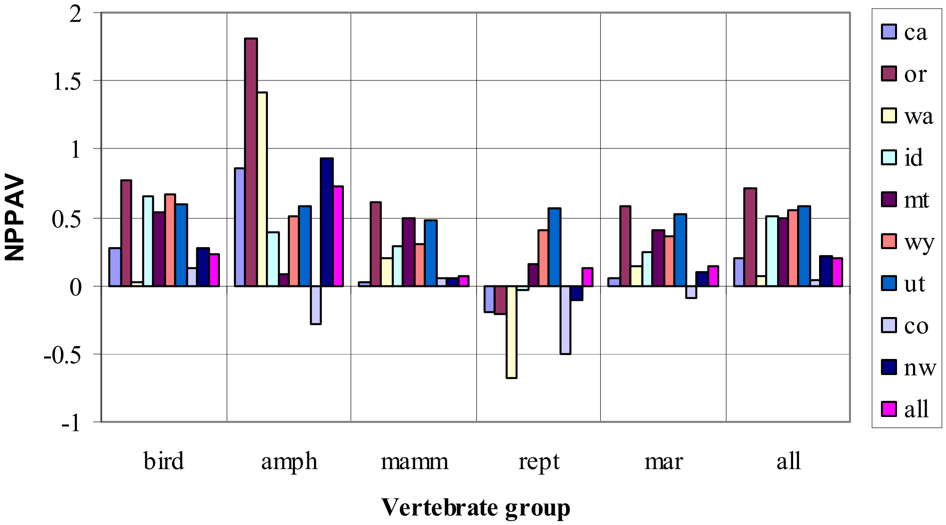


Figure 18: NPPAV slope summarized over vertebrate group, by state.

In examining this series of graphs for all variables, two features become evident. First, in reptile models, the coefficients are often of a different sign than those for the other vertebrate groups. For TDA (H1b) and TVNA (H2), both reptiles and amphibians have the opposite sign from birds and mammals. Figure 19 illustrates this for TDA coefficients. For the H2 variables TRA, TFS, and PTSR, and the H3 variables GLCV, GLCF, and ELD, the coefficient signs for reptiles are opposite those of all other vertebrate classes. Figure 20 shows this for the GLCV coefficients.

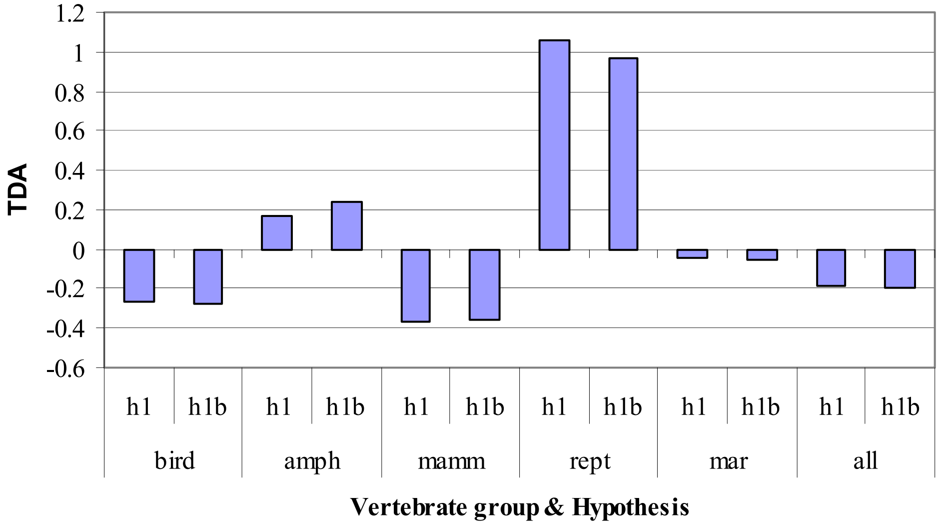


Figure 19: TDA slope summarized over vertebrate groups and hypothesis.

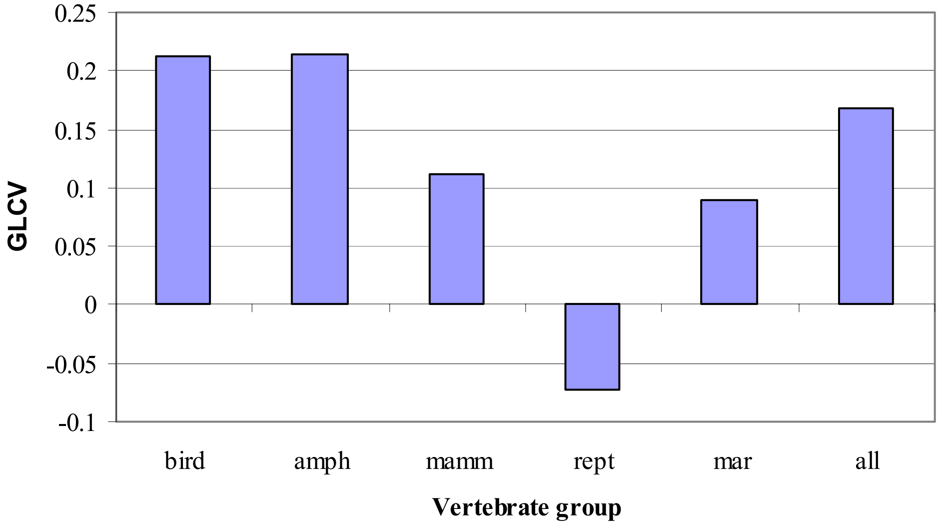


Figure 20: GLCV slope summarized over vertebrate groups.

Second, the coefficients for the variables for the H1a hypothesis (NPPAV), and for the H3 hypothesis (GLCV, GLCF, and ELD), all show markedly more stability and consistency across dimensions than the variable coefficients for H1b and H2. Graphs for TVNA (which is included in the two-variable and all-variable H2 models), provides a typical example of this instability; figure 21 shows how the coefficients vary significantly across states, within each vertebrate group, and figure 22 shows the converse: how much the coefficients vary across vertebrates, within states. Neither paint an encouraging picture of a consistent effect of TVNA on richness. Graphs of the H1b variable (TDA), and the other H2 variables (TRA, TFS, and PTSR) are similar in nature.

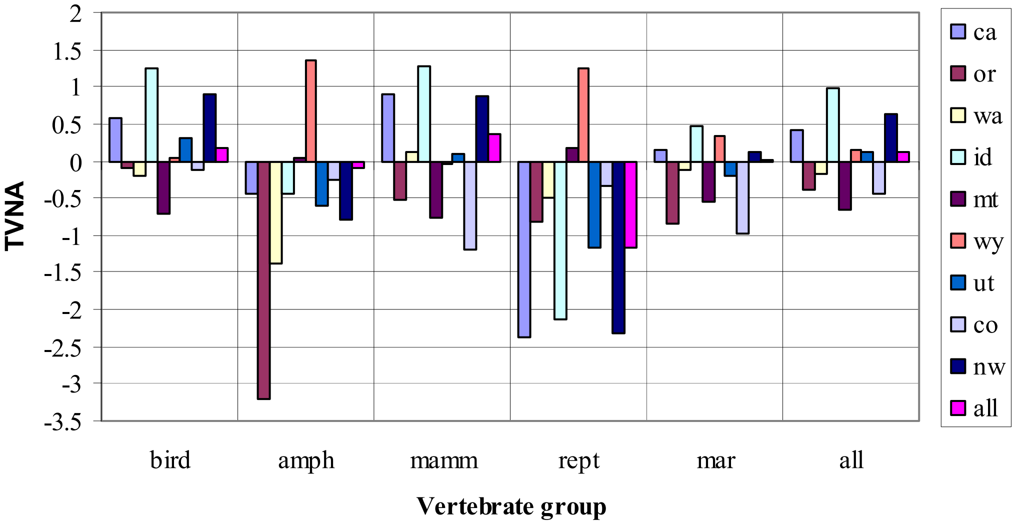


Figure 21: TVNA slope summarized over vertebrate group.

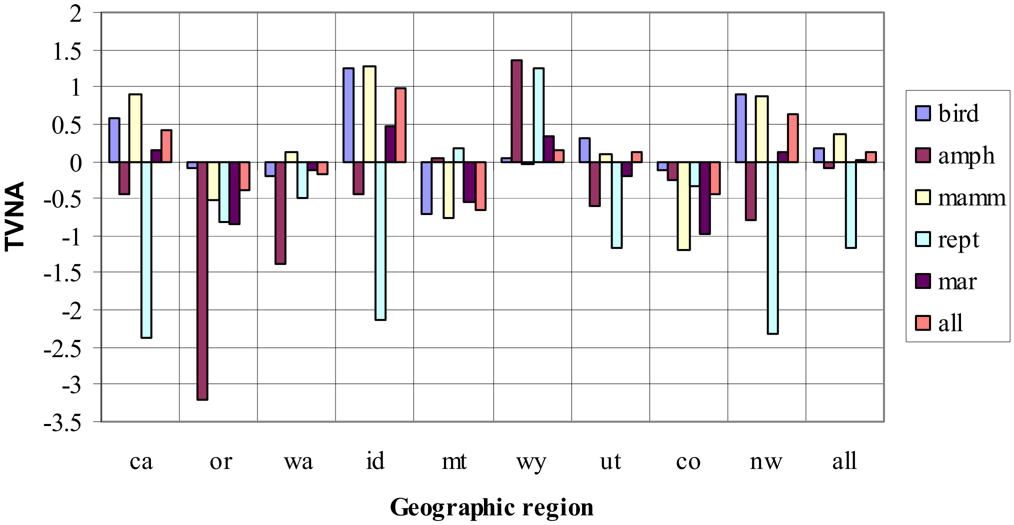


Figure 22: TVNA slope summarized over region.

Conversely, the H1a and H3 variables show a higher degree of consistency across these dimensions. The H1a (NPPAV) graphs are presented above in figures 17 and

18, and show much more consistent behavior of the coefficients. GLCF's coefficients are also less erratic – see figures 23 and 24. This pattern holds for all three H3 variables, and thus may indicate that productivity (H1a) and habitat heterogeneity (H3) are capturing a more fundamental relationship to species richness than climatic energy (H1b), or climatic variability (H2).

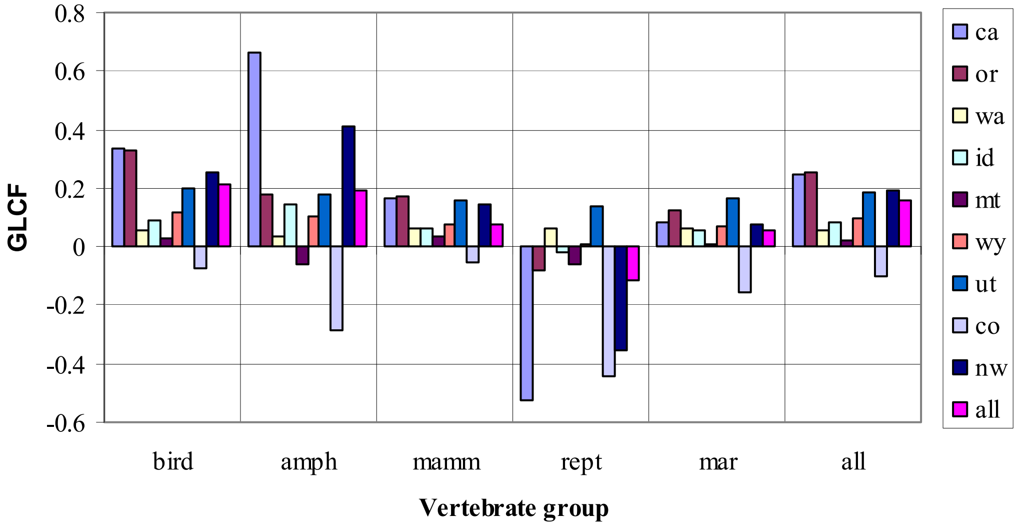


Figure 23: GLCF slope summarized over vertebrate group.

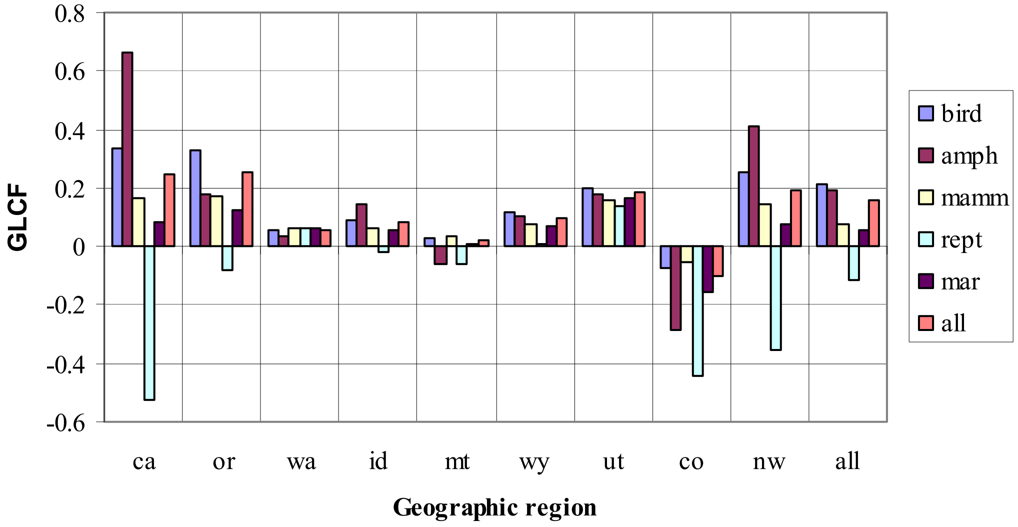


Figure 24: GLCF slope summarized over geographic region.

In addition to the consistency of coefficient signs and magnitudes, the absolute sign of the coefficients is important for understanding if the variables support the hypotheses in the expected way. As with the consistency of the coefficients, the signs generally conform to the hypotheses' mechanism for the H1a (productivity), and H3 (habitat heterogeneity) hypotheses, but the results are less conclusive for the other hypotheses.

For the energy/productivity hypotheses (H1, H1a, H1b), the supposition is that higher amounts of energy lead to increased species richness. As shown above, NPPAV is almost always positive, except for reptiles in some states, where the signs are mixed (figure 18). Aside from those reptile cases, this accords well with the hypothesized mechanism. For TDA, the results are mixed, with generally negative coefficients for birds and mammals, positive for reptiles, and somewhat mixed, depending on state, for amphibians. For all vertebrates, the sign is also negative – not surprising, since the all vertebrates group is heavily dominated by birds and mammals.

For the habitat heterogeneity hypotheses (H3), the coefficients for GLCV, GLCF, and ELD are mostly positive for birds, mammals, and amphibians, and mostly negative for reptiles (figures 20, 23, and 24 show this for GLCV and GLCF). A positive coefficient agrees with the hypothesized mechanism, by which more heterogeneity leads to higher species richness. In this case, the coefficient for all vertebrates and for the non-volant (MAR) is also positive, and so in agreement with the mechanism.

For the climatic variability hypothesis, the signs are mixed for all four variables. TVNA has mostly negative signs for amphibian and reptile models, mostly positive for birds, and mixed signs for mammals. The other three variables, TRA, TFS, and PTSR, all have the same signs for all vertebrate groups except reptiles. For TRA and PTSR these are negative generally but positive for reptiles, and for TFS the signs are positive generally, but negative for reptiles. Thus there appears to be some consistency between the effect of a variable on the different vertebrate groups, with the effect on reptile richness usually being the opposite of the effect on the other groups. However, for the mechanism of the H2 hypothesis, the effect of a variable on richness is not clearly defined, because the mechanism can work in two alternate ways: higher variability leading to higher species richness by partitioning the environment into more specialized niches; or lower variability leading to higher species richness because it allows species to subdivide an equable environment more finely, as there is less need to maintain access to a broader base of resources. Both mechanisms may be at work here, and thus it is difficult to conclude, from coefficient sign alone, whether the H2 hypothesis is, or is not, well supported.

5.3.1.4. Likelihood ratio comparisons

The results from the likelihood ratio tests also support the conclusion that H1 is the best supported hypothesis. The likelihood ratio value is an indication of the difference between a specific sub-model of interest and the overall model, which includes all candidate variables. Higher values of the ratio indicate larger differences with the overall model. Thus when a set of variables representing a given hypothesis are removed from the overall model, a high value of the resulting likelihood ratio indicates that the removal of those variables has significantly impacted the model's performance. A χ^2 test on these ratio values produces a p-value indicating the significance of the difference between the two models; p-values less than 0.05 indicate the reduced model is performing substantially worse than the overall model, and thereby infers that the removed variables have a significant explanatory effect, even in the company of the other hypotheses' variables. Figure 25 summarizes the LR results for all dimensions of the analysis (spatial scale, vertebrate class, geographic region), for the cases of one-variable, two-variables, and all-variables models. This figure shows the summed p-values for all models, across all dimensions; the best performing hypotheses are those with the lowest sums of p-values. For both the one-variable and two-variables cases, H1 performs best; in the one-variable case, either H1a or H1b perform better than H2 or H3. For the case of all-variables, H2 marginally performs better than H1. However, note that the differences are much more substantial for the one and two-variable cases.

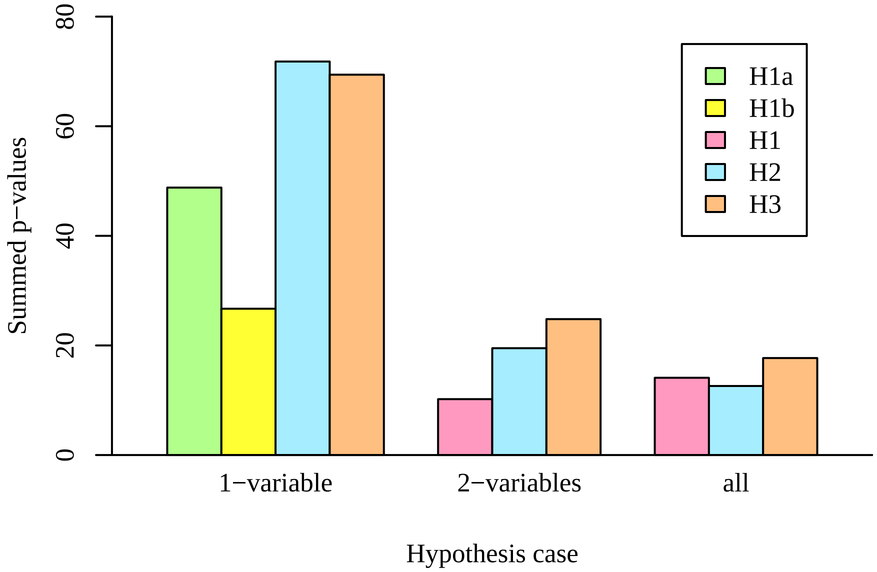


Figure 25: Likelihood ratio test results summarized over all dimensions (spatial scale, vertebrate group, and geographic region), for one-variable, two-variable, and all-variable cases of each hypothesis. Lower p-value sums indicate better supported hypothesis.

Focusing on the case of two-variable models, as the most relevant comparison, figure 26 shows how the likelihood ratio results break out by state and vertebrate grouping. As the figure shows, H1 is the best overall performer for all states except Washington, which is better served with H2. Note, however, that even so, mammals are better modeled by H1, and other individual discrepancies exist, such as amphibians in Idaho, Montana, and Washington being better modeled with either H2 or H3.

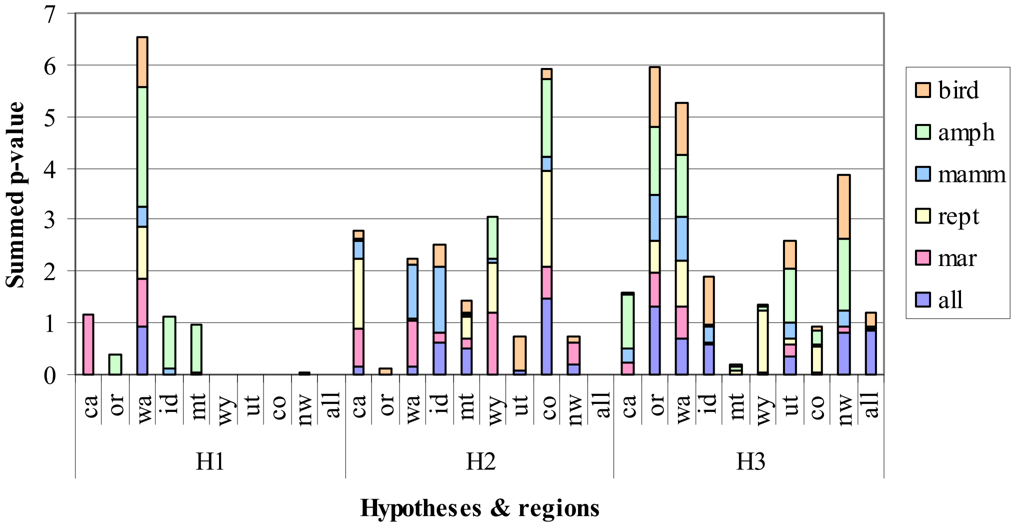


Figure 26: Likelihood ratio results by state. p-values summed, for two-variable case only. Lower sums indicate better supported hypothesis.

Figure 27 shows how the results break out by scale. This provides some useful diagnostic information. As is clear from the figure, most of the differences in LR results for the two-variable case derive from the 32 and 64 km spatial scales; at 8 and 16 km, the performance of the three hypotheses is very similar. Figure 28 shows the same plot, but for the one-variable case. In this case, H1a and H1b do appear to perform better at the 32 and 64 km scales. H3 performs as well as the 8 and 16 km scales, but suffers significant performance degradation at the coarser 32 and 64 km scales.

Overall, the likelihood ratio results support the general conclusion that the H1 hypothesis is better supported by the data, although for the two-variable case, which is of most interest, this conclusion appears limited to the 32 and 64 km spatial scales. At the 8 and 16 km scales, these likelihood ratio tests do not provide substantial supportive evidence for any of the two-variable hypotheses. For the one-variable case, H1a and H1b perform about as well as H3, overall, for the 8 and 16 km scales, and all perform better than H2. At the coarser 32 and 64 km scales, H1 (H1a or H1b) perform better than H3.

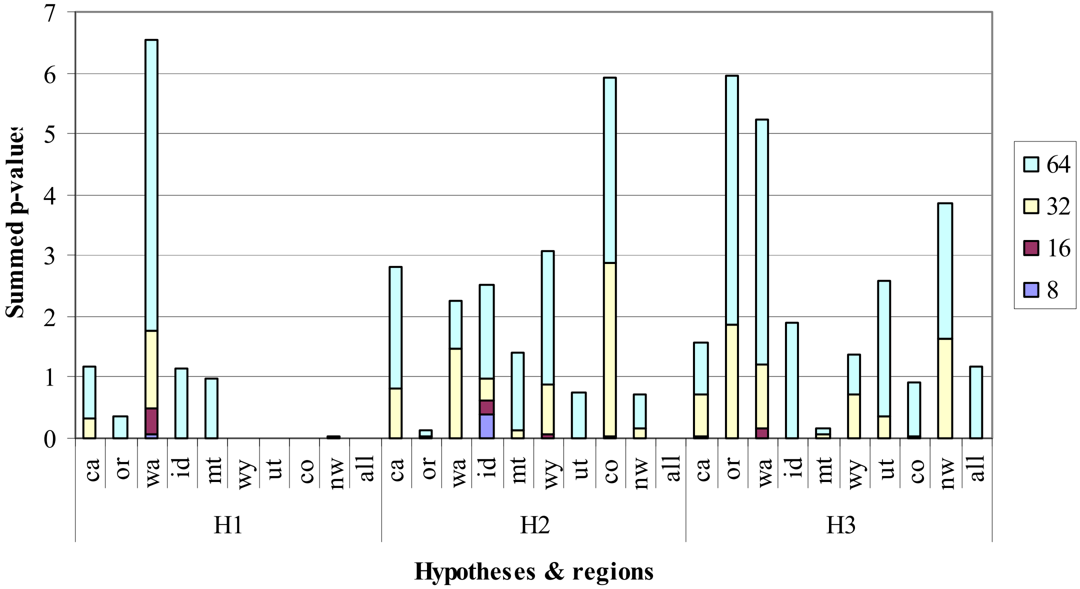


Figure 27: Likelihood ratio results for two-variable models, split out by region and scale.

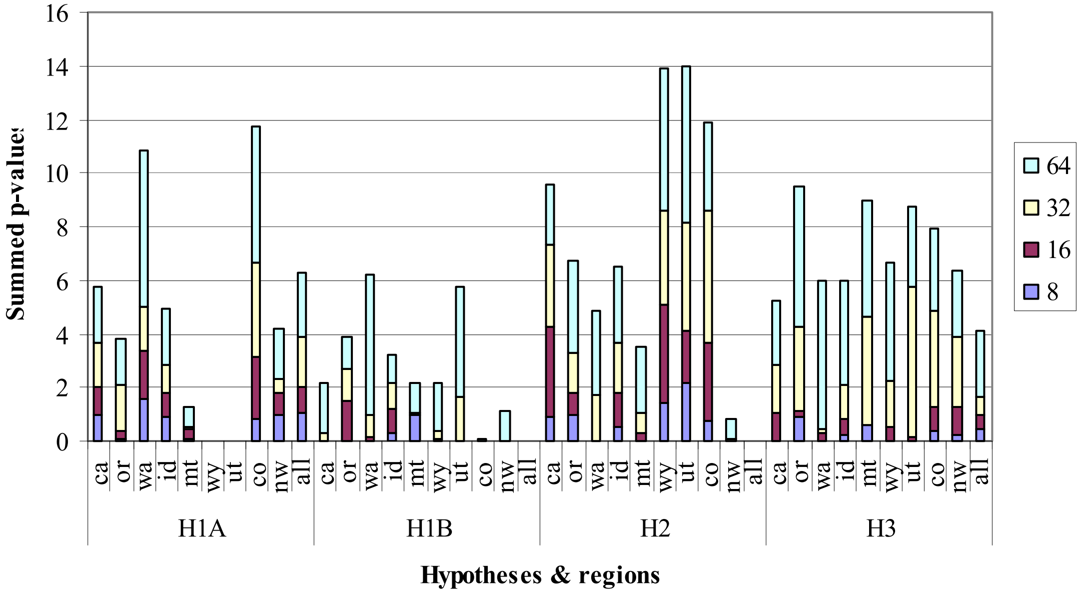


Figure 28: Likelihood ratio results for one-variable models, split out by region and scale.

5.3.1.5. Summary: hypothesis support

Encouragingly, the AIC, R^2, coefficient evaluation, and likelihood ratio tests all provide similar conclusions about which hypotheses are best supported by the data. The clear primary conclusion is that H1 is the best supported hypothesis. The AIC and R^2 methods then suggest that either H2 or H3 are the second best supported, with neither consistently or overwhelmingly preferred, and that within H1, H1b performs better than H1a. However, the consistency and sign of the coefficients indicate that H1a may be preferred over H1b, and that H3 may be preferred over H2. Thus there are some differences in the second-order details of which hypotheses are best supported, but fairly good agreement about the overall message: the energy/productivity hypothesis appears to be best supported by the data over most of the eight states considered here. It also appears to be best for most individual vertebrate classes, as well as all vertebrates, and for the non-volant vertebrates (MAR).

Examining the results more closely, there are clearly many exceptions to these general conclusions, and some cases that seem to follow different relationships. The results for Colorado, for example, are often different that those for most of the other states. Washington is also frequently differently characterized, as are reptiles. After H1, the relative performance of H2 and H3 varies somewhat by vertebrate class; birds and mammals are both best supported by H1, followed by H3, and with H2 clearly as worst performing hypothesis. Reptiles are best supported by H1, followed by H2, and with H3 clearly as worst performing hypothesis. Amphibians are best supported by H1 only in the two-variable case, and by H2 in the all-variables and one-variable cases. H3 performs worst by a clear margin for amphibians.

5.3.2. Atypical model performance

Aside from identifying the best supported hypotheses, the modeling results also provide additional useful insight into the hypotheses and the data considered here to evaluate them.

5.3.2.1. Regionally

When broken out by region (figures 12 and 14), the AIC and R^2 results both indicate that for the state of Washington, the H2 hypothesis offers better performance than either H1 or H3, for either all-variable, two-variable, or one-variable models. Examining coefficients by state, those for Colorado are often generally of a different sign than those for all other states. This is certainly true for NPPAV, TDA, TRA, TFS, GLCV, GLCF, and ELD. No other state stands out with consistently different coefficients as does Colorado.

Spatially, figures 29 – 34 illustrate, for each vertebrate class, the richness predicted by each OLS two-variable hypothesis model, for both all states combined (“Combined”), and for each state individually (“Individual”), at the 16 km spatial scale. The figures also show the spatial distribution of residuals, and have boxplots showing the comparative distribution of residuals between the different maps. These graphs support and illustrate the results presented in the previous section on hypothesis support; they show the residuals for the H1 models, for individual states, have the least spatial pattern, and the smallest distribution (e.g., are closest to zero). These graphs also highlight the differences for certain states—Colorado in particular, which in the combined case always has anomalously high residuals (indicating under-prediction). The significances of the models for all states are seriously reduced by the inclusion of

Colorado; in most cases, as the residual maps make clear, the all-state models show high either positive or negative residuals for Colorado. For the H3 reptiles, the combined all-state model, which contains GLCF and ELD, explains very little variance, at least at the 16 km scale. The residual histograms also clearly show that the all-state models greatly increase the range of residuals. The residuals for the H1 and H3 hypotheses are the most distributed, with the least residual spatial pattern. Although not shown, the residuals from the corresponding CAR models have no detectable spatial pattern.

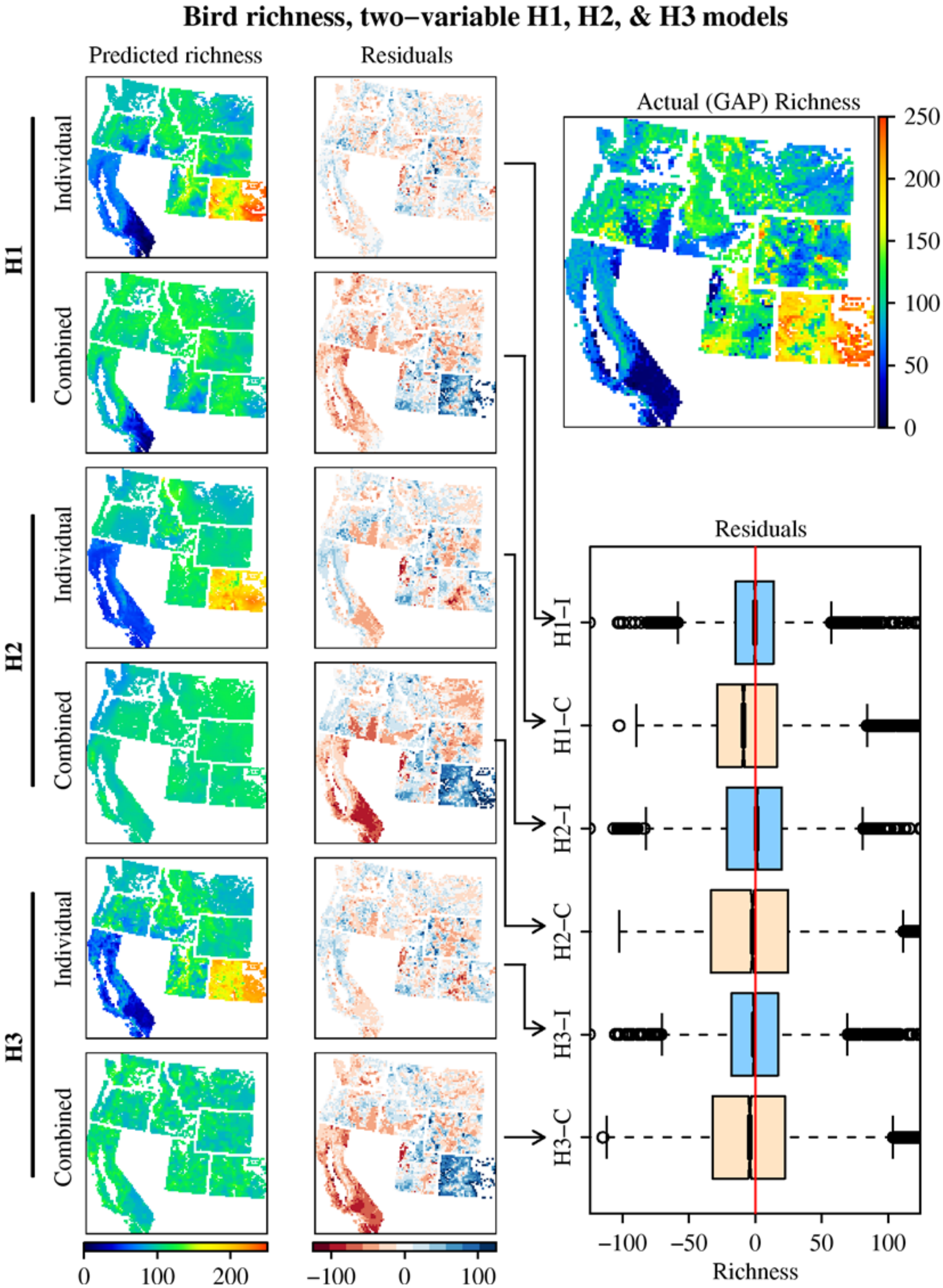


Figure 29: Predicted bird richness, for H1, H2, and H3 two-variable hypotheses, 16 km scale. “Individual” rows show individual models per state, mosaicked for display. “Combined” show the model for all states combined. Boxplot of residuals shows the distribution of residuals for each row; blue boxes indicate individually modelled states, while beige indicate combined state models. Not all outliers (circles) are shown, in order to focus on central quartiles, which are delimited by the box. Black vertical line inside box indicates median. “Whiskers” (vertical lines at end of dashed line) are located at 1.5 × interquartile range (length of box).

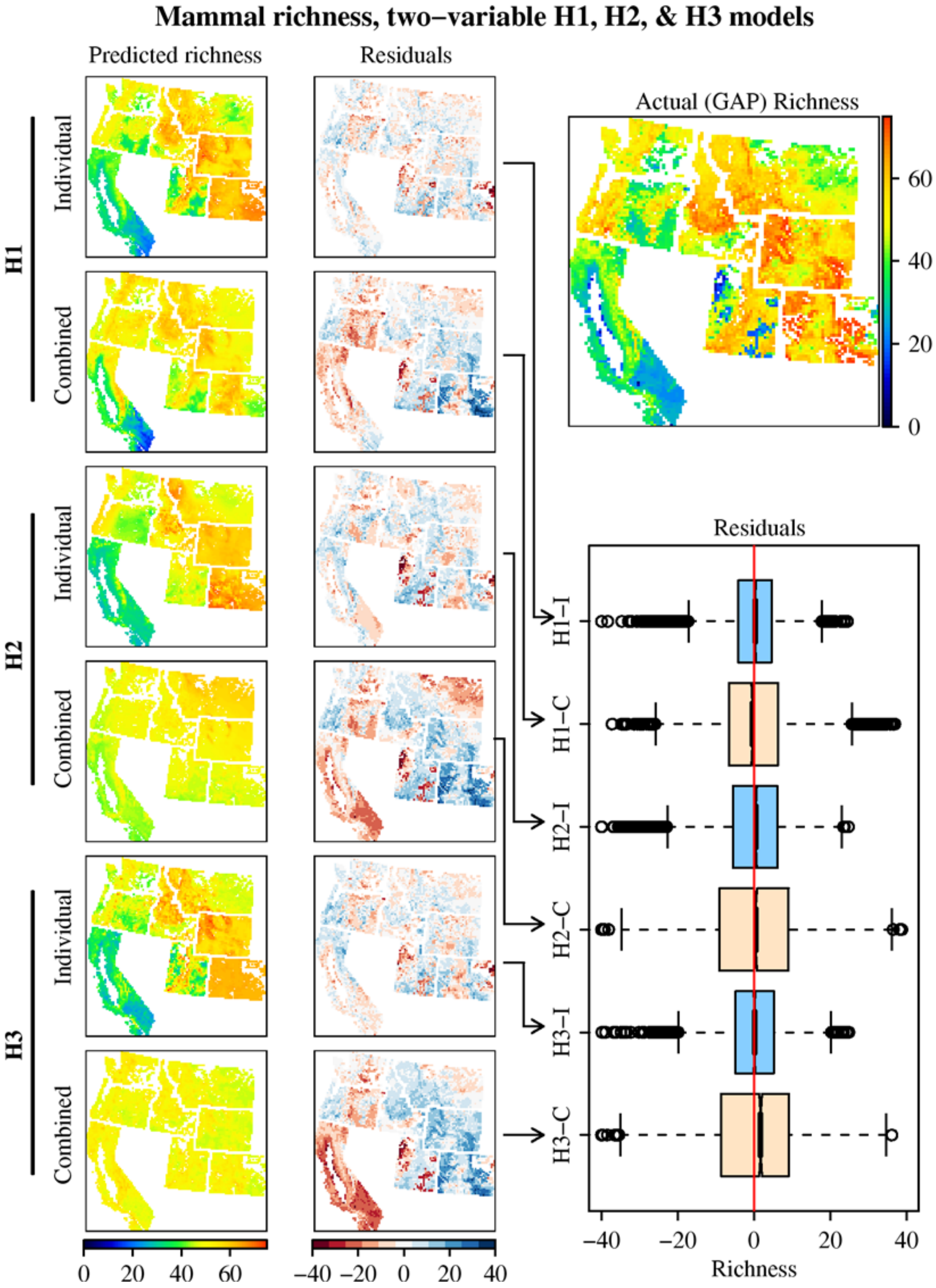


Figure 30: Predicted mammal richness, for H1, H2, and H3 two-variable hypotheses, 16 km scale. "Individual" rows show individual models per state, mosaicked for display. "Combined" show the model for all states combined. Boxplot of residuals shows the distribution of residuals for each row; blue boxes indicate individually modelled states, while beige indicate combined state models. Not all outliers (circles) are shown, in order to focus on central quartiles, which are delimited by the box. Black vertical line inside box indicates median. "Whiskers" (vertical lines at end of dashed line) are located at 1.5 × interquartile range (length of box).

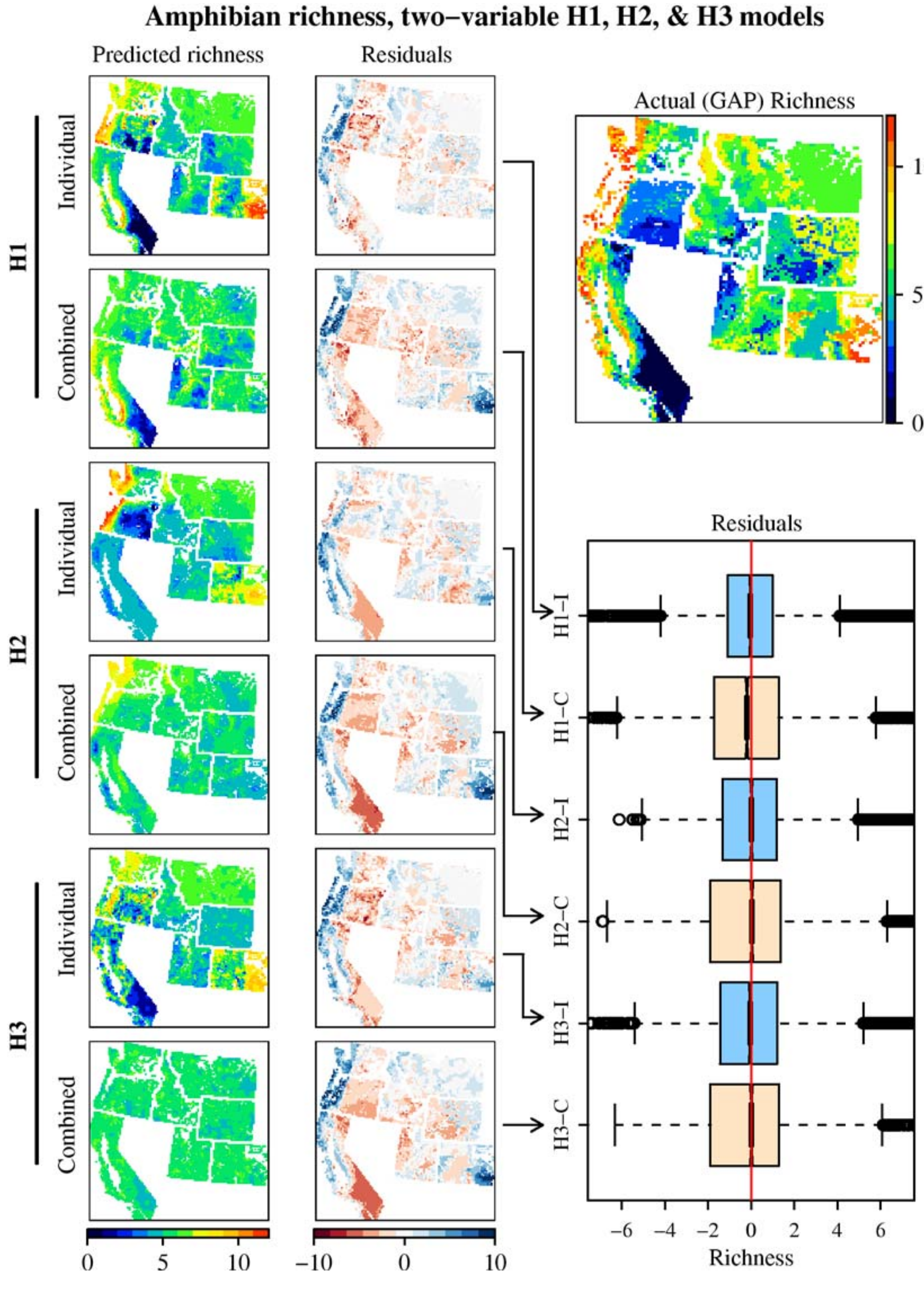


Figure 31: Predicted amphibian richness, for H1, H2, and H3 two-variable hypotheses, 16 km scale. "Individual" rows show individual models per state, mosaicked for display. "Combined" show the model for all states combined. Boxplot of residuals shows the distribution of residuals for each row; blue boxes indicate individually modelled states, while beige indicate combined state models. Not all outliers (circles) are shown, in order to focus on central quartiles, which are delimited by the box. Black vertical line inside box indicates median. "Whiskers" (vertical lines at end of dashed line) are located at 1.5 × interquartile range (length of box).

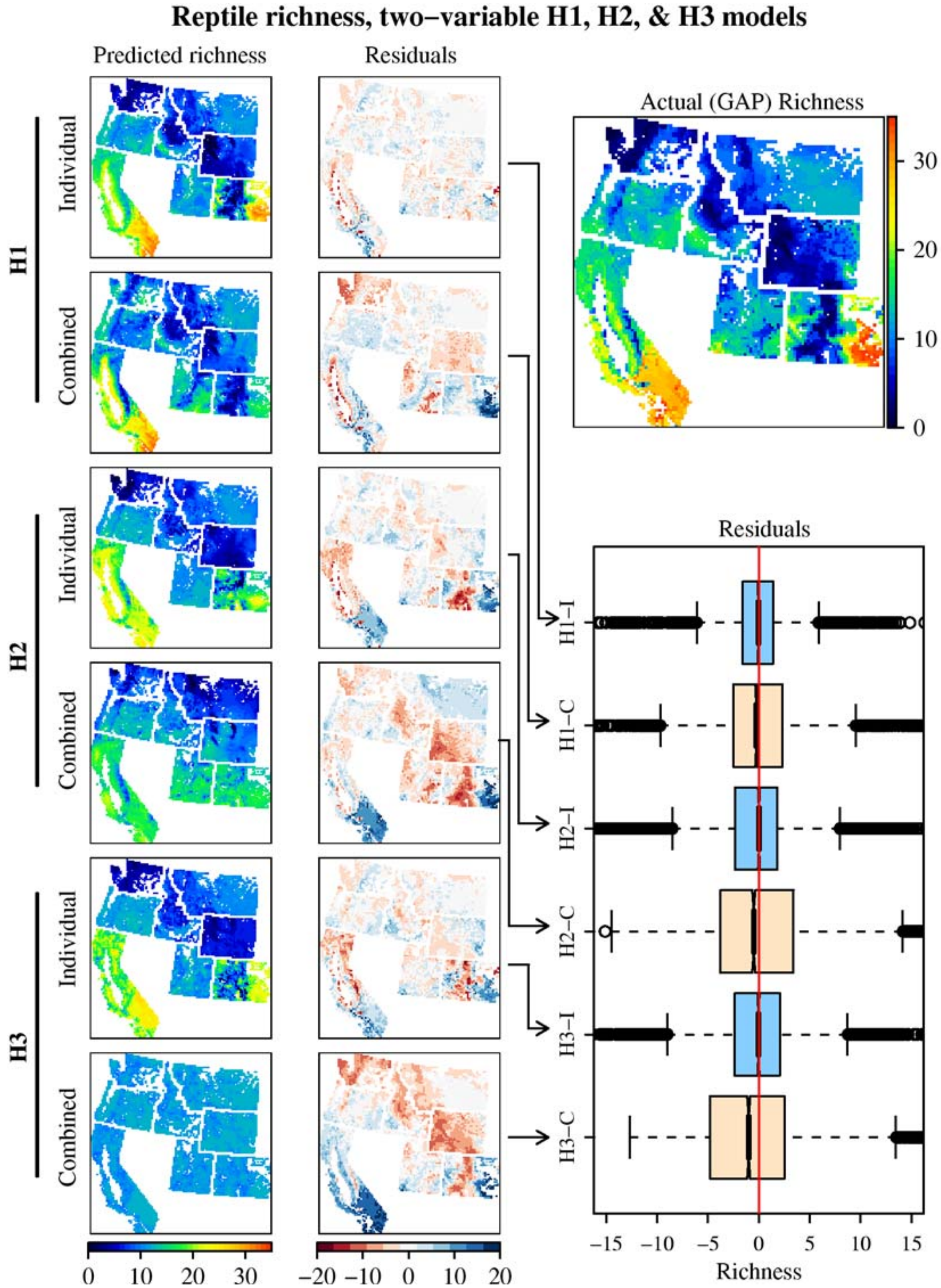


Figure 32: Predicted reptile richness, for H1, H2, and H3 two-variable hypotheses, 16 km scale. "Individual" rows show individual models per state, mosaicked for display. "Combined" show the model for all states combined. Boxplot of residuals shows the distribution of residuals for each row; blue boxes indicate individually modelled states, while beige indicate combined state models. Not all outliers (circles) are shown, in order to focus on central quartiles, which are delimited by the box. Black vertical line inside box indicates median. "Whiskers" (vertical lines at end of dashed line) are located at 1.5 × interquartile range (length of box).

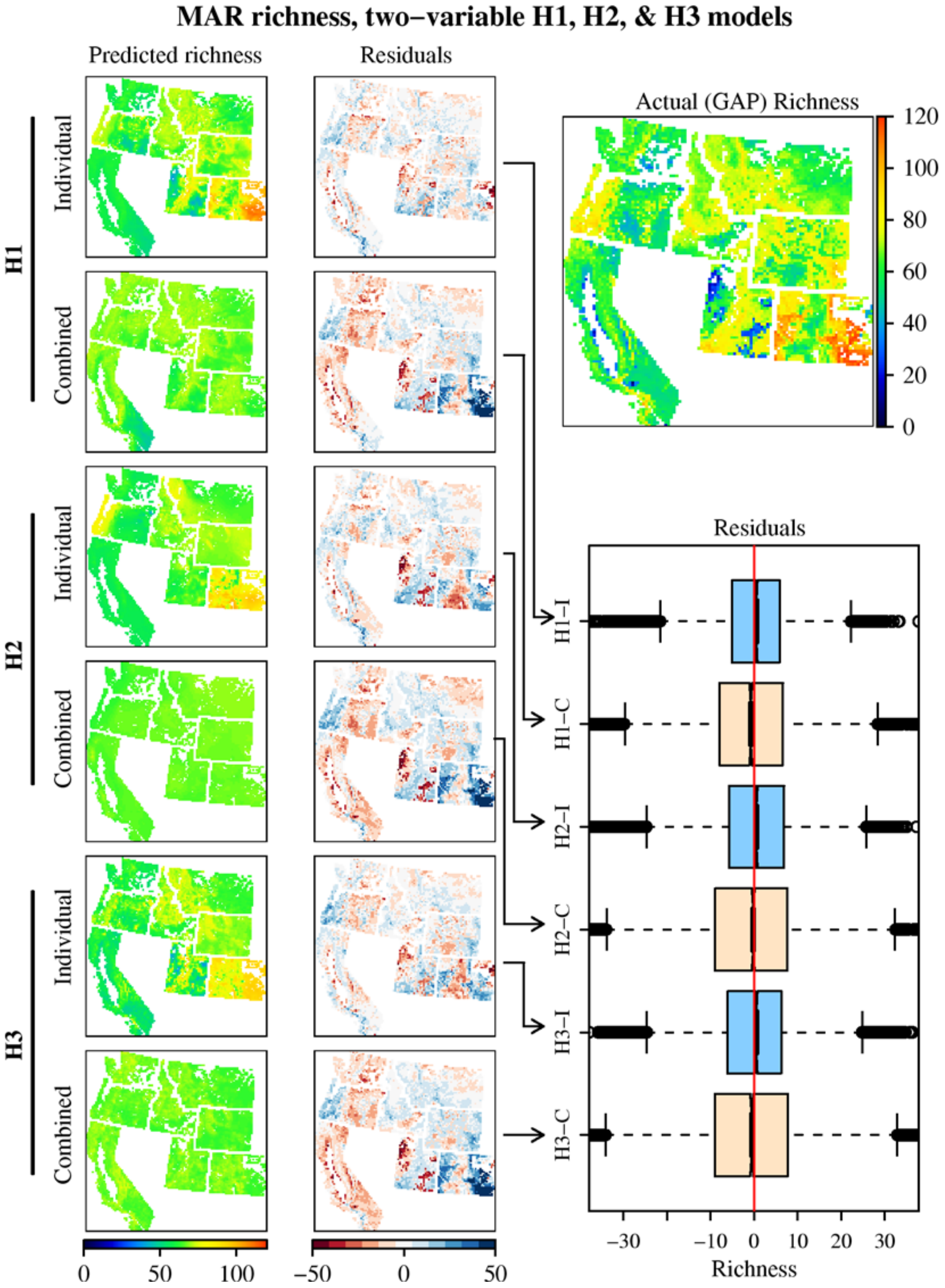


Figure 33: Predicted MAR richness, for H1, H2, and H3 two-variable hypotheses, 16 km scale. "Individual" rows show individual models per state, mosaicked for display. "Combined" show the model for all states combined. Boxplot of residuals shows the distribution of residuals for each row; blue boxes indicate individually modelled states, while beige indicate combined state models. Not all outliers (circles) are shown, in order to focus on central quartiles, which are delimited by the box. Black vertical line inside box indicates median. "Whiskers" (vertical lines at end of dashed line) are located at 1.5 × interquartile range (length of box).

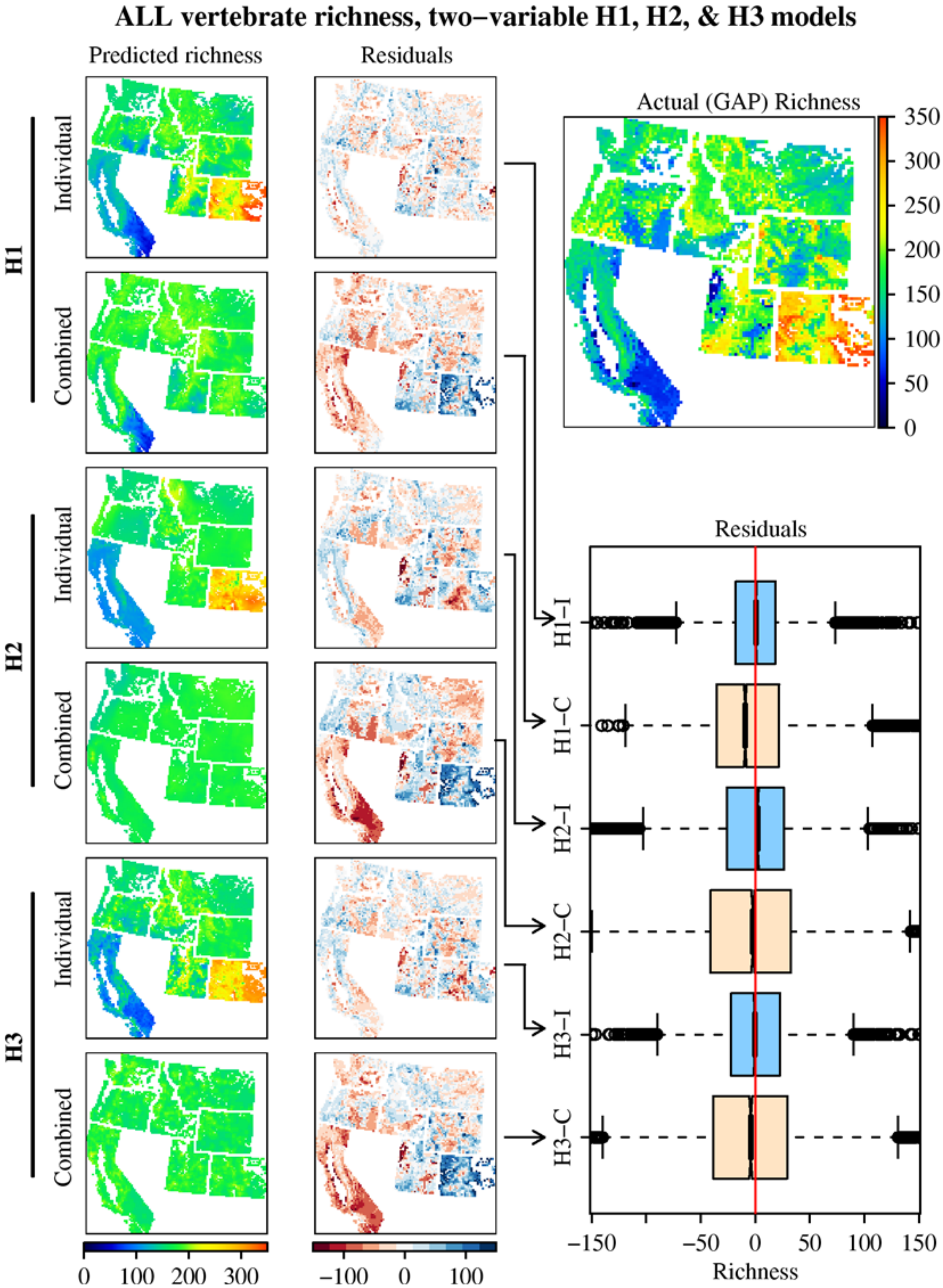


Figure 34: Predicted ALL vertebrate richness, for H1, H2, and H3 two-variable hypotheses, 16 km scale. "Individual" rows show individual models per state, mosaicked for display. "Combined" show the model for all states combined. Boxplot of residuals shows the distribution of residuals for each row; blue boxes indicate individually modelled states, while beige indicate combined state models. Not all outliers (circles) are shown, in order to focus on central quartiles, which are delimited by the box. Black vertical line inside box indicates median. "Whiskers" (vertical lines at end of dashed line) are located at 1.5 × interquartile range (length of box).

5.3.2.2. Vertebrate grouping

Examining the results by vertebrate grouping, the most consistent anomaly occurs with reptiles, as previously mentioned. The sign of the coefficients for NPPAV, TDA, TRA, TFS, GLCV, GLCF, and ELD in reptile models is opposite the sign for all other individual vertebrate classes. For TVNA, the coefficients for both reptile and amphibians models are opposite that for mammals and birds, and of much greater magnitude. This indicates that TVNA is much more important for amphibians and reptile models than for mammals and birds. As TVNA is a measure of the day-to-day variability in temperature, its importance for reptile and amphibian diversity is perhaps not surprising, since both groups are generally ectothermic and thus much more dependent on ambient temperature, and its fluctuations, than birds or mammals, which have more developed abilities for body temperature regulation.

Chapter 6: Discussion

6.1. General approach

The question of what drives species richness across large geographic regions has been unresolved within the ecological community for some time now, without, if the recent contentious literature is any indication, any immediate resolution in sight. The current study attempts to shed some light on the issue in a variety of ways. For one, the datasets considered here were specifically chosen to provide better and more representative measures of both important environmental variables as well as species richness itself. The methodology was also developed in response to some of the more obvious deficiencies commonly seen in published work on this topic. Typically, issues of scale and spatial autocorrelation are not adequately addressed, nor the many dangers of collinearity and variable selection procedures. Perhaps more importantly, most studies in the literature evaluate or develop only a single hypothesis. While this may be perfectly appropriate for the specific question of interest, it does little to inform the larger debate over which hypothesis is best supported. In this work, a different approach was taken, in which several standard hypotheses were evaluated simultaneously, with common methods, and with the goal of identifying the best supported hypotheses.

A considerable difficulty exists with evaluating the types of rather general hypotheses considered in this study. The hypotheses are general in that they often specify only a broad mechanism, which may be well supported theoretically, and often even empirically in some cases. But there is much less, if any, guidance, on specifics for parameterizing a model of the hypothesis. The most obvious example here is the climatic variability hypotheses; the mechanism is fairly well defined (environmental variability affecting the size and distribution of species' niches, and thereby species richness), but the very nature of the hypothesis, as generally stated, precludes an exact specification of the relevant drivers: 'environmental variability' encompasses a wide range of possibilities, and these may differently affect different groups of species. Which manifestation of variability should then be chosen for testing a hypothesis? The most relevant variability is the obvious answer, but that might change with area, scale, species groups, or other factors. The question then arises as to whether *any* generalization can be made, applicable to all species. Whether justified or not, those types of generalizations underlie most, if not all, research on this topic, the present study included, and likely contribute much to the current confusion.

A related issue arises with spatial scale; when hypotheses are evaluated at a specific scale, the data must be appropriately generated at that scale. Furthermore, the meaning of physical variables can mutate as the scale changes, as they begin to convolve with their environment—the classic mixed pixel problem. For example, the meaning of NPP over a hectare of homogenous forest is obvious. But what is the meaning of NPP over a 32 km pixel? The "data" answer is that it is the average of the underlying component NPP values. But in the context of a driver of species richness, which is more important, the average NPP, or the NPP at the given locations, where specific organisms reside and where speciation occurs? If the latter, then it will also depend on the size and sensible environment of the creature; a salamander and a bird will likely be responsive areas differing in size by an order of magnitude or more. Variables measuring habitat heterogeneity, which is even more directly tied to scale, present yet more substantial issues of this nature.

Despite these sometimes daunting problems, the present study was conducted in the hopes of being able to draw out some useful conclusions about the scope of validity of the hypotheses across a relatively large area. A key feature of the current analysis in this regard is the reliance on the summary results from many variant models for each hypothesis. Using these variants—constructing models for four different spatial scales, for six different vertebrate groupings, and for ten different regions—specifically attempts to avoid unrepresentative results dependent on any particular model configuration or methodological choice, and so provide more generally applicable conclusions. In the final, eight-state analysis that forms the basis for the conclusions of this study, the outputs from over 5000 individual OLS models were collated in various ways to derive the conclusions presented here; approximately 2000 models were necessary for the R^2, AIC, and coefficient comparison evaluation methods, and another 3000 models for the likelihood ratio method. An equally large number of models were used in the preliminary work to develop that final methodology, in order to test the use of CAR models, variable selection methods, variable transformations, and the other methodological choices discussed in Chapter 4. Certainly, one issue is whether the selected dimensions are genuinely expansive enough to provide the general conclusions claimed; replicating the models over trivial dimensions could lead to apparently more general conclusions, but this would be misleading. As the results here show, there is often significant variation across those dimensions, and this should serve to validate their inclusion as a useful tool to quantify the generality of the hypotheses. Undoubtedly, other dimensions might also be useful in this regard, and future work might consider other possibilities.

6.2. Species richness hypotheses

As the full, eight-state analysis results show (section 5.3), the productivity/energy hypothesis (H1) is clearly the best supported, whether the evaluation is done by examining R^2 values, AIC values, a likelihood ratio comparison to omnibus models, or the consistency of estimated model coefficients. The climatic variability (H2) and habitat heterogeneity (H3) hypotheses almost always underperform H1, but their rank varies by vertebrate group and region.

These results are fairly unequivocal on this support for H1, which was somewhat surprising, considering some of the study's earlier results—over the four NW states, with all-variable models—which often suggested H2 to be best supported. In retrospect, it is clear that three features of the final analysis were critical: the decision to combine the productivity (H1a) and energy (H1b) hypotheses into a single energy/productivity (H1) hypothesis and model; the additional dimension added to iterate the number of predictor variables in each model (retaining all, or limiting to just one or two variables); and the selection of the variables used in H2 to a set of four containing very little multicollinearity.

Combining H1a and H1b into H1 generates a model combining both the energy and productivity aspects of the overall hypothesis. The fact that this H1 model usually performs substantially better than either H1a or H1b alone reflects the fact that the constituent variables are not substantially correlated; their r^2 values are well under 0.1. Thus, combining both variables in a single model will, if either variable has any relevance on its own, produce a more effective model. Including both also admits the possibility synergistic interaction effects. In either case, the combined H1 model performs substantially better than either H1a or H1b alone.

Restricting hypothesis comparisons to models with the same number of predictor variables also had a significant impact on the overall results. Although this may seem to be an obviously important constraint, particularly in retrospect, it was not considered earlier partly because there was a strong interest in finding an optimal characterization of climatic variability, and using just a few of the available variables might not be sufficient to properly isolate the relevant "variability" entity. And, indeed, this may still be the case. However, given the nature of the datasets and the concerns with collinearity, there is clear reason to be uneasy with such an approach. Most of the environmental variables are correlated to one another to some degree, simply because they reflect conditions on a geographic surface—the earths surface—that are generally related to one another. This correlation means, in a practical sense, that when one variable is found to be correlated to a response variable, there will likely be other similar variables that will also be correlated. In an OLS modeling context, those additional variables may appear significant when added to a model, even though their additional contribution is relatively minor.

The principal components variable selection methodology used for the final analysis was chosen partly because it was able to directly address this concern about collinearity. In previous iterations, a larger set of variables was used for the H2 hypotheses, and they exhibited significantly more collinearity that the four chosen for the final analysis; reviewing VIFs of each model (and dropping variables as necessary) was then obligatory, and with the number of models considered here, quite a laborious task. Thus, the set of variables chosen for the final analysis are fairly optimal from the perspective of minimal collinearity, and it is likely that this is the primary reason the H2 hypothesis did not perform as well as it had in the preliminary studies, where the set of variables was more correlated.

The full analysis's geographic expansion to cover four additional states (MT, UT, WY, CO), beyond the four initially considered in the preliminary NW studies, also provides support for these conclusions. That preliminary work gave more weight to the H2 hypothesis (Washington state is always best supported by H2, and H2 performs well for Oregon for the all-variables and one-variable cases), but these four additional states clearly indicate H1 is better supported. The preliminary work also suffered from more collinearity among the H2 variables, which likely also inflated the importance of H2 relative to the other hypotheses. In both cases, the addition of four more states helps confirm the robustness of the conclusions.

As noted in the results chapter, the H2 and H3 hypotheses vary in their position as next best supported hypothesis, after H1. By R^2 predictive ability, H3 outperforms H2 for the two-variable case, with H2 outperforming H3 for both the all-variable and one-variable cases. However, the difference in the performance of the two hypotheses, in aggregate (over all states, or all vertebrates), is not large. Within state or vertebrate class, the differences can be much more substantial. The most obvious pattern is that H2 performs notably better than H3 for amphibians and reptiles, while H3 performs notably better for birds and mammals. This might reflect two obvious differences between these groups: amphibians and reptiles are ectothermic, and so more dependent upon the variability (or lack thereof) of environmental conditions, while they are also comparatively small-ranging animals, and so may be less affected by the relatively coarse measures of habitat heterogeneity measured by the landcover variety variables (GLCV and GLCF) in H3. The strong negative coefficients for H2's TVNA (day-to-day temperature variability) for both amphibians and reptiles supports the former suggestion.

Note that H3 suffers a significant handicap, in its potential performance ability, in comparison to H2, based on the source and definition of the predictor variables. For H2, many different measures of climatic variability were initially considered (see the 13 variables in table 6), and these climatic variables might be expected to fairly well characterize the pixel of interest (albeit with local variation due to topography). For the GLCF and GLCF measures of landcover variety for H3, however, it is less clear that counting the number of landcover types, either within a pixel of interest or surrounding it, is necessarily a condition 'felt' by animals within that pixel.

Overall, the superior support for the energy/productivity hypotheses appears substantial and clear. This is obviously, as the results chapter demonstrated, the case across the various dimensions of the analysis. It is also clear even though similar iterations of this research tentatively led to alternate, and less convincing, conclusions; the current methodological approach effectively addresses those previous oversights, and is supported by the consistency of the results. More specific observations about the results and methodology are also informative, and will be discussed next.

6.3. Additional observations about species richness

The Results chapter summarizes many of the idiosyncrasies observed in the final analysis, but the most noteworthy will be reviewed here. Spatially, Colorado appears to have rather different relationships to richness than are observed in most other states, although it still supports H1 as the best hypothesis. But within those results, the coefficients for variables in Colorado models are almost consistently of a different sign than the majority of states. Other states may show such inconsistencies too, but they are less pervasive. This suggests something unusual about Colorado. The species richness maps of Colorado support this too; as the figures in the Appendix I show, Colorado richness is often substantially higher than that of neighboring states, particularly for birds. A closer examination of that state's GAP project methodology and results might help in understanding this anomaly. But it is reassuring that even so, the Colorado models come to the same general conclusions as this study as a whole.

The other obvious anomaly is for the models of reptiles. Similar to the case for Colorado, the coefficients of the predictor variables in reptile models are also often of the opposite sign than the coefficients for non-reptile models. Sometimes amphibians join reptiles in this sign switching, suggesting a difference between ectothermic and endothermic animals. Ectothermic animals – and their distributions – will be more sensitive to temperature variability, as well as to temperature extremes (particularly cold conditions), and this is likely responsible for many of the observed differences between the models for reptiles and amphibians, and those for mammals and birds. Within ectotherms, the differences between amphibians and reptiles might simply reflect that they inhabit very different habitats; amphibians are found in places of equable climates where plenty of birds and mammals also coexist, and so have some commonality in environmental conditions with those animals. Reptiles, on the other hand, often inhabit arid areas, with few mammals, birds, or, especially, amphibians. Thus, although both reptiles and amphibians are ectothermic, they have very different environmental affinities, and this may be reflected in the differing coefficient signs. Because of their basic physiological differences with endotherms, it may well be expected that the climatic variability that ectotherms respond to is substantially different than that affecting endotherms. This difference may highlight one of the fundamental shortcomings of the H2 hypothesis, as parameterized in this study: the climatic variability that a given organism (or class of vertebrates) is responsive to may simply be very different than the variability that other groups are responsive to. In such a case, the most appropriate H2 hypothesis model might require a composite of models, in which each vertebrate class is modeled separately, with individually selected sets of predictor variables, or a non-linear model that would produce the same effect. The same types of conclusions could be extracted from such models as have been developed here, but it would require a major change in the modeling approach.

Another notable issue with the reptile models is that in the one-variable case, productivity (H1a) performs very poorly. On its face, this is unsurprising, as reptiles frequently inhabit arid areas with little vegetation. However, this may again point out that the hypotheses considered here—or, at least, how they have been parameterized here—may not be general enough to fully model all richness. This does not invalidate the hypotheses under scrutiny here, but may simply indicate that, for example, the productivity that is relevant to desert reptiles may simply not be well measured by the

NPPAV variable used here; it may require a more carefully constructed variable, perhaps capturing the rainy-season flush of vegetation, to properly parameterize those reptile models. This is essentially the same point as discussed in the previous paragraph; non-linear models, or more complex models with interactions, may be necessary to best accommodate a more effective set of predictor variables.

The results also presented a surprise concerning the impact of spatial scale: there appear to be no substantial differences related to scale. One of the assumptions of this work was that such dependencies would be important, perhaps even critical, to deriving more general conclusions. That it is not does not mean that scale is not important, but may simply indicate that scale was properly addressed in the generation of the datasets. Considerable attention was paid to this element of data preparation, in particular for scaling the species ranges into richness at the four analysis spatial scales. Scale's relative lack of importance may also simply reflect that the other sources of variation (state, vertebrate group, number of predictor variables) generally overshadow scale's effect, in this study.

"Scale" as used in this study is perhaps more accurately referred to as grain size, and defines the spatial unit of measure for analysis. The other major component of scale is the spatial extent of the study, and this, as the per-state ("region") results have shown, is clearly important. Unfortunately, due to the issues with the contiguity of the richness datasets across state borders, it is difficult to determine if the differences seen by state are due to issues of spatial extent – a true regional effect – or are simply artifacts of differences in the GAP richnesses. The often strikingly different results for Washington and Colorado suggest the latter may be responsible for the oddities observed in models for those states. The representativeness of these results globally, or just for North America as a whole, is difficult to assess purely from this study, but within the context of other North American studies, these results do not appear unusual, as the next section will discuss.

6.4. Results in context

As reviewed in the Background chapter, much of the literature does not explicitly examine multiple hypotheses, as has been done here. Nevertheless, the results of this research support what appears to be a growing consensus in the literature for the principal role of productivity and/or energy in determining species richness.

For overall vertebrate richness, Currie's (1991) results also supported the role of productivity at continental scales, through the proxy measure of PET. Kerr & Packer (1997) found topographic heterogeneity more important when restricting the analysis to only the continental US. The present study found habitat heterogeneity frequently important, but rarely more so than energy/productivity, and of overall similar importance as climatic variability. Its importance was much more substantial for birds and mammals than for amphibians and reptiles. Fraser's (1998) study is interesting because he used the same GAP datasets for species richness, but he found environmental heterogeneity much more important than measures of energy. However, his study only covered the state of Wyoming. Although the present work does show variation among states, it does not agree with Fraser's study, as it fairly consistently finds H1 to be the best supported hypothesis in Wyoming, although H3 always performs better than H2. The only instances where H3 is the best hypothesis are: the all-variable case; for mammals overall; and for Utah and Montana.

Most of the other continental studies examine either bird or mammals richness, and most come to similar conclusions. Badgley & Fox (2000) found PET most important for North American mammals, while for South American mammals both Ruggiero & Kitzberger (2004) and Tognelli & Kelt (2004) found productivity the best predictor. The former also found elevation variability to be an important predictor—in accord with what is reported here. Curiously, however, the Tognelli & Kelt (2004) study found little support for ambient energy. This is in contrast to this work, which found environmental energy, as measured by growing degree days (TDA), to be the more powerful correlate within the H1 hypothesis. For sub-Saharan birds, Jetz & Rahbek (2002) found productivity and habitat heterogeneity most important – again, very similar to the results here. Rahbek & Graves's 2001 study emphasized the importance of topographic variability for South American birds.

Although the results here often align with those in the recent literature, there are some consistent differences. As just noted, most recent studies seem to find more support for productivity itself, than this study has (as indicated in the one-variable case); more often, environmental energy was found to be the more important predictor. Also, only a few studies, such as H-Acevedo & Currie (2003) and Kerr & Packer (1997) appear to explicitly support climatic variability. One potential explanation for these differences is the spatial extent of this study. Most of the previously cited studies examined richness at continental scales—either North America, South America, or sub-Saharan Africa. The present study examines richness over a significant portion of the western US—a region containing much environmental and ecological diversity—but nonetheless this study area is substantially less than continental in extent; the results reported here may, to some degree, be specific to the western US, and may not, necessarily, be expected to hold for either the remainder of North America, or globally. Thus, the spatial extent of this study may be responsible for some of these differences. It is perhaps more notable that despite of the lack of a dataset of continental extent the conclusions here are in such general agreement with those found in the literature.

It is difficult to accurately compare results from studies that use differing methods and datasets, but insofar as one can compare, the conclusions that this research have come to are not, in the main, very different than those reached in other published studies, although certainly some of the details and second-order effects do vary. In view of the literature, the contribution here is most significant as an additional confirmation of the importance of the energy/productivity hypothesis at continental scales, particularly as this derives from a comparison of all three major hypotheses using several different evaluative criteria.

6.5. Statistical methods

A great deal of time and effort was spent in this study on the variable selection problem. Since many different variables are available to characterize different aspects of climatic variability, selecting for the H2 hypothesis was particularly difficult. Somewhat surprisingly, the existing methods that extract information from all possible combinations of variables – all subsets and hierarchical partitioning – did not, in the end, provide useful selections of variables. For this study, this may have largely derived from the fact that four different response variables were involved – the four vertebrate class richnesses – and some of these exhibit substantially different relationships to the environmental variables than do others. The principal components approach avoided this problem by simply selecting variables that best filled the data space of relevance to the vertebrate richness. Perhaps fortuitously, the same principal component data space, consisting of the first two PCs, was relevant to each vertebrate class.

Another substantial and related statistical issue was that of multicollinearity. The correlations among predictor variables was likely partially responsible for the conflicting results returned by the all-subsets and hierarchical partitioning methods. In the preliminary studies over CA and NW, multicollinearity was a common issue with the H2 models, and thus required more careful screening during backward elimination. The use of the set of variables selected by the PC method dramatically reduced these problems, greatly simplifying the construction of models for the final, eight-state analysis.

It is important to note that the preliminary work over the NW states, which allowed five somewhat collinear predictor variables for the H2 hypothesis, found H2 as the best supported hypothesis. Three adjustments to the methodology were fundamentally responsible for the emergence of H1 as the best supported hypothesis in the final, eight state analysis: combining H1a and H1b into H1 (although even for the one-variable case, H1a or H1b almost always outperform H2 or H3); limiting the number of predictor variables to two (although in the final results, H1 almost always outperforms H2 and H3 in the all-variable case); and, most significantly, the selection of an essentially uncorrelated set of predictor variables for use in the H2 hypothesis.

At the onset of this study, it was somewhat expected that the use of CAR models might rather fundamentally affect the results. This does not appear to be the case. Although CAR models are justified in this context, due to the strong autocorrelation of most of the input variables, they did not, in the end, significantly affect the results. Partly, this is because they only occasionally (approximately 10% of models) identified redundant variables in the OLS models – variables that were included only because of their spatial correlation to other variables – and so did not affect the final summaries, over the many model dimensions. For purely predictive applications, it is much preferred to use CAR models, because the spatial component of the models can contribute significantly to the model's predictive performance. However, the increase in performance is only due to the spatial component, and thus does not provide any new information about the contribution of the individual predictor variables, or, therefore, the relative performance of the different hypotheses.

One issue that was not investigated in this study is the possible effect of higher order interactions between variables on model performance. These may be important, and some studies do include interaction effects. However, including interactions increases the complexity of the modeling by an order of magnitude, and greatly exacerbates any issues with multicollinearity and variable selection. Under such circumstances, including interaction effects should only be pursued if there is strong theoretical support. For the hypotheses under study here, there is no such support. The likelihood ratio method of evaluating hypotheses does, however, allow some interactions – first order interactions between variables of the different hypotheses. The results for this method did not vary meaningfully from the results of the other hypothesis evaluation methods, so first order interactions do not appear to be a significant effect.

Finally, it is important to note that the methodology used in this study was not designed to produce the best or most accurate predictions of species richness. At an earlier stage of this work, that additional goal was pursued, and with notable success; models incorporating variables from all three hypotheses could frequently explain from 60-75% of richness, and if CAR models were used, this could reach 85%. However, a best-performing, pan-hypothesis model does not aid in resolution of the question under scrutiny here, and so these additional models were eliminated. It may, however, perhaps indicate that, as might be expected, all of these hypotheses have some, perhaps substantial, validity, for certain vertebrate groups, spatial scales, or regions. The use of path analysis or structural equation models (see next section) would be one possible way of better understanding such combined models.

6.6. Future work

The results from this study appear fairly convincing, and given the many permutations and paths this research has followed, there is good reason to believe they are fairly robust. Nevertheless, this study should only be considered as a basis for future work on the topic, as there are many areas for possible improvements or extensions.

The GAP datasets used for species richness continue to be produced, and to be refined, by the various state projects. There are currently several regional “re-GAP” projects underway, in which a group of states’ GAP analyses are regenerated, using a common methodology and producing a single dataset. For example, the five southwestern states of Arizona, New Mexico, Colorado, Utah, and Nevada are currently being redone (see http://fws-nmcfwru.nmsu.edu/SWREGAP/default.htm). Incorporating these new results will help to allay concerns about discontinuities of richness across state borders, and allow a more effective examination of the relationships across larger regions. As the results here show, Colorado was often an oddity, providing different relationships to environment, and clearly showing different levels of raw richness. Using the new data products may help provide added coherence to the results. Additionally, the original goal of this work was to evaluate these hypotheses over much of the continental United States. However, the data for most projects was not, at that time, available. Currently, most states' analyses are done, or will be in 2006. Incorporating more states to cover a larger area would be useful to better address the hypotheses directly at more continental scales. The main problem with such an expanded study will be the discontinuities across state lines, but this may be less of a concern with the various re-gap projects underway. And although these issues were substantial in the present study for Colorado, it is unknown how prevalent the issue would be between other states.

Another area for improvement is the initial selection of variables. Specifically, other measures of habitat heterogeneity might provide additional support to that hypothesis, if evaluated. Both GLCV and GLCF are fairly crude measures of heterogeneity, based upon a landcover classification. Other measures that take into consideration metrics such as patch size and shape complexity might provide a more relevant correlate of richness. Finding the best such measure and computing it over the varied states included here will be a challenge, but perhaps one worth pursuing.

The set of variables used for the H2, climatic variability, hypothesis might also be reconsidered. The candidate variables chosen from the Daymet dataset are all based on either 18-year means or standard deviations of those annual means. But, it is possible that the timescale of climatic variation that is relevant to richness processes is different.

As other studies have shown, it is often very informative to examine relationships to more specific groups of vertebrates, breaking out groups by either body size, range size, habits, or other relevant factors (Badgley and Fox 2000; Jetz and Rahbek 2002; Marquet, Fernández et al. 2004). Doing so requires, of course, more detailed information about each species, and for this reason was not attempted in this work. But if such information could be extracted from databases of species, which are becoming increasingly available (for example, see the US Geologic Survey's Integrated Taxonomic Information System, http://www.itis.usgs.gov), the results might provide useful insight into the relevance of particular hypotheses and predictor variables.

The preliminary California study suggested fundamentally different relationships between richness and environmental variables in heavily human-modified agricultural areas. For the purposes of this study, these different relationships were not very relevant, and so not pursued further. However, it would be useful to know if there is any regular relationship between richness and environment over agricultural regions, since agriculture does dominate a significant portion of the earths land surface.

One fundamental limitation to the type of correlational analysis conducted here is that it does not inform on causality. Correlation can never prove causality in this type of ‘natural experiment’, in which data manipulations are not feasible. Some work done at the plot scale has found causal connections between diversity and productivity (Tilman, Wedin et al. 1996), but such experiments are obviously not replicable on continental scales. In terms of the present study, it has sometimes been suggested that richness might be influencing either productivity (Waide, Willig et al. 1999), or habitat heterogeneity (Rosenzweig and Abramsky 1993). Recently developed methods such as path analysis and structural equation modeling do allow some inference of causality from non-experimental data (Li 1975), and they have been used in related ecological analyses in California (Williams, Seabloom et al. 2005). Using the current analysis as a traditional regression reference for a study employing structural equation modeling to examine these issue more closely could be informative.

Another method that might provide additional insight is geographically weighted regression (GWR) (Fotheringham, Brunsdon et al. 2002). This method is an alternative to spatial regression models, such as CAR, but operates on a fundamentally different basis. Whereas spatial regression models, like CAR, assume that there is a global relationship that is sought, GWR instead assumes that the relationship can vary locally; the regression coefficients are not fixed, but allowed to vary over space. This method was not used here because the objective was to evaluate hypotheses of universal extent, but it might nevertheless be useful for more fully understanding the variability of the modeled relationships.

Finally, as is clear from the literature (Waide, Willig et al. 1999; Mittelbach, Steiner et al. 2001), the relationship between productivity and richness is often hypothesized to be hump shaped, with highest richness levels found at intermediate levels of productivity. This study did not specifically examine this issue, and, by using linear models, assumed that the relationship would be linear—either positive or negative. Initial exploratory work suggested that such linear models would be sufficient, and using linear models certainly helped to maintain a consistent methodology. However, although unreported here, the data over the full eight-state region do show a possible humped pattern, although it is unclear if the hump is a true feature of the data, or simply an artifact from the much higher number of data points at intermediate productivity levels; their expected variance might be causing the slight observed hump. In any case, this issue could benefit from a closer look in future studies, and perhaps even evaluating whether a non-linear model might better characterize any actual hump-shaped relationship. If a valid humped relationship is found, the resulting nonlinear models would only increase the support for H1, and H1a in particular, and might even explain the generally inferior performance of productivity compared to environmental energy found in this study—perhaps the productivity models are being handicapped by the linear nature of the OLS models.

6.7. Conclusion

This research has followed a long and sometimes rather torturous path to come to the conclusions presented here. The difficulties with conducting this type of correlational analysis became increasingly apparent as the analysis progressed, and, if overlooked, could easily have led to spurious and non-representative results. Whether these types of issues have contributed to the divergence of opinion on this question in the published literature is unknown, but it does suggest that methodological considerations are more important than they often appear. This is particularly true as the data sources and tools for conducting these types of analyses become increasingly available.

Methodological considerations aside, the research presented here provides clear support for the contention that the energy/productivity hypotheses better explain patterns in richness, at least for vertebrates in the western United States. With that basis, there are several clear paths for future related work to follow, to either provide additional support for this conclusion, extending it to other geographic regions, or to perhaps point out its limitations or oversights. As biodiversity is under threat from human activity over most of the globe, any additional insight into the processes that drive and support it will be particularly useful, especially if and when policy decisions are based on science.

Appendix I: Maps of Variables

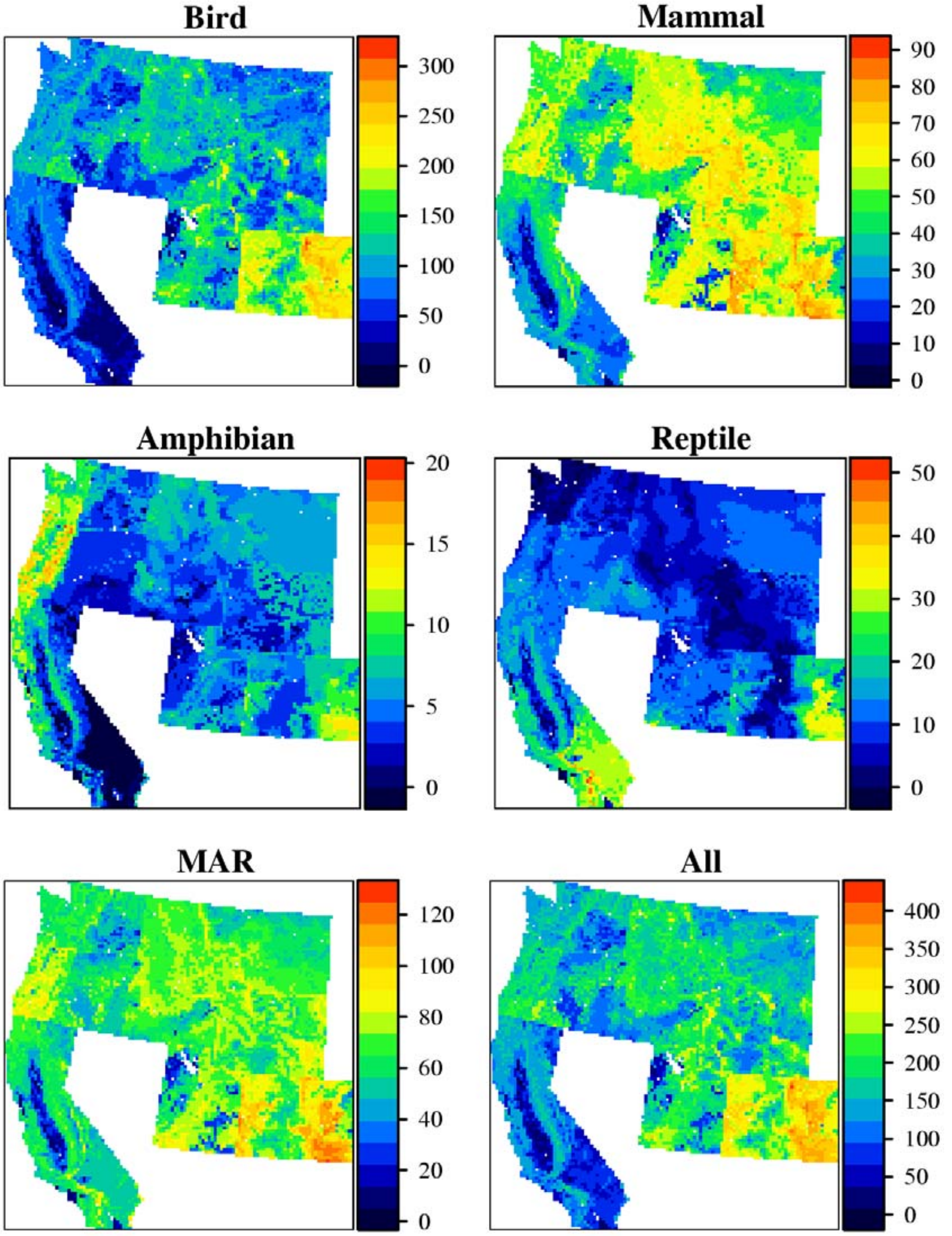


Figure 35: Species richness, 16 km. MAR = mammals, amphibians, and reptiles. All = bird + mammal + amphibian + reptile.

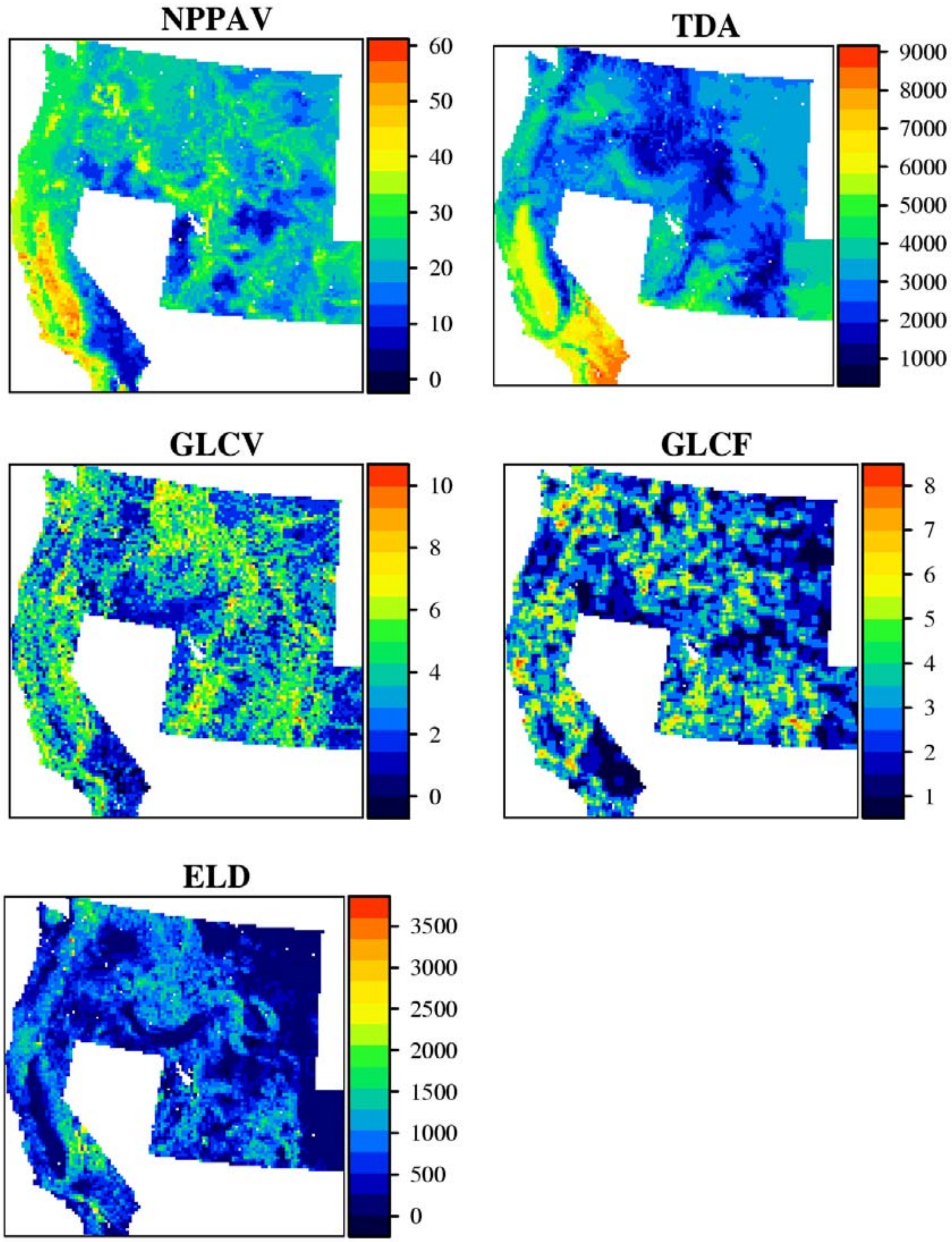


Figure 36: Variables for H1 and H3 hypotheses, 16 km. Units: NPPAV in g C/m^2/year; TDA (growing degree days) in Celsius degree days; GLCV (landcover variety, within pixel) and GLCF (landcover variety of surrounding pixels), are unitless indices, with maximum values of 11 and 8 respectively; ELD (elevation difference within cell) in meters.

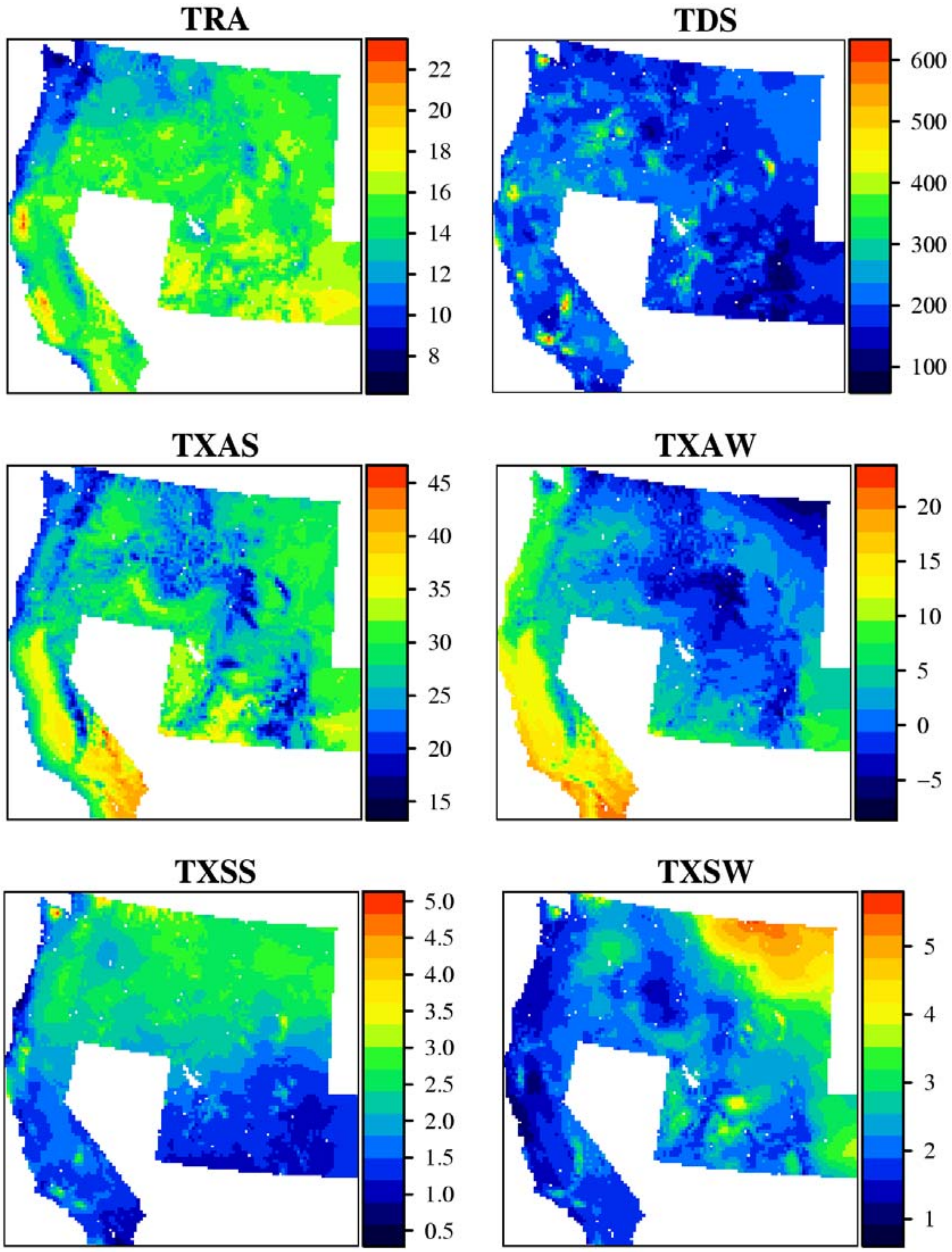


Figure 37: Temperature variables for H2 hypothesis, 16 km. Units: TRA (annual max – min daily T range), TDS (SD of annual growing degree days), TXAS (July mean T), TXAW (January mean T), TXSS (SD of July T), TXSW (SD of January T) all in °C.

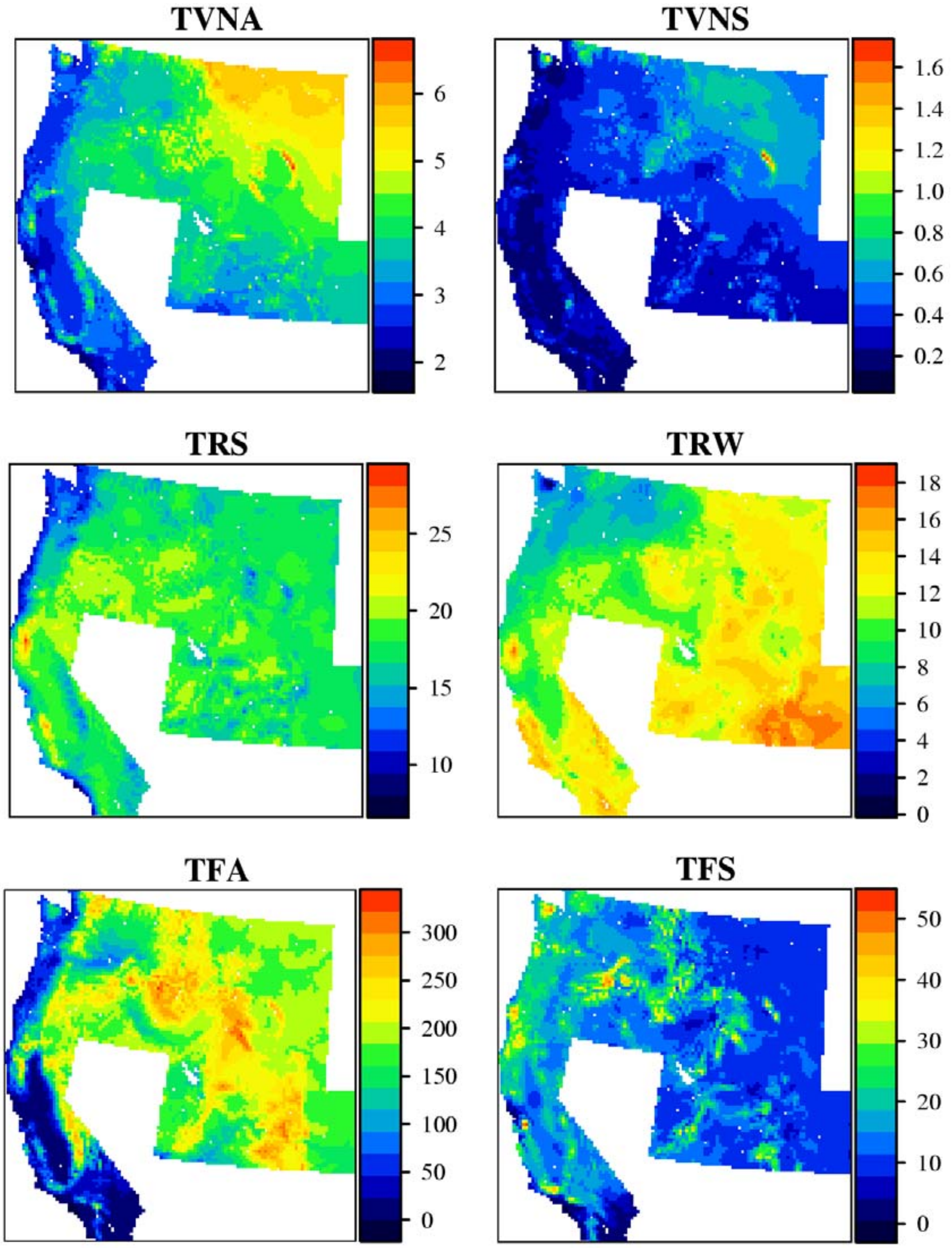
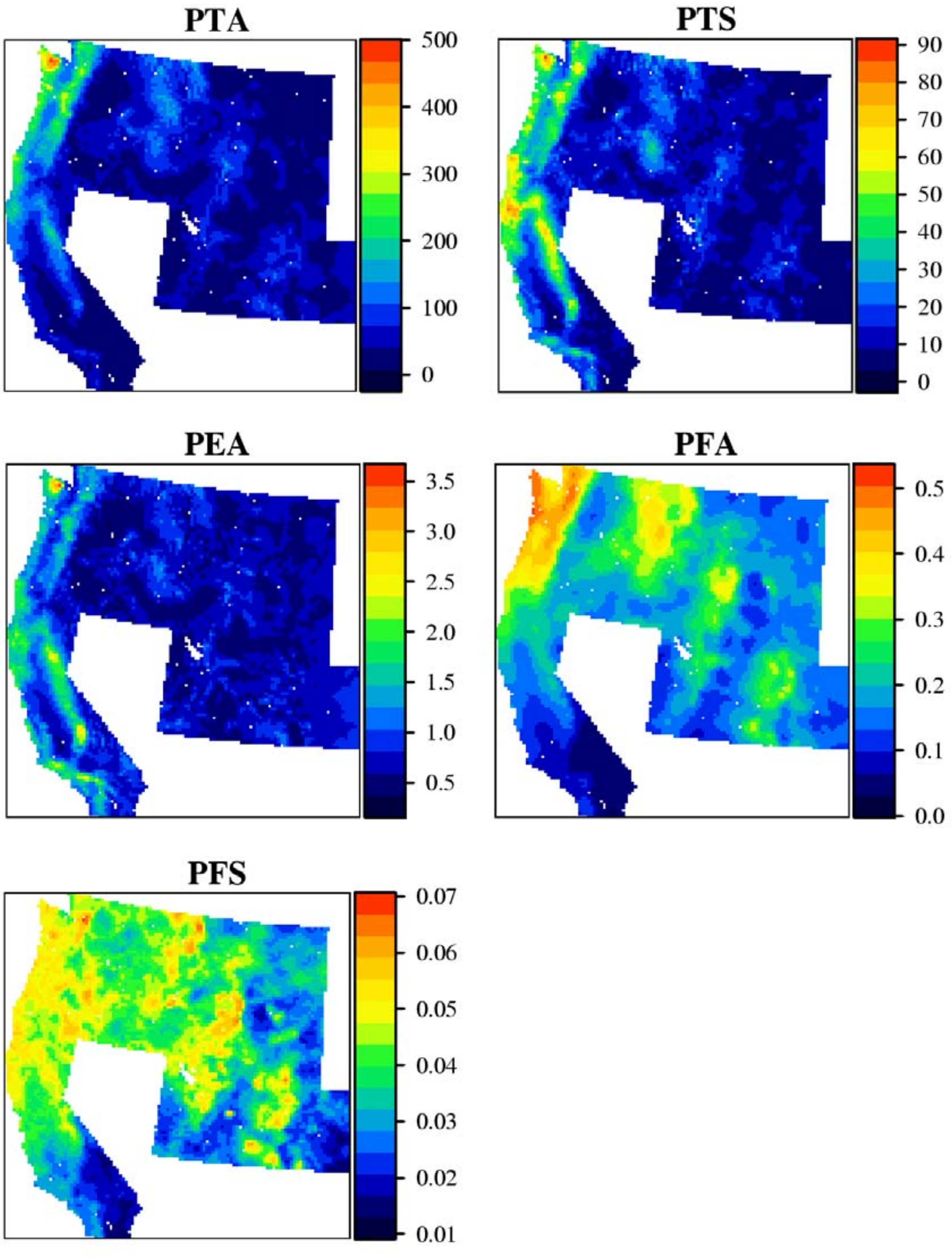


Figure 38: Temperature variables for H2 hypothesis, continued, 16 km. Units: TVNA (annual mean day-to-day T variation), TVNS (SD of day-to-day mean T variation), TRS (July max – min daily T range), TRW (January max – min daily T range) all in °C. TFA (number of days with frost), and TFS (SD of number of days with frost) in days.


PTA

PTS

PEA

PFA

PFS

Figure 39: Precipitation variables for H2 hypothesis, 16 km. Units: PTA (mean annual precipitation), PTS (SD of annual precipitation) in cm; PEA (precipitation event size) in cm/day; PFA (annual precipitation frequency) and PFS (SD of precipitation frequency) are in proportions (PFA of 0.4 means 40% of days in a year have some precipitation).

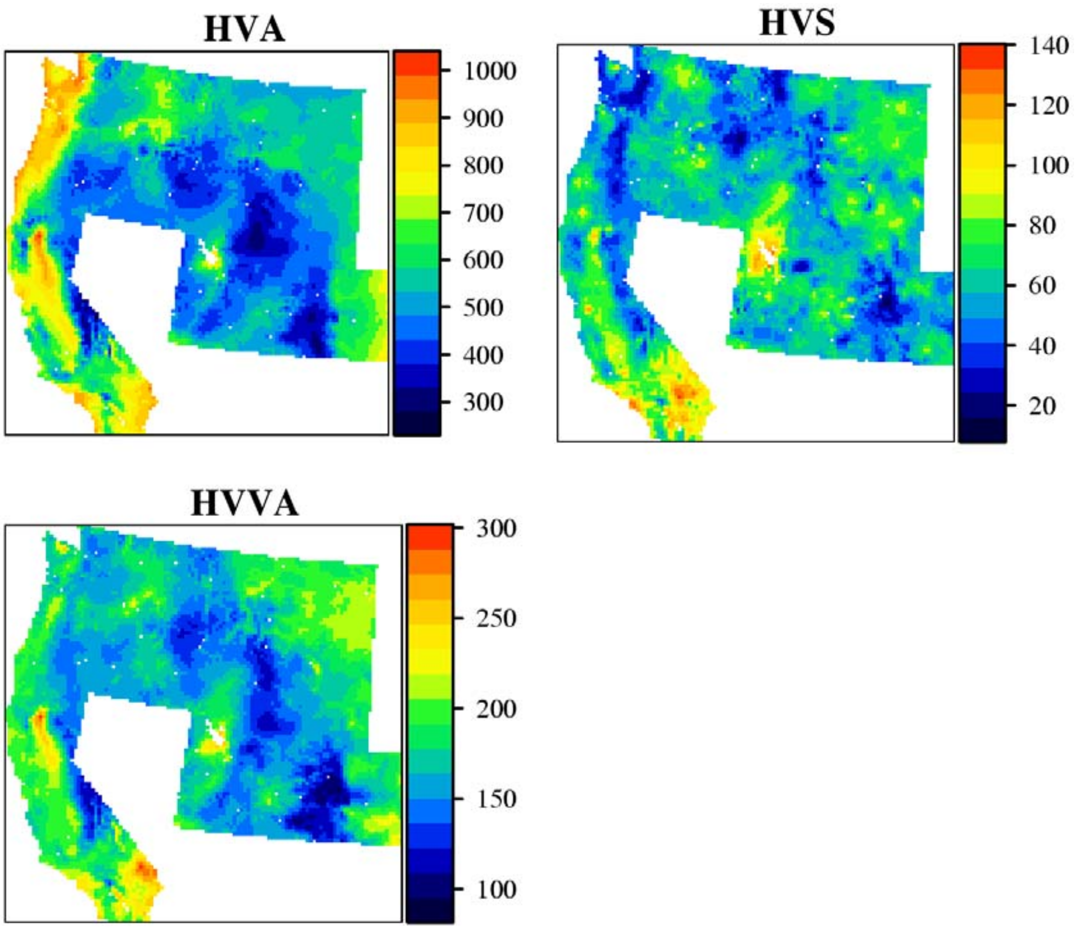


Figure 40: Humidity variables for H2 hypothesis, 16 km scale. Units: HVA (daily average water vapor pressure), HVS (SD of daily water vapor pressure), and HVVA (day-to-day variability in water vapor pressure) all in Pa.

Appendix II: Histograms of Variables

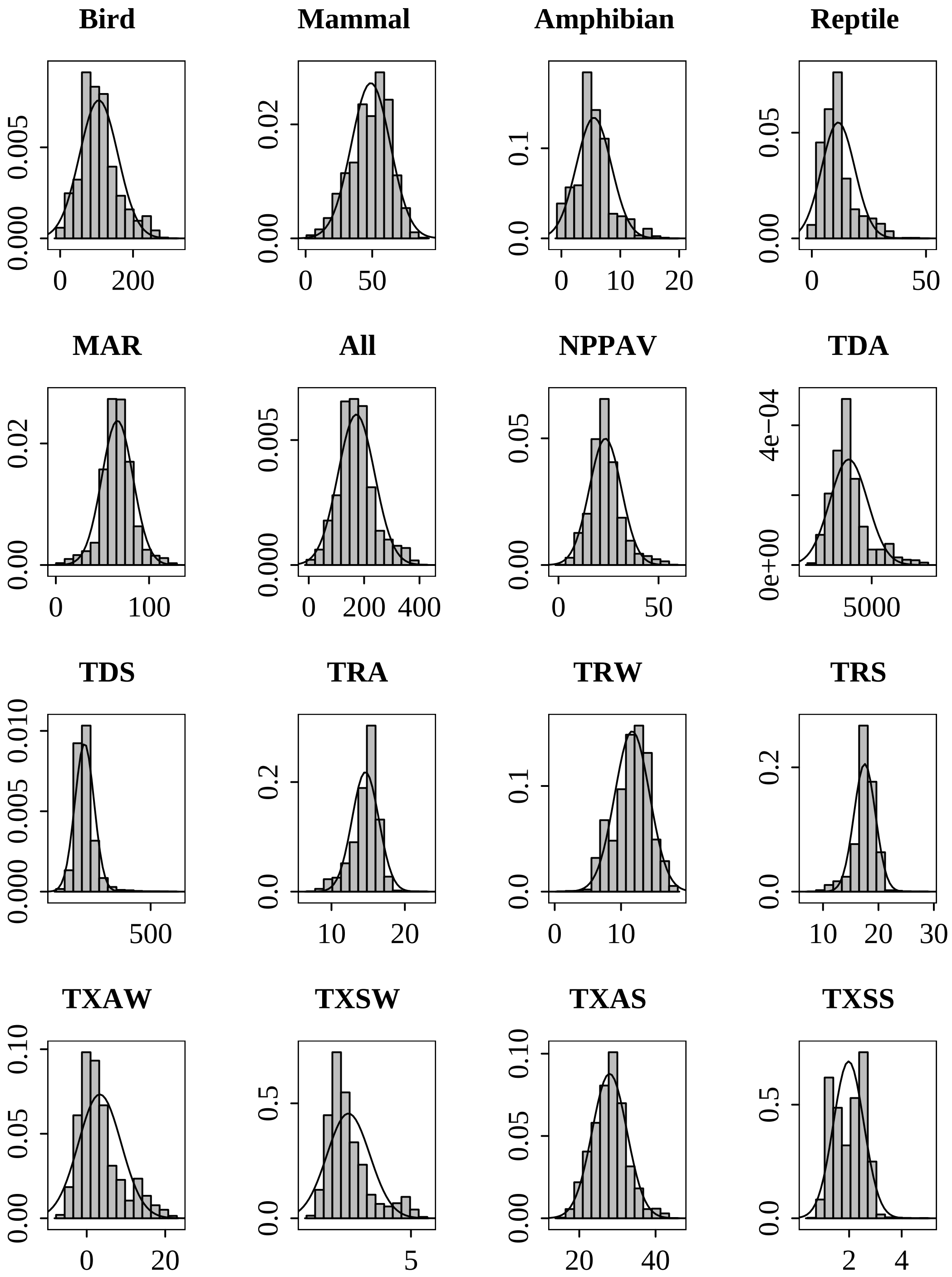


Figure 41: Histograms of response and predictor variables, part 1, 16 km. Y-axis is density (total area inside histogram is 1, in units of variable). X-axis is in units of the specific variable (Bird, Mammal, Amphibian, Reptile, MAR, All in richness; NPPAV in gC/m^2/year; TDA in degree days; all others in °C). Over-plotted line is a Gaussian curve fit to the data; significant deviations from this line indicate non-normality of the data.

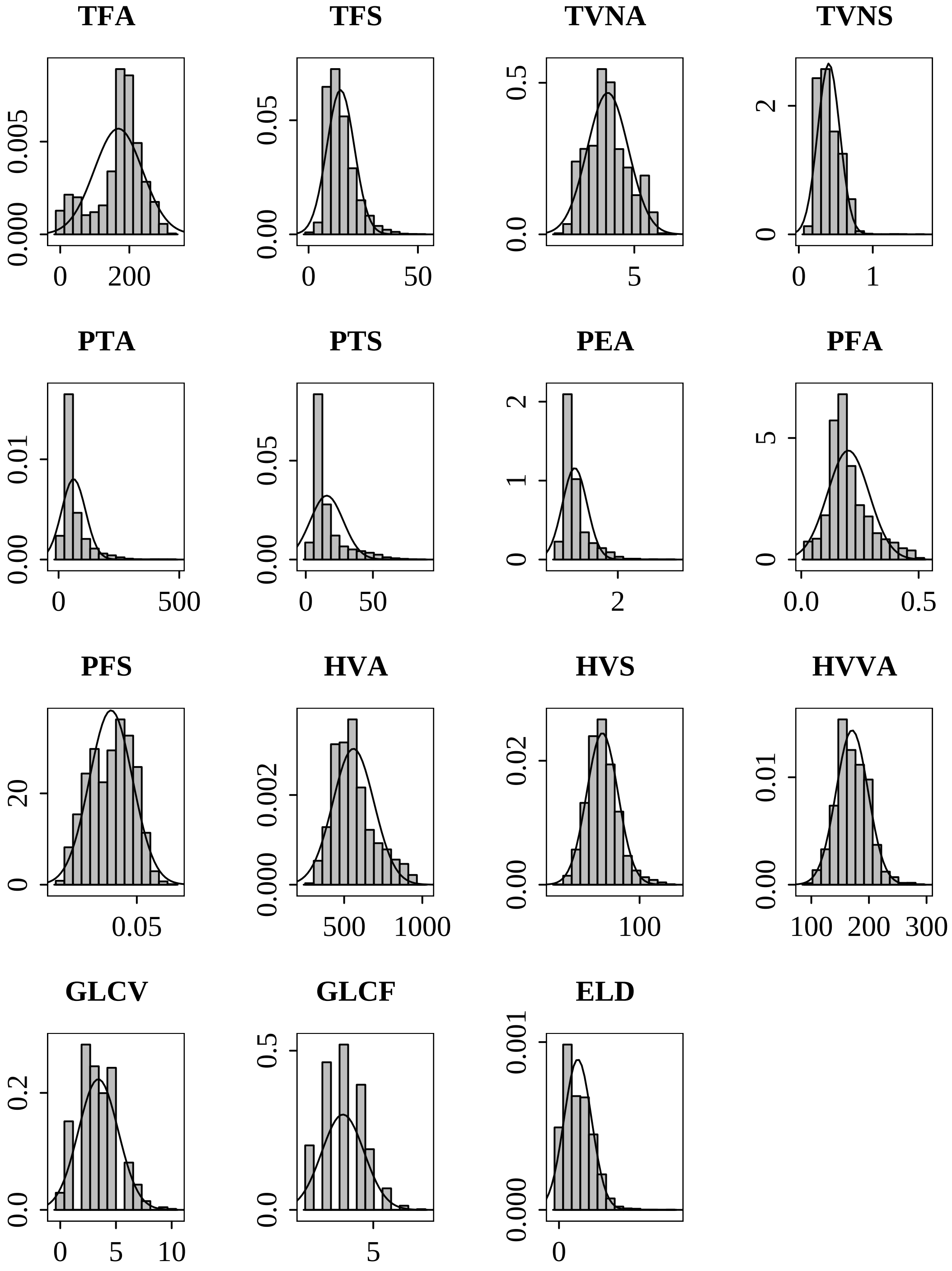


Figure 42: Histograms of response and predictor variables, part 2, 16 km. Y-axis is density (total area inside histogram is 1). X-axis is in units of the specific variable (TFA, TFS, TVNA, TVNS in °C; PTA, PTS in cm; PEA in cm/day; PFA, PFS in proportion; HVA, HVS, HVVA in Pa; GLCV, GLCF unitless; ELD in meters. Over-plotted line is a Gaussian curve fit to the data; significant deviations from this line indicate non-normality of the data.

References

Adams, J. M. and F. I. Woodward (1989). "Patterns in tree species richness as a test of the glacial extinction hypothesis." Nature **339**: 699-701.

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. Second International Symposium on Information Theory, Akademiai Kiado, Budapest.

Allen, A. P., J. H. Brown, et al. (2002). "Global biodiversity, biochemical kinetics, and the energetic-equivalence rule." Science **297**: 1545-1548.

Badgley, C. and D. L. Fox (2000). "Ecological biogeography of North American mammals: species density and ecological structure in relation to environmental gradients." Journal of Biogeography **27**: 1437-1467.

Begon, M., J. L. Harper, et al. (1990). Ecology: individuals, populations, and communities. Boston, Blackwell.

Bisby, F. A. (1995). Characterization of biodiversity. Global Biodiversity Assessment. V. H. Heywood. Cambridge, Cambridge University Press**:** 21-106.

Blackburn, T. M. and K. J. Gaston (1996). "A sideways look at patterns in species richness, or why there are so few species outside the tropics." Biodiversity letters **3**: 44-53.

Boone, R. B. and W. B. Krohn (2000). "Partitioning sources of variation in vertebrate species richness." Journal of Biogeography **27**: 457-470.

Box, G. E. P. and D. R. Cox (1964). "An analysis of transformations." Journal of the Royal Statistical Society, Series B **26**: 211-252.

Brown, J. H. and M. V. Lomolino (1998). Biogeography. Sunderland, Sinauer.

Burnham, K. P. and D. R. Anderson (2002). Model selection and multimodel inference: a practical information-theoretic approach. New York, Springer.

Cao, M., S. D. Prince, et al. (2004). "Remotely sensed interannual variations and trends in terrestrial net primary productivity 1981-2000." Ecosystems **7**: 233-242.

Carroll, S. S. and D. L. Pearson (2000). "Detecting and modeling spatial and temporal dependence in conservation biology." Conservation Biology **14**(6): 1893-1897.

Chevan, A. and M. Sutherland (1991). "Hierarchical Partitioning." The American Statistician **45**(2): 90-96.

Chown, S. L. and K. J. Gaston (2000). "Areas, cradles and museums: the latitudinal gradient in species richness." TREE **15**: 311-315.

Cliff, A. D. and J. K. Ord (1981). Spatial processes: models & applications. London, Pion Limited.

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

Colwell, R. K. and D. C. Lees (2000). "The mid-domain effect: geometric constraints on the geography of species richness." TREE **15**(2): 70-76.

Connell, J. H. and E. Orias (1964). "The ecological regulation of species diversity." American Naturalist **98**: 399-414.

Crawley, M. J. (2002). Statistical computing: an introduction to data analysis using S-Plus. Chichester, Wiley.

Cressie, N. A. C. (1993). Statistics for spatial data. New York, J. Wiley.

Csuti, B. and P. Crist (2000). Methods for developing terrestrial vertebrate distribution maps for Gap Analysis, USGS Gap Analysis Program. **2000**.

Currie, D. J. (1991). "Energy and large-scale patterns of animal- and plant-species richness." American Naturalist **137**(1): 27-49.

Currie, D. J., A. P. Francis, et al. (1999). "Some general propositions about the study of spatial patterns of species richness." Ecoscience **6**(3): 392-399.

Currie, D. J. and V. Paquin (1987). "Large-scale biogeographical patterns of species richness of trees." Nature **329**: 326-327.

Dynesius, M. and R. Jansson (2000). "Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations." Proceedings of the National Academy of Sciences USA **97**: 9115-9120.

ESRI (1982-2005). ArcInfo Workstation, version 9.1. Redlands, CA, Environmental Systems Research Institute, Inc.

Foody, G. M. (2004). "Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna." Global Ecology and Biogeography **13**: 315-320.

Fotheringham, A. S., C. Brunsdon, et al. (2002). Geographically Weighted Regression: the analysis of spatially varying relationships. Chichester, England, Wiley.

Fox, J. (1997). Applied regression analysis, linear models, and related methods. Thousand Oaks, CA, Sage.

Fox, J. (2002). An R and S-Plus companion to applied regression. Thousand Oaks, CA, Sage.

Fraser, R. H. (1998). "Vertebrate species richness at the mesoscale: relative roles of energy and heterogeneity." Global Ecology and Biogeography Letters **7**: 215-220.

Goetz, S. J., S. D. Prince, et al. (1999). "Satellite remote sensing of primary production: an improved production efficiency modeling approach." Ecological Modelling **122**: 239-255.

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

Gross, K. L., M. R. Willig, et al. (2000). "Patterns of species density and productivity at different spatial scales in herbaceous plant communities." Oikos **89**: 417-427.

H-Acevedo, D. and D. J. Currie (2003). "Does climate determine broad-scale patterns of species richness? A test of the causal link by natural experiment." Global Ecology and Biogeography **12**: 461-473.

Haining, R. P. (1990). Spatial data analysis in the social and environmental sciences. Cambridge, Cambridge University Press.

Hansen, M. C., R. S. DeFries, et al. (2000). "Global land cover classification at 1 km spatial resolution using a classification tree approach." International Journal of Remote Sensing **21**(6 & 7): 1331-1364.

Hawkins, B. A., R. Field, et al. (2003). "Energy, water, and broad-scale geographic patterns of species richness." Ecology **84**(12): 3105-3117.

Hawkins, B. A. and E. E. Porter (2001). "Area and the latitudinal diversity gradient for terrestrial birds." Ecology Letters **4**: 595-601.

Hawkins, B. A. and E. E. Porter (2003). "Relative influences of current and historical factors on mammal and bird diversity patterns in deglaciated North America." Global Ecology and Biogeography **12**: 475-481.

Heywood, V. H. and R. T. Watson, Eds. (1995). Global biodiversity assessment. Cambridge, Cambridge University Press.

Hughes, L., E. M. Cawsey, et al. (1996). "Geographic and climatic range sizes of Australian eucalypts and a test of Rapoport's rule." Global Ecology and Biogeography Letters **5**: 128-142.

Hurlbert, A. H. and J. P. Haskell (2003). "The effect of energy and seasonality on Avian species richness and community composition." American Naturalist **161**(1): 83-97.

Huston, M. A. (1994). Biological diversity: the coexistence of species on changing landscapes. Cambridge, Cambridge University Press.

Huston, M. A. (1999). "Local processes and regional patterns: appropriate scales for understanding variation in the diversity of plants and animals." Oikos **86**(3): 393-401.

Hutchinson, G. E. (1959). "Homage to Santa Rosalia or why are there so many kinds of animals?" American Naturalist **93**: 145-159.

Insightful Corporation (1988-2002). S-PLUS version 6.1 for Windows (computer software). Seattle, Washington, Insightful Corporation.

Jackson, J. E. (2003). A User's Guide to Principal Components. Hoboken, New Jersey, Wiley.

Jetz, W. and C. Rahbek (2002). "Geographic range size and determinants of Avian species richness." Science **297**: 1548-1551.

Kaluzny, S. P., S. C. Vega, et al. (1998). S+SpatialStats: user's manual for Windows and UNIX. New York, Springer.

Kareiva, P. and Anderson (1988). Spatial aspects of species interactions: the wedding of models and experiments. Community ecology. A. Hastings. Berlin, Springer-Verlag**:** 35-50.

Kerr, J. T. (1999). "Weak links: 'Rapoport's rule' and large-scale species richness patterns." Global Ecology and Biogeography **8**: 47-54.

Kerr, J. T. and D. J. Currie (1999). "The relative importance of evolutionary and environmental controls on broad-scale patterns of species richness in North America." Ecoscience **6**: 329-337.

Kerr, J. T. and L. Packer (1997). "Habitat heterogeneity as a determinant of mammal species richness in high-energy regions." Nature **385**: 252-254.

Kleidon, A. and H. A. Mooney (2000). "A global distribution of biodiversity inferred from climatic constraints: results from a process-based modelling study." Global Change Biology **6**: 507-523.

Latham, R. E. and R. E. Ricklefs (1993). "Global patterns of tree species richness in moist forests: energy-diversity theory does not account for variation in species richness." Oikos **67**(2): 325-333.

Laurance, W. F. and R. O. Bierregaard, Eds. (1997). Tropical Forest Remnants: Ecology, Management, and Conservation of Fragmented Communities. Chicago, University of Chicago.

Leakey, R. E. and R. Lewin (1995). The sixth extinction: patterns of life and the future of humankind. New York, Doubleday.

Legendre, P. and L. Legendre (1998). Numerical Ecology. Amsterdam, Elsevier.

Lennon, J. J. (2000). "Red-shifts and red herrings in geographical ecology." Ecography **23**: 101-113.

Li, C. C. (1975). Path analysis: a primer. Pacific Grove, CA, Boxwood Press.

Lichstein, J. W., T. R. Simons, et al. (2002). "Spatial autocorrelation and autoregressive models in ecology." Ecological Monographs **72**(3): 445-463.

Lumley, T. and A. Miller (2005). leaps: regression subset selection.

Lyons, S. K. and M. R. Willig (1999). "A hemispheric assessment of scale dependence in latitudinal gradients of species richness." Ecology **80**(8): 2483-2491.

MacArthur, R. H. and E. O. Wilson (1967). The theory of island biogeography. Princeton, NJ, Princeton University Press.

MacNally, R. (2002). "Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables." Biodiversity and Conservation **11**: 1397-1401.

Marquet, P. A., M. Fernández, et al. (2004). Diversity emerging: toward a deconstruction of biodiversity patterns. Frontiers of biogeography: new directions in the geography of nature. M. V. Lomolino and L. R. Heaney. Sunderland, MA, Sinauer**:** 191-209.

McGlone, M. S. (1996). "When history matters: Scale, time, climate and tree diversity." Global Ecology and Biogeography Letters **5**: 309-314.

Miller, A. J. (2002). Subset selection in regression. Boca Raton, Chapman & Hall / CRC.

Mittelbach, G. G., C. F. Steiner, et al. (2001). "What is the observed relationship between species richness and productivity?" Ecology **82**(9): 2381-2396.

Nelder, J. A. and R. W. M. Wedderburn (1972). "Generalised linear models." Journal of the Royal Statistical Society, Series A **135**: 370-384.

Noss, R. F. (1990). "Indicators for monitoring biodiversity: a hierarchical approach." Conservation Biology **4**(4): 355-364.

O'Brien, E. M. (1998). "Water-energy dynamics, climate, and prediction of woody plant species richness: an interim general model." Journal of Biogeography **25**: 379-398.

O'Brien, E. M. (2000). "Climatic gradients in woody plant (tree and shrub) diversity: water-energy dynamics, residual variation, and topography." Oikos **89**(3): 588-600.

O'Connor, R. J., M. T. Jones, et al. (1996). "Spatial partitioning of environmental correlates of avian biodiversity in the conterminous United States." Biodiversity Letters **3**: 97-110.

Palmer, M. W. (1994). "Variation in species richness: toward a unification of hypotheses." Folio Geobotanica & Phytotaxonomica **29**: 511-530.

Pianka, E. R. (1966). "Latitudinal gradients in species diversity: a review of concepts." American Naturalist **100**(910): 33-46.

Purvis, A. and A. Hector (2000). "Getting the measure of biodiversity." Nature **405**: 212-219.

Quinn, J. F. and S. P. Harrison (1988). "Effects of habitat fragmentation and isolation on species richness: evidence from biogegraphic patterns." Oecologia **75**: 132-140.

R Development Core Team (2005). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.

Rahbek, C. (1997). "The relationship among area, elevation, and regional species richness in neotropical birds." American Naturalist **149**: 875-902.

Rahbek, C. and G. R. Graves (2001). "Multiscale assessment of patterns of avian species richness." Proceedings of the National Academy of Sciences **98**(8): 4534-4539.

Richerson, P. J. and K.-L. Lum (1980). "Patterns of plant species diversity in California: relation to weather and topography." American Naturalist **116**(4): 504-536.

Ricklefs, R. E., R. E. Latham, et al. (1999). "Global patterns of tree species richness in moist forests: distinguishing ecological influences and historical contingency." Oikos **86**(2): 369-373.

Ricklefs, R. E. and D. Schluter, Eds. (1993). Species diversity in ecological communities: historical and geographical perspectives. Chicago, University of Chicago Press.

Ripley, B. D. (1981). Spatial statistics. Hoboken, NJ, Wiley-Interscience.

Rohde, K. (1992). "Latitudinal gradients in species diversity: the search for the primary cause." Oikos **65**(3): 514-527.

Rohde, K. (1998). "Latitudinal gradients in species diversity. Area matters, but how much?" Oikos **82**(1): 184-190.

Rosenzweig, M. L. (1995). Species diversity in space and time. Cambridge, Cambridge University Press.

Rosenzweig, M. L. and Z. Abramsky (1993). How are diversity and productivity related? Species Diversity in Ecological Communities: Historical and Geographical Perspectives. R. E. Ricklefs and D. Schluter. Chicago, University of Chicago**:** 52-65.

Rosenzweig, M. L. and E. A. Sandlin (1997). "Species diversity and latitudes: listening to area's signal." Oikos **80**(1): 172-176.

RSI (2005). IDL. Boulder, CO, Research Systems, Inc.

Ruggiero, A. and T. Kitzberger (2004). "Environmental correlates of mammal species richness in South America: effects of spatial structure, taxonomy and geographic range." Ecography **27**: 401-416.

Scheiner, S. M., S. B. Cox, et al. (2000). "Species richness, species-area curves and Simpson's paradox." Evolutionary Ecology Research **2**: 791-802.

Scott, J. M., F. Davis, et al. (1993). "Gap analysis: a geographical approach to protection of biological diversity." Wildlife Monographs **123**: 1-41.

Simpson, G. G. (1964). "Species density of North American recent mammals." Systematic Zoology **13**(2): 57-73.

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

Stoms, D. M. (1994). "Scale dependence of species richness maps." Professional Geographer **46**(3): 346-358.

Taylor, P. H. and S. D. Gaines (1999). "Can Rapoport's Rule be rescued? Modeling causes of the latitudinal gradient in species richness." Ecology **80**: 2474-2482.

Thornton, P. E., H. Hasenauer, et al. (2000). "Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria." Agricultural and Forest Meteorology **104**: 255-271.

Thornton, P. E., S. W. Running, et al. (1997). "Generating surfaces of daily meteorological variables over large regions of complex terrain." Journal of Hydrology **190**: 214-251.

Tilman, D. (1999). "The ecological consequences of changes in biodiversity: a search for general principles." Ecology **80**: 1455-1474.

Tilman, D., D. Wedin, et al. (1996). "Productivity and sustainability influenced by biodiversity in grassland ecosystems." Nature **379**: 718-720.

Tognelli, M. F. and D. A. Kelt (2004). "Analysis of determinants of mammalian species richness in South America using spatial autoregressive models." Ecography **27**: 427-436.

Turner, J. R. G. (2004). "Explaining the global biodiversity gradient: energy, area, history and natural selection." Basic and Applied Ecology **5**: 435-448.

Turner, J. R. G., C. M. Gatehouse, et al. (1987). "Does solar energy control organic diversity? Butterflies, moths and the British climate." Oikos **48**(2): 195-205.

USGS EROS Data Center (2000). HYDRO1k Elevation Derivative Database, LP DAAC. Accessed online 28-Mar-2001 via http://edc.usgs.gov/products/elevation/gtopo30/hydro/namerica.html

Venables, W. N. and B. D. Ripley (1999). Modern applied statistics with S-PLUS. New York, Springer.

Waide, R. B., M. R. Willig, et al. (1999). "The relationship between productivity and species richness." Annual Reviews of Ecology and Systematics **30**: 257-300.

Walsh, C. and R. MacNally (2005). hier.part: Hierarchical Partitioning.

Whittaker, R. J., K. J. Willis, et al. (2001). "Scale and species richness: towards a general, hierarchical theory of species diversity." Journal of Biogeography **28**: 453-470.

Williams, J. W., E. W. Seabloom, et al. (2005). "Anthropogenic impacts upon plant species richness and net primary productivity in California." Ecology Letters **8**(2): 127-137.

Wright, D. H. (1983). "Species-energy theory: an extension of species-area theory." Oikos **41**: 496-506.

Wright, D. H., D. J. Currie, et al. (1993). Energy supply and patterns of species richness on local and regional scales. Species diversity in ecological communities: historical and geographical perspectives. R. E. Ricklefs and D. Schluter. Chicago, University of Chicago**:** 66-74.