AN ABSTRACT OF THE DISSERTATION OF

Mary E. Cablk for the degree of Doctor of Philosophy in Forest Resources presented on *June 13, 1997*. Title: Modeling Vertebrate Diversity in Oregon Using Satellite Imagery.

Signature redacted for privacy.

Abstract approved: _____
William J. Ripple

Vertebrate diversity was modeled for the state of Oregon using a parametric approach to regression tree analysis. This exploratory data analysis effectively modeled the non-linear relationships between vertebrate richness and phenology, terrain, and climate. Phenology was derived from time-series NOAA-AVHRR satellite imagery for the year 1992 using two methods: principal component analysis and derivation of EROS data center greenness metrics. These two measures of spatial and temporal vegetation condition incorporated the critical temporal element in this analysis. The first three principal components were shown to contain spatial and temporal information about the landscape and discriminated phenologically distinct regions in Oregon. Principal components 2 and 3, 6 greenness metrics, elevation, slope, aspect, annual precipitation, and annual seasonal temperature difference were investigated as correlates to amphibians, birds, all vertebrates, reptiles, and mammals. Variation explained for each regression tree by taxa were: amphibians (91%), birds (67%), all vertebrates (66%) reptiles (57%), and mammals (55%). Spatial statistics were used to quantify the pattern of each taxa and assess validity of resulting predictions from regression tree models. Regression tree analysis was relatively robust against spatial autocorrelation in the response data and graphical results indicated models were well fit to the data.

Modeling Vertebrate Diversity in Oregon Using Satellite Imagery

by

Mary E. Cablk

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed *June 13, 1997*
Commencement *June 1998*

ACKNOWLEDGEMENTS

Due to the scope and breadth of this research, I have many acknowledgments. There are many more people I would like to recognize but space is limited so I will name the key players formally and state that there are others whose contributions do not go unrecognized. I sincerely thank the Biodiversity Research Consortium for supporting this research, especially Mary Santelman who freed funds for its completion. Tom Loveland and EROS Data Center provided funding and data. Bradley Reed at EROS Data Center provided digital data and technical advice. Jeannie Sifneos and Denis White made it possible to transfer this work to paper and their other support is greatly appreciated. Denis White made the regression trees and maps not only possible, but in color mapped to the spectrum! He also contributed valuable insight into the analyses. Lisa Ganio and Manuela Huso contributed statistical assistance. Chaur-Fong Chen offered thoughtful and valuable insight into orthogonal transformations and life in general. Jennifer Dwyer and Mary Rasmussen were supportive friends and officemates. Suzanne Remillard, Irene Dumkow, and Vince Gonor had a direct effect on the completion of this research and their outstanding belaying is much appreciated. Dick Waring taught me much about the ecology of the Pacific Northwest. I thank Susan Stafford, who helped me to perservere and was a wonderful mentor and my family who were always supportive. I thank my advisor Bill Ripple and the efforts and support of my committee. Most importantly, I offer a sincere thank you to Ross Kiester for his advice and humor throughout this entire project, ranging from intellectual discourse on cellular automata to Madonna's role in *Evita.*

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

DEDICATION

This work is dedicated to my grandparents who immigrated from Czechoslovakia. I also dedicate this to my grandfather's father who was tired of shining his familys' shoes. He ran away, became a cabin boy, and spent his fifteenth birthday off the coast of Russia shining the Captain's shoes.

Modeling Vertebrate Diversity in Oregon Using Satellite Imagery

1.0 Chapter 1: Introduction

We are faced with a global biodiversity crisis in that species are disappearing from the face of the earth at an alarming rate. While there is consensus among the scientific community of the intrinsic value of biodiversity, there is no agreement on the appropriate scale at which to measure diversity. What is clear is that if we are to protect the remaining biotic diversity, or at least slow the rate at which species go extinct, we must develop an understanding of the mechanisms which shape and drive the patterns of diversity on the landscape. There is neither enough time nor adequate resources to concentrate exclusively on individual species. To effectively quantify the processes which may be related to biodiversity, more specifically to vertebrate species richness, efforts must concentrate on larger groupings of individuals, such as entire taxa. Mechanisms which affect entire taxa occur over broad scales and mechanistic variations may be detected only over long time periods. The research presented here quantified the relationships between vertebrate species richness and potential correlative mechanisms over a broad spatial scale and included the critical temporal axis of measurement. My study specifically addressed the question, *what controls the number of species in a given area?* for the state of Oregon.

This research was part of a pilot study conducted by the Biodiversity Research Consortium (BRC), a consortium of researchers from universities and government agencies nationwide. There were three parts to this BRC pilot project: the first was to evaluate biodiversity using landscape indices, the second was to investigate biodiversity in the wake of anthropogenic stresses, and the third was to quantify patterns of biodiversity using indices derived directly from satellite imagery. My dissertation research was the third part of the pilot project which investigated the use of satellite imagery for biodiversity assessment.

How can we quantify the relationship between vertebrate species richness and potential mechanistic processes? One method is to categorize the landscape into landcover classes where each class represents either a unique cover type or unique sets of criteria. Species richness can then be correlated with either occurrence in cover types or

correlated with some index of spatial configuration or composition. When a significant relationship is found, the interpretation can be relatively straightforward and discussed in terms of the criteria which defined the landscape classes. While this type of analysis may produce straightforward results, it is inherently limited for a number of reasons.

The most obvious and important limitation is that the results from this type of analysis depend directly on how the landscape was categorized. For example, the relationship between diversity and a landscape comprised of 30 categories will be different from the relationship between diversity and the same landscape classified into 10 categories. Differences will exist regardless of whether or not the 30 categories are telescopically regrouped into 10 classes. These differences exist because of the way landscape indices are calculated. Indices such as contagion, dominance, and diversity are directly related to the number of landscape categories. These indices reflect and quantify a landscape based on the heterogeneity and pattern of that landscape. This is the reason why a 30 class landscape will have a different relationship with diversity than a 10 category landscape. The numeric indices calculated on those landscapes will be different. No two independent expert assessments will interpret a given landscape in exactly the same fashion, because there is no universally accepted standard or set of rules which define landcover types. Even given a set of criteria, there will exist variability in landcover interpretation due to human based differences. The second difficulty with using landcover types as a basis for quantifying diversity is the assumption that fauna interpret or view the landscape in the same way as humans. Do fauna view, for example, a sagebrush landscape differently from a sagebrush/juniper landscape? How do we determine the different cover types fauna distinguish, whether all fauna respond to all landscapes in similar or different manners, or exactly what attributes differentiate the land into cover types? Finally, landcover types typically do not contain temporal criteria. There are no descriptions, quantitative or qualitative, of the temporal dynamics of the landscape despite temporal variations which are obvious in many landscapes.

Based on these considerations, three questions were addressed: 1) can temporal variation of a landscape be measured with satellite imagery to cover a very large area; 2) if so, can this temporal variability be quantified parametrically rather than as categorical; and 3) are temporal variations of the landscape, or temporal dynamics, significantly

related to vertebrate species richness? This research involved two main components to address these three questions. The first component addressed questions (1) and (2) above, which involved quantifying the temporal changes in vegetation over the course of one year with time series satellite imagery using non-classificatory methods. The second component of this research used indices of spatial and temporal vegetation dynamics (derived from the first component of this work) to quantify the pattern of vertebrate species richness in Oregon. The indices of spatial and temporal vegetation dynamics, or vegetation phenology, were modeled with climate and elevation data to quantify the relationship of vertebrate species richness with environmental variability in space and time.

In this study I addressed the question, *what controls the number of species in Oregon,* by taking a different approach from standard statistical methods. Once vegetation phenology was derived from time series satellite imagery as a continuous measure, the second step was to quantify the relationships between this environmental heterogeneity and vertebrate taxa. Rather than use classical statistics to force the data into a pre-defined relationship, I used an exploratory data analysis method to quantify the hierarchical structure inherent in the ecological data and to graphically visualize the relationships (which were potentially complex) between pattern and process. To do this, I used what I termed a despatialization-respatialization framework, where the spatial data were quantified using spatial statistics, then despatialized for analysis with the results respatialized for interpretation.

The results of this study, using the despatialization-respatialization method, indicated that there were significant relationships between diversity and environmental heterogeneity in space and time. Measures of vegetation phenology, which represented temporal dynamics of the landscape, were important processes or mechanisms and were highly significant for all vertebrate taxa in Oregon. Climate and elevation measures were also significant for explaining patterns of vertebrate diversity. Each vertebrate taxa (mammals, reptiles, birds, amphibians, and all vertebrates combined) had different sets of mechanisms or processes which explained the respective richness patterns on the landscape. This research quantified the relationship between biodiversity and landscape variability. The approach of this study, which incorporated a parametric analysis with

exploratory methods, was shown to be an alternative to traditional analyses involving one or few species and atemporal categorical data. While this research was not intended to provide the definitive answer to biodiversity issues, it did result in a comprehensive view of possible pattern-process relationships at a broad scale and resulted in generation of additional hypotheses which may be further tested.

2.0 Chapter 2: Ecological utility of time series NOAA-AVHRR satellite imagery

Principal component analysis (PCA) was calculated on 21 biweekly composites of NOAA-AVHRR 1km^2 pixel data for the state of Oregon for the years 1991-1993 to capture the greatest spatial and temporal variability in the 21 scene subset. The objectives were: 1) to determine interannual relative consistency of the normalized difference vegetation index (NDVI) biweekly composite data, and 2) to derive phenological information for the state of Oregon for incorporation into future analyses. For the years 1991, 1992, and 1993, the first principal component (PC1) explained 83%, 87%, and 74%, respectively, of the intra-annual variance in NDVI. PC1 showed the spatial representation of the data over the landscape, including changes in solar azimuth illumination over the course of a year. For these same years principal component 2 (PC2) explained 6%, 5%, and 10%, respectively, of the variance in NDVI and indicated a strong relationship with annual cycles of greenup and senescence. Principal component 3 (PC3) explained 4%, 3%, and 2%, of the intra-annual variance in NDVI for 1991, 1992, and 1993, respectively. PC3 indicated a trend temporally shifted from PC2 such that the peak positive contribution was from winter months indicating a relationship with non-growing season baseline photosynthesis. These results were consistent with previous research for other continents. Relative information content was consistent among years and confidence in the data for ecological application was high.

2.1 INTRODUCTION

There are many ways to measure biophysical properties of Earth. Over large landscapes such as the state of Oregon, remote sensing is an efficient means of collecting continuous measurements of these properties. Transformed digital reflectance is used to calculate vegetation indices such as the NDVI (see page 8), from which biophysical properties such as plant productivity can be estimated. These estimates, with ancillary or additional data, can be used further to characterize the landscape using either a parametric or categorical approach. With parametric methods there is little or no subjectivity in analyses and multiple dimensions in space and time are easily represented. Classification

methods compress multiple criteria into discrete categories but typically these criteria are only spatial, not temporal. Temporal variability exists for all landscapes and may be an important factor, but is difficult to effectively quantify.

Both categorical and parametric methods are suitable for digital image and spatial analyses, but have different scopes of application. To compare analyses of multiple datasets derived from either a parametric or categorical approach, there must be some measure of confidence in the level of comparability. The ecological utility of satellite imagery is, therefore, the consistency, accuracy, and comparability of biophysical variables derived from remotely sensed data. Utility in this sense, is empirically based.

In this research, I wanted to quantify the spatial and temporal dynamics of the Oregon landscape. These dynamic processes, which represent environmental heterogeneity, were quantified for incorporation into a larger biodiversity assessment. The objective of this initial study was to capture environmental heterogeneity in Oregon using non-classificatory methods. Taking a parametric approach, I derived annual vegetation phenology from time series biweekly composited National Oceanic Atmospheric Administration (NOAA)-Advanced Very High Resolution Radiometer (AVHRR) data for the state of Oregon. This approach was an alternative to standard landcover classification with the purpose of capturing temporal events and variability in vegetation reflected in measures of the NDVI. The objectives of this paper were 1) to derive phenology as a continuous measurement for the state of Oregon; and 2) to evaluate the ecological utility of biweekly composited imagery for the years 1991, 1992, and 1993, where utility was defined as consistency, accuracy, and comparability among the three years.

2.1.1 *Two approaches to describe a landscape: continuous versus categorical*

Single date satellite imagery can be classified into meaningful categories. These categories or thematic classes are based on a set of criteria which are often point-in-time parameter measurements. The rationale is that there are more uses for data organized into meaningful categories than there are uses for unprocessed raw data. Criteria for forest categories, for example, may include species type, age class, tree size, slope, and/or aspect. Raw imagery may be classified directly or may be pre-processed, such as

conducting a tasselled cap transformation to obtain indirect measures of landscape features such as forest maturity or structure (Cohen and Spies, 1992). Classification may also involve analyses of multi-date imagery in conjunction with ancillary data such as the Conterminous U.S. and global land-cover characteristics databases developed by the U.S. Geological Survey's (USGS) EROS Data Center (EDC) (Brown et al., 1993; Loveland et al., 1991). These two databases were created from multisource data such as maximum value biweekly composited imagery (MVC), climate, and terrain data.

While a classified dataset may be useful, relatively straightforward, and compact, some applications may be better suited for continuous measurements of biophysical parameters. Maximum value composited imagery, for example, has been used for time series analyses in Africa (Eastman and Fulk, 1993), global land monitoring (Gutman and Ignatov, 1995), and in potential fire danger/behavior systems (Burgan and Hartford, 1993). MVC imagery offers one source of data which can be analyzed to produce relatively continuous measures of biophysical phenomenon on the ground. The question remains, however, how consistent, accurate, and comparable results from these same analyses are for different datasets collected at different times. This is a concern for multi-date or multi-year comparisons, such as long term ecological assessments of vegetation phenology. It is important to have confidence in the ecological utility of the data.

2.1.2 *Rationale for PCA-derived phenology to capture temporal dynamics*

There are alternatives to classification for quantifying the temporal variability of a landscape. With the objective of eliminating subjective interpretation in describing the landscape, I used methods involving continuous variable measurements such as principal component analysis. Principal component analysis reduces data dimensionality by maximizing variation in successive independent orthogonal bands. The result is a linear combination where the original dataset is rotated against new axes to maximize variation. Each successive orthogonal axis explains variability in a dataset independent of all other axes. Principal component analysis using NOAA-AVHRR data has been shown to be an effective method for time series analysis (Eastman and Fulk, 1993), data compression (Fung and LeDrew, 1987), and classification (Townshend et al., 1987) at broad (i.e.

continental) scales. The NOAA-11 AVHRR is a polar orbiting satellite which provides daily global coverage resampled to 1.1 km^2 pixels (NOAA, 1990). The AVHRR sensor measures reflected radiation in the red visible and near-infrared, and emitted radiation in the thermal infrared parts of the spectrum. The normalized difference vegetation index (NDVI) is easily calculated from reflected visible and near-infrared wavelengths. The equation for NDVI is

$$\frac{(NIR - red)}{(NIR + red)}$$

where NIR is the measured reflectance in the near-infrared wavelength and red is the reflectance in the red wavelength. For the NOAA-11 AVHRR the corresponding wavelengths are (0.72 - 1.1 μm) and (0.58 - 0.68 μm), respectively (Lillesand and Kiefer, 1994). NDVI has been correlated with green leaf biomass, net primary productivity, and vegetation conditions (Di and Hastings, 1995; Townshend et al., 1987; Goward et al., 1985). The NDVI can be thought of as a measure of the amount of green, potentially photosynthesizing vegetation. When NDVI is calculated over time, it is a measure of the change in vegetation condition based on reflectance properties related to photosynthetic capacity. Healthy vegetation will generate a stronger signal and consequently will have a higher value of NDVI than the same vegetation type under stress. The NDVI calculated from satellite imagery is also a spatial measure having constant location. Therefore the use of NDVI from satellite imagery is a direct measure of the spatial and temporal variability of the vegetation or the larger landscape.

This combination of spectral and spatial resolution with high temporal frequency makes the AVHRR sensor ideal for time-series analyses of terrestrial vegetation at broad scales. There are, however, some considerations for using an entire annual dataset inclusive over one year. Three hundred sixty-five images may be available, but not all will be cloud free. Weather interference will vary by geographical region. In the Pacific Northwest it is difficult to acquire cloud-free scenes from November to March because most of the annual precipitation from this region falls during the winter and spring months. Gaps or unequal sampling intervals due to atmospheric interference (i.e. clouds) must be accounted for and this is likely inconsistent within and among years. Three hundred sixty-five images contain a tremendous amount of data to manage, analyze, and

store. Physical system requirements may inhibit functionality of databases, especially if large subsets or entire AVHRR scenes are used. Finally, at such a coarse spatial scale, questions arise regarding temporal sampling frequency. One must decide whether changes in NDVI will vary enough to merit daily sampling or conversely, given availability of cloud-free scenes, if the phenomenon of interest was captured in the data.

2.1.3 *Pros and cons of maximum value biweekly composites*

With the potential for enormous data volume from AVHRR daily satellite overpasses, a more compact, usable dataset must be selected which will effectively capture landscape temporal dynamics without including extraneous noise (i.e. clouds, atmosphere, etc.). This means that to accurately assess variability in vegetation over time, recorded reflectance must be from vegetation or ground conditions with minimal interference from external sources. EDC adopted a method for compositing daily AVHRR scenes into biweekly composites based on the MVC technique to provide a dataset relatively free of cloud contamination and atmospheric interference (Holben, 1986). NDVI is calculated and evaluated on a per pixel basis over 14 day time intervals for the course of one year. The scene containing the highest calculated NDVI value is selected and contributes corresponding pixels for all other spectral bands to create a biweekly composited image. The argument for the maximum value NDVI algorithm is that by selecting the maximum NDVI value, vegetation is more likely selected over haze or clouds because maximum NDVI translates into more reflective vegetation (Stoms et al., in press; Davis and Stoms, 1994; Townshend et al., 1987). The MVC criteria also has advantages related to reflective properties of the earth in certain wavelengths. For Lambertian targets, reflected radiance is proportional to the cosine of Sun elevation. Sun elevation differences are reduced by using NDVI because ratios cancel out these and similar topographic effects (Townshend et al., 1987).

Biweekly compositing has been critically reviewed based on a number of concerns such as sensor noise, bi-directional effects, atmospheric haze, and the compositing algorithm itself. Stoms et al. (in press) determined that the MVC algorithm selected pixels with high off-nadir view angles and cited regional inconsistencies in zenith angle. Cihlar

et al. (1994) tested alternative compositing methods to the MVC and reported similar results. As a result of increasing off-nadir view angles, NDVI values were found to be artificially inflated. Eklundh (1995) on the other hand, found that estimated noise was decreased by the smoothing effect of larger footprints. However, it was also suggested that individual composited images contained enough noise to warrant questionable utility for per pixel comparisons. EROS Data Center has since imposed a look angle threshold of 42° to eliminate scenes with most extreme look angles from the compositing process

2.1.4 *Ecological Considerations for Using Biweekly Composites*

Biweekly composited AVHRR imagery provides coverage over large areas with a high temporal frequency relative to other sensors. With the intent of using biweekly composited imagery for capturing temporal landscape dynamics, certain considerations had to be addressed to ensure the phenomenon of interest (vegetation phenology) was detectable within the collection time frame and that the integrity of the phenomenon was preserved in the compositing process. Theoretically in a biweekly composited image, two adjacent pixels can be selected from scenes acquired 14 days apart. This may or may not affect time-series analyses based on vegetation phenology. Fourteen days is a short amount of time relative to the course of a year, but the months of December and January were composites of 31 dates of imagery. As these were winter months it could be hypothesized that longer compositing periods were acceptable because little occurred phenologically and more dates were needed to acquire cloud-free pixels. During the growing season, however, temporal consideration was critical. The selection of pixels from multiple dates at time period extremes could have resulted in a spatially heterogeneous landscape that was somewhat artificial due to temporal sampling. The landscape would have been artificial not because of inaccurate pixel values, but because the heterogeneity resulted directly from the temporal sampling interval. Likewise, the phenomenon captured by NDVI may have been the result of different phenological stage, vegetation type, or a combination of both stage and type. Although corresponding image dates and viewing geometry are available on a per pixel basis for each composited image,

it is difficult to quantify cumulative effects of the compositing process itself within an ecological context and to understand these effects spatially.

Environmental heterogeneity in this study was measured in both time and space. Given this, there may have been implications for these time-series analyses which involved a spatial component, such as aggregation of data across scales. Where analyses involve two different spatial scales, such as pixel data at 1 km^2 and 646 km^2 hexagons, outliers or high heterogeneity of pixel values may have affected statistics or landscape indices calculated within a hexagon. The results of this aggregation will then affect analyses involving both temporal and spatial variability on the landscape.

Stoms et al. (in press) and Cihlar et al. (1994) evaluated alternative compositing strategies and compared these alternatives to MVC. In both studies alternative compositing algorithms based on temperature and/or zenith angle were compared to the NDVI based MVC. Cihlar et al. (1994) found mixed results in Canada. While the MVC algorithm selected off-nadir pixels with consistently higher NDVI values than temperature-based algorithms, none of the alternative compositing methods produced composites which were consistent or as accurate as a single date scene. Stoms et al. (in press) reported higher quality composites were created by incorporating temperature and weighted scan angle for an area of southern California, USA. These comparisons were conducted on imagery composited before the look angle threshold was imposed. Accuracy of composites in both studies were checked against other raw or classified digital data.

2.1.5 *Why transform the biweekly composited imagery?*

Because vegetation phenology is a temporal measure, time-series biweekly composites were an obvious choice for deriving phenology for Oregon as a continuous variable measure due to the high temporal frequency and the coarse resolution of the imagery. Based on previous evaluations of biweekly composited imagery discussed above, however, there was concern regarding the utility of the data for multi-year comparisons. Previous research showed no single compositing method produced a perfect dataset and that there was some level of subjectivity in other compositing methods.

Theoretically and in practice, compositing produced a dataset which was "better" than a single date scene for the same time period, that is, one that was relatively free of cloud and atmospheric contamination. The 365 AVHRR scenes in an annual dataset are composited to reduce such "noise" and the NDVI values in time-series biweekly composited imagery could be used as a measure of phenology. Like the tasselled-cap transformation, principal component analysis has been shown to further refine a large dataset down to the most meaningful data. For this study, time series NDVI data represents seasonal vegetation characteristics. Repeatedly, seasonal differences and temporal variation which do not appear in single date or biweekly composites have been highlighted against orthogonally transformed axes (Cicone and Olsenholler, 1997; Eastman and Fulk, 1993; Fung and LeDrew, 1987; Tucker et al., 1985).

This research was intended to quantify temporal dynamics of vegetation, termed phenology, for the state of Oregon using parametric or non-classificatory methods. In doing so, this study also supported the utilization of data generated from 1992 NOAA-AVHRR biweekly composites in a larger project. To test the ecological utility of 1991-1993 biweekly composites for time-series analysis of vegetation, principal component analysis was conducted for each year independently. Results from this analysis provided information about the consistency, accuracy, and comparability of composite-derived phenology.

2.2 METHODS

Three years of biweekly composited imagery were acquired from EROS Data Center for 1991, 1992, and 1993. The state of Oregon was subset to political boundaries and water bodies were masked from further analysis. The NDVI band was subset from each biweekly composite for each of the three years resulting in three new datasets comprised of 21 bands. In these new datasets, each band corresponded to the NDVI of timeperiod 1-21 rather than distinct spectral channels.

Standardized principal component analysis was run on each of the three consecutive years of NDVI data to minimize the effects of atmospheric conditions or sun angles (Fung and LeDrew, 1987) and to improve the signal to noise ratio resulting from the atmosphere

and from optics or detectors (Singh and Harrison, 1985). Standardized principal components were calculated from the correlation matrix, rather than covariance matrix, which forced each timeperiod to have equal weight in the transformation (Eastman and Fulk, 1993). Eigenvalues, loadings, and reconstruction of transformed principal component bands from resulting linear combinations were calculated for dataset comparisons.

2.3 RESULTS

Eigenvalues for the first three principal components for each of the three years 1991-1993 and percent variance explained are shown in Table 2.1. The first three principal components combined account for 93.2% of the total variation in NDVI for 1991, 94.1% for 1992, and 88.3% for 1993. Loadings for the first three principal components for each of the years are plotted in Figure 2.1. The corresponding loading values are given in Tables 2.2, 2.3, and 2.4. Principal components 1, 2, and 3 for each of the three years showed nearly identical plots and values with similar signs and magnitudes, evidenced in loading plots and eigenvalues. Principal components 4 through 21 were highly variable

Table 2.1 Eigenvalues for the first three principal components for each of the three years 1991-1993 and percent variance explained.

1991			1992			1993		
PC#	**eigenvalue**	**%**	**PC#**	**eigenvalue**	**%**	**PC#**	**eigenvalue**	**%**
1	4322.95	83.1	1	4090.23	87.0	1	2608.88	74.1
2	326.82	6.3	2	212.83	4.5	2	349.29	9.9
3	200.29	3.9	3	120.66	2.6	3	151.63	4.3

among years with little consistency in loadings and eigenvalues. PC1 was positively loaded for all timeperiods with slightly higher values for timeperiods 8 to 18. PC2 was positively loaded for the first half of each year with the exception of timeperiod 1, and was

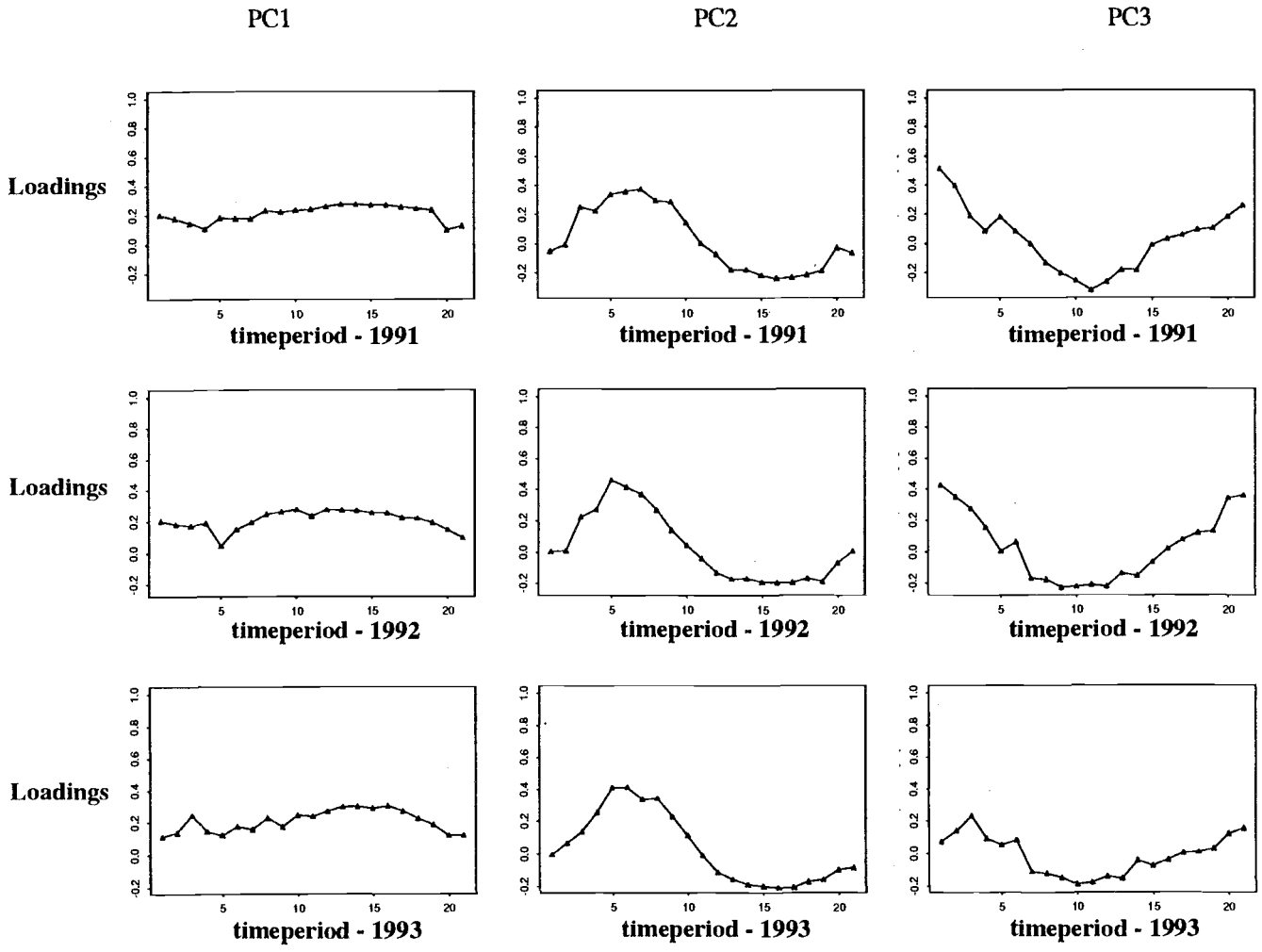
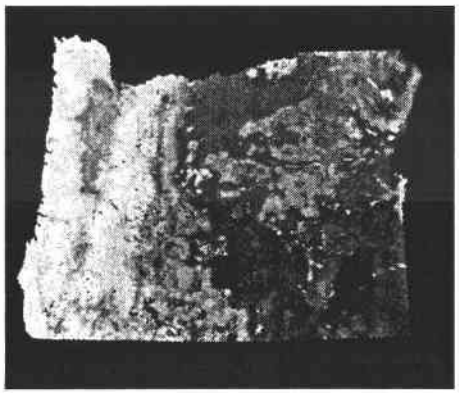
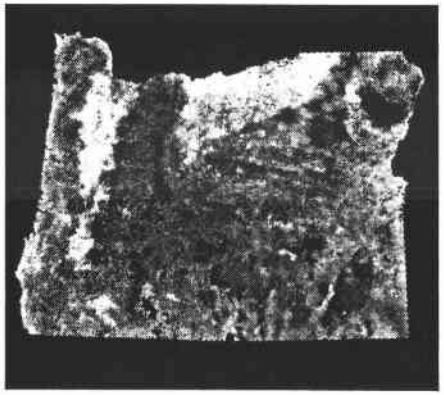


Figure 2.1 Loading plots for the first three principal components for each year. Loadings are positive for the first PC. In the second PC, loadings become negative halfway through the year. The third PC is positively loaded during the winter months and negatively loaded during the summer, or growing season months.

PC1



PC2



PC3

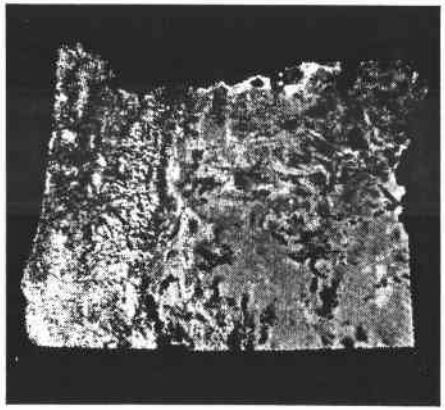


Figure 2.2 Linear transformations of principal components 1, 2, and 3 of 21 for 1992. Higer values are bright and lower values are dark.

Table 2.2 Loadings for first three principal components for 1991.

	1991		
timeperiod	PC1	PC2	PC3
1	0.2008780	-0.051081194	0.5162717722
2	0.1790190	-0.006753765	0.3992308364
3	0.1458926	0.250341280	0.1893627050
4	0.1074954	0.224605777	0.0880015872
5	0.1832189	0.339224088	0.1832902793
6	0.1792885	0.358720453	0.0862677295
7	0.1778293	0.374498121	-0.0008291459
8	0.2320088	0.295058194	-0.1316474308
9	0.2231159	0.285184072	-0.2012368745
10	0.2354409	0.143949217	-0.2572652750
11	0.2412045	0.001138298	-0.3215574026
12	0.2605073	-0.070442681	-0.2646989433
13	0.2767662	-0.178200273	-0.1813845063
14	0.2767662	-0.178200273	-0.1813845063
15	0.2707554	-0.216608524	-0.0111824946
16	0.2710535	-0.239008460	0.0313595335
17	0.2593850	-0.231236627	0.0580164702
18	0.2476461	-0.213326753	0.0957260051
19	0.2374127	-0.185474452	0.1060773810
20	0.1028147	-0.028150981	0.1825234656
21	0.1296655	-0.064183648	0.2587701062

negatively loaded from approximately timeperiod 11 through the end of each year. For the third PC, negative loadings were temporally shifted 8 weeks earlier in the year and had positive loadings during what is typically considered the non-growing season.

The original data were transformed through the linear combinations which resulted in an image reconstructed as the new PC bands. The first three PC bands for 1992 are

Table 2.3 Loadings for first three principal components for 1992.

	1992		
timeperiod	PC1	PC2	PC3
1	0.20406622	0.005760274	0.428834239
2	0.18419818	0.010905439	0.351387647
3	0.17392605	0.223173972	0.276482783
4	0.19090221	0.271621721	0.156251929
5	0.04890647	0.460888533	0.007380893
6	0.15282185	0.414837586	0.063752943
7	0.19549334	0.368145541	-0.168012261
8	0.24802805	0.269731814	-0.176753768
9	0.26296316	0.142065882	-0.227092376
10	0.27765828	0.045630696	-0.223224622
11	0.23713426	-0.039833042	-0.210808840
12	0.27753067	-0.132710206	-0.221213211
13	0.27616875	-0.173445992	-0.137142131
14	0.27137068	-0.171165526	-0.152938914
15	0.25800447	-0.193167120	-0.062359433
16	0.25671966	-0.195543897	0.022607859
17	0.22449726	-0.195856146	0.080457739
18	0.22262244	-0.168941757	0.123802028
19	0.19453542	-0.188286551	0.134174163
20	0.15039239	-0.068524363	0.342815180
21	0.09974118	0.004903927	0.359836531

shown in Figure 2.2 and are representative of each of the three years. A feature space plot was generated to assist in interpretation of these components. Figure 2.3 shows PC1 plotted against PC2 where color intensities correspond with frequency. Erdas Imagine® Image Processing software enabled these plots to be linked to the PC band data displayed

Table 2.4 Loadings for first three principal components for 1993.

	1993		
timeperiod	PC1	PC2	PC3
1	0.1142480	-0.002693546	0.074839007
2	0.1395446	0.067883333	0.141219614
3	0.2457103	0.139143339	0.232469922
4	0.1452457	0.256648651	0.092953623
5	0.1224141	0.411604197	0.055733400
6	0.1748566	0.412310569	0.085016346
7	0.1582511	0.337440144	-0.112460350
8	0.2291306	0.344190234	-0.123024355
9	0.1747699	0.230924529	-0.147832635
10	0.2449343	0.115745921	-0.190273387
11	0.2380658	-0.008932400	-0.177243910
12	0.2702962	-0.112964701	-0.141251343
13	0.2989741	-0.156903654	-0.153139109
14	0.3009983	-0.188588161	-0.041023183
15	0.2883089	-0.200001481	-0.072539637
16	0.3029078	-0.210236958	-0.035557956
17	0.2703627	-0.207291719	0.006044025
18	0.2240006	-0.171625909	0.011431770
19	0.1870850	-0.159236024	0.033747253
20	0.1215003	-0.097239339	0.124645658
21	0.1213749	-0.084964760	0.155502380

as images (Figure 2.2). With crosshairs, pixels in the original PC band image were simultaneously identified with corresponding pixel values in the frequency plots. This allowed real time display of relationships between the different PC band combinations. To interpret the continuous data from the principal components, supervised classification was

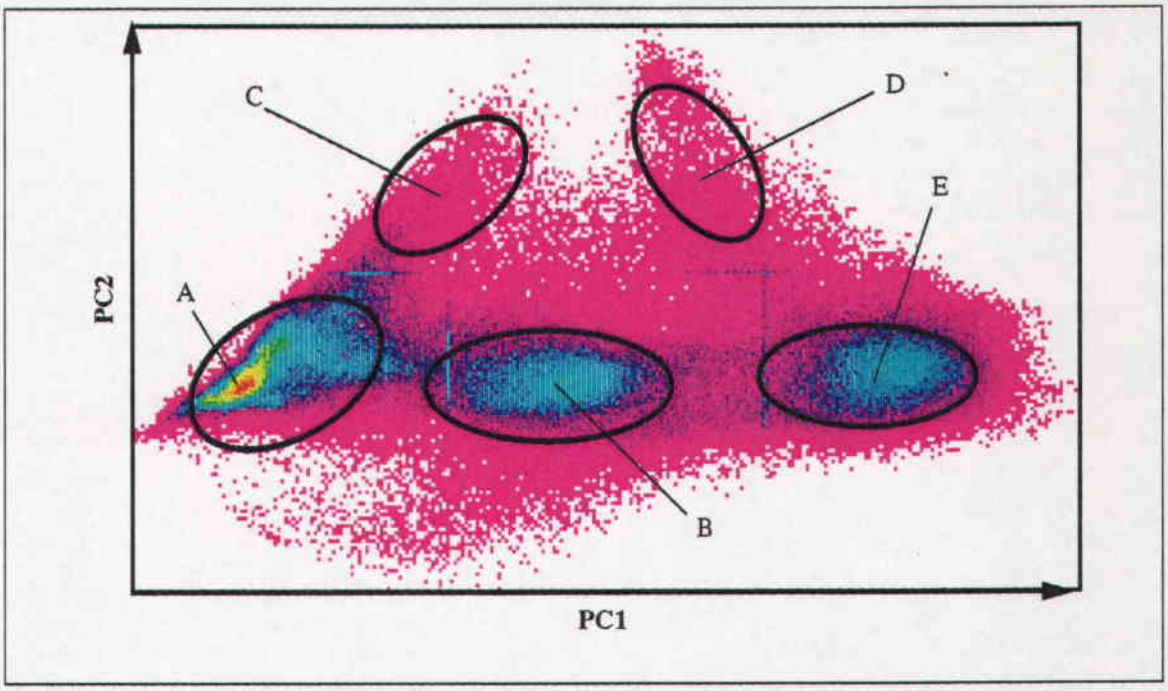


Figure 2.3 Feature space plot of PC1 and PC2 for 1992 with phenologically similar regions identified.allowed real time display of relationships between the different PC band combinations. To interpret the continuous data from the principal components, supervised classification was used on the feature space plots to generate signatures for clustering pixels in PC band Frequencies are mapped following the color spectrum where red/ orange/yellow indicate highest counts and green/blue/magenta are lowest frequencies. PC1, on the x-axis, indicated an east-west gradient of increasing brightness. PC2, on the y-axis, represented a seasonal gradient. The identified areas are as follows: (A) high desert, (B) eastside mountains, (C) eastside irrigated agriculture, (D) westside agriculture and (E) Cascade and Coast Mountain Ranges.

used on the feature space plots to generate signatures for clustering pixels in PC band images. Signatures were generated from groups of pixels identified in the feature space plot. These signatures were used to identify corresponding regions in the reconstructed PC images. Results from these investigative classifications are shown in Figure 2.4. Spatially conterminous areas which were mapped in Figure 2.4 corresponded to ecologically similar pixels. These corresponded with the high desert, eastside mountain ranges, eastside irrigated agriculture, westside agriculture, and Cascade and Coast Ranges.

These two methods, plotting band loadings and interactive feature space classification, facilitated identification of phenologically similar and spatially contiguous regions derived from the time-series NDVI data. Signature generation was used to

interpret the continuous principal component data; classification was not the end product. Results of the signature generation resulted in identification of different parts of the Oregon landscape which had similar temporal variability over the course of one year.

The high desert area east of the Cascades corresponded to the lower left corner of the feature space plot. This area also showed highest pixel frequencies, as the majority of Oregon itself is east of the Cascade Mountains. Mountainous regions east of the Cascades were clearly distinguishable from non-mountainous eastern Oregon and western Oregon along the PC1 axis and also showed high frequencies. Both the Columbia Gorge agricultural region and Willamette Valley had highest values of PC2, and were separable from each other by PC1 value. Feature space plots for 1991 and 1993 were identical to that of 1992 shown in Figure 2.3. The regions described above plotted out similarly for all three years.

2.4 DISCUSSION

What is important about the plot of PC1 versus PC2 (Figure 2.3) in terms of temporal dynamics or phenology, is the organization of the plot and resulting contiguous groups of pixels (Figure 2.4). The two axes represented the spatial organization of the temporal pattern of the Oregon landscape. The initial data were spatially explicit NDVI values which related to biomass or productivity. The regions generated from signatures in the feature space plot therefore corresponded to ground conditions which exhibited similar patterns through time and in space. This means that the two axes represented the spatial and temporal component of the landscape. If the entire PC1 versus PC2 plot were used to generate a classified landcover map it would differ from other classified satellite imagery-derived maps by the incorporation of the temporal component. The result would be a vegetation map based on temporally dynamic processes. This might offer an advantage for discriminating ecological cover types such as eastern and western mountains and eastern and western agriculture, but would not assist in discriminating temperate coniferous forests. Clearly the axes did not represent species level information, as different grasses were not distinguished nor were different conifer species discriminated in the Cascades and Coast Ranges. This may also have been related to the relatively coarse

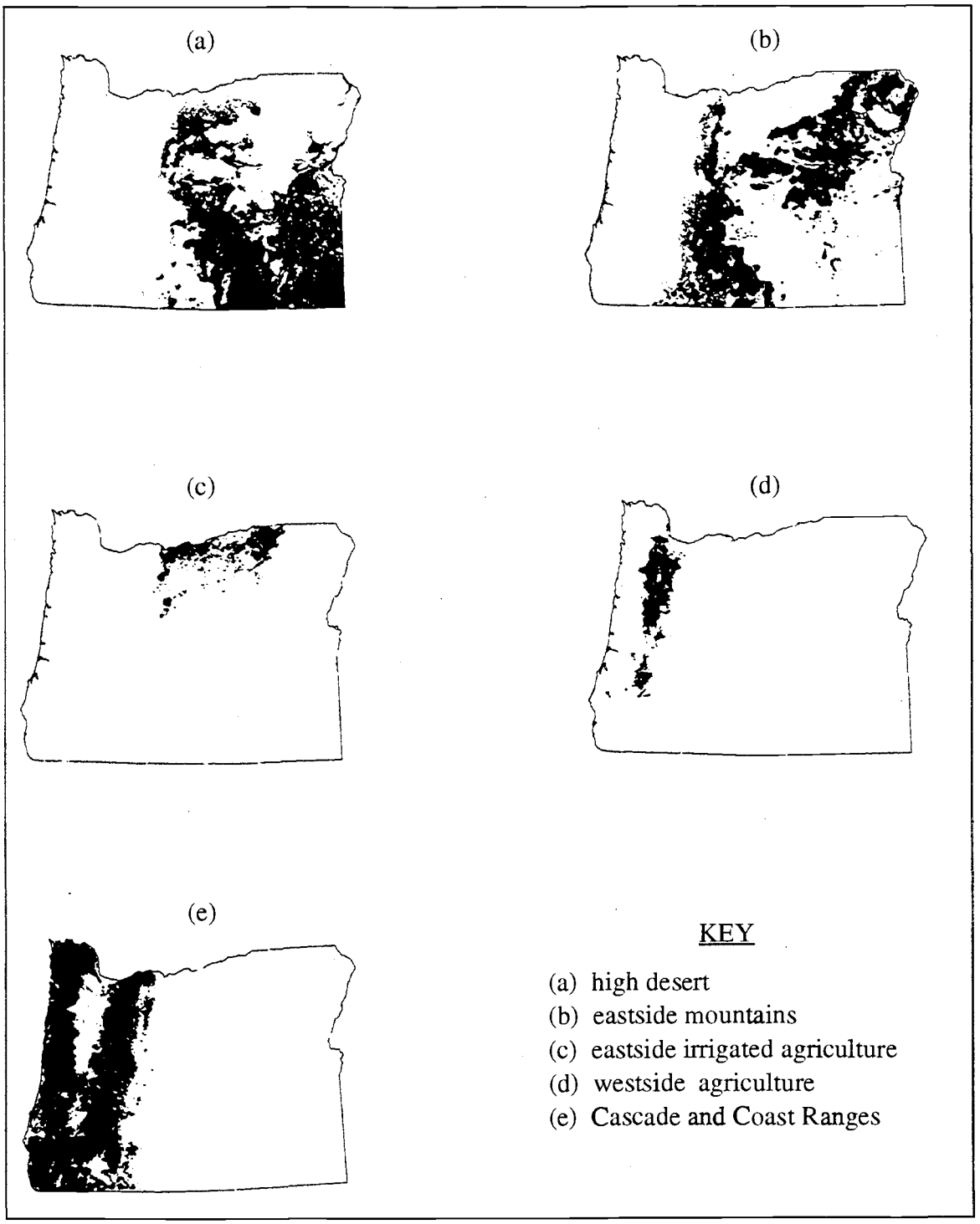


Figure 2.4 from feature space signature generation. Groups of pixels were selected directly from the feature space plot of PC1 and PC2 for signature generation. Clustering based on these signatures identified the different regions shown in black. Each region has similar temporal variability and represents the different seasonal components of the Oregon landscape.

scale of the imagery. The lack of species discrimination was also due to the fact that what we think of as similar species (firs, pines, etc.) exhibit similar phenology.

If noise were a major component of the three datasets (1991, 1992, 1993) it would have been expected to show up in the first few principal component bands. Noise in an NDVI time-series dataset would be very apparent as either aberrant values relative to temporal neighbors or spatially as neighborhood outliers. This can be visualized in two ways, as plots of loadings (Figure 2.1) and by viewing the resulting PCA bands reconstructed into images (Figure 2.2). Visually, the PC band plots did not appear particularly noisy, especially when compared to any given individual biweekly composite scene. Noise resulting from sun angle, atmospheric conditions, or sensor properties would also have been expected to vary among the three years. Therefore if noise were dominating the scenes, not only would it have shown up in the first few components, but the loadings and scores themselves would have been expected to have different magnitudes and/or signs. Resulting eigenvalues, loadings, and plots from PCA were similar for each successive year, implying high signal to noise ratios consistent for the three datasets.

Loadings and plots were consistent with results for Africa (Eastman and Fulk, 1993; Tucker et al., 1985), South America (Townshend et al., 1987), Asia (Cicone and Olsenholler, 1997), and Canada (Fung and LeDrew, 1987) where the first two to four principal components in NDVI time-series analyses were found to relate to vegetation conditions and phenology. Eastman and Fulk (1993) found additional information relating to large scale, long term climate patterns in additional components. This was not found in subsequent PC bands for Oregon, a fact which may have been related to the relatively small area being investigated. It has been suggested that time series trends from NDVI may indicate global climate change patterns (Di and Hastings, 1995) and this would further support the results from Eastman and Fulk (1993), but the manifestation of these effects in vegetation may not be detectable at a statewide scale.

2.4.1 *Principal Component #1*

Oregon can be divided into two distinct regions. Temperature fluctuations and precipitation patterns correspond to very lush, productive forests west of the Cascades and semi-arid pine, brush, or grass dominated high desert east of the Cascades. These biogeographical regions were easily identified with feature space plots and feature space clustering in PC band images. From loading and feature space plots, PC1 represented the spatial distribution of greenness and gave the appearance of variation in illumination from east to west. There is no actual variation from east to west in sun illumination, but there were illumination differences over Oregon as a whole which changed subtly over the course of the year with solar azimuth. This annual variation was apparent by positive loadings for all bands and by the fact that the magnitude of the loadings increased in the summer months, when sun angle was highest. Cicone and Olsenholler (1997) identified PC1 for Asia "cumulative greenness" or total leaf biomass.

An east-west gradient was apparent in PC1 by the regions identified in feature space and based on feature space clustering. There is an explanation for the appearance of this variability in illumination. This gradient started in eastern Oregon, which also has minimum topographic variability relative to the major mountain ranges, and is therefore more uniformly illuminated when the satellite overpasses. East side mountainous regions such as the Wallowas, Klamath, and eastern Cascade Ranges, have considerably more terrain variability than the non-forested high desert regions, although not to the extent of the central Cascades and Coast Ranges which had the highest values of PC1. This gradient might also be related to physical properties previously discussed, where pixels farther west tend to be farther off-nadir and therefore measure artificially high NDVI resulting from a larger column of air between the satellite and the Earth's surface. Although larger footprints will tend to average surface properties and anomalies, more land is physically incorporated in the reflected signal. However, we are reminded by Singh and Harrison (1985) that principal components are linear combinations of variables which are artificial constructs of the original data. Units of measure of PCA are not NDVI values and the transformed components themselves cannot be directly measured.

2.4.2 *Principal Component #2*

Tucker et al. (1985) plotted results from PCA on time series NDVI for Africa in feature space and found similar results to those shown in Figure 2.3. In Oregon, three major groups (high desert, eastside mountain ranges, and west side mountains) were immediately apparent based on color intensities. These areas stood out because color intensity was mapped to frequency and therefore the color of the points in the feature space plot also represented numbers of pixels in the corresponding PC images. The majority of pixels in Oregon (i.e. the majority of the landmass) are east of the Cascade Range and correspond to the primarily red and yellow pixel cluster in the lower left of the plot. Feature space plots showed organization of pixels into spatially contiguous groups which also exhibited temporally similar values or dynamics. This meant that spatially and ecologically through time the response of the vegetation was consistent by region.

Plots of principal component loadings showed highest contribution of NDVI during the spring to early summer months. Seasonally, this corresponded with the end of the winter rains in the west and onset of warmer temperatures throughout the state. These two factors correspond with the spring greenup or leaf-out. As summer progresses and water becomes limited, particularly in eastern Oregon, vegetation scenesces or decreases productivity. Temperatures also begin to drop to freezing levels at night in parts of eastern Oregon by the end of the summer and in early autumn. Loading plots showed this seasonal pattern phenomenon as decreasing values which become negative towards the end of the growing season.

The exact nature of the seasonal component was difficult to identify and has been termed "summer greenness flux" by Cicone and Olsenholler (1997). Length of growing season or water availability were hypothesized to be the seasonal component of PC2. For either of these to be true, greater differentiation between eastern and western Oregon would be expected based on PC2 value alone. This was not the case, however, as regions east and west of the Cascades had the same PC2 value and were discriminable only in conjunction with PC1. PC2 might be considered to have a topographic component, except that the region designated 'eastside irrigated agriculture' is of similar elevation as the 'high desert' region and elevationally is much higher than the Willamette Valley. Terrain

is a slightly different measure of topography which can be considered, but the variability which exists between the high desert region and the western Oregon mountains is as great as the variation between the Willamette Valley and the western Oregon mountains.

Separability based on human alteration is apparent by the PC2 axis. Those regions which have been converted from their original 'natural' systems to intensively managed agriculture were easily discernible from areas which remain largely in the original 'natural' landcover type. This is not to say that there has not been dramatic and intensive management in other ecological regions. Grazing and timber production are predominant throughout Oregon however, a smoothing effect is possible at a 1 km^2 pixel resolution (or greater with increasing look angle). More likely than smoothed pixels is an axis label which relates to growing season duration, number of greenup events, or rate of greenup-senescence events. Over the course of one year the growth of more 'natural' ecosystems can be described by a curve which approximates a bell-shaped distribution. Vegetation remains dead, dormant, or minimally photosynthesizing until spring. For harvested regions, vegetation may be present or growing one day and harvested the next, dramatically reducing the NDVI response in a relatively short time span. Sub-seasonal count or number of leaf-out events is another possible explanation following this logic. It is possible that PC2 represents the number of times a dramatic increase in NDVI occurs. Although productivity can occur throughout the entire year in parts of western Oregon, there is typically only one leaf-out event so coniferous regions would have one greenup time annually.

2.4.3 *Principal Component #3*

The seasonal information in PC3 is independent of the seasonal information of PC2. Loading plots and eigenvalues of this component indicate highest contribution in the non-growing season. Therefore the NDVI contribution from winter months may indicate productivity or photosynthesis occurring during the non-growing season. This was supported by highest values of PC3 corresponding to coniferous forests of the Klamath and Siskiyou regions in southwestern Oregon. From late fall to early spring coniferous

dominated forests in Oregon remain green. West of the Cascade mountains temperatures remain mild enough to allow productivity throughout the entire year.

Eastman and Fulk (1993) reported a discontinuity in similar plots of PC3 around November for Africa which were not apparent in Oregon principal components. They concluded excessive dust was responsible for artificially high NDVI values which disappeared with the onset of rain in winter months. In all three years for Oregon there is a small spike coinciding with timeperiod 6, followed by a drop in values. Field burning is prevalent in parts of Oregon but the corresponding drop in NDVI cannot be explained by onset of rain because Oregon summers are fairly free of precipitation. Wildfire season in Oregon is in late summer and lasts until the rains begin but the lack of a spike in the late summer months do not support this as an influential factor. The data used in this analysis are more recent than those of other works cited here. Potential problems encountered from sensor look angles may have been minimized with imposed look angle constraint. Additionally the relatively small study area also likely reduces what would otherwise be noticeable variation. Cicone and Olsenholler (1997) called PC3 in Asia "early season greenness flux" and suggested this component represented vegetation productivity occurring earlier than the summer growing season.

2.5 CONCLUSIONS

The objectives of this research were to assess the ecological utility of NOAA-AVHRR biweekly composite data as a continuous variable measure of vegetation phenology. Principal component analysis indicated consistent results for each of the three years. These results were also consistent with previous research and were interpreted ecologically in terms of vegetation phenology or temporal dynamics on the landscape. This study was not intended to provide a new regionalization method, but the possibility of doing so may exist using extended methodologies from those presented here. This study showed that principal component analysis of time-series AVHRR NDVI data is an acceptable method for measuring environmental heterogeneity in space and time for Oregon.

Off-nadir pixels may have influenced the first PC, but the most extreme look angles have been eliminated from biweekly composites by imposing threshold restrictions on pixel selection. This dataset may have had such consistency between datasets and overall success because Oregon is a relatively small area compared to continental scale datasets. Based on the first three principal components, confidence level in the data is high for use in future research.

3.0 Chapter 3: Regression tree analysis of vertebrate species richness using NOAA-AVHRR satellite imagery, DEM, and climate data

Vertebrate species richness was modeled for the state of Oregon, USA, using regression tree analysis. A total of 27 vegetation phenology, terrain, and climate variables were investigated as correlates. Vegetation phenology was derived from two methods: 1) time-series NOAA-AVHRR satellite imagery as principal components and 2) as greenness metrics from a method developed at EROS Data Center. Semivariance and Moran's I were calculated on taxonomic response data of mammals, birds, reptiles, amphibians, and all vertebrates and were calculated on model residuals as a test of goodness of fit. Amphibians had the most straightforward spatial pattern of the taxa and as a result generated the shortest tree which explained 91% of the variation in amphibian diversity. This was followed by birds (67%), all vertebrates (66%), reptiles (57%), and mammals (55%). Taxa which had a less complex spatial pattern, quantified with spatial statistics, had better fitting models than taxa with complex spatial patterns. Graphical results of the regression trees indicated significant relationships between phenology, elevation and climate for each taxa. Amphibian richness was related entirely to phenology and climate. Climate was the sole predictor for bird richness west of the Cascade Mountains while elevation was a dominant factor in eastern Oregon. Reptiles were modeled primarily with phenology derived from principal component analyses while mammals richness was highly correlated with spatial heterogeneity.

3.1 INTRODUCTION

Much of ecology over the last 30 years has been devoted to the identification and understanding of mechanistic processes which drive the pattern-process relationship between biodiversity and the environment. Biodiversity, the number of species in a given area, has been studied at many scales, ranging from the smallest ecological plot to the planet. Our perspective of biodiversity and landscape pattern varies with spatial and temporal scale, as patterns of biotic diversity are complex and the relationship of diversity mechanisms with scale are difficult to quantify. Species are going extinct rapidly

worldwide. Unfortunately, exclusive attention to individual species will not preserve the remaining biotic diversity, nor will focusing solely on single species slow extinction rates. To slow rates of extinction or preserve remaining fauna or flora we must develop an understanding of the mechanisms or processes which drive the patterns of richness on the landscape. To develop this understanding, we must study diversity by groups of species, such as by taxa, and at broad spatial scales. My premise was that we can develop an understanding of the mechanisms which shape the pattern of species diversity from environmental diversity, quantified with remotely sensed data. In this study, environmental diversity was defined as all variability in space and time (beta diversity), and species diversity was defined as vertebrate species richness. This was a study of the pattern-process relationships between vertebrate species richness in Oregon and environmental heterogeneity in space and time.

People categorize the natural world for many reasons. These categorizations are important for communicating about and managing the landscape and can be used to study biodiversity. There are problems with taking a categorical approach to addressing the question: *what mechanisms shape the patterns of biodiversity?* For example, the relationship between diversity and a landscape comprised of 30 categories will be different from the relationship between diversity and the same landscape classified into 10 categories. No two independent expert assessments will interpret a given landscape in exactly the same fashion, because there is no universally accepted standard or set of rules which define landcover types. Even given a set of criteria, there will exist variability in landcover interpretation due to human based differences. The second difficulty with using landcover types as a basis for quantifying diversity is the assumption that fauna interpret or view the landscape in the same way as humans. Do fauna view, for example, a sagebrush landscape differently from a sagebrush/juniper landscape? How do we determine the different cover types fauna distinguish, whether all fauna respond to all landscapes in similar or different manners, or exactly what attributes differentiate the land into cover types? Finally, landcover types typically do not contain temporal criteria. There are no descriptions, quantitative or qualitative, of the temporal dynamics of the landscape despite temporal variations which are obvious in many landscapes.

The question then became, are temporal variations of the landscape, or temporal

dynamics, significantly related to vertebrate species richness when taken as continuous variable measurements? This research used non-categorical indices of spatial and temporal vegetation dynamics to explain potential process-pattern relationships of vertebrate species richness in Oregon. The indices of spatial and temporal vegetation dynamics, or vegetation phenology, were derived directly from time series satellite imagery as objective measures of vegetation dynamics. These indices, with climate and elevation data, were modeled to quantify the relationship of vertebrate species richness with environmental variability in space and time.

This study investigated the relationships between vertebrate species richness of birds, mammals, reptiles, and amphibians with plant phenology, terrain, and climate for the state of Oregon. In this research I attempted to maximize the number of species investigated and to directly correlate vertebrate species richness with meaningful variables. The goal of this study was to determine whether 1 km^2 National Oceanic and Atmospheric Administration Advanced Very High Resolution Radiometer (NOAA-AVHRR) data, 1 km^2 digital elevation model (DEM) data, and climate variables could be used directly to explain vertebrate species richness.

3.2 A BRIEF HISTORY OF VERTEBRATE RICHNESS AT BROAD SCALES

The study of biodiversity at broad scales is not new. Researchers have looked at biotic and abiotic factors to explain the patterns of richness on the landscape at regional and continental scales. Simpson (1964) related topographic factors and other gradients to mammalian richness for North America. MacArthur and Wilson (1967) proposed island biogeography theory which related population size and persistence to habitat area and distance from source populations. Kiester (1971) conducted a similar analysis to the work of Simpson (1964) for herpetofauna, mammals, and birds of the continental United States. He related vertebrate taxon distributions to topography and latitudinal gradient. Wright (1983) expanded MacArthur and Wilson's area-island theory to include energy, which he called species-energy theory. Owen (1990) correlated mammal richness with elevation, temperature, and productivity in Texas. Stoms (1995) related species richness by taxa in California to habitat heterogeneity, productivity, and dynamic processes following the

organization of Diamond's (1988) "QQID". Diamond's (1988) "QQID" was a proposed organizational framework for factors controlling species diversity where Q's = resource quantity and resource quality, respectively, I = species interactions, and D = dynamic processes.

Researchers have consistently identified the amount of energy available to a system, measured as evapotranspiration (ET) or net primary productivity (NPP), as a parameter which explained much of the variability in both plant and animal diversity at coarse scales (Currie, 1991; Owen, 1990; Wright, 1983). Climate has been correlated with species richness (Turner et al., 1988; Scheibe, 1987; Schall and Pianka, 1978) as climate plays an important role in shaping vegetation and habitat (Richerson and Lum, 1980; Daubenmire, 1978).

Spatio-temporal aspects of ecosystems are important elements of ecological theory. Diverse communities are created when organisms aggregate in heterogeneous areas rich in biotic and ecologic processes (Legendre and Fortin, 1989). Ecological processes, and thus ecological data, are spatially autocorrelated and occur in patches, as gradients, or exist in other spatial forms (Rossi et al., 1992; LeGendre and Fortin, 1989; Boots and Getis, 1988). In classical statistics, correlated data is a problem. In this study spatial pattern was the phenomenon of interest and I attempted to describe the environmental processes which relate and contribute to such patterns on the landscape. Variables related to available energy and vegetation phenology derived from time series satellite imagery were investigated as possible correlates to vertebrate species richness. I will refer to 'vegetation phenology' hereafter as simply 'phenology', defined as the growth of vegetation over time. While habitat type and landscape metrics have been investigated (Davis and Stoms, 1994), phenology as a continuous measure over time has not.

3.3 EXPLORATORY DATA ANALYSES - WHY I USED A MACROECOLOGIC APPROACH TO STUDY BIODIVERSITY

Computers have greatly expanded our data collection, storage, and processing capabilities. With this additional analytical power, new statistical theory and methods were developed which allow analysts to expand analyses beyond the traditional limitations

of statistical assumptions and physical processing power. These new statistical methods substitute computer algorithms for more traditional mathematics (Efron and Tibshirani, 1991). Exploratory data analysis is one of these emergent statistical fields. With exploratory techniques that are data-driven, nonparametric and computer-intensive, non-linear structures in data can be quantified as alternatives to traditional linear models (Miller, 1994). Exploratory analysis can be contrasted to Neyman-Pierson hypothesis testing. The computational intensity of large datasets at one time limited researchers' abilities for analysis, but with modern computers data volume is no longer a limitation. With exploratory techniques we can apply more complicated statistical estimators to explore and describe data and to draw valid stastical inferences on very large datasets (Efron and Tibshirani, 1991). Exploratory methods allows us to uncover and quantify the structure inherent in data, free of traditional bell-shaped curve assumptions.

The process-pattern relationship between vertebrate richness and environmental heterogeneity was expected to have a hierarchical structure of complexity. Interactions between variables were unknown. I was interested in quantifying this hierarchical structure inherent in the data to provide insight for understanding process-pattern relationships rather than fit the data to a pre-defined curve. With the use of exploratory data analysis and graphical display tools, these complex structures can be described and interpreted.

Exploratory data analysis also generates hypotheses. The resulting informed, empirically based hypotheses can be tested with standard statistical methods. In this manner, complex relationships and interactions between vertebrate taxa and the environment provide direction and refinement of future studies. Exploratory data analysis is not experimental, but as Brown (1995) noted, there are ways other than experimental ecology to conduct rigorous science. If we are to understand the effects of human population and technology on the environment and attempt to reduce negative effects, we must examine these effects at appropriately large scales. This research was designed to explore the relationships between vertebrate species richness by taxon with ecological mechanisms to guide future efforts of biodiversity research at broad scales.

Consistent significant correlative relationships between species richness and environmental variables remain weakly defined and three reasons might be cited for this:

1) quantitative statistical and modeling methods have not yet advanced enough to explain complex higher level interactions between autocorrelated richness data and mechanistic variables, 2) important variables were not included in analyses, and/or 3) significant correlative relationships with selected variables exist but were undetected at the investigated scale. I employed an exploratory data analysis approach to account for these issues. To approximate the spatial statistical tools which are necessary to conduct this type of analysis, I employed what I referred to as a 'despatialization-respatialization' framework methodology. In this method I quantified the richness patterns spatially, despatialized the response and predictor variables for analysis, respatialized the predicted results and residuals, and quantified them spatially. A parametric approach to capture vegetation phenology was used to include potentially important but as yet uninvestigated correlative mechanisms. Spatial statistics were used to address issues of scale.

There are two axes of measurement for environmental diversity: space and time. This raised the question: *are phenological variations expressed in terms of vegetation changes important?* Phenology may capture the temporal variability of the landscape in a more meaningful manner than an index or classification scheme generated from a one-time measurement and like landscape indices, phenology can also be evaluated spatially. Phenology is a continuous measurement free from subjective categorization. Variables which measure vegetation phenology may be correlated with species diversity. If this held true and phenology could be accurately characterized, the implications are far reaching for the advancement of biodiversity research around the globe, particularly in remote or inaccessible regions.

3.4 NDVI EXPLAINED AND HOW IT MEASURES ENVIRONMENTAL HETEROGENEITY

Many variables in this study were based directly or indirectly on the normalized difference vegetation index (NDVI) calculated from NOAA-AVHRR data for the state of Oregon. The NOAA-11 AVHRR is a polar orbiting satellite which provides daily global coverage resampled to approximately 1 km^2 pixels (NOAA, 1990). The AVHRR sensor measures reflected radiation in the red visible and near-infrared, and emitted radiation in

the thermal infrared parts of the spectrum. The NDVI is easily calculated from this data using the following equation:

$$\frac{(NIR - red)}{(NIR + red)}$$

where NIR is the measured reflectance in the near-infrared wavelength and red is the reflectance in the red wavelength. For the NOAA-11 AVHRR, the corresponding wavelengths are (0.72 - 1.1 μm) and (0.58 - 0.68 μm), respectively (Lillesand and Kiefer, 1994). NDVI has been correlated with green leaf biomass, net primary productivity, and vegetation conditions (Di and Hastings, 1995; Townshend et al., 1987; Goward et al., 1985). The NDVI can be thought of as a measure of the amount of green, potentially photosynthesizing vegetation. When NDVI is calculated over time, it is a measure of the change in vegetation condition based on reflectance properties related to photosynthetic capacity. Healthy vegetation will generate a stronger signal and consequently will have a higher value of NDVI than the same vegetation type under stress. The NDVI calculated from satellite imagery is also a spatial measure having constant location. Therefore the use of NDVI from satellite imagery is a direct measure of the spatial and temporal variability of the vegetation or the larger landscape.

3.5 STEP ONE: IS THERE A PATTERN TO BE EXPLAINED?

The extent of this study was the state of Oregon, defined by political boundaries. Was it correct to assume that vertebrate species richness exhibited any quantifiable pattern within state lines? If patterns of vertebrate richness exist, were they effectively captured by the grain of the species richness data? If there were no quantitative pattern to the dependent variables and the data were random, an attempt to predict a random pattern using spatially autocorrelated data would be an exercise in futility. Spatial statistical methods have been developed to identify spatial distributions and model spatial dependence in ecological data (Rossi et al., 1992; Cohen et al., 1990; Legendre and Fortin, 1987). Classical statistical methods often assume independence, an assumption violated by ecological data exhibiting spatial and/or temporal autocorrelation.

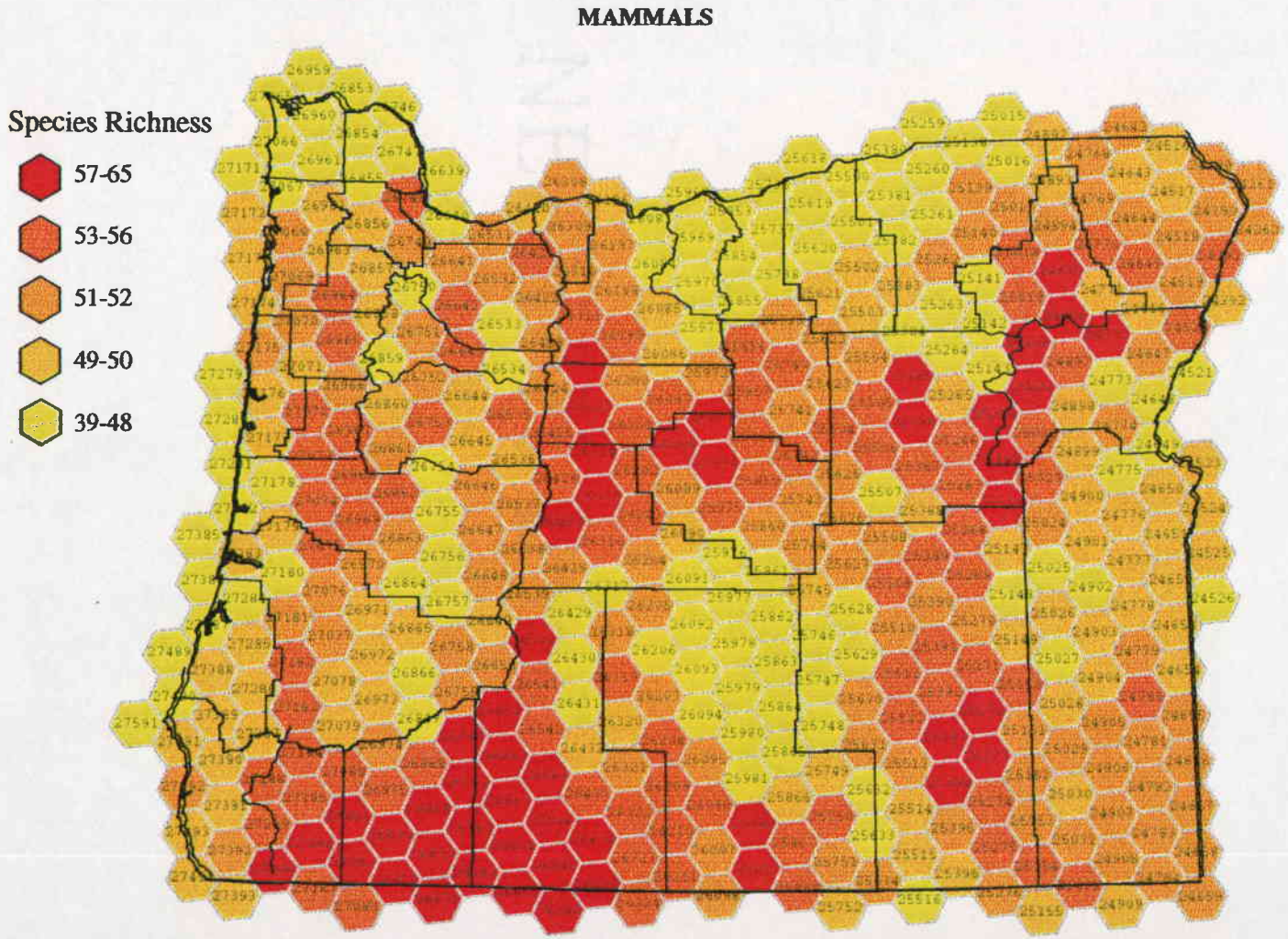


Figure 3.1 Mammal richness for the state of Oregon.

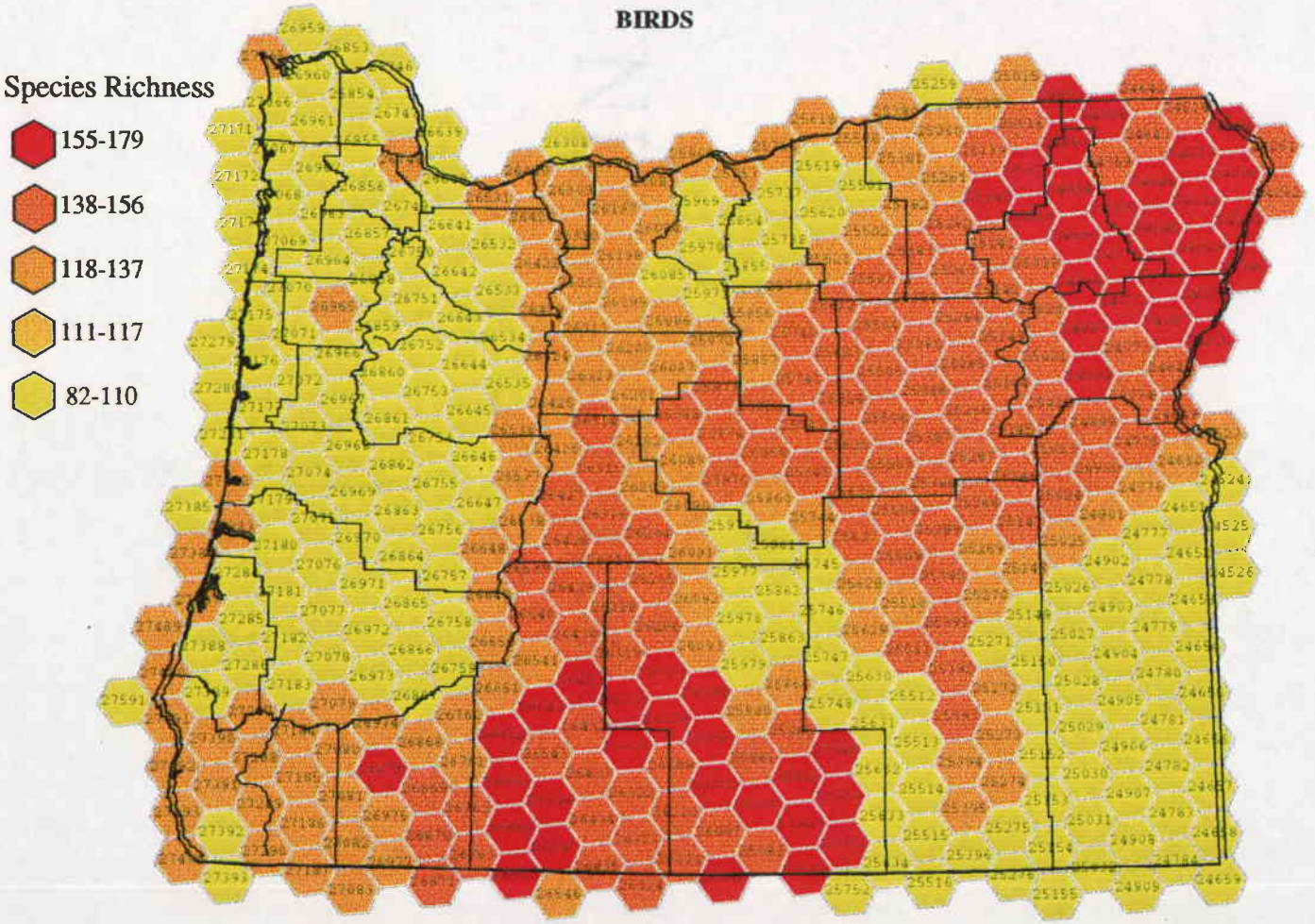


Figure 3.2 Bird richness for the state of Oregon.

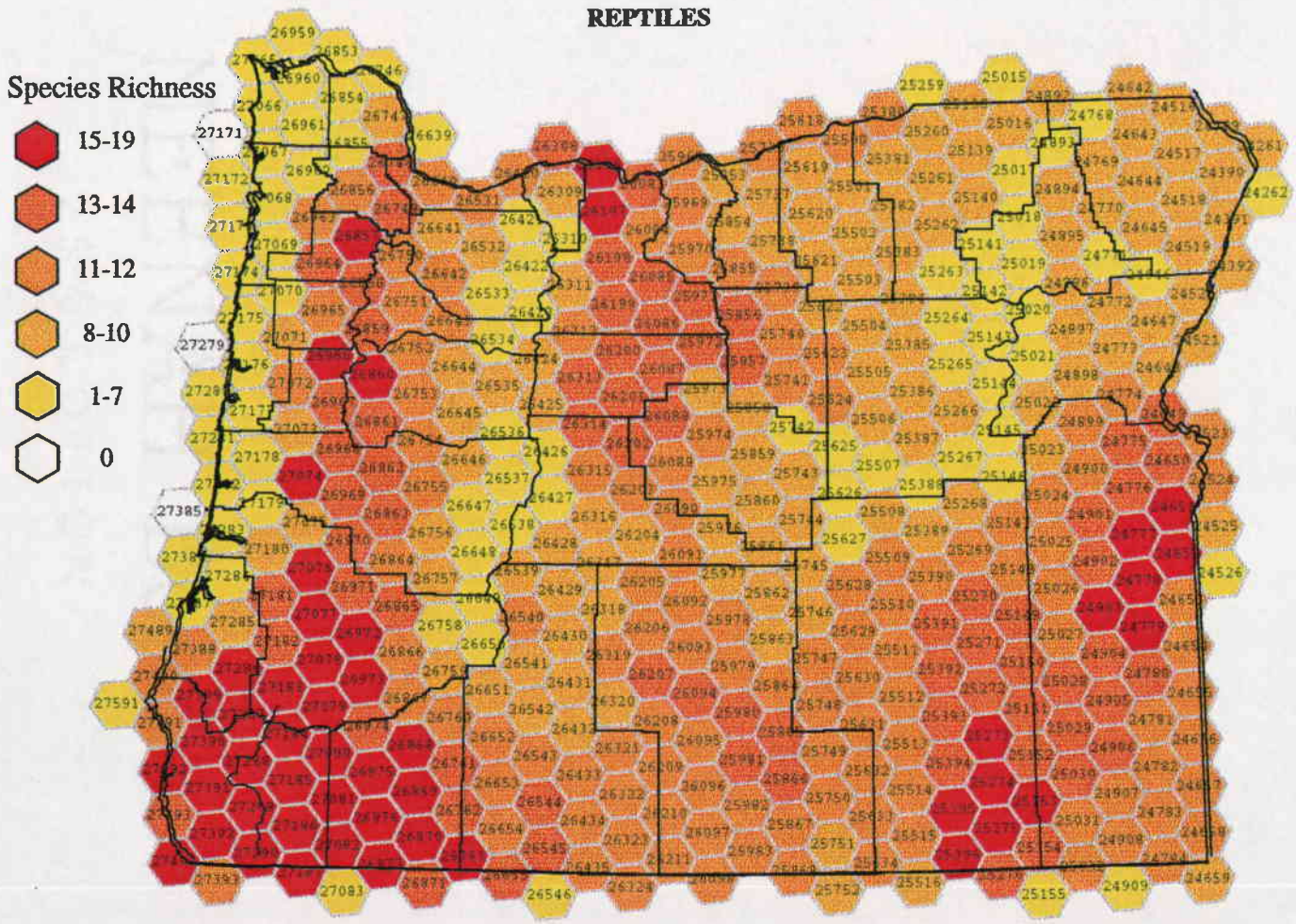


Figure 3.3 Reptile richness for the state of Oregon.

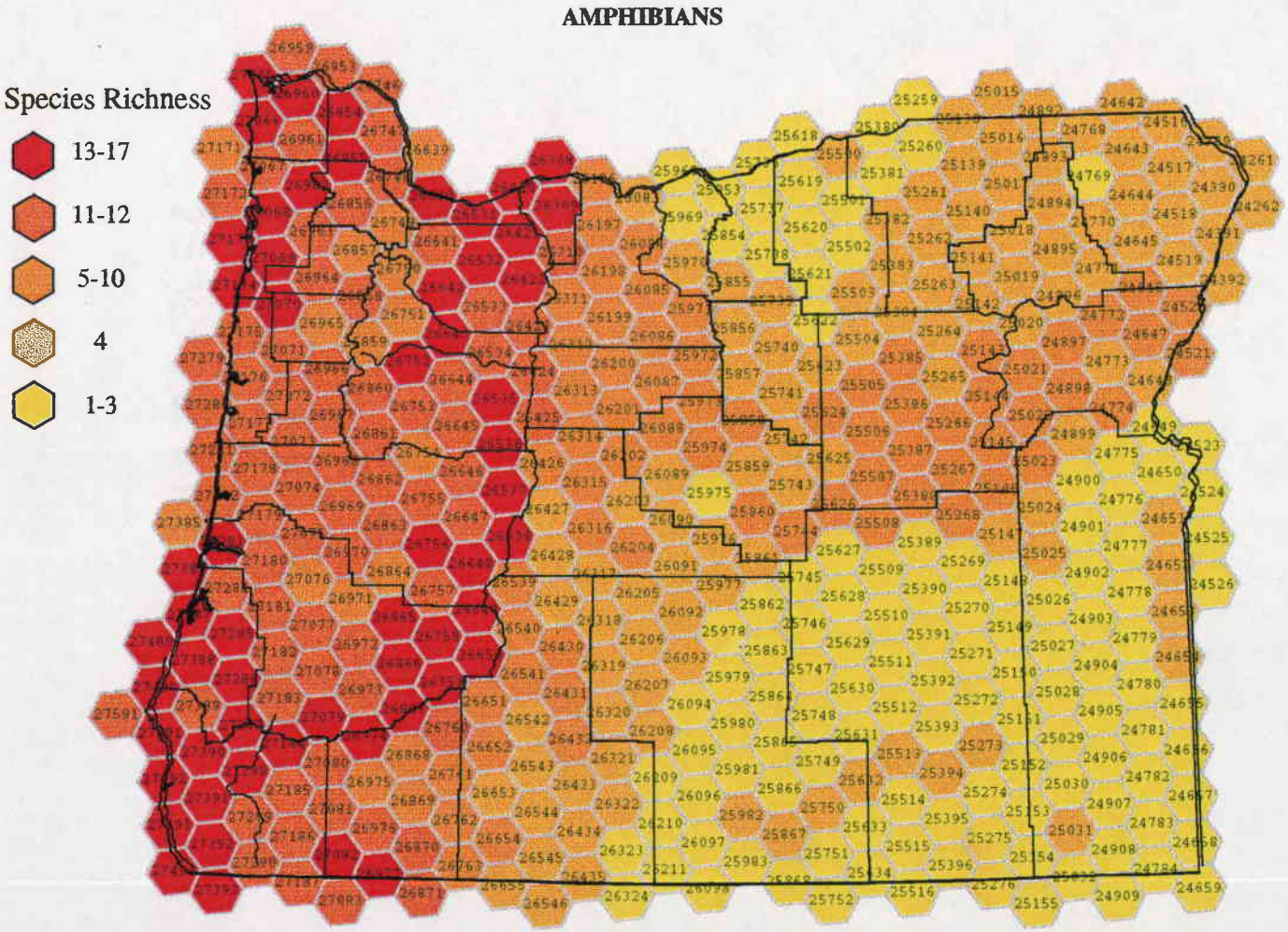


Figure 3.4 Amphibian richness for the state of Oregon.

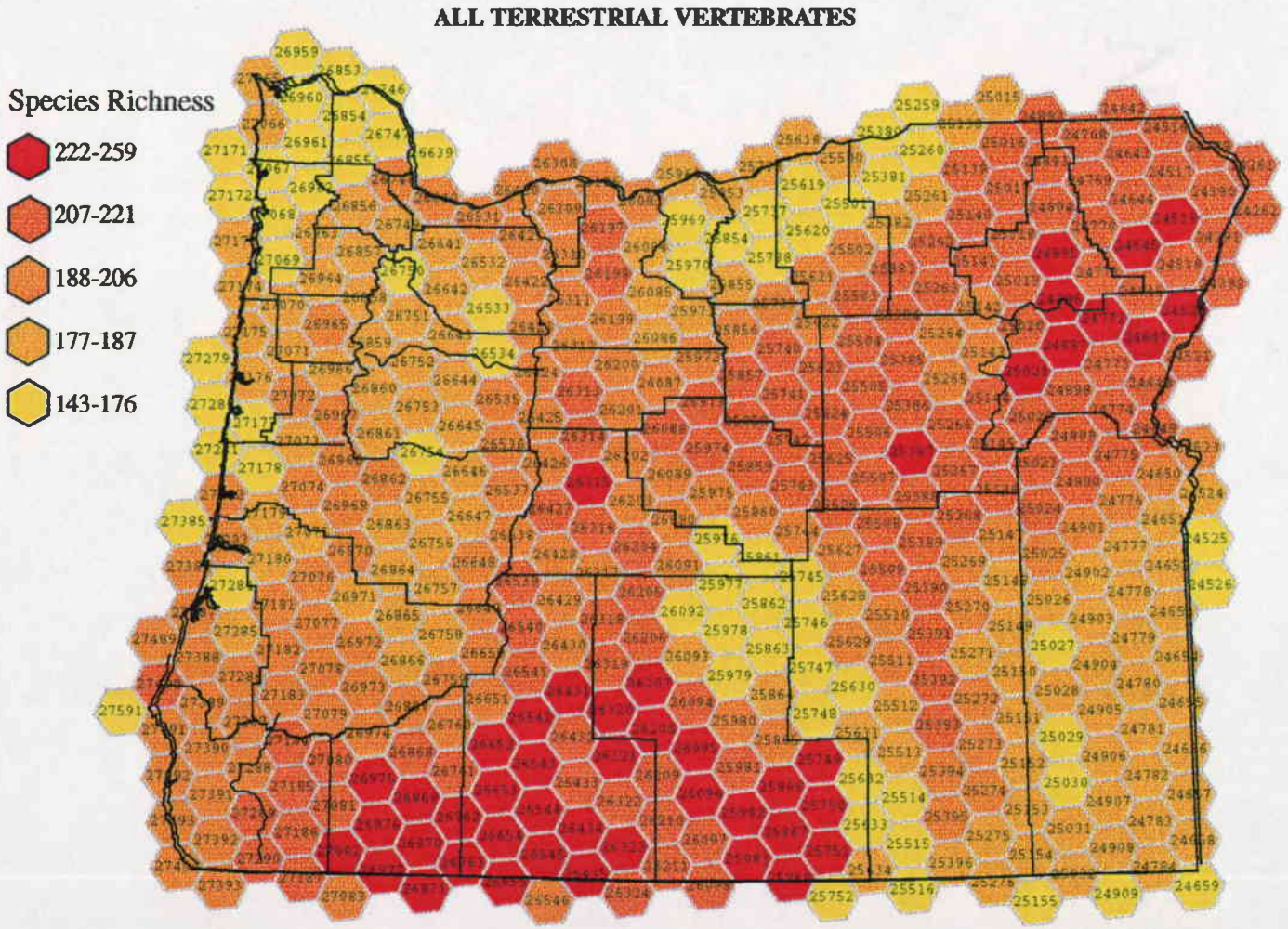


Figure 3.5 All terrestrial vertebrate richness for the state of Oregon.

Spatial statistics such as semivariance and Moran's I can be used as methods for model validation in a manner similar to classical statistics using residual plots in linear regression. Semivariance calculated on residuals of a spatial model can be used as a surrogate to determine randomness of residuals in space. I used semivariance and Moran's I to quantify what, if any, spatial pattern existed in the richness data and to check the validity of the results from my statistical analyses.

3.6 METHODOLOGY

3.6.1 *Species data - dependent variables*

The study area of the project was the state of Oregon. Species richness data was compiled on a statewide tessellation of 646 km^2 hexagons by The Nature Conservancy (TNC) for native mammals (Figure 3.1), breeding birds (Figure 3.2), reptiles (Figure 3.3), and amphibians (Figure 3.4) and reviewed by experts throughout the state. The distribution of all vertebrates combined is shown in Figure 3.5. All vertebrates were the sum of all vertebrate taxa within each hexagon. Each species was assigned an occurrence probability ranking for each hexagon. Recorded sighting and specimen collection locations were registered within the hexagon grid. These hexagons were termed "confirmed" and assigned a rank of 96-100% certainty that a given species occurred in that hexagon. The second ranking, "probable", was defined as 80-95% confidence that a given species occurred in that hexagon. Based on habitat and expert opinion, a species was assigned to a hexagon with the "probable" ranking if there was not a recorded specimen in that hexagon. This analysis included species given either "confirmed" or "probable" ranking.

3.6.2 *Explanatory variables*

Eleven indices derived from 1992 satellite imagery, three DEM variables, and two climate variables were analyzed for correlation with vertebrate species richness (Table 3.1). Eight of these variables were seasonal metrics which quantified the shape of the NDVI curve on a per pixel basis over the course of one year (Reed et al., 1994). At a pixel

level, three variables were very highly correlated (> 0.90) with two or more other variables and were excluded from the analysis. These were time integrated new growth (ting), value of NDVI at the end of the growing season (endv), and principal component 1 (pc1).

Table 3.1 Variables evaluated for correlation to vertebrate species richness

Variable		Definition/Interpretation
total integrated NDVI[1]	tot	net primary production during the growing season
time integrated new growth[1]	ting	primary production of new photosynthetic material
NDVI value at start of growing season[1]	onv	level of potential photosynthetic activity at beginning of growing season
NDVI value at end of growing season[1]	endv	level of potential photosynthetic activity at end of growing season
maximum NDVI value during growing season[1]	maxv	level of maximum photosynthetic activity
rate of greenup[1]	sup	acceleration of photosynthesis
rate of senescence[1]	sdn	deceleration of photosynthesis
range of NDVI[1]	range	range of annual photosynthetic activity
principal component 1	PC1	spatial representation of vegetation
principal component 2	PC2	seasonal pattern of phenology
principal component 3	PC3	non-growing season photosynthesis
elevation	elev	elevation
slope	slope	slope
aspect	asp	cardinal direction (aspect)
seasonal temperature difference	seas	difference between monthly mean July and January temperatures
precipitation	precip	rainfall

[1] *Adapted from Reed et al. (1994)*

The second and third principal components calculated on 1992 time series NDVI data were included as independent seasonal measures of vegetation dynamics. Three DEM derived variables were evaluated: elevation, slope, and aspect. Climate data

included precipitation and seasonal temperature differences (July minus January). Two datasets were constructed from the remaining 13 variables. One dataset included principal components 2 and 3 and the other did not include principal components 2 and 3.

3.6.2.1 *Seasonal metrics*

The seasonal metrics derived from AVHRR time-series NDVI satellite imagery quantitatively characterized seasonal phenological phenomena annually. The eight metrics evaluated for this analysis were chosen based on phenological relevance, units of measure (NDVI values), and whether or not each was a continuous versus discrete measure. These metrics quantified an annual NDVI curve thus capturing ecosystem dynamics (Reed et al., 1994).

Total integrated NDVI was defined as the net primary production during the growing season, calculated as the non-growing season baseline plus new growth during the growing season. This is an important variable in areas where baseline NDVI is consistently high, such as coniferous forests. In these systems productivity is likely occurring at NDVI levels below the standard 0.30 and the total integrated NDVI metric captures this phenomenon (Reed et al., 1994). Time integrated new growth only included productivity above the non-growing season baseline and is considered "new growth" or primary production of new photosynthetic material. Onset value of greenness and end value of greenness indicated the level of photosynthetic activity at the beginning and end of the growing season, respectively, in units of NDVI. Maximum value of NDVI was the greatest value of NDVI achieved over the course of the growing season and represented maximum photosynthetic activity. Rate of greenup and rate of senescence were interpreted as acceleration and deceleration of photosynthesis, respectively. Rate of greenup was the straight line slope of the curve between the onset of the growing season and the maximum NDVI. Likewise, rate of senescence was the straight line slope of the curve from maximum NDVI to the end of the growing season. Range of NDVI values was the difference between maximum NDVI value and end value and was interpreted as the range in photosynthetic activity (Reed et al., 1994). Quantitatively characterizing annual phenology into these metrics compressed a large dataset of 21 biweekly "raw" data layers

into 8 meaningful data layers. Collectively, these represented different facets of productivity on the landscape as interpretable measures. Composite totals of productivity can be misleading without examination of temporal curve distributions. For example, two landscapes with different vegetation might have similar total annual productivity appearing equally productive, yet have different temporal productivity distributions. In this situation the temporal productivity pattern would be necessary to discriminate ecological system differences.

3.6.2.2 *Phenology derived from time-series AVHRR biweekly composites*

Principal component analysis (PCA) was conducted on 1992 time series NDVI data. PCA was calculated for 21 biweekly composites of NOAA-AVHRR 1 km^2 pixel data for the state of Oregon for the year 1992 to capture the greatest spatial and temporal variability in the 21 scene dataset. This type of analysis is well documented by Cicone and Olsenholler (1997), Eastman and Fulk (1993), Fung and LeDrew (1987), and Tucker et al., (1985). Principal component 2 (PC2) represented seasonal vegetation growth independent of the first principal component and accounted for 4.5% of the annual variation in NDVI. Principal component 3 (PC3) was interpreted to be baseline photosynthesis occurring in primarily coniferous dominated or evergreen broadleaf forests during the winter months and as non-growing season vegetation characteristics.

3.6.2.3 *DEM and climate variables*

Terrain variables were calculated from 1 km^2 DEM obtained from the U.S. Geological Survey (USGS) EROS Data Center (EDC). Variables included were elevation (meters), slope (degrees), and aspect (degrees). Climate data were acquired for the state of Oregon as a subset from a larger database for the conterminous U.S. January and July temperatures and annual precipitation were compiled to a 1 km^2 rectangular grid. January and July mean temperature data were modeled and compiled using the method of Marks (1990). Monthly averages of forty year means, from approximately 1948 to 1988 were calculated at approximately 1200 stations in the Historical Climate Network database.

These values were first corrected to potential temperatures at a reference air pressure of 1000 Mb using the station elevations and assuming a normal adiabatic lapse rate. The potential temperatures were then interpolated to the 1 km^2 grid using a linear model. These interpolated values were then converted to estimated actual temperatures from the adiabatic lapse rate correction using corresponding elevation values at each grid point. Annual precipitation data were compiled from the 10 km resolution data to 1 km (Daly et al., 1994). Mean precipitation and seasonal differences defined as (mean July - mean January temperatures) calculated on a per pixel basis were summarized by hexagon as mean and variance. Seasonal difference and precipitation were also characterized by range, defined as (mean maximum value - mean minimum value) within a hexagon.

3.6.3 *Summarization of explanatory data to hexagons*

Data were collected at two different grains: the 1 km^2 pixel data including AVHRR-derived parameters and DEM variables, and climate and richness at 646 km^2 hexagons. The statistical grain of the study was therefore 646 km^2 hexagons and pixel data were aggregated up to the hexagon scale. Two summary statistics, median and variance, were calculated for all variables within each of the 434 hexagons in Oregon. Median was not calculated for aspect. Median values were not available for climate variables so mean values were used. Diversity, the number of different pixel values within a hexagon, was calculated for elevation, slope, and aspect as an additional measure of terrain variability. The final dataset of variables summarized to the hexagon level are shown in Table 3.2.

3.6.4 *Spatial statistical methods*

Two spatial statistical methods were used: semivariance and Moran's I. A semivariogram shows autocorrelation of data as a function of distance and when plotted represents spatial variability (Cohen et al., 1990; LeGendre and Fortin, 1989). Moran's I is a spatial autocorrelation coefficient which indicates significant patch size pattern in spatially distributed data. The two methods are complementary for evaluating spatial structure of autocorrelated data. Standardized semivariance and Moran's I with 95%

Table 3.2 Twenty-seven variables included in datasets for correlation with vertebrate richness and the summary statistic used to aggregate each up to the hexagon level.

	PC	Greenness	DEM	Climate
median and variance:	pc2 [1] pc3 [1]	maxv tot onv sdn range sup	elev slope	
variance:			aspect	seas
mean:				seas precip
diversity:			elev slope aspect	
total	4	12	8	3

[1] *included only in first dataset*

confidence intervals were calculated for each of the taxa and for residuals from each regression tree model.

Semivariance, or $\gamma(d)$, is a model of the average degree of similarity between observations as a function of distance (Rossi et al., 1992) and is calculated as

$$\gamma(d) = \frac{([1/2n_d) \Sigma [(y_{(i+d)} - y_{(i)})^2])}{\sigma}$$

where n_d is the number of pairs of points separated by distance *d*, *y* is the point value, and σ is the overall standard deviation (Legendre and Fortin, 1989). Standardization of semivariance, division by σ, allowed between dataset comparisons. Semivariance values range between 0, which indicates complete autocorrelation, and ∞, which indicates complete randomness in the data. Figure 3.6 shows the shape of a spherical

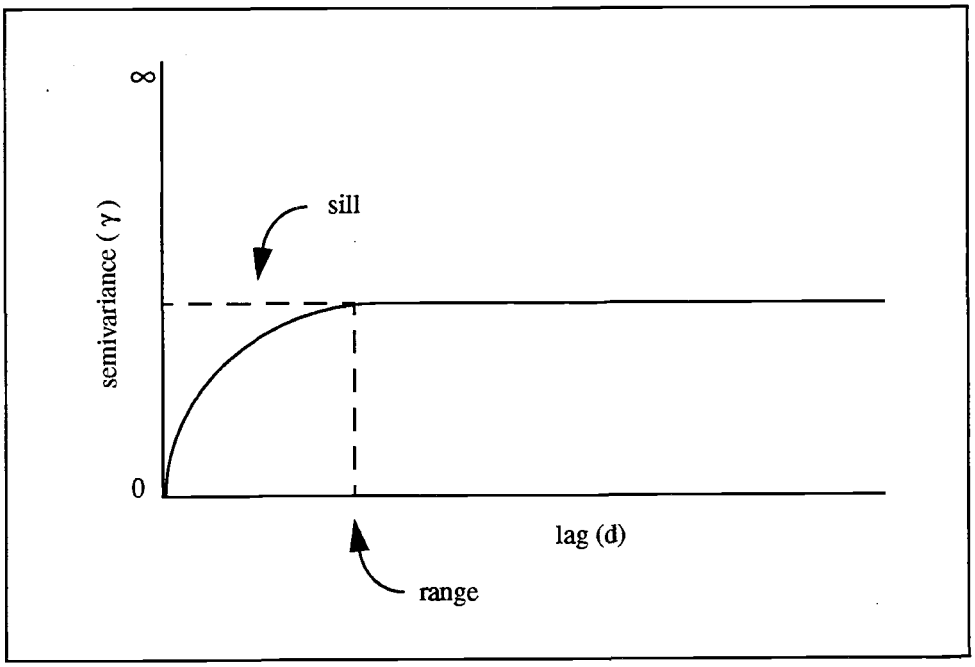


Figure 3.6 Theoretical semivariogram and the two variables used to describe spatial pattern in the data (sill and range). The sill is the value of semivariance where the curve levels off and range is the distance at which curve levels off. Lag (d) is the distance between pairs of points and semivariance is the degree of spatial autocorrelation.

semivariogram and the two parameters used to describe spatial pattern of the data. These two parameters are the *sill*, which is the value at which the curve levels off, and the *range*, which is the lag distance corresponding to the sill. The plot of Moran's I, or *I*(d), is called a correlogram. Moran's I is calculated as

$$I(\mathrm{d}) = \frac{[n \sum\sum w_{ij} (y_i - y)(y_j - y)]}{[W \sum (y_i - y)^2]}$$

where n is the number of data points, y is the point value, $w_{ij} = 1$ when the pair (i,j) pertains to distance class d or is 0 otherwise, and W is the sum of the wij's. Values of I(d) range between -1 and 1 with positive values corresponding to positive autocorrelation, zero indicating randomness, and negative values representing negative autocorrelation. Positive significant values indicate a patch size that is larger than the lag distance, or distance between pairs of points. Negative significant values show the distance between peaks and troughs. Patchiness is thus indicated by a series of significant positive and negative values (Legendre and Fortin, 1989). A 95% confidence interval computed at each lag distance was a test for significance where the null hypothesis was that the coefficient at a given lag distance was not significantly different from zero.

3.6.5 *Exploratory statistics - Regression tree analysis (RTA)*

Regression tree analysis (RTA) is a tree-based exploratory data analysis method which has been shown to be effective in identifying and estimating complex hierarchical relationships in multivariate data such as satellite-derived indices and DEM data (Michaelsen et al., 1994). The dataset is repeatedly partitioned into homogenous subsets using binary recursive partitioning until the entire dataset has been evaluated. Node splits are determined by deviance, where a split occurs at a given node such that the change in deviance is maximized and the variance of all resulting subsets is minimized. RTA is useful where there are expected but unknown non-linear or non-additive predictor-response interactions (Michaelsen et al., 1994). Regression tree analysis is a data-driven, non-parametric, and computer intensive method (Miller, 1994) which selects variables and values for splitting which best discriminate among responses (Efron and Tibshirani, 1991). Because this method results in overfit models, i.e. all data are categorized, regression trees must be pruned to select the most parsimonious subset which best explains the relationship between response and predictors. Cross-validation was used to select subtrees that have optimal predictive performance with lowest complexity. In a cross-validation, regression trees are computed on 9/10ths of the data and are checked with the 1/10th of the data withheld (Miller, 1994). Three cross-validation methods were used to determine the appropriate number of terminal nodes for the pruned trees: prune

method, one standard error rule, and adjusted minimum deviation. Each method was run on 10 iterations where the suggested number of terminal nodes was calculated from each method. For each iteration, the methods are as follows. The prune method divides the original dataset into 10 mutually exclusive data subsets and calculates respective deviances to determine optimal number of terminal nodes (Chambers and Hastie, 1993). The one standard error method is the calculated value of the cost-complexity parameter which is a linear combination of the cost of the tree and its complexity, plus one standard error. The greater the number of terminal nodes, the more complex the tree. The relative cost of adding an additional terminal node to a large tree is small (Breiman et al., 1993). The adjusted minimum deviation method applies a complexity penalty to the weighted linear combination of risk and complexity. This method results in selection of a less complex subtree without substantially higher deviance (Miller, 1994). Cross validation was run 10 times for each of the taxa. For each taxa, the median value of the 10 cross validations was calculated. Tree length was based on median number of suggested nodes from the 10 iterations. Further analysis as to why such differences existed in tree lengths was beyond the scope of this study.

Analyses resulted in numeric and graphical means for evaluating final RTA models. Final models from each RTA were plotted to show node splits, values at node splits, predicted average number of species per hexagon at terminal nodes, and number of hexagons at each terminal node. Semivariance and Moran's I were calculated on residuals from each RTA model by taxa as a means to evaluate randomness in residuals similar to using a residual plot to check the fit of a regression model.

3.7 RESULTS

Pairwise scatterplots of richness data by taxa is shown in Figure 3.7 and corresponding Spearman rank coefficients are given in Table 3.3. Birds (0.90) and mammals (0.62) had highest correlations with the all vertebrate dataset. Among the taxa, very weak negative correlations with both birds (-0.11) and amphibians (-0.10). birds had the highest correlation with mammals (0.36) and a negative correlation with amphibians (-0.20). Reptiles were weakly correlated with mammals (0.23) and showed very weak

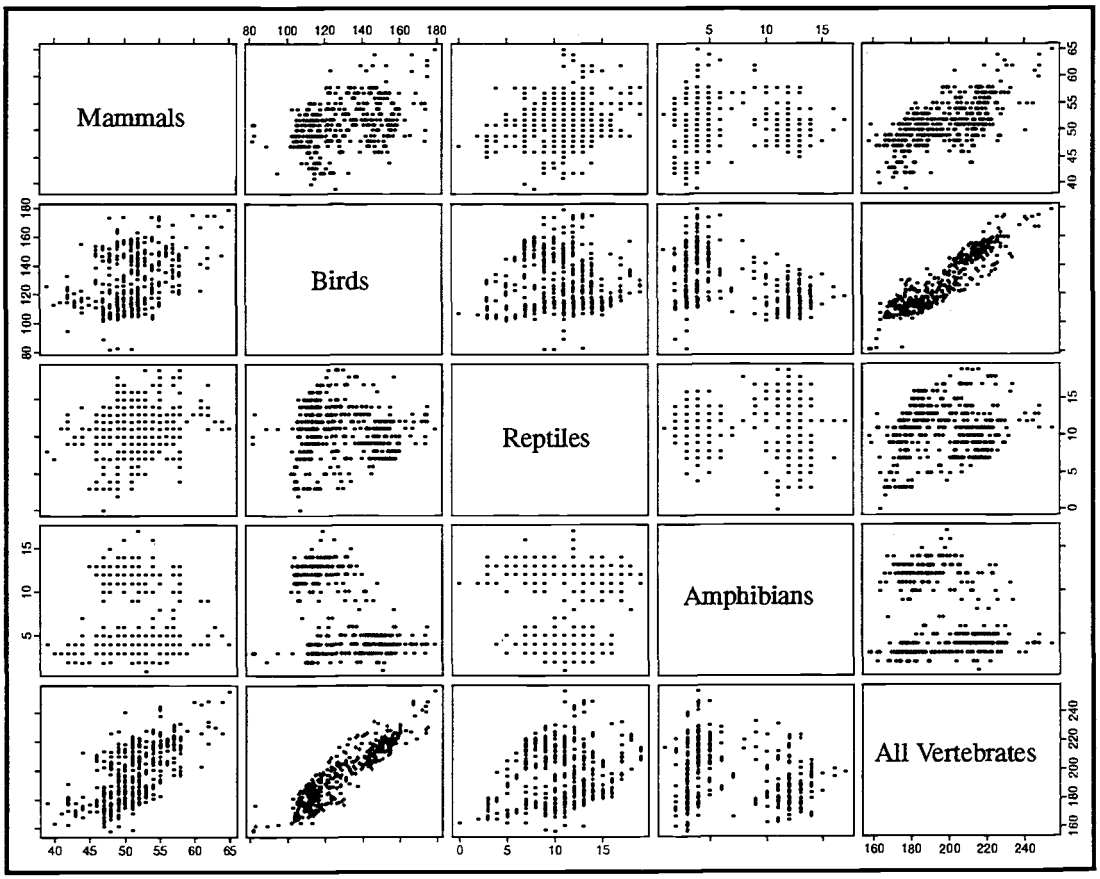


Figure 3.7 Pairwise scatterplots for response data: mammals, birds, reptiles, amphibians and all vertebrates combined.

negative correlations with both birds (-0.11) and amphibians (-0.10). Pariwise scatterplots for explanatory variables summarized to the hexagon level, at total of 27 variables, are shown in Appendix 1a - 1h. Corresponding correlation coefficients are given in Appendix 2a - 2h. Example results from the three cross-validation methods for the full dataset are given in Appendix 3.

3.7.1 *Regression tree analysis*

RTA explained between 55% and 91% of the variation in taxon richness for vertebrates in Oregon. Deviance explained for each regression tree model by taxon and optimal number of terminal nodes for each tree are given in Table 3.4. With the exception

Table 3.3 Spearman rank correlation coefficients for response data.

	mammals	**birds**	**reptiles**	**amphibians**	**all verts**
mammals	1.00	0.36	0.23	0.09	0.62
birds		1.00	-0.11	-0.20	0.90
reptiles			1.00	-0.10	0.11
amphibians				1.00	0.00
all verts					1.00

of amphibians, the RTA models from the reduced dataset (without principal components) were pruned to have nearly twice the number of terminal nodes as RTA models which inluded principal components. The deviance explained was higher for taxa modeled with the reduced dataset, however trees from the reduced dataset were nearly twice as complex, again with the exception of amphibians. The RTA model for amphibians was the same

Table 3.4 Deviance explained for regression tree models for full and reduced datasets. Number of terminal nodes for each dataset are the minimum suggested tree size from cross-validation.

	deviance explained				
	mammals	birds	reptiles	amphibians	all verts
full dataset	.55	.67	.57	.91	.66
reduced dataset	.62	.77	.75	.91	.79
	number of terminal nodes				
full dataset	11	6	9	5	7
reduced dataset	17	13	21	5	16

regardless of dataset and explained the most variance of all the RTA models. Regression trees become increasingly difficult to interpret as they grow with the addition of terminal

nodes. The three cross validation methods consistently indicated shorter trees with lower explained deviance for the full dataset. Deviance explained would have increased for these models if they were not pruned back as far. Likewise, the trees generated from the reduced dataset would have had lowered deviance explained were they pruned back farther. I selected the shorter trees for interpretation and present only the results from trees generated from the full dataset.

3.7.1.1 *Mammals*

The regression tree for mammals derived from the full dataset is shown in Figure 3.8 and is the largest model of the taxa. Figure 3.8 also shows prediction results of the tree plotted spatially. Seven different variables and eleven terminal nodes explained 55% of the variance of mammal richness across the state. Principal component 2 was the first variable split at median value of PC2 of 101.5. The mean, minimum, and maximum values of median PC2 were 91, 53, and 149, respectively. The right side of the tree was relatively simple, where mean seasonal temperature difference and median elevation predicted between an average 43 and 50 species in a total of 70 hexagons. The left side of the tree was more complicated. Two greenness variables, maximum value of NDVI and onset value of NDVI, two climate measures, precipitation and mean seasonal temperature difference, and principal component 3 predicted between an average of 44 and 60 species per hexagon. The median or mean summary statistic appeared 7 times while variance occurred only 3 times. Most terminal splits resulted in a difference of at least 5 species per hexagon. In the case of variance of PC3 < 39.8 there was a difference of only 2 predicted species per hexagon. The mean, minimum, and maximum values of PC3 were 57, 0, and 330, respectively.

The largest group of hexagons (111), which had an average predicted mammal richness of 51, was correlated with median PC2 less than 101.5, variance in maximum value of NDVI (maxv.var) less than 11.5, and mean precipitation greater than 283.5 mm. Mean, minimum, and maximum of maxv.var was 0, 28, and 267, respectively, and for precipitation was 210, 900, and 3720, respectively.

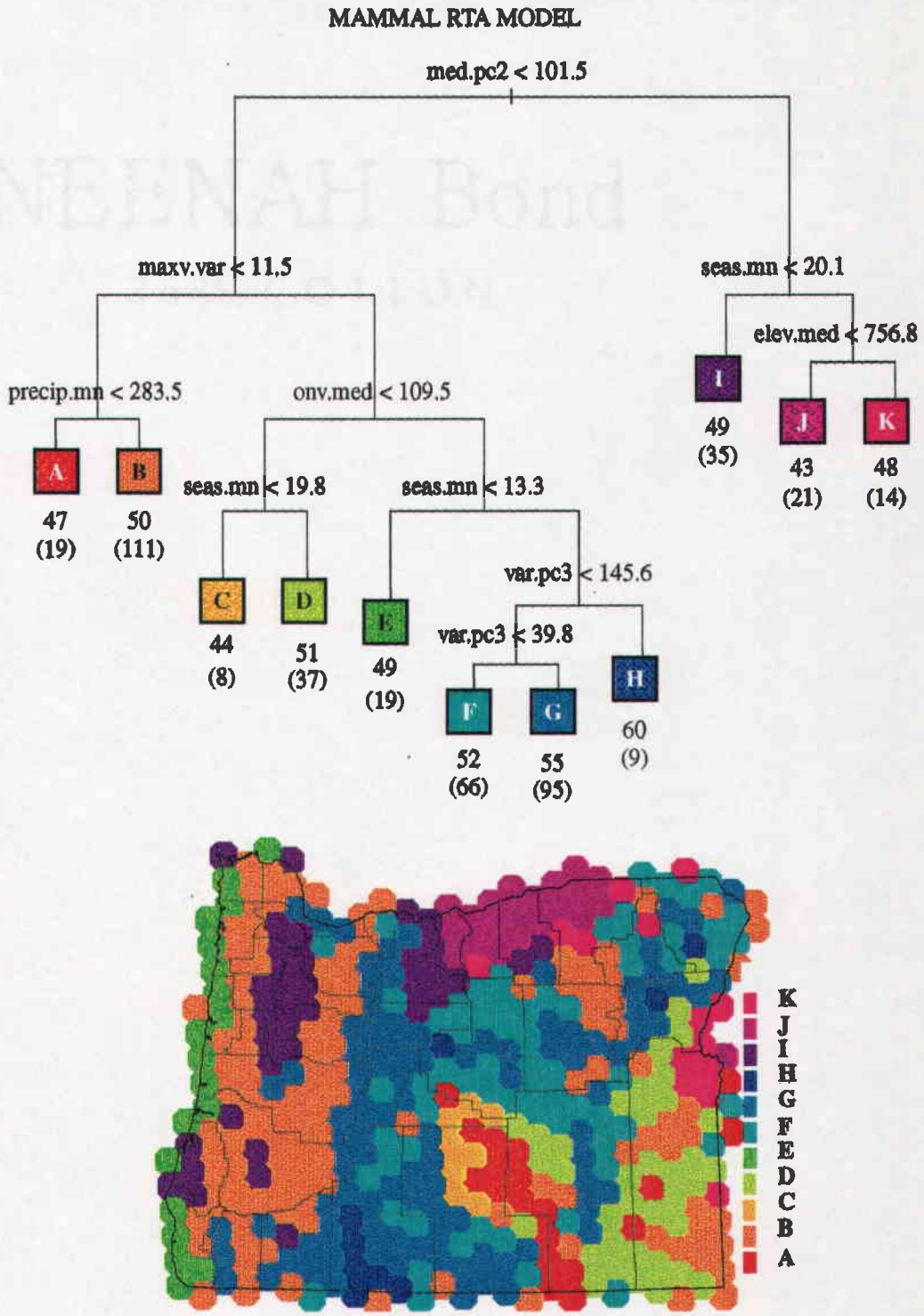


Figure 3.8 Mammal regression tree and spatial plot of regression tree results. Average number of predicted species are shown below lettered boxes and number of cases are given in parentheses.

3.7.1.2 *Birds*

The RTA model for bird richness (Figure 3.9) had 5 variables with 6 terminal nodes and explained 67% of the variance in richness in Oregon. Figure 3.9 also shows the predicted number of bird species plotted spatially based on the RTA bird model. The first split was mean seasonal temperature difference at 16.6 °C. One hundred twenty three hexagons were correlated with only the mean seasonal temperature difference less than 16.6 °C. These hexagons had lowest average predicted richness of 114 species. Variance of onset value of NDVI (onv.var) of 5.5 was the second split of the tree. The mean value of onv.var in Oregon was 23, with a minimum and maximum of 0.17 and 173, respectively. Where variance of onset value of NDVI was less than 5.5, variance in slope differentiated between average predicted richness of 115 species and 150 species per hexagon. Where variance of onset value of NDVI was greater than 5.5, elevation and net primary productivity during the growing season (tot) predicted between an average of 129 and 149 species per hexagon The largest group of hexagons (136), where predicted average richness was 149 species, were correlated with variance of onset value of NDVI greater than 5.5, median elevation greater than 954 m, and median annual productivity was greater than 42.5.

3.7.1.3 *Reptiles*

The reptile tree was the second most complex of the RTA models, second to mammals, with 9 nodes and 6 variables. Predicted reptile richness and the reptile RTA model is shown in Figure 3.10. Deviance explained for the reptile tree, also shown in Figure 3.10 was 0.57. Five of eight nodes were split on principal components starting with the first split of the tree at median value of PC3 < 75.5. The mean, minimum, and maximum values of median PC3 in Oregon was 56.5, 0, and 331, respectively. A very small number of cases were correlated entirely with principal components, where an average of between 11 and 16 species were predicted with median values of PC3 and PC2. The largest group of hexagons (156 cases), which had an average of 12 species predicted per hexagon, was correlated with median PC3 between 75.5 and 87.5, median range in NDVI less than 32.5, and variance in PC2 less than 63.5. Where the first split at median

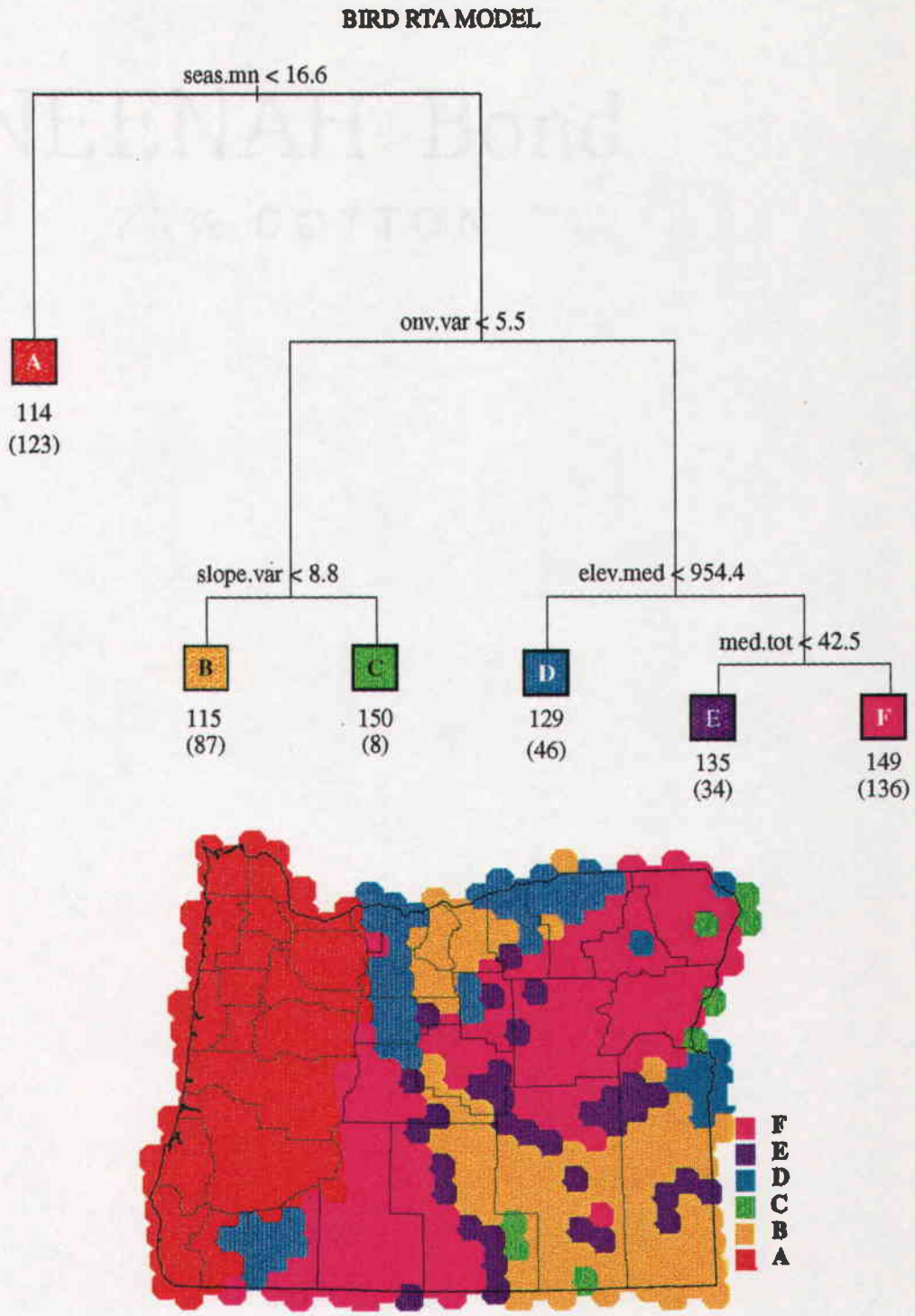


Figure 3.9 Bird regression tree and spatial plot of regression tree results. Average number of predicted species are shown below lettered boxes and number of cases are given in parentheses.

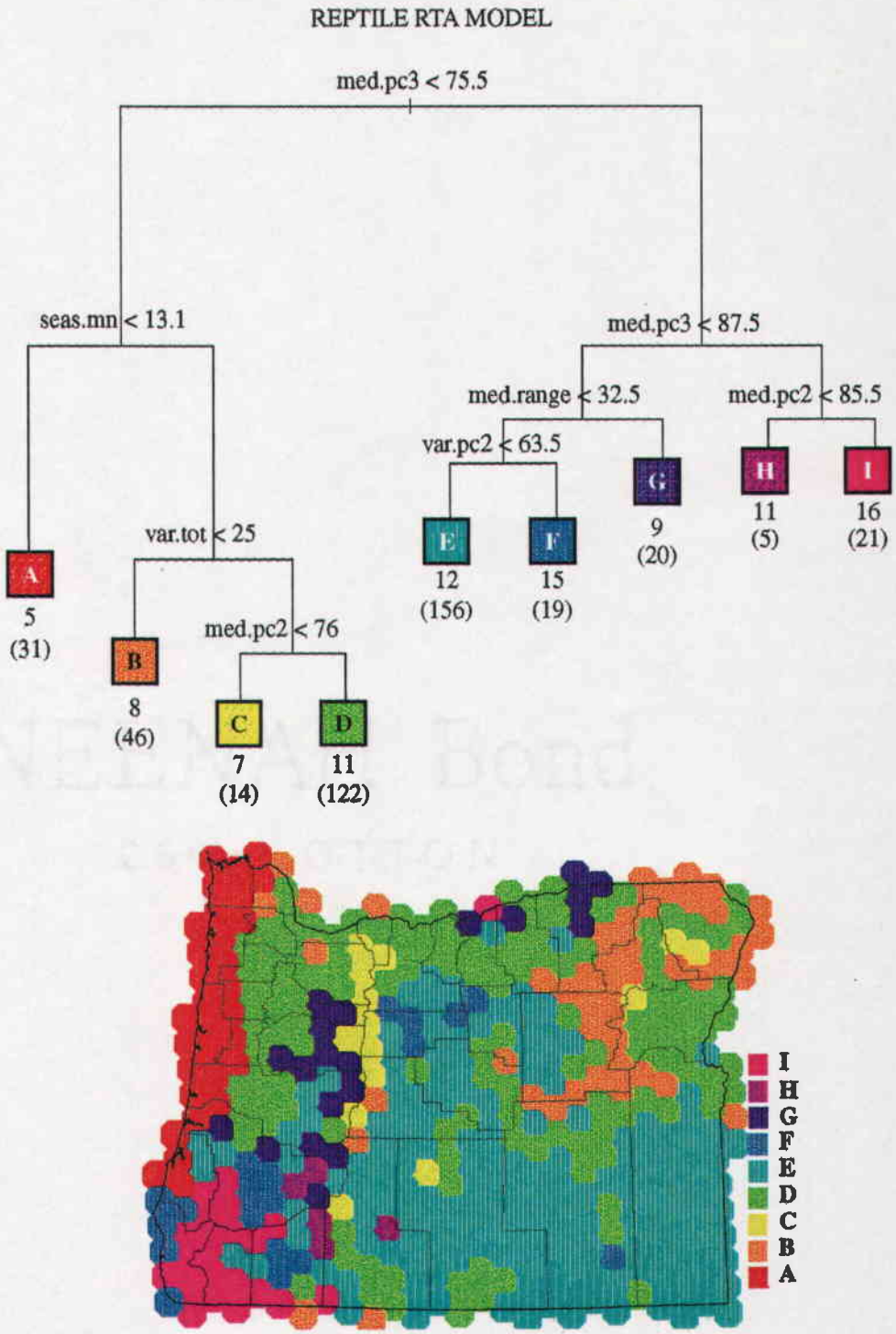


Figure 3.10 Reptile regression tree and spatial plot of regression tree results. Average number of predicted species are shown below lettered boxes and number of cases are given in parentheses.

PC3 was less than 75.5, the largest group of hexagons (122) were correlated with mean seasonal temperature difference greater than 13 °C, variance in net primary productivity during the growing season (var.tot) greater than 25, and median value of PC2 greater than 3.6.5. The mean, minimum and maximum values of var.tot statewide were 57, 0 and 304, respectively.

3.7.1.4 *Amphibians*

The amphibian model, shown in Figure 3.11, was the shortest RTA model and explained the most variability (0.91) of all the models. Prediction results of the amphibian RTA model are plotted spatially in Figure 3.11 as well. Identical trees were produced from both the full and reduced datasets. Five nodes and 4 variables correlated with amphibian richness in Oregon. Mean and median were the only summary statistics in this RTA model.

Mean seasonal temperature difference was overwhelmingly the most significant split at 17.8 °C. Where the mean seasonal temperature difference was greater than 17.8 °C, median onset value of NDVI (onv.med) split 283 hexagons into two groups. Of these 283 cases, 274 hexagons had an average of 4 predicted species where the median onset value of NDVI was less than 126.5. The mean, minimum, and maximum values of median onset value of NDVI statewide were 12, 105, and 149, respectively.

The remaining 151 hexagons were correlated with mean seasonal temperature difference was less than 17.8 °C. The greatest number of these hexagons, 111 cases, were correlated with median value of maximum NDVI greater than 142.5 and mean precipitation greater than 1217.9 mm. Thirty two cases correlated with mean precipitation less than 1217.9 mm.

3.7.1.5 *All vertebrates*

The tree for all vertebrates is shown in Figure 3.12 and explained 66% of all vertebrate variability with 7 terminal nodes and 5 variables. Figure 3.12 also shows the prediction results of the all vertebrate RTA model plotted spatially. The variance summary

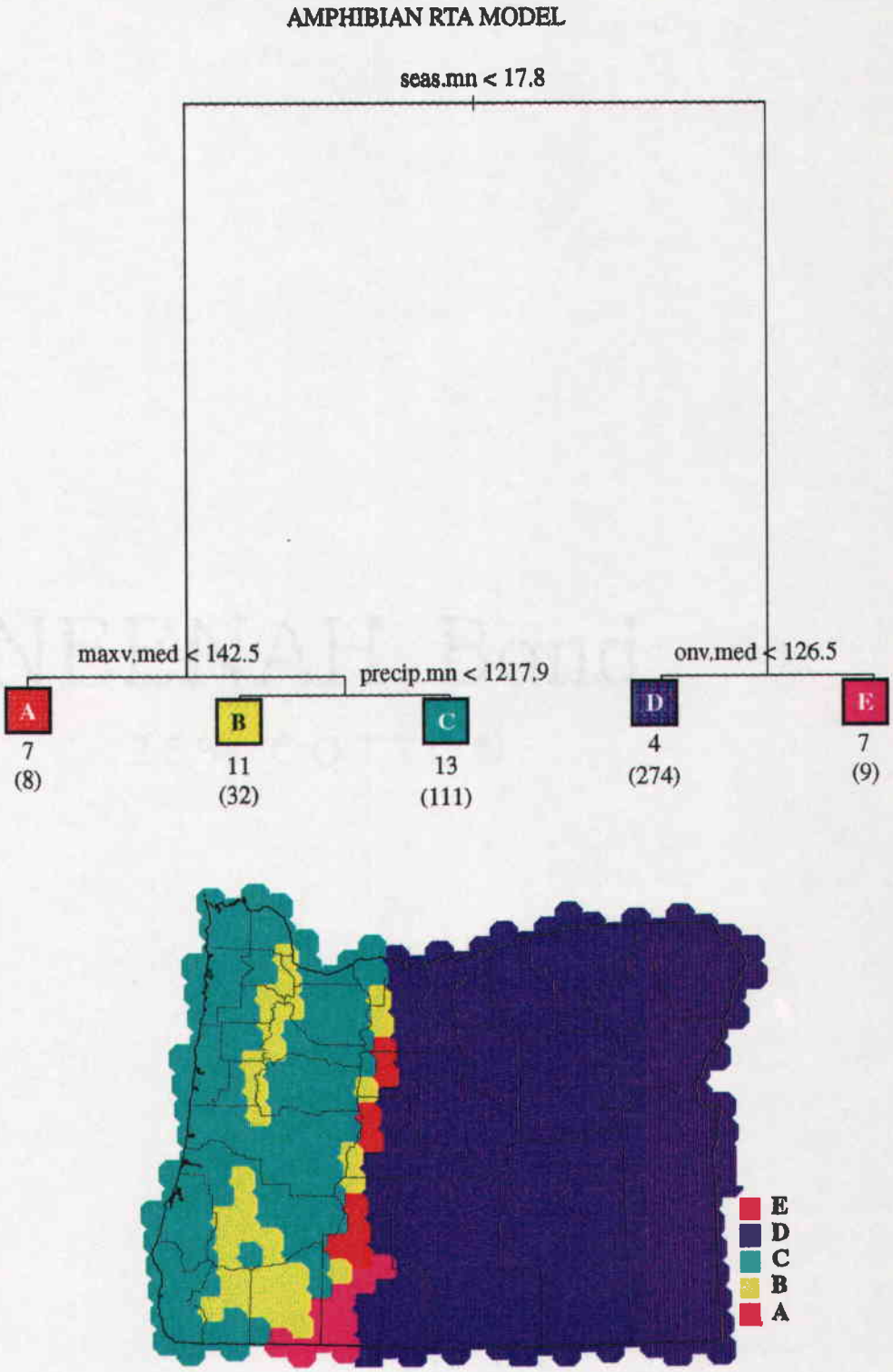


Figure 3.11 Amphibian regression tree and spatial plot of regression tree results. Average number of predicted species are shown below lettered boxes and number of cases are given in parentheses.

statistic appeared at 4 nodes. Median elevation less than 559.8 meters was the sole predictor of an average of 182 vertebrates in 105 hexagons. Statewide, the mean, minimum, and maximum elevation medians by hexagon were 1028, 19, and 2033, respectively. The largest group of hexagons (134 cases), which had 213 average predicted species, were correlated with median elevations greater than 559.8 m, variance of onset value of NDVI, mean seasonal temperature difference, variation of slope, and variance of PC3. Variance in slope appeared at two different nodes and predicted between 181 and 214 average number of vertebrates per hexagon. Of these, only 8 hexagons had 214 predicted species.

3.7.2 *Spatial statistical results*

Semivariance and Moran's I were calculated for the richness data by taxon and for the residuals from each regression tree model. This was done to check how well the regression trees fit the data and to determine how well the models dealt with autocorrelated richness data. Like residual plots from classical regression methods, no pattern was expected from models which fit the data well. Lines on semivariograms and correlograms in all figures are visual best-estimate curve approximations.

3.7.2.1 *Mammals*

Semivariograms and correlograms for mammal richness data (A, C) and for residuals (B, D) from the mammal tree are shown in Figure 3.13. There was significant spatial autocorrelation in the mammal data to a lag of 180 km or an approximate neighborhood of 6 hexagons. There appeared to be a potential second sill which might have been significant had the calculations been carried out beyond 480 km. Moran's I was significant and positive to a lag of 60 km, indicating a patch size of 2 neighboring hexagons. At the maximum lag distance of 480 km there was some indication of significant negative autocorrelation. The range of the semivariogram for mammal RTA residuals was 120 km and the correlogram indicated a weak positive significance at a lag

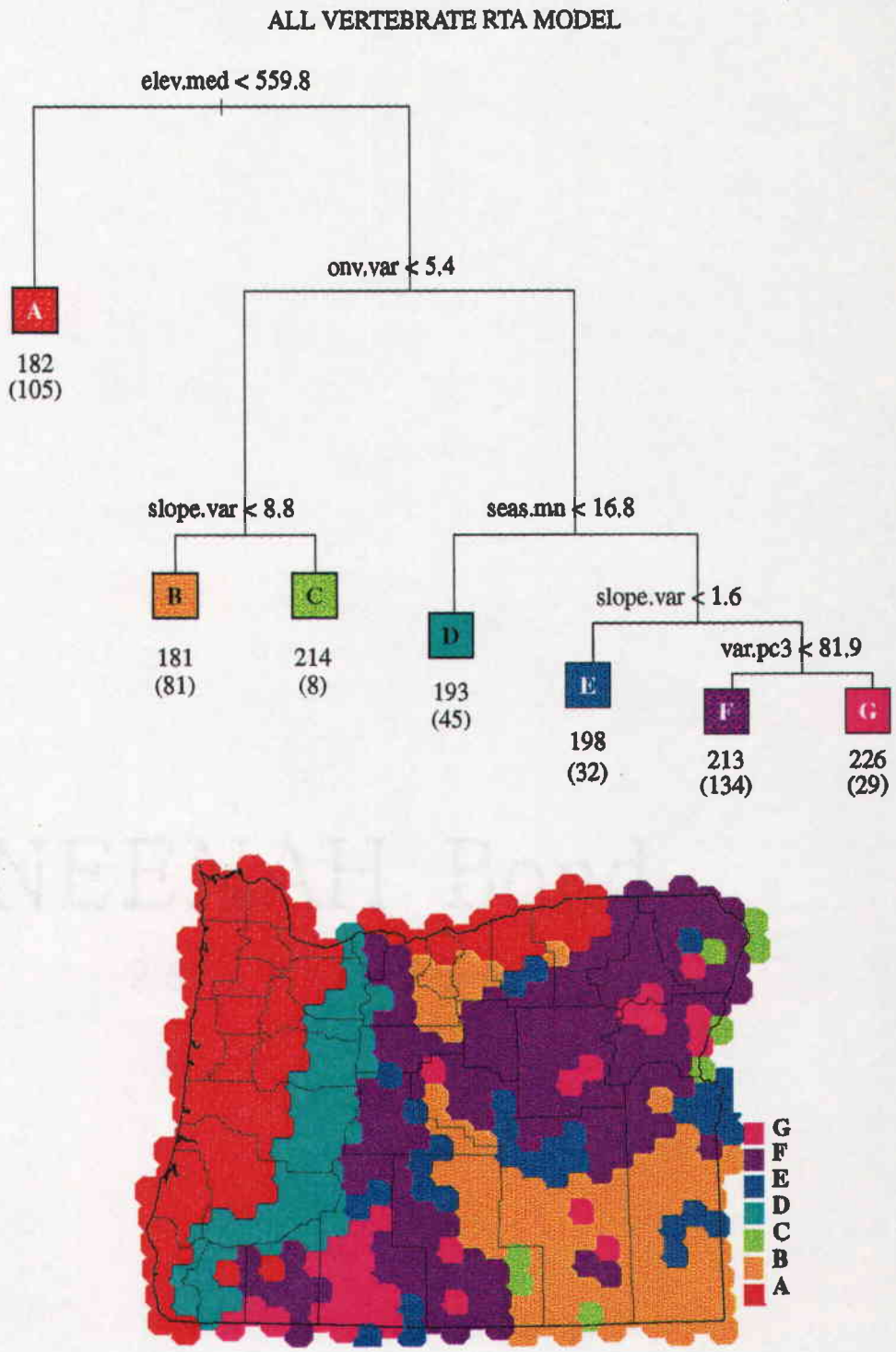


Figure 3.12 All vertebrate regression tree and spatial plot of regression tree results. Average number of predicted species are shown below lettered boxes and number of cases are given in parentheses.

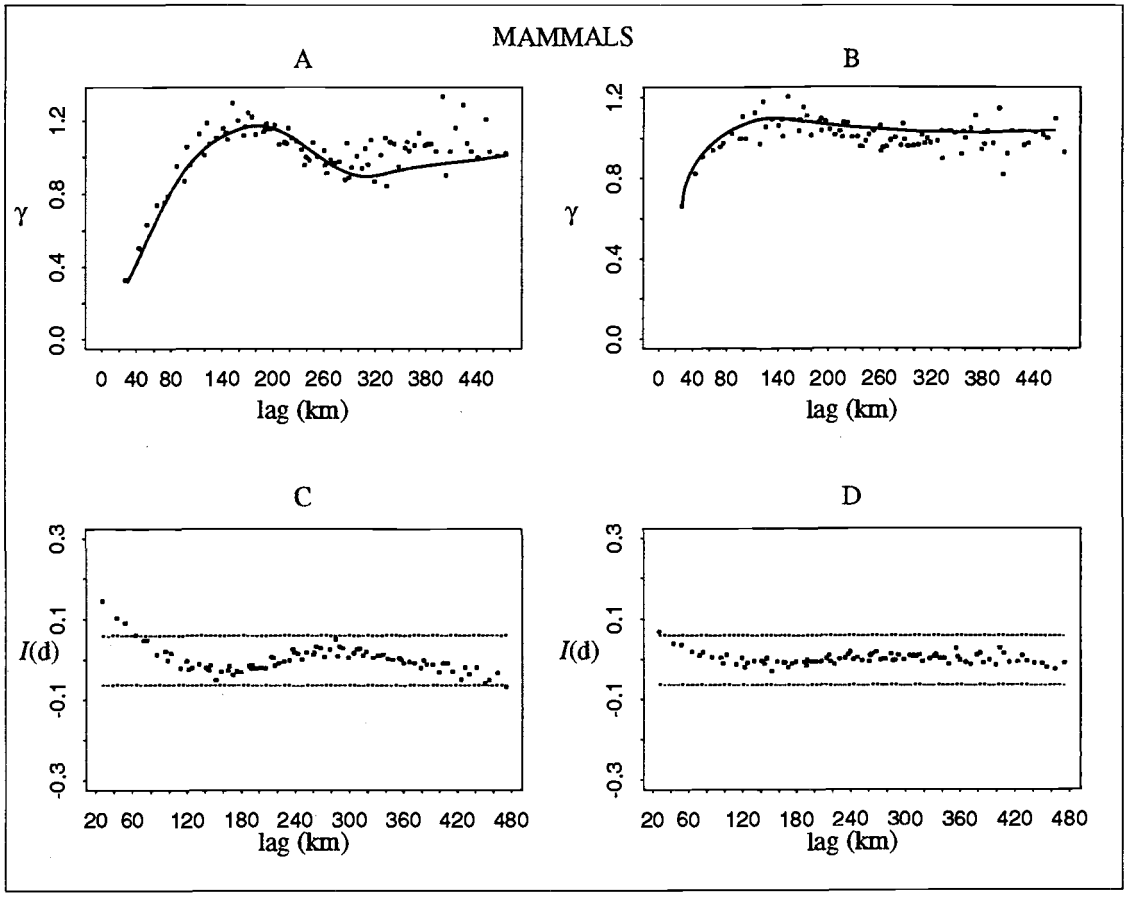


Figure 3.13 Semivariogram (A) and correlogram (C) for mammal richness in Oregon. Semivariogram (B) and correlogram (D) for residuals from mammal RTA model. Dashed lines on correlograms indicate 95% confidence intervals.

equivalent to 1 hexagon. The amplitude seen in the plots of richness data was noticeably dampened in the residuals.

3.7.2.2 *Birds*

Bird richness and bird RTA residual semivariograms and corresponding correlograms are shown in Figure 3.14. Bird richness exhibited spatial autocorrelation to 340 km, or nearly 12 neighboring hexagons. Moran's I coefficients were significant and positive to a lag of 100 km. The residuals from the bird RTA tree semivariogram (A) and correlogram (C) showed autocorrelation to lag distances of 200 km and 27 km, respectively. Some coefficients on the correlogram were weakly and negatively significant at a 95%

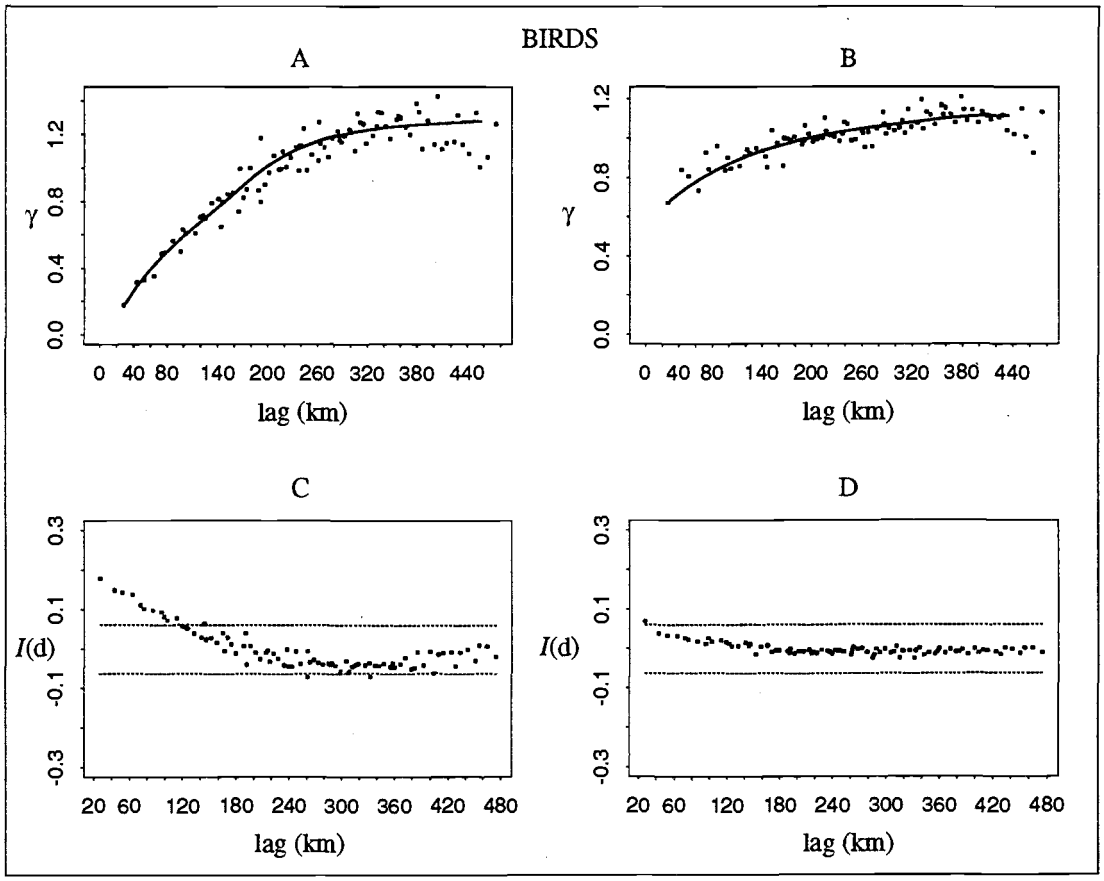


Figure 3.14 Semivariogram (A) and correlogram (C) for bird richness in Oregon. Semivariogram (B) and correlogram (D) for residuals from birds RTA model. Dashed lines on correlograms indicate 95% confidence intervals.

confidence interval. This was not apparent in the correlogram of bird RTA model residuals.

3.7.2.3 *Reptiles*

The semivariograms and correlograms for reptile richness data and RTA model residuals are plotted in Figure 3.15. Semivariance of reptile richness (A) had a range of 240 km and Moran's I coefficients were positive and significant to a lag of 80 km, or 3 neighboring hexagons. The RTA residual semivariogram (B) had a range of 140 km. The corresponding correlogram (D) indicated RTA residuals were positive and significant to 1

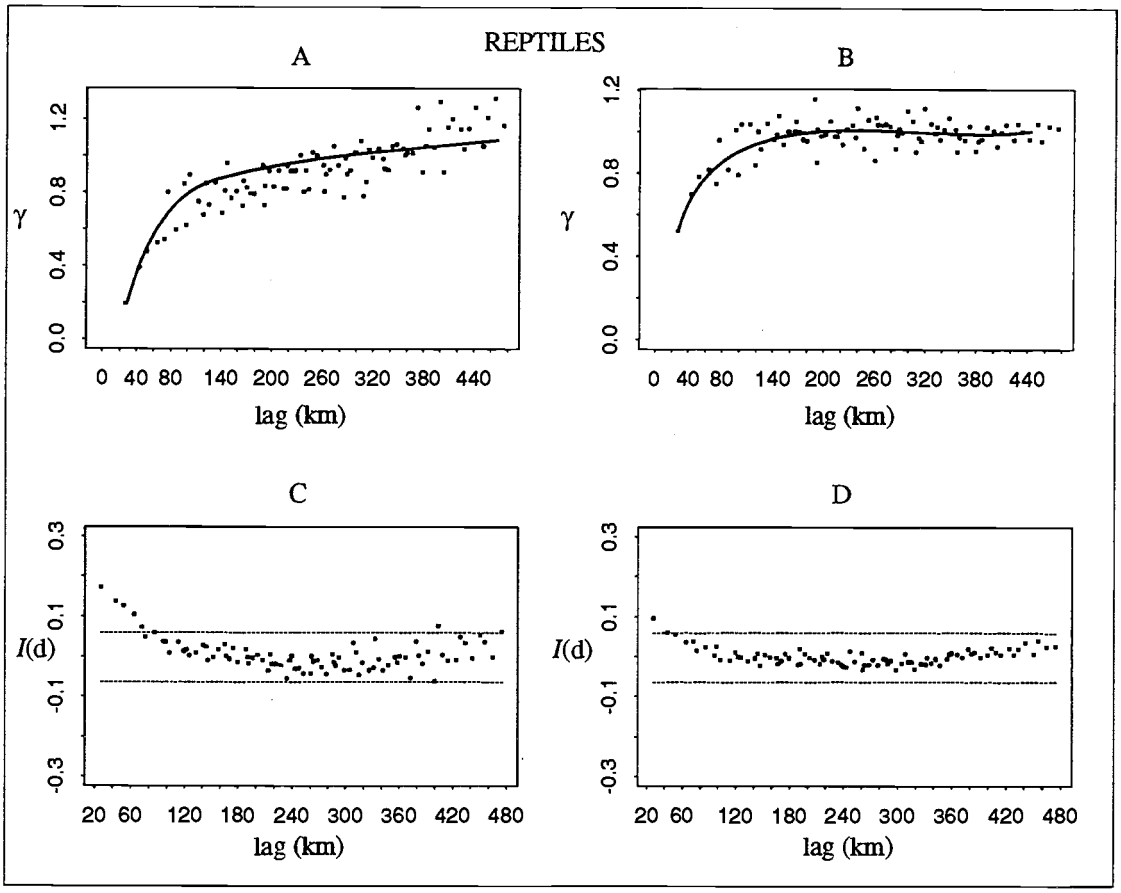


Figure 3.15 Semivariogram (A) and correlogram (C) for reptile richness in Oregon. Semivariogram (B) and correlogram (D) for residuals from reptile RTA model. Dashed lines on correlograms indicate 95% confidence intervals.

neighboring hexagon. The increasing trend in semivariance which was apparent in the richness data was not found in the semivariogram of the reptile RTA model residuals. Moran's I coefficients for reptile residuals were consistently closer to 0 than the coefficients for the corresponding reptile richness data.

3.7.2.4 *Amphibians*

The semivariogram for amphibians (A), shown along with the corresponding correlogram (C) for taxa and for residuals in Figure 3.16, is called an exponential semivariogram. This type of semivariogram does not have a sill and likewise lacks a range. This indicated a spatial gradient, which in the case of amphibians in Oregon runs

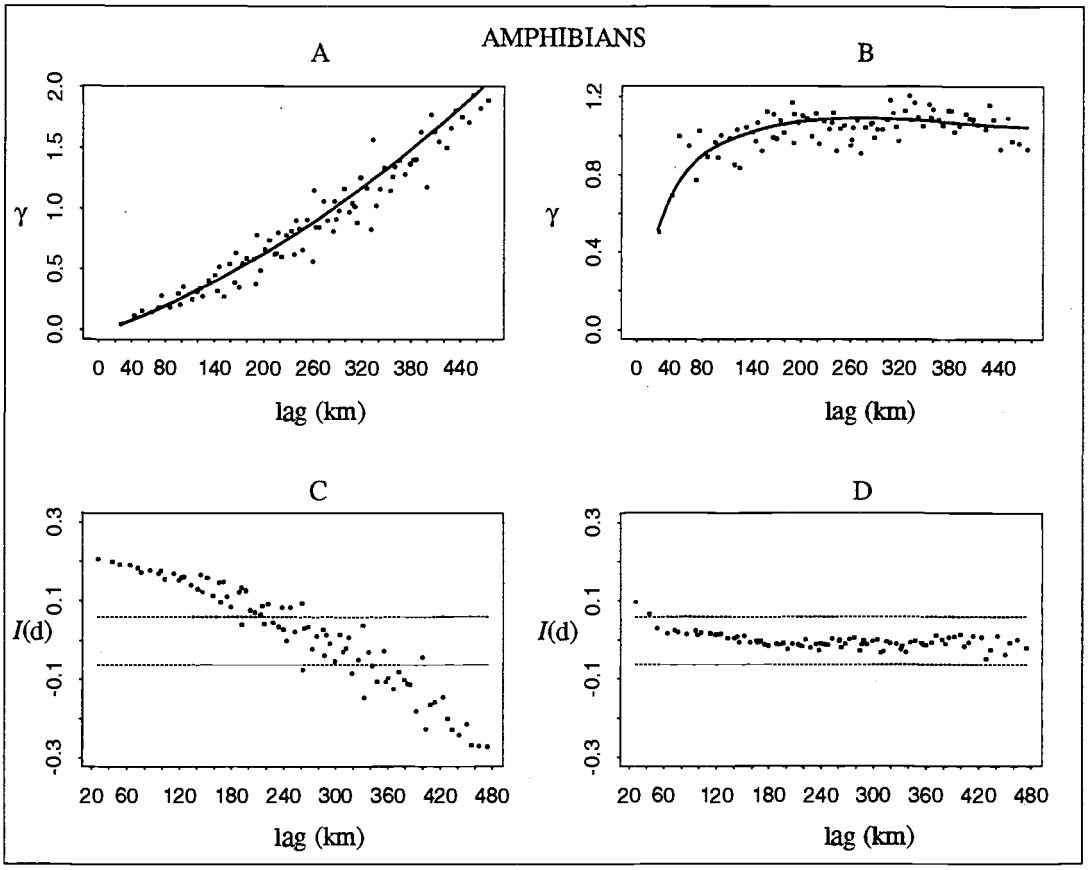


Figure 3.16 Semivariogram (A) and correlogram (C) for amphibian richness in Oregon. Semivariogram (B) and correlogram (D) for residuals from amphibian RTA model. Dashed lines on correlograms indicate 95% confidence intervals.

east-west. This was also supported by the significant positive and negative coefficients in the corresponding correlogram (C). Significant positive correlations exist to a lag of 260 km and significant negative correlations begin at a lag of 320 km. The semivariogram (B) and correlogram (D) of RTA residuals showed no indication of spatial trend. The correlogram indicated significant positive correlation at smallest lag distances of 27 km.

3.7.2.5 *All vertebrates*

Results from calculation of semivariance and Moran's I for all vertebrates are shown in Figure 3.17. The semivariogram for all vertebrates (A) had a range of 320 km and the Moran's I coefficients (C) were significant and positive to a lag of 80 km. The

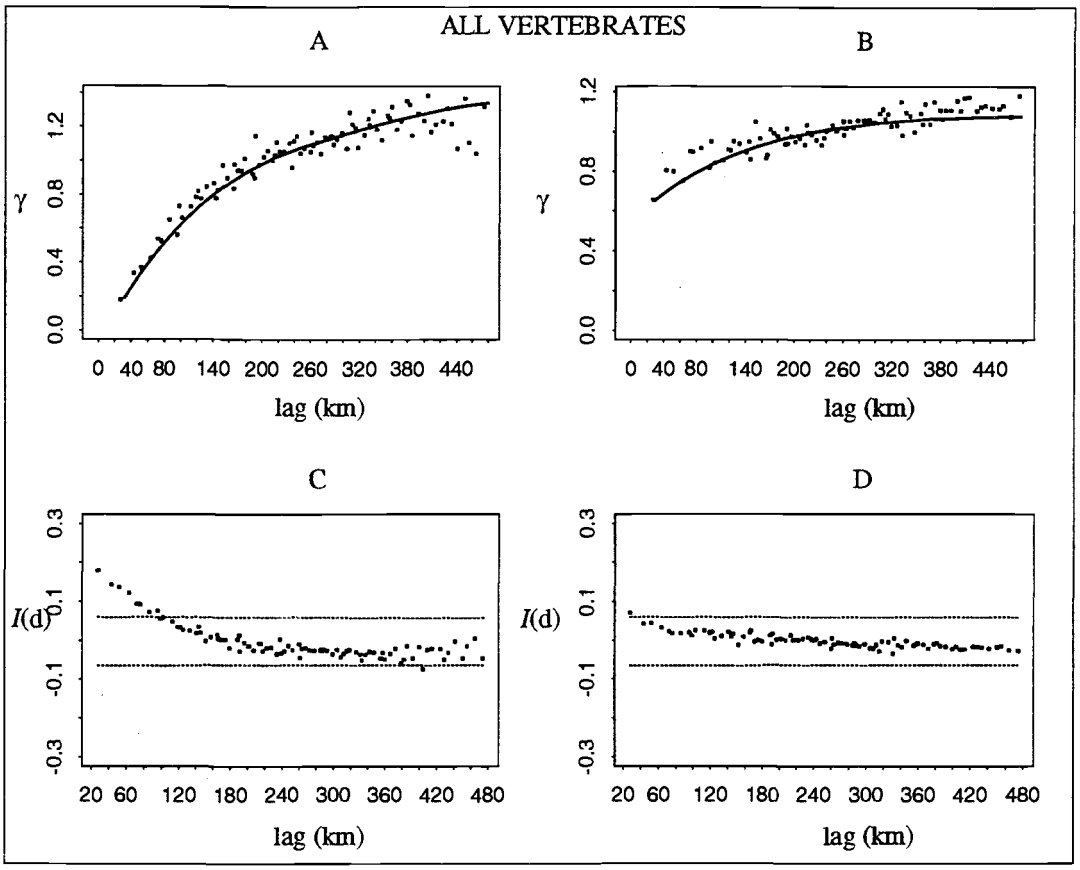


Figure 3.17 Semivariogram (A) and correlogram (C) for all vertebrate richness in Oregon. Semivariogram (B) and correlogram (D) for residuals from all vertebrate RTA model. Dashed lines on correlograms indicate 95% confidence intervals.

semivariogram and correlogram for residuals indicated a significant decrease in autocorrelation. The semivariogram (B) and correlogram (D) for all vertebrates resembled the semivariogram and correlogram of birds. This was not surprising given the overwhelming numbers of bird species relative to the other taxa (70% of total vertebrates).

3.8 DISCUSSION

I was interested in how well measures of environmental heterogeneity (phenology, climate, and elevation) correlated with or explained modern patterns of vertebrate species richness. The first step in this analysis was to match sample sizes between 1 km^2 pixels and 648 km^2 hexagons. Two options were considered for selecting a grain size for this analysis. The first was to resample the species richness maps down from hexagons to

pixels, which was not ecologically justifiable. The second option was to summarize the pixel data up to the hexagon level. Resampling pixels to hexagons was not feasible because of the mixed mapping unit shapes (squares to hexagons).

The spatial autocorrelation of the species data itself presented an interesting challenge for statistical fitting, and RTA effectively modeled much of the spatial variability and autocorrelation of the original data. For any given explanatory variable by hexagon, $x_{(i,j)}$, the value in a neighboring hexagon, $x_{(i+n,\ j+m)}$, was expected to have a similar value. In most cases, RTA models captured this spatial variability. Model residuals reflected low level noise and were spatially random. This was supported by the deviance explained in combination with interpreted semivariograms and correlograms of the individual taxon distributions. Taxa with more complex spatial patterns and greater autocorrelation were more difficult to model and this was reflected in corresponding lower values of deviance explained. Statewide amphibian distribution, for example, was the simplest pattern, had the most variability accounted for in the final model (0.91), and a random spatial residual plot. All taxa had significant positive autocorrelation and RTA effectively modeled these groups evidenced by random semivariograms and non-significant values of Moran's I of model residuals. Mammal and reptile distributions exhibited the most complex spatial patterns and were likewise the most difficult to model. RTA explained a little over half of the variability in mammal (0.55) and reptile (0.57) richness data.

How well the current spatial patterns of vertebrates and their predictors can be related to events from an evolutionary time scale is difficult to quantify. The number and variety of physiographic provinces in Oregon presented quite a challenge for such an interpretation, however, as discussed above, some patterns were readily apparent and explanation of these relationships are discussed.

3.8.1 *Taxonomic models*

3.8.1.1 *Mammals*

At large scales, such as the continental U.S. or North America, mammals have an east-west richness distribution gradient where more mammal species are found in the west

than in the east (Currie, 1991; Kiester, 1971; Simpson, 1964). This does not hold true for smaller areas such as Oregon (see Figure 3.1) or Texas (Owen 1990). Mammal distribution in Oregon does follow Currie's (1991) statement that 'homeotherms have highest species richness in high energy, dry, mountainous areas'. Both Kiester (1971) and Owen (1990) found positive relationships between mammals and elevation, but elevation occurred only once at a terminal split in this analysis (Figure 3.8). Owen (1990) also found negative correlations with temperature and mean seasonal temperature difference which occurred at three different nodes on both sides of the mammal tree. He (Owen) was suspicious of productivity as a correlate to mammalian richness citing contradictory results in literature to that time. However, results reported by Wright (1983) and Currie (1991) provided strong evidence to the contrary, that available energy is correlated with vertebrate richness, and results from this study concur.

The first split of the mammal tree was median value of PC2, which represented seasonal variation over one year. The first split at PC2 produced two very different subtrees: a complex set of branches which predicted an average richness ranging between 44 and 60 species per hexagon, and a simple set of branches which predicted between 43 and 48 species per hexagon. While the mapped results of the RTA in Figure 3.8 appeared somewhat noisy, there are some very contiguous patches. For example, much of the mountains of western Oregon, mapped as (B) had low average seasonality, relatively homogenous spatial distribution of productivity, and high rainfall. There were other regions of Oregon (B) which made less sense, such as the hexagons east of the Cascades. As with any classifier, there were errors in classifications. Another area which appeared to have been modeled reasonably well were the hexagons along the Columbia River (J). These hexagons were predicted to have 43 species, which was similar to actual species counts. Three variables described conditions in those hexagons: high seasonality, high mean seasonal temperature difference, and relatively low median elevation.

Highest mammal richness in Oregon (G and H), was related to spatial heterogeneity of vegetation conditions. This could be taken further to say that mammal richness was correlated with habitat heterogeneity if different habitats are defined as having different phenology. With this interpretation, results were consistent with Owen (1990) who interpreted variation in elevation as greater numbers of habitat types. Given the area

covered by these hexagons it might also be interpreted that it was patch heterogeneity defining habitats rather than different vegetation compositions. The semi-arid and arid conditions of eastern Oregon and the coarse scale of pixel data would have reflected very differently depending on vegetation densities and interpatch conditions. The two branches which predicted highest richness had 5 and 6 explanatory variables, of which at least two were principal components. More than half of the explanatory variables for highest mammal richness were variance summaries, which indicated a relationship with spatial heterogeneity. Highest mammal richness was correlated with low seasonality, high variation in spatial configuration of NDVI, high values of NDVI at the beginning of the growing season, higher average seasonal temperature differences, and spatial heterogeneity of non-growing season productivity.

Along the Oregon coast, 19 hexagons were predicted to have 49 species per hexagon (E) on average. For the most part, the adjacent hexagons were predicted to have only 1 additional species (B), described above. Coastal hexagons were correlated with low seasonality, high spatial variation in NDVI, high onset value of NDVI, and low mean seasonal temperature difference. These variables describe a very stable and mild climate in terms of vegetation where productivity is also quite high. Given the variability of non-coniferous forest along the coast, such as estuarine vegetation, it is not surprising that these hexagons were correlated with different variables than were those of the adjacent hexagons to the east. What was confusing about these results was the inconsistency with the bird RTA results for hexagons along the coast. Two different sets of variables were used to discriminate between a difference of 1 species for mammals but for birds, which had a potential difference of 40 or 50 species, the same variables did not differentiate hexagons. Results from bird RTA are discussed below.

The spatial pattern of mammals in Oregon, quantified in Figure 3.13, was the most complex pattern of the taxa. This may have contributed to the low percentage of variance explained (55%). In fact, the mammal RTA model had the lowest deviance explained of all the taxa. Results from calculated semivariance and Moran's I indicated the model reduced significant autocorrelation in the richness data, but there still existed some pattern to the residuals. Unlike the amphibian distribution, mammals exhibited correlation at two detectable patch sizes and as a result was the most difficult pattern to explain. In essence,

simple patterns on the landscape were easier to explain than more complex patterns. Mammals had a more complex pattern and were therefore more difficult to model.

3.8.1.2 *Birds*

Highest bird richness in Oregon occurs in southern Oregon just east of the Klamaths and in the mountainous northeast region bordering Idaho. The regression tree for birds effectively modeled these two regions (F) in Figure 3.9. Mean seasonal temperature difference, which was the first splitting variable for the bird data, discriminated eastern from western Oregon. Highest bird richness (F) in eastern Oregon corresponded to higher terrain and was highly correlated with two greenness variables: onset value of NDVI and annual productivity. Higher values of these two variables predicted higher bird richness. For those areas of eastern Oregon which were relatively flat, namely the northern irrigated region along the Columbia river and the southeast corner of the state, lowest bird richness was predicted by lowest onset values of NDVI and low terrain variability. For these areas (B), 115 species were predicted for each hexagon. The hexagons mapped in yellow (B) described conditions which had high seasonal fluctuations but had low productivity and were spatially homogeneous. Hexagons (E), which predicted an average of 135 species, captured the boundary between the highest and lowest richness of eastern Oregon. The difference between these hexagons and the magenta (F) hexagons was productivity.

More than one quarter of the bird richness in Oregon was captured by mean seasonal temperature difference alone. These hexagons (A) represented the west side of the Cascades. This was interesting because given the topography of the Cascades and the Coast Ranges, a terrain variable might have been expected to be significant for these areas. There is a relationship between productivity and elevation, however, and with this data productivity was the parameter which best explained the variation in the data.

In general, RTA underpredicted bird richness, particularly in western Oregon. Actual bird richness is as high as 195 species but the highest number of predicted species on average by the RTA model was 150, and this was only in 8 cases. However, with this analysis I was interested in the ecological processes and not the numerical prediction accuracy. Most of the high values of bird richness were modeled with seasonal

temperature difference, phenology, and elevation (136 hexagons with 149 species predicted). The model explained 67% of the variation in bird richness and was the second-best fitting model of the individual taxa. By visual accounts of the patterns in Figure 3.9, the RTA model showed similar patterns to actual bird richness. The spatial statistics also indicated randomness in residuals. Based on these assessments the RTA for birds, while imperfect, was not a bad model.

3.8.1.3 *Reptiles*

Although the deviance explained for reptiles was 0.57, the results of RTA were consistent with actual patterns of reptiles in Oregon (Figure 3.3). Seven of the eight variables in the reptile richness tree were phenological variables. The lack of elevation variables in the reptile tree somewhat agrees with Owen (1990) who found no relationship between herps and elevation. Currie (1991) stated that 'poikilotherms show greatest species richness in high energy, moist, low elevation areas'. In Oregon, highest reptile richness occurs in the southwestern corner of the state which borders California. This area has terrain variability but as a whole is not exceptionally high in terms of elevation above sea level. It is interesting to note that concentrations of highest mammal and reptile richness occur at the same latitudes in the same general region of the state, but mammals have highest richness just east of highest reptile richness. Kiester (1971) suggested a complementarity between reptiles and mammals and suggested historical competitive exclusion between the two taxa. The low correlation coefficient (0.23) lends evidence to this in Oregon. While the scope of this data was limited to Oregon, it does at least suggest that mammals and reptiles are thriving under similar conditions.

Lowest reptile richness was predicted for the coast of Oregon (A), which agreed with actual reptile richness data counts. These hexagons had lower non-growing season productivity and mild annual temperature differences. The largest group, which had a predicted average of 12 species per hexagon for 156 hexagons in Oregon (E), was highly related to phenology. These hexagons (E), primarily in eastern Oregon, had intermediate non-growing season productivity, lower average range of NDVI, and lower seasonality. The next largest group of hexagons (122) had a difference of only 1 species per hexagon

(D) from (E) and corresponded to the other main branch of the tree, where non-growing season productivity was lower. These hexagons (D) had a higher seasonal temperature difference, higher spatial variation in total productivity, but similar seasonality captured by PC2 for (E).

Highest reptile richness (I and F) corresponded with actual reptile data counts and were related entirely to phenology. Highest richness, in fact, was best explained entirely with principal components. Hexagons with 15 predicted species on average were related to the same processes as (E), discussed above, with the difference being greater richness where there was greater spatial variation in seasonality.

Available energy has been presented as food sources (Wright, 1983) and as atmospheric for thermoregulation (Currie, 1991). In Oregon for reptiles, phenology as a surrogate for available energy could be defined as either of the above or a combination of both. With warmer temperatures photosynthesis may occur throughout the entire year. This also provides a potential food source in the form of insects and other reptile prey. Available energy may not be concentrated visible light as we think of it on a bright sunny day, but radiation does penetrate cloud cover which in turn serves to hold heat in the atmosphere. Reptiles, which are not homeotherms, require incident energy to operate. This same energy provides fuel for other fauna and flora.

The Klamath mountain region in southwest Oregon is an interesting area unlike any other region in the state. Glacial isolation played a large role in shaping the diversity of vegetation and soils of that region, which was cut off from more northern regions allowing for divergence. Faunal migration occurred from the south, making the region a sink for species. Serpentine soils, high in magnesium, low in calcium, and depauparite in plant nutrients are common. Although there is high annual precipitation in the region typical of the Pacific Northwest, the landscape is semi-arid, as soils possess poor water holding capacity, creating complex microhabitats. As a result of moisture and nutrient unavailability and potential mineral toxicity, some plant species which are common to other areas occur here in stunted form. Other plant species occur only in this region of the state. Questions arise such as why is reptile richness greatest in this region while other vertebrate taxa are more diverse elsewhere? What advantages do reptiles possess that allow them to exploit this region more than other taxa? For example, highest amphibian

diversity encompasses, but is not limited to, this region. Historically, some amphibians migrated to Oregon from the east coast through boreal forest which extended farther south from present day latitudes, when climate was dramatically different. These amphibian species colonized the coast range southward from Washington state down into northern California.

Herpetofauna are not extremely mobile and typically possess smaller home ranges than mammals and birds, so individual herps are much more dependent on local conditions, whether they are stable or highly variable. Mammals and birds, being relatively mobile and better suited for temperature extremes, may have been better able to outcompete reptiles in other regions. The reverse may be true as well, that herps outcompete mammals and birds in this region.

3.8.1.4 *Amphibians*

RTA results from the amphibian dataset suggested that amphibian richness in Oregon is related solely to climate and phenology. Hexagons with lowest amphibian richness had high mean seasonal temperature differences and low onset values of NDVI. These hexagons were areas of extreme temperature differences with hot summers and cold winters east of the Cascades. Vegetation was likely either distributed sparsely or produced the majority of annual biomass during a short growing season. Conversely, high amphibian richness occurred where seasonal temperature differences were minimal and annual productivity and precipitation was highest in Oregon. West of the Cascades the Willamette Valley and the Klamath Mountains were differentiated based on precipitation, although the difference in number of species is only 2.

The distribution of amphibian richness was the simplest pattern of all the taxa and consequently the final model was the most simple as well. Amphibians have some very obvious and specific requirements, such as available water and/or appropriate temperatures. This was clearly captured in the RTA model, which indicated that there are more amphibian species in highly productive, wet, mild areas. Fewer amphibians occur where there is little suitable habitat and harsher climatic conditions. Certainly there are microclimatic conditions which cannot be measured at this coarse scale and which would

require inclusion of additional species-specific factor in analysis. However, these results clearly indicate that there are mechanisms which act at broad scales which relate to amphibian diversity.

3.8.1.5 *All vertebrates*

The all vertebrate response dataset was very similar to the bird richness dataset because birds overwhelmingly dominate in terms of numbers. There is a maximum of 195 bird species whereas the next highest taxa count is mammals with 65 species. It is not surprising, therefore, that the plot of the RTA results, the shapes of the trees, and some of the significant variables themselves, are very similar between the two. Deviance explained for the all vertebrate RTA model was 0.67 and was 0.66 for birds.

Elevation alone predicted 105 hexagons with lowest predicted vertebrate richness (182 species per hexagon) shown in red (A). These regions included most of western Oregon and along the Columbia River Gorge. The Cascade range had 193 species predicted on average (D). These hexagons had higher elevation, spatial heterogeneity of NDVI, and lower mean seasonal temperature differences. What was mapped as (F) in the bird tree corresponded closely with (F) in the all vertebrate tree, and both share three common variables (median elevation, variance of onset value of NDVI, and mean seasonal temperature difference). With the exception of (D) in the Cascades, the results of the all vertebrate RTA were like the results of the bird RTA. Semivariograms and correlograms for the response datasets and model residuals were nearly identical as well.

3.8.2 *Results in the context of biodiversity theory*

Perhaps the most interesting case of the taxa investigated here was amphibians. One might have expected amphibians to have the poorest RTA model fit because of the natural history and specific requirements of the individual species. Amphibians are not highly mobile and have perhaps the strictest criteria for survival given their water and temperature requirements. This study showed that processes measured at coarse scales can provide information about the occurrence of a taxon which might otherwise have been

thought to be governed by local microprocesses. This might point to an umbrella set of conditions which sets the environmental limits to richness in Oregon. Within these limits, microprocesses may govern individual abundances and concentrations of co-existing species.

It is difficult to compare the results from this analysis to prior work because this was the first work on terrestrial vertebrates which provided graphical means for assessing the fit of the regression tree models. We can learn about the goodness of fit from regression models by residual plots and r-squared values but we cannot quantify the spatial accuracy of linear models. Also, other studies have not dealt with the aggregation problem (pixels to hexagons) which presented different issues from interpolation and the effects of which are unknown. There could easily be correlations between response and predictor variables in this dataset which were not explored because of the summary statistics chosen. There are also differences in how the explanatory variables were collected. So although it can be said that r-squared values or percent variance explained are similar or dissimilar, there is little comparability between results of this study and previous work because we have not seen steps (mainly prediction) beyond deriving relationships between vertebrate richness and environmental variables.

How much does this study contribute to biodiversity knowledge and literature? There were some significant relationships found between phenology, climate, and elevation and vertebrate richness. Phenology as a continuous measure other than mean NDVI over the course of a year was previously unexplored. Satellite derived phenology as a surrogate to available energy is more easily obtained or calculated than calculating biomass estimates in the field. Satellite data also affords continuous spatial coverage within a very short time frame (i.e. within the same day). The results from this study do not imply cause and effect relationships but do suggest relationships between diversity of taxa and specific phenologic, climatic, and topographically related forces. It is impossible to determine what factor(s) directly caused the patterns of diversity on the landscape today for several reasons, the most obvious being that we are restricted to retrospective studies given the relatively short life span of humans on an evolutionary time scale. This creates lively debate as there is no experimental repeatability, evidenced by the ongoing discussion regarding Rapoport's Rule as an explanation for latitudinal gradients in

biodiversity (Rohde, 1996). We are also somewhat encumbered by our need to aggregate and characterize, evidenced in this study by the fact that none of the semivariograms of the taxa passed through the origin. This indicated an undetected pattern at a different spatial scale due to inappropriate sample size. Regardless of grain and extent, there will always be interpretation of the initial response variable (i.e. number of species per given area) and the explanatory variables (i.e. interpolated climate or averaged reflectance). These interpretations may create artifacts before any analysis is conducted. To effectively manage and interpret data and results from biodiversity analyses, however, we must aggregrate, interpolate, and categorize.

There are some unresolved questions and issues which resulted from this study. The algorithms used in the RTA and the methods for pruning trees could be investigated further as suggested numbers of terminal nodes varied by cross validation method. The effect of removing 4 variables from the full dataset was marked; RTA models from the reduced datasets were twice as complex and as a result explained more variation in the richness data. However, complex models are difficult to interpret. There are, therefore, questions about how much additional non-redundant information the principal components added to the analysis. Principal components represented temporal measures of vegetation growth and condition over the course of a year. It may be possible that the variables in the reduced dataset which replaced principal components were better correlates of richness and the inclusion of principal components masked these relationships algorithmically. The flip side, however, is that principal components did add important and significant information which could not be fully explained by the remaining set of variables. This would explain why the trees from the reduced datasets were so much more complex than those from the full dataset.

This study showed some general trends which support findings of previous work, but also added new insight because of the methods used to assess suitability of the regression trees for predicting vertebrate richness. The development of spatial statistical methods, also called spatial statistics, has created tremendous opportunity for researchers to quantify and explain patterns in nature, but these new tools must be used with caution and thoughtfulness. For example, I used three methods to cross-validate the regression trees generated for each taxa. Each method suggested trees of different complexity and the

interpretation of the calculated relationships would have varied accordingly. With these new methods, researchers must question more that just the ecological or just the statistical significance of results. When expert judgement is required to select "the best" result, subjective interpretation plays an increasingly important role. Ecological interpretation of two closely "best" fitting models may be very different. We must have faith in our own knowledge to select the model that is closest to "truth". In this light, the results from this study showed some agreement with previous work, and generated more questions to be investigated.

3.9 CONCLUSIONS

Results from this analysis agree with the idea that energy or productivity affects patterns of vertebrate diversity. While significant relationships were found between the different taxa and phenology, climate, and elevation, results varied with complexity of spatial pattern of the taxa. RTA effectively modeled relationships in autocorrelated data which were likely nonlinear, although autocorrelation was not completely eliminated. Satellite imagery-derived measures of phenology were important correlates to vertebrate richness, indicated by the length of the regression trees and the resulting plots of model prediction results. Continuing this analysis to encompass a larger geographic area would be the next test of scale for investigating strength of the relationships derived in this study. The other natural continuation would be to conduct the same analysis in a very different ecosystem, such as the Piedmont of the east coast or bayous and swamps of the deep south. Nevertheless, results clearly indicated relationships between vertebrate species richness and continuous variable measurements derived from satellite imagery.

4.0 Chapter 4: Summary

Biodiversity in Oregon was shown to be related to phenology, climate, and elevation, although relationships varied by taxa. Remotely sensed imagery provided important time series data for estimating phenology. Biweekly composites from the NOAA-AVHRR satellite were evaluated for data quality for the state of Oregon using a method tested in other parts of the world. Imagery from 1991, 1992, and 1993 were evaluated to provide a level of confidence in these data because of recent criticism of the compositing process.

Results of the principal component analysis of time-series AVHRR biweekly composites were consistent with previous work and provided a perspective on Oregon's vegetation not included in standard landcover classifications. The time-series derived data was not intended to replace the use of landcover maps. In this study continuous time-series data was more appropriate for estimating phenology over the course of a year. The parametric approach eliminated subjective bias inherent in a landcover classification and subsequent calculation of landscape indices. It was shown that the combination of spatial and temporal axes were effective in discriminating different vegetation groups in Oregon based on temporal pattern of NDVI. Most of the mountainous forests of western Oregon had similar signatures, and these signatures were very different from the primarily agricultural Willamette Valley. The Willamette Valley was also discriminable from agriculture and grasslands in eastern Oregon. Pine dominated mountains in eastern Oregon exhibited a different phenological pattern than fir dominated western Oregon. This part of the study left many questions to be answered from an ecological and a remote sensing standpoint. Results support the difficulty researchers have discriminating between conifer species and might indicate that productivity of these stands are similar over the course of one year. However, this can be questioned due to the coarse scale of the sensor. There is a lot of ground contributing to the spectral signature within a 1 km^2 pixel and it may be that this scale is inappropriate for testing a productivity hypothesis for conifers.

Phenology was derived using principal component analysis and using a method developed by Brad Reed at EROS Data Center which results in greenness metrics. Greenness metrics and principal components were shown to be related by spatial

component but not by temporal component, indicating two different measures of phenology or different processes relating to phenology. This data was investigated with terrain and climate data for correlation to vertebrate species richness. DEM and climate variables had been investigated in previous work at similar scales with success but phenology had not. An exploratory data analysis was used to allow for non-linear relationships in the data. Spatial statistics, semivariance and Moran's I, were used to quantify the pattern in the response data (taxa). All taxa were autocorrelated and exhibited spatial pattern. Regression tree analysis was shown to be relatively robust to autocorrelation. Residuals were tested for randomness using semivariance and Moran's I much as residual plots are used in standard regression methods of classical statistics.

Determining the best size of final regression trees was somewhat subjective. Three methods were used to assess the most parsimonious numbers of nodes, but results varied by method. There is much to be learned about regression tree analysis and it should be used with prudence and an understanding of the application science.

Ecologically, amphibian diversity was the easiest to model because the pattern was the most straightforward: many amphibian species occur in western Oregon and few occur in eastern Oregon. Birds were the next successfully modeled taxon. General patterns were effectively predicted where highest richness occurred east of the Cascades in the mountains of eastern Oregon. Lowest richness was west of the Cascades. Bird richness was highly related to climate variation. All vertebrate results were similar to the bird results because birds were the dominant taxonomic group. Reptile diversity was highly related to phenology and mammal diversity was related to spatial heterogeneity of the landscape.

This research indicated that the incorporation of temporal measures of the landscape were important to explain patterns of vertebrate diversity in Oregon. Results cannot be extrapolated to other geographic regions but should be tested in other areas to determine if relationships at this scale are consistent. Results also indicate that an overall measure of diversity which includes multiple taxa might not be the best measure of richness due to the variety of processes specific to taxa. This study showed that temporal variation of a landscape can be measured with satellite imagery to cover a very large area, that this temporal variability be quantified parametrically rather than as categorical, and that

temporal variations of the landscape, or temporal dynamics, are significantly related to vertebrate species richness.

BIBLIOGRAPHY

Boots, B.N. and A. Getis. 1988. Point pattern analysis. Sage publications, Inc. Newbury Park, CA.

Bradshaw, G.A. and S.L. Garman. 1994. Detecting fine-scale disturbance in forested ecosystems as measured by large-scale landscape pattern. in *Environmental information management and analysis: Ecosystems to global scales*, W.K. Michener, J.W. Brunt, and S.G. Stafford (eds). Taylor and Francis LTD. Bristol, PA. pp. 535-550.

Breiman, L., J.H. Friedman, R.A. Olshen, C.J. Stone. 1993 Classification and regression trees. Chapman and Hall. NY, NY. 358 pp.

Brown, J.F., T.R. Loveland, J.W. Merchant, B.C. Reed, and D.O. Ohlen. 1993. Using multisource data in global land-cover characterization: Concepts, requirements, and methods. *Photogrammetric Engineering and Remote Sensing*. 59:977-987.

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

Burgan, R.E. and R.A. Hartford. 1993. Monitoring vegetation greenness with satellite data. General Technical Report INT-297. USDA Forest Service Intermountain Research Station.

Chambers, J.M. and T.J. Hastie. 1993. Statistical models in S. Chapman and Hall, Inc., London. 608pp.

Cicone, R. C. and J.A. Olsenholler. 1997. A summary of Asian vegetation using annual vegetation dynamic indicators. Geocarto International. 12:13-25.

Cihlar, J., D. Manak, and M. D'Iorio. 1994. Evaluation of compositing algorithms for AVHRR data over land. *IEEE Transactions on Geoscience and Remote Sensing*. 32:427-437.

Cohen, W.B. and T.A. Spies. 1992. Estimating structural attributes of douglas-fir/western hemlock forest stands from Landsat and SPOT imagery. *Remote sensing of Environment.* 41:1-17.

Cohen, W.B., T.A. Spies, and G.A. Bradshaw. 1990. Semivariograms of digital imagery for analysis of conifer canopy structure. *Remote Sensing of Environment.* 34:167-178.

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

Daly C, RP Neilson, DL Phillips. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology.* 33:140-158.

Daubenmire, R. 1978. Plant geography (with special reference to North America). Academic Press. NY, NY.

Davis, F.W. and D.M. Stoms. 1994 Synoptic national assessment of comparative risks to biological diversity and landscape type: Pilot studies for southern California. Report submitted to Environmental Protection Agency. Corvallis, OR.

Di, L. and D.A. Hastings. 1995. Temporal stability of some global NDVI products derived from NOAA/AVHRR GVI. *International Journal of Remote Sensing.* 16:3569-3583.

Diamond, J. 1988. Factors controlling species diversity: overview and synthesis. *Annals of the Missouri Botanical Garden.* 75:117-129.

Eastman, J.R. and M. Fulk. 1993. Long sequence time series evaluation using standardized principal components. *Photogrammetric Engineering and Remote Sensing.* 59:991-996.

Efron, B. and R.Tibshirani. 1991. Statistical data analysis in the computer age. *Science.* 253:390-395.

Ehlundh, L.R. 1995. Noise estimation in NOAA AVHRR maximum-value composite NDVI images. *International Journal of Remote Sensing.* 16:2955-2962.

Fung, T. and E. LeDrew. 1987. Application of principal components analysis to change detection. *Photogrammetric Engineering and Remote Sensing.* 53:1649-1658.

Goward, S.A., C.J. Tucker, and D. Dye. 1985. North American vegetation patterns observed with NOAA-7 Advanced Very High Resolution Radiometer. *Vegetatio.* 64:3-14.

Gutman, G. and A. Ignatov. 1995. Global land monitoring from AVHRR: potential and limitations. *International Journal of Remote Sensing.* 16:2301-2309.

Holben, B. N. 1986. Characteristics of maximum value composites from temporal AVHRR data. *International Journal of Remote Sensing.* 7:1417-1437.

Kiester, A.R. 1971. Species density of North American amphibians and reptiles. *Systematic Zoology.* 20:127-137.

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

Lillesand, T. M. and R. W. Kiefer. 1994. Remote sensing and image interpretation, third ed. John Wiley and Sons, Inc. NY, NY. 750pp.

Loveland, T.R., J.W. Merchant, D.O. Ohlen, and J.F. Brown. 1991. Development of a land-cover characteristics database for the conterminous U.S. *Photogrammetric Engineering and Remote Sensing*. 57:1453-1463.

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

Marks, D. 1990. The sensitivity of potential evapotranspiration to climate change over the continental United States. Pages IV-1 to IV-31 in H. Gucinski, D. Marks, and D.P. Turner, editors. Biospheric feedbacks to climate change: the sensitivity of regional trace gas emissions, evapotranspiration, and energy balance to vegetation redistribution. EPA/600/3-90/078. Corvallis, OR.

Miller, T.W. 1994. Model selection in tree-structured regression. in *Proceedings of the ASA Statistical Computing Section*:158-163.

NOAA. 1990. NOAA polar orbiter data user's guide. U.S. Department of Commerce, NOAA, NESDI, NCDC, and the Satellite Data Service Division, Washington, D.C.

Owen, J.G. 1990. Patterns of mammalian species richness in relation to temperature, productivity, and variance in elevation. *Journal of Mammalogy*. 71:1-13.

Reed, B.C, J.F. Brown, D. VanderZee, T.R. Loveland, J.W. Merchant, and D.O. Ohlen. 1994. Measuring phenological variability from satellite. *Journal of Vegetation Science*. 5:703-714

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

Rohde, K. 1996. Rapoport's Rule is a local phenomenon and cannot explain latitudinal gradients in species diversity. *Biodiversity Letters*. 3:10-13.

Rossi, R.E., D.J. Mulla, A.G. Journal, E.H. Franz. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*. 62:277-314.

Schall, J.J. and E.R. Pianka. 1978. Geographical trend in the number of species. *Science*. 201:679-686.

Scheibe, J.S. 1987. Climate, competition, and the structure of temperate zone lizard communities. *Ecology*. 68:1424-1436.

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

Singh, A. and A. Harrison. 1985. Standardized principal components. *International Journal of Remote Sensing.* 6:883-896.

Stoms, D.M. 1995. Landscape measures of heterogeneity, productivity, and dynamic processes in relation to vertebrate species richness in California. unpublished manuscript.

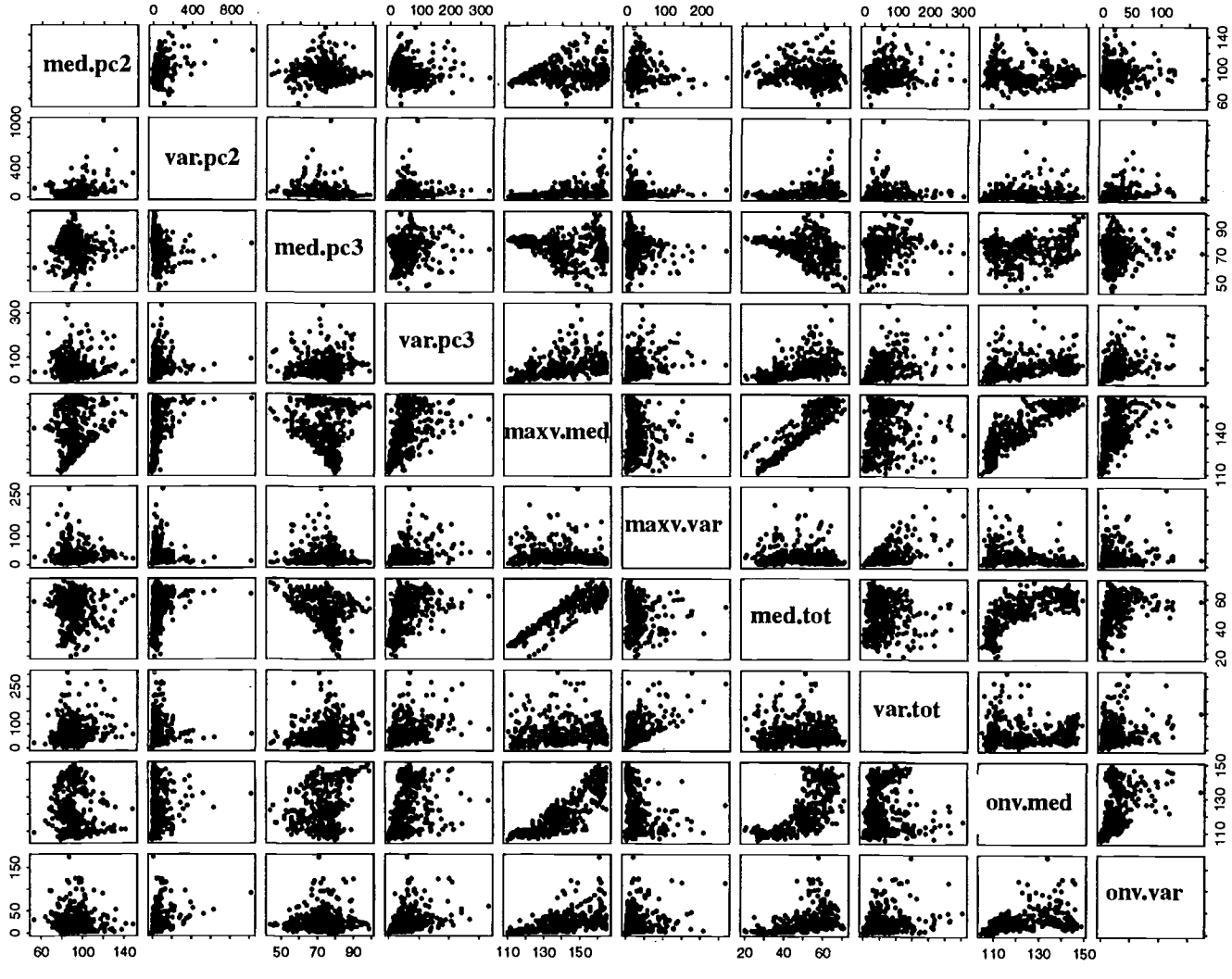
Stoms, D.M., M.J. Bueno and F.W. Davis. 1997. Viewing geometry of AVHRR image composites derived using multiple criteria. *Photogrammetric Engineering and Remote Sensing.* in press.

Townshend, J.R.G., C.O. Justice, and V. Kalb. 1987. Characterization and classification of South American land cover types using satellite data. *International Journal of Remote Sensing.* 8:1189-1207.

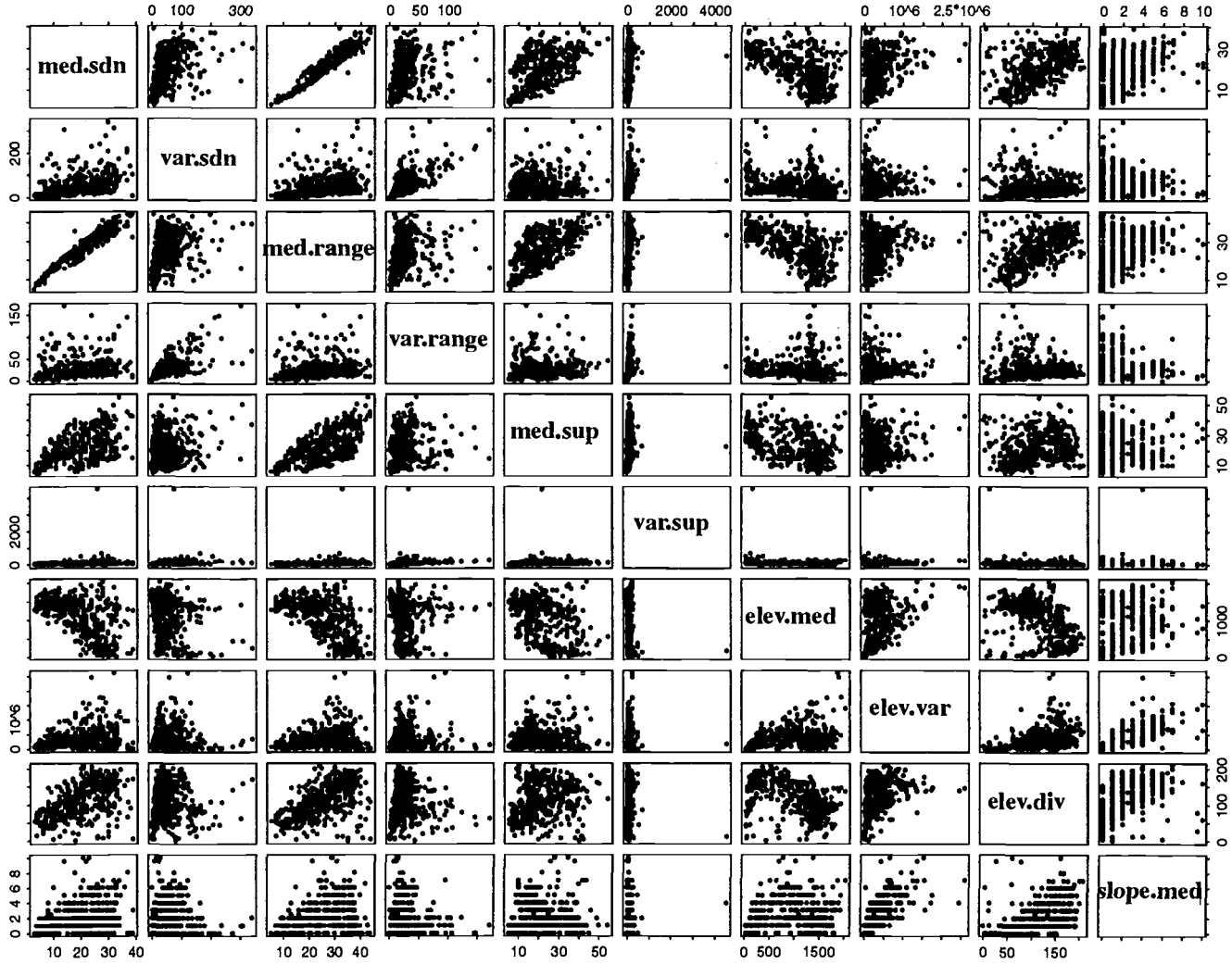
Tucker, C.J., J.R.G. Townshend, T.E. Goff. 1985. African land-cover classification using satellite data. *Science.*227:369-375.

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

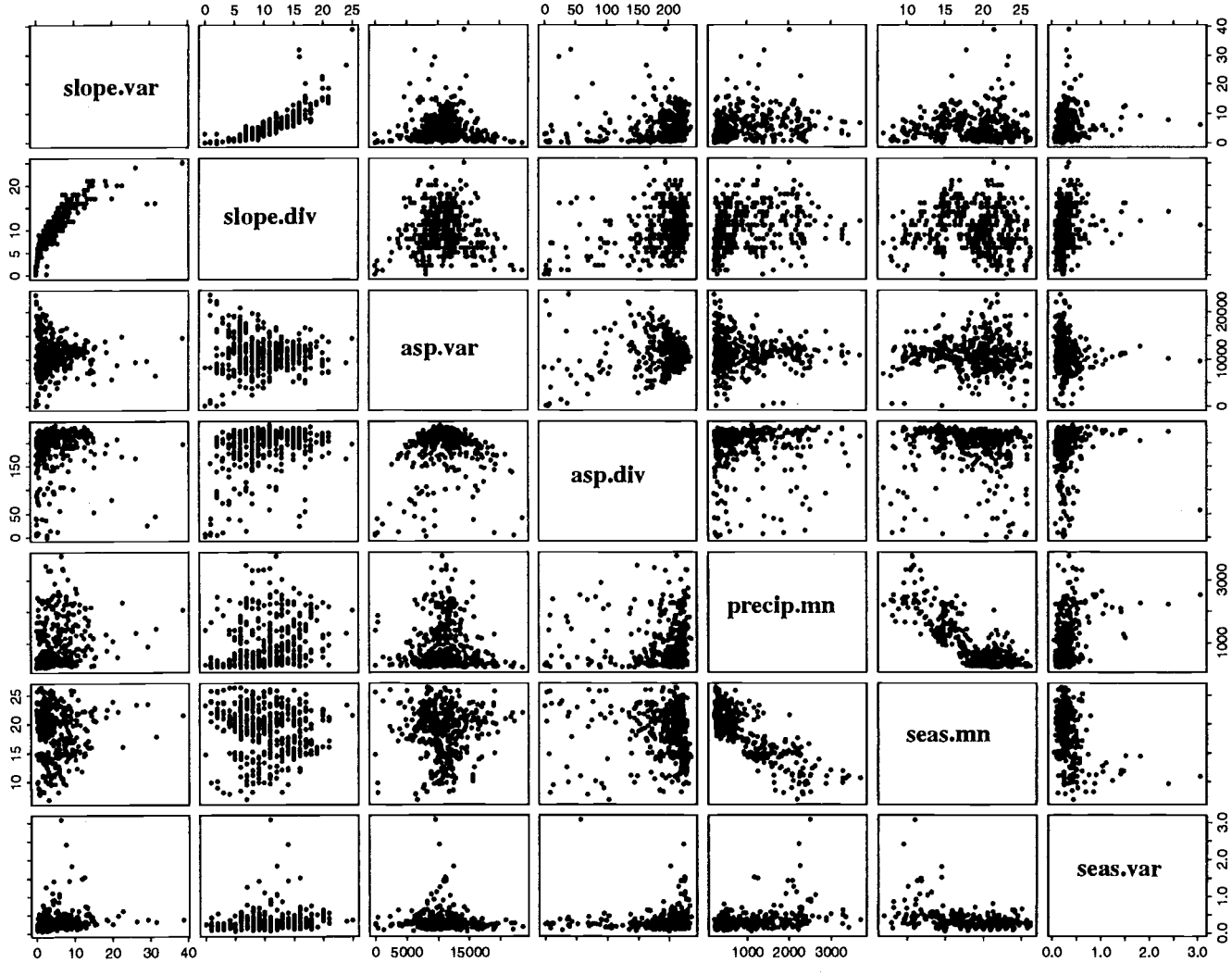
APPENDICES



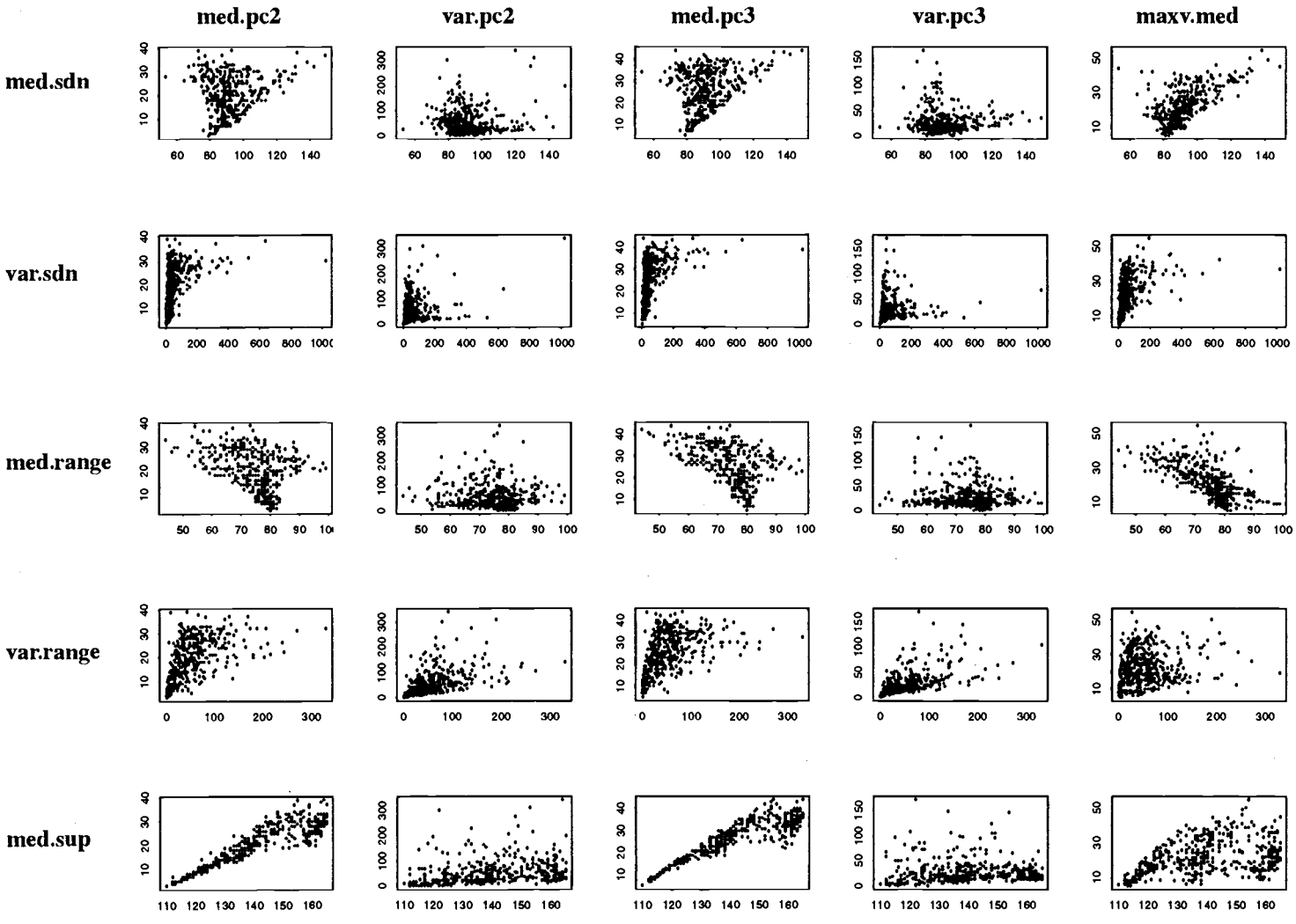
Appendix 1a. Scatterplots for explanatory variables summarized to the hexagon level.



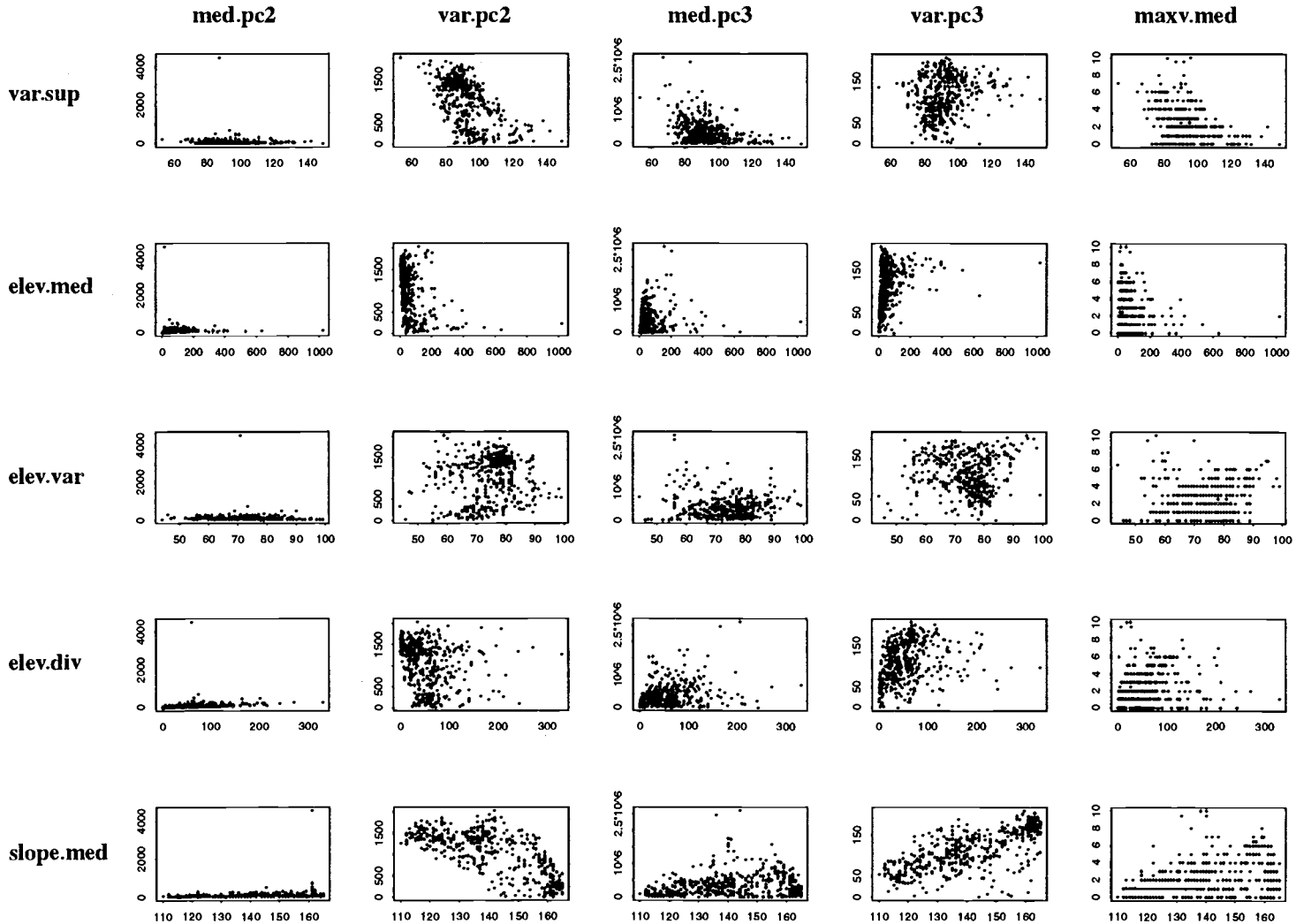
Appendix 1b. Scatterplots for explanatory variables summarized to the hexagon level.



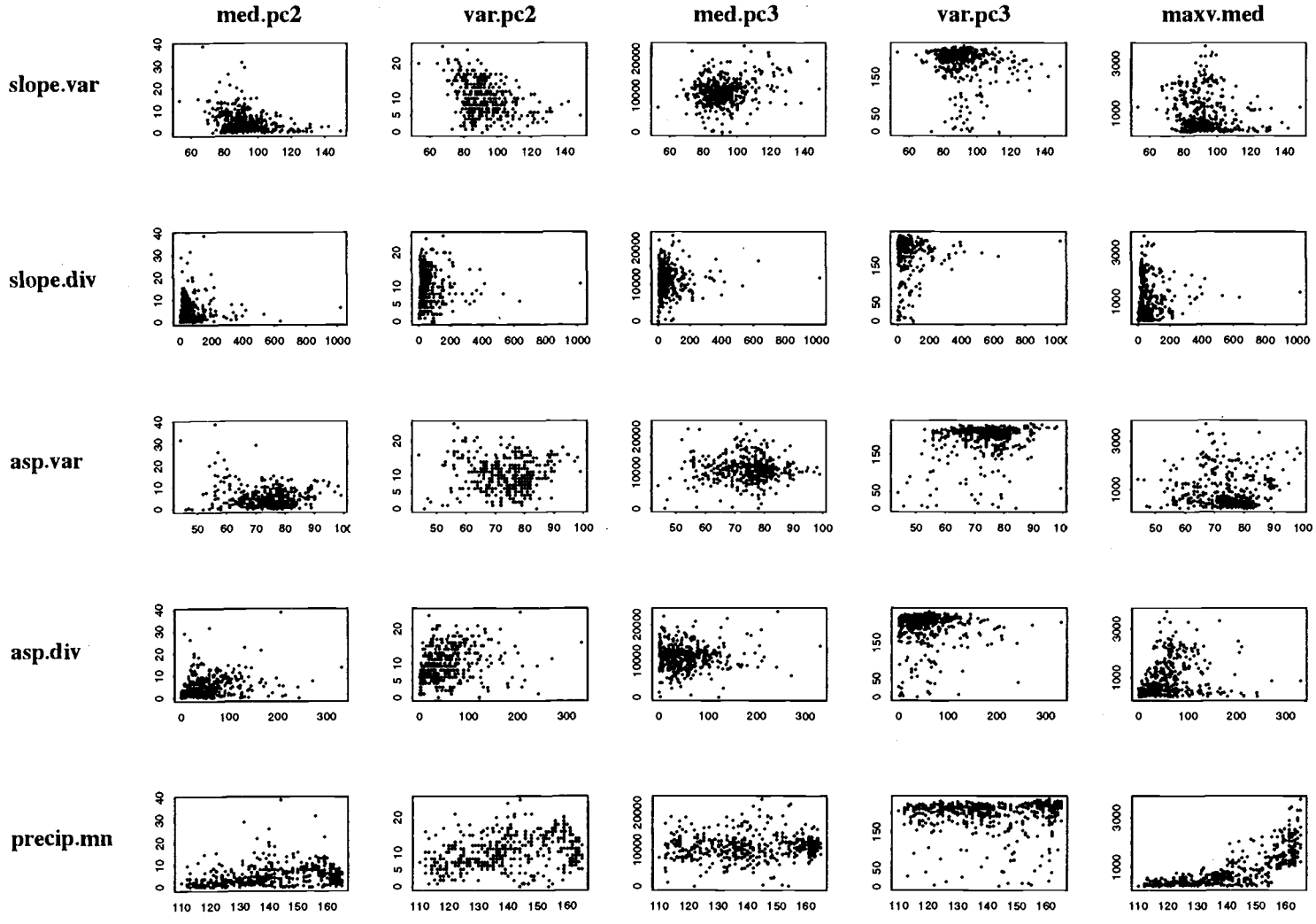
Appendix 1c. Scatterplots for explanatory variables summarized to the hexagon level.



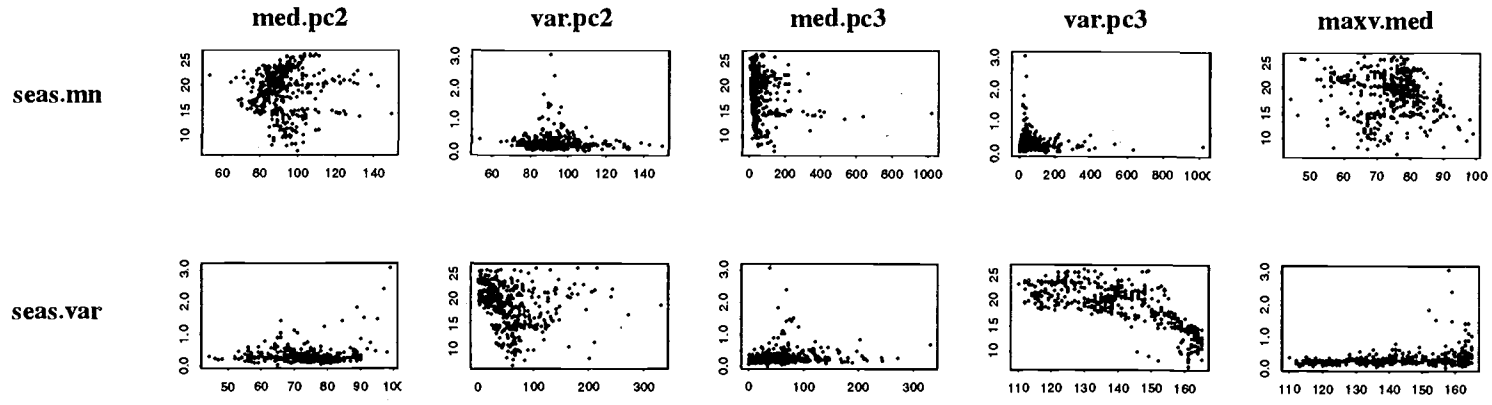
Appendix 1d. Scatterplots for explanatory variables summarized to the hexagon level.



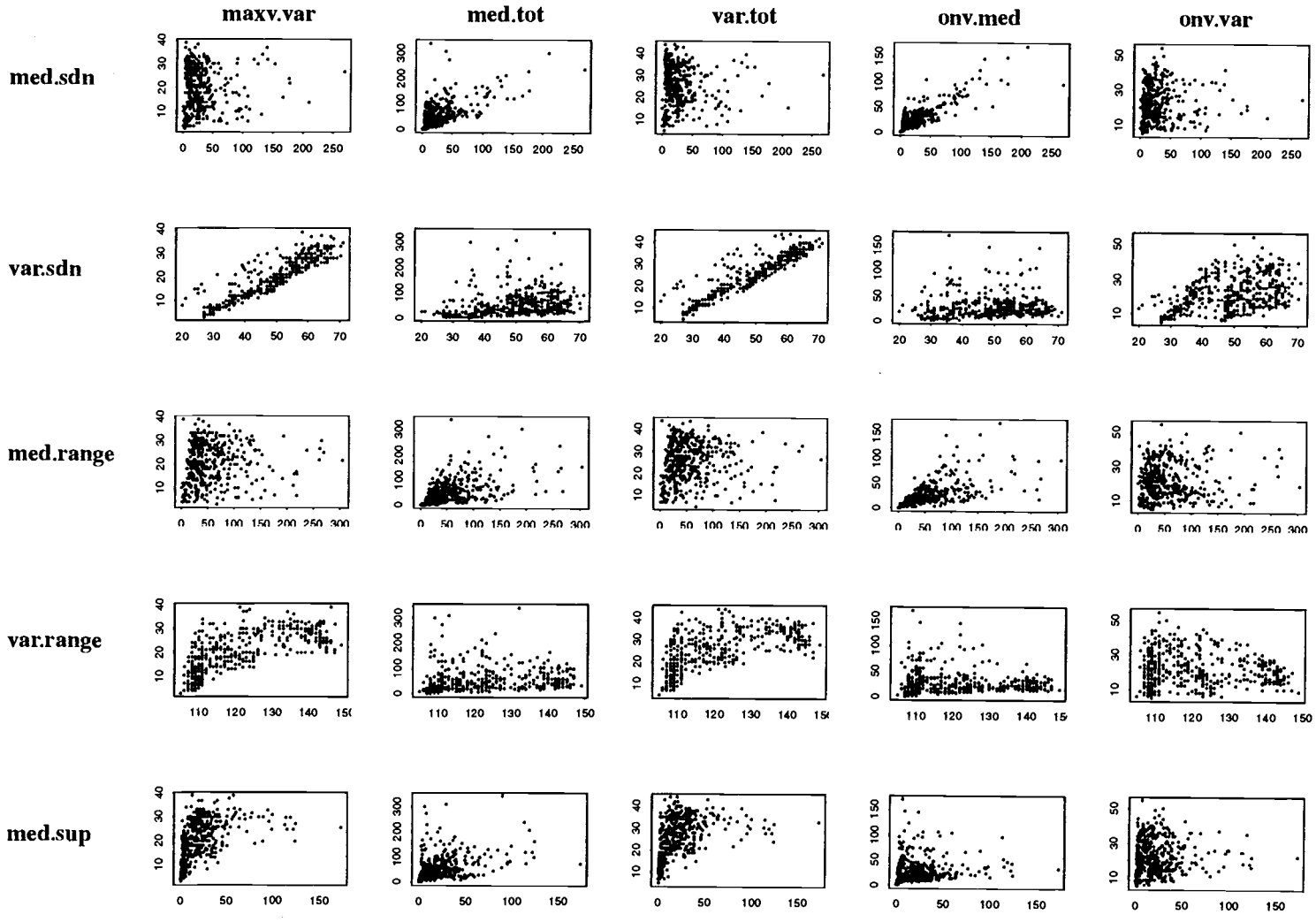
Appendix 1e. Scatterplots for explanatory variables summarized to the hexagon level.



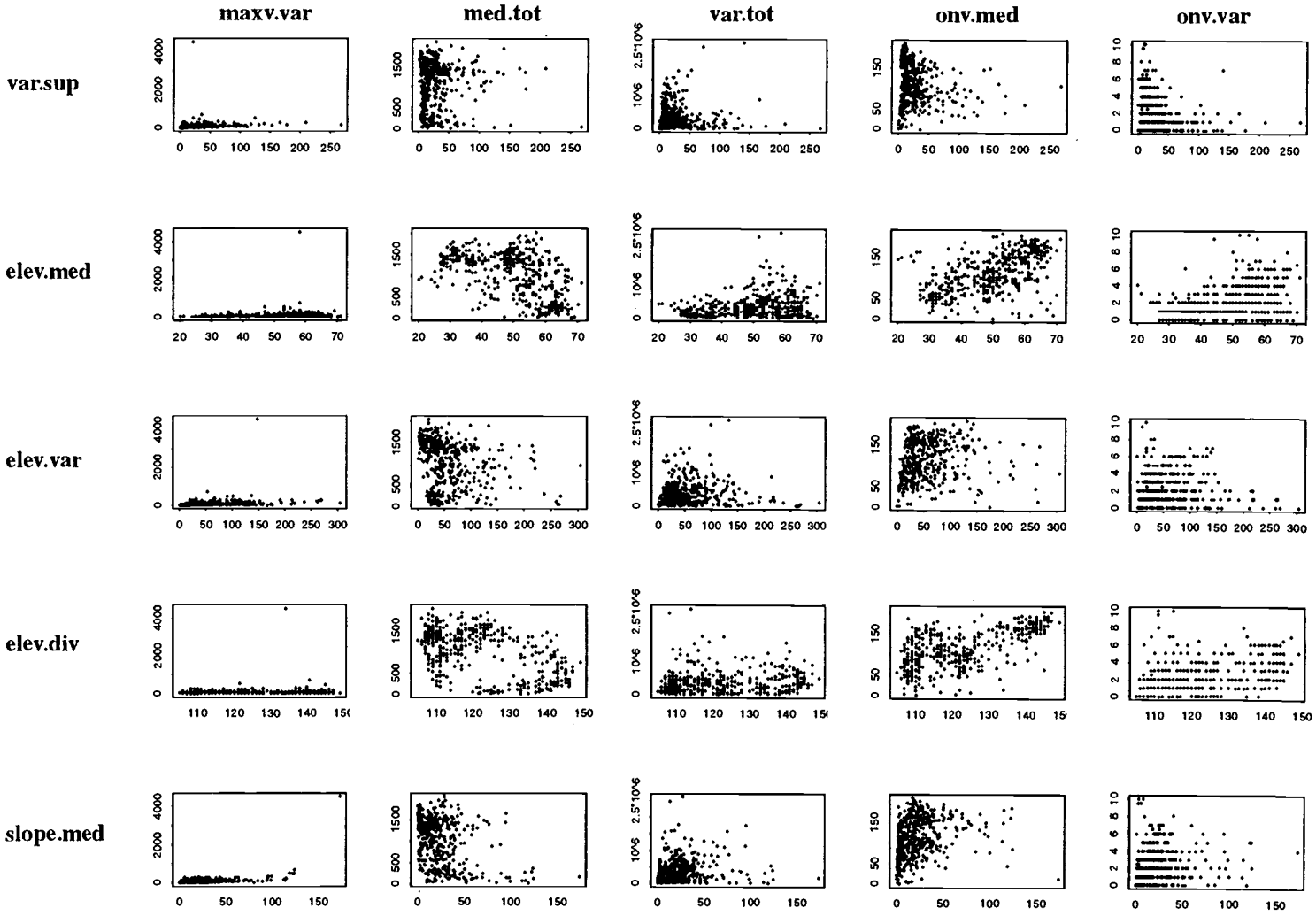
Appendix 1f. Scatterplots for explanatory variables summarized to the hexagon level.



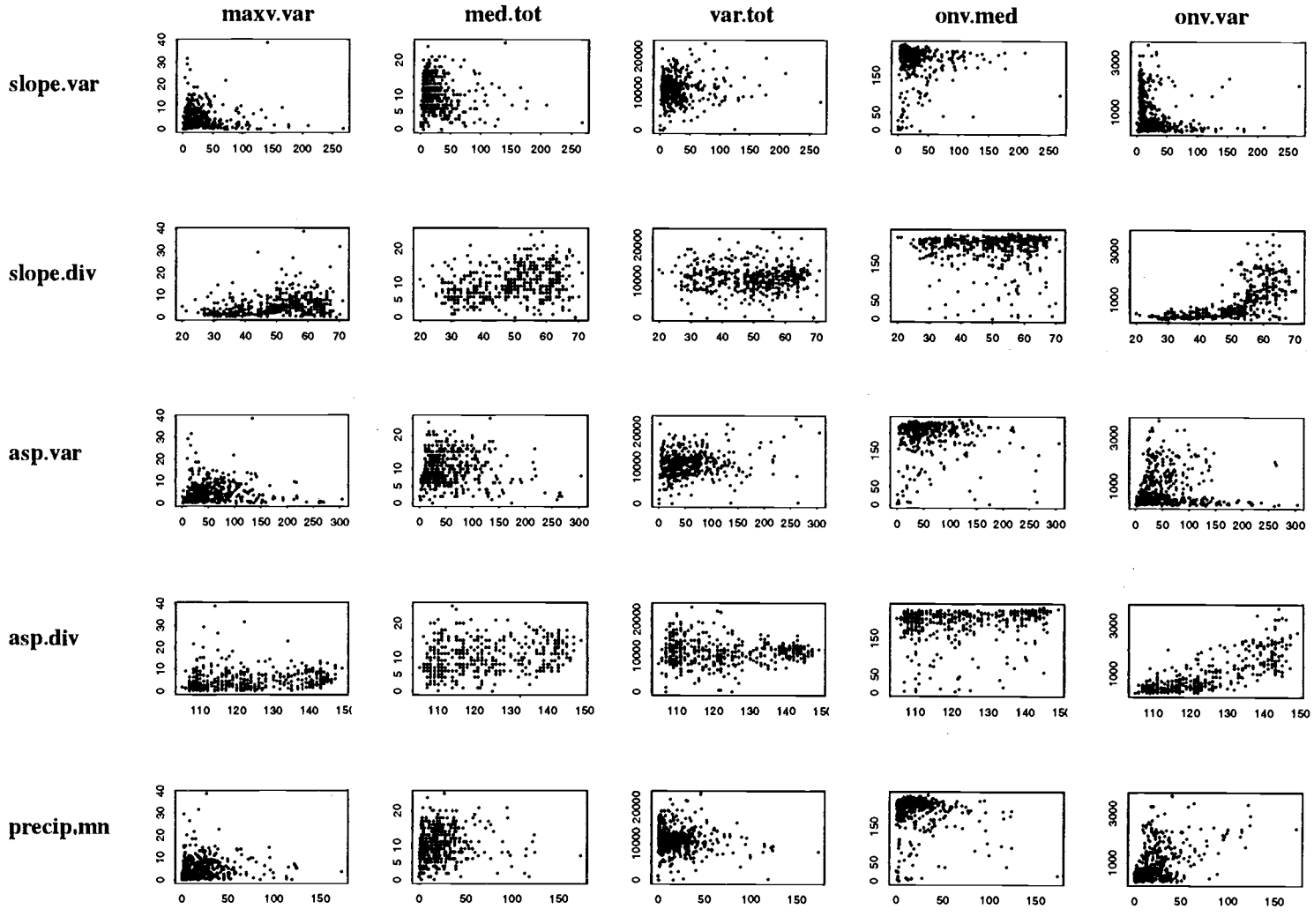
Appendix 1g. Scatterplots for explanatory variables summarized to the hexagon level.



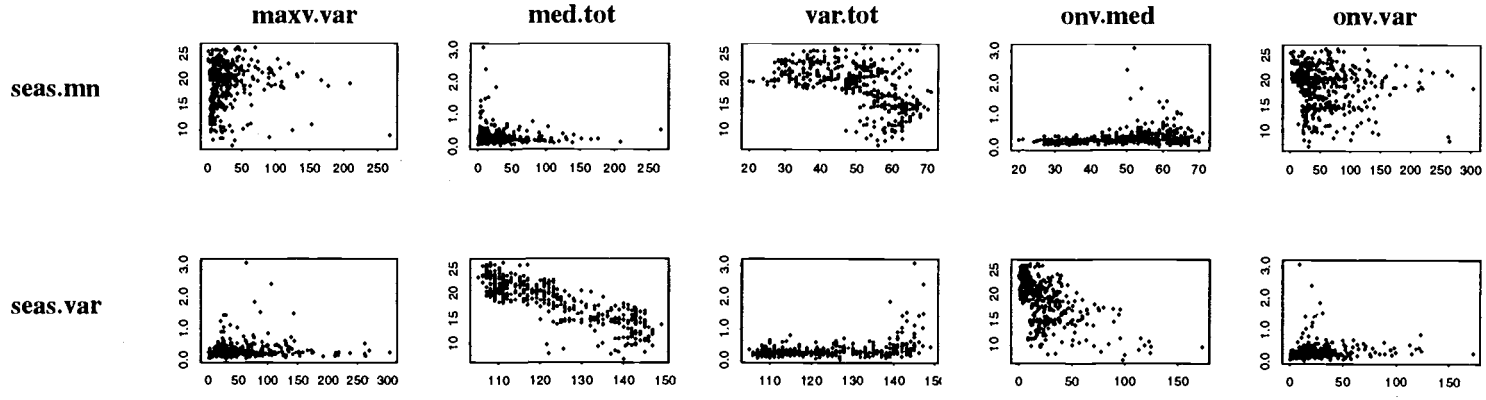
Appendix 1h. Scatterplots for explanatory variables summarized to the hexagon level.



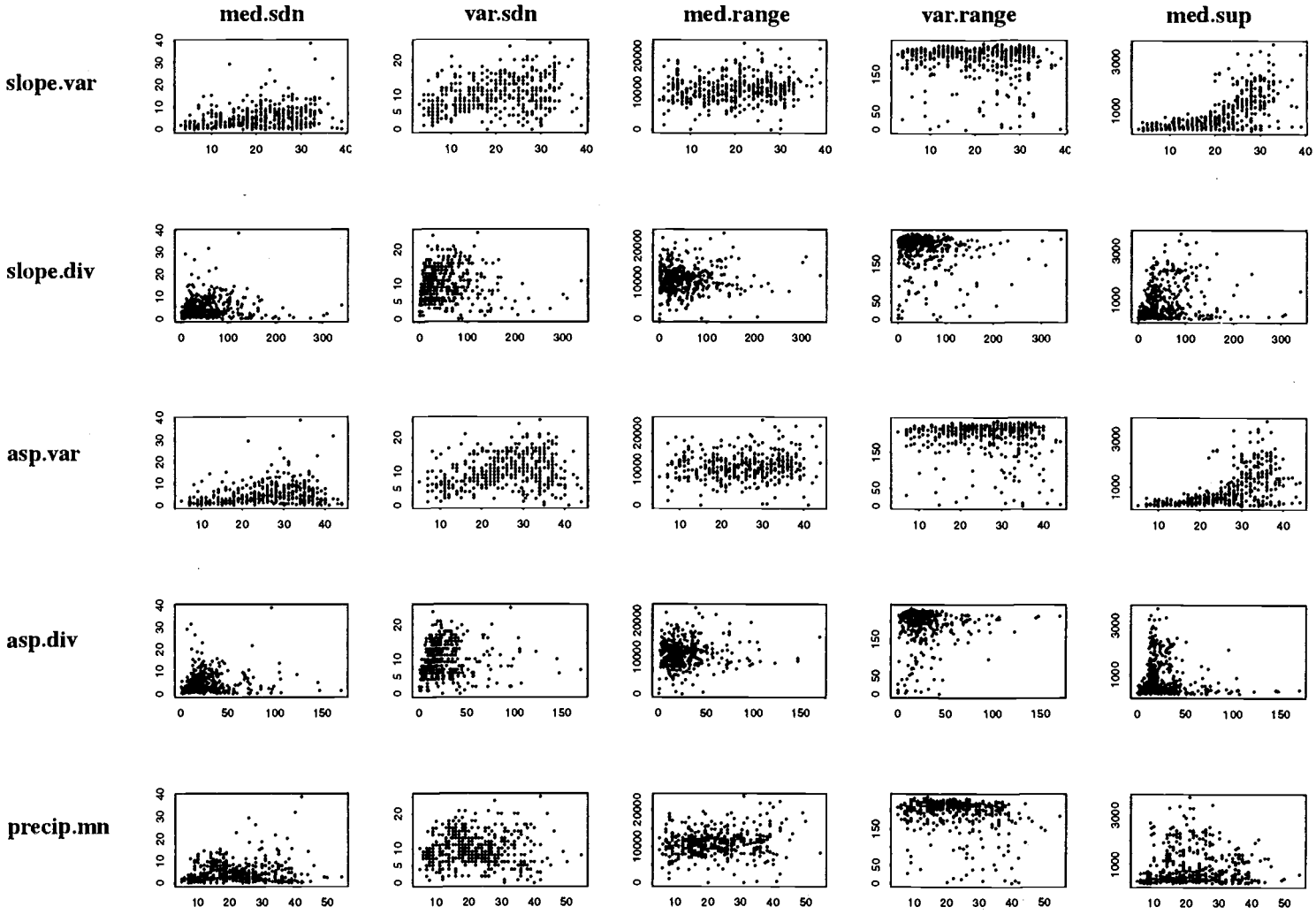
Appendix 1i. Scatterplots for explanatory variables summarized to the hexagon level.



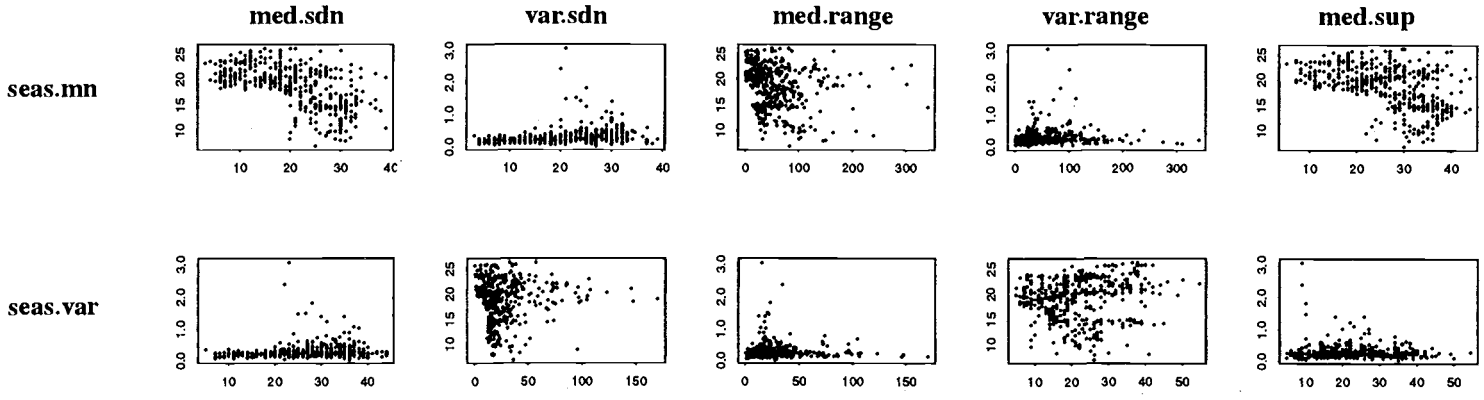
Appendix 1j. Scatterplots for explanatory variables summarized to the hexagon level.



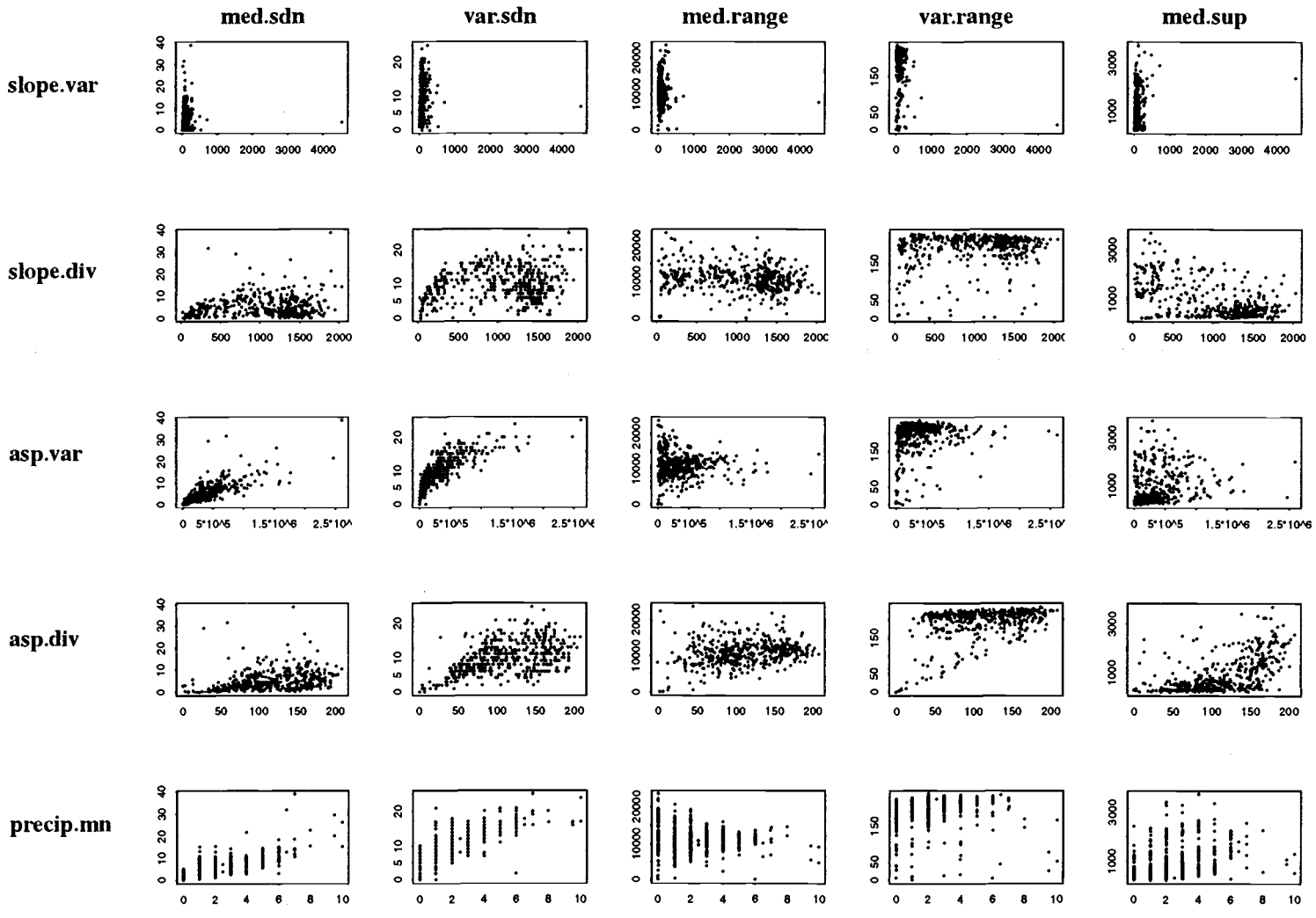
Appendix 1k. Scatterplots for explanatory variables summarized to the hexagon level.



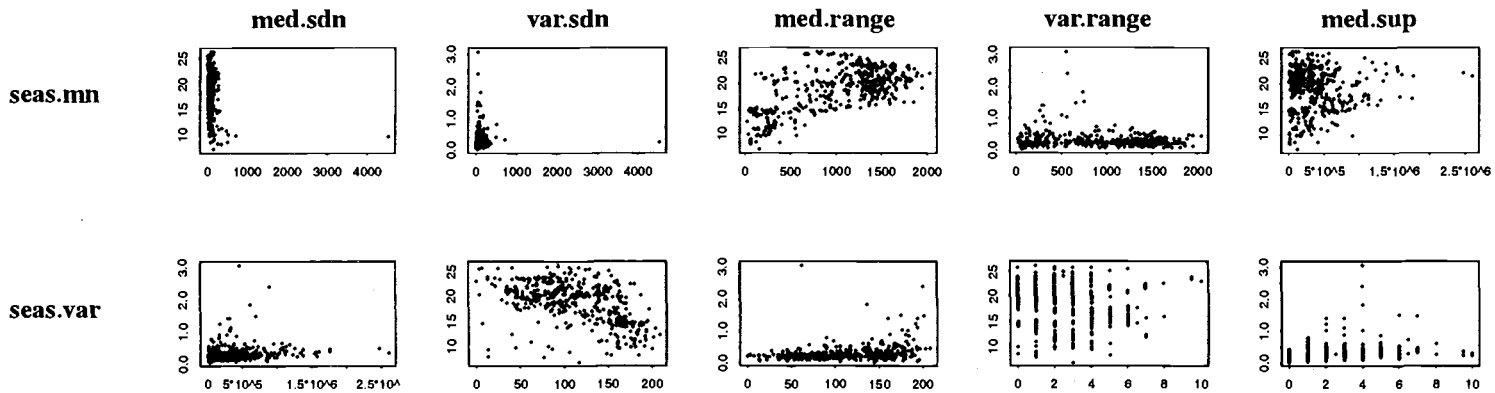
Appendix 11. Scatterplots for explanatory variables summarized to the hexagon level.



Appendix 1m. Scatterplots for explanatory variables summarized to the hexagon level.



Appendix 1n. Scatterplots for explanatory variables summarized to the hexagon level.



Appendix 1o. Scatterplots for explanatory variables summarized to the hexagon level.

Appendix 2a. Spearmean rank correlation coefficients for explanatory variables summarized to the hexagon level.

	med.pc2	var.pc2	med.pc3	var.pc3	maxv.med	maxv.var	med.tot
med.pc2	1.00	0.41	-0.15	-0.01	0.21	0.09	0.05
var.pc2	0.41	1.00	-0.41	0.37	0.51	0.25	0.46
med.pc3	-0.15	-0.41	1.00	-0.07	-0.26	-0.13	-0.44
var.pc3	-0.01	0.37	-0.07	1.00	0.61	0.25	0.59
maxv.med	0.21	0.51	-0.26	0.61	1.00	-0.22	0.92
maxv.var	0.09	0.25	-0.13	0.25	-0.22	1.00	-0.20
med.tot	0.05	0.46	-0.44	0.59	0.92	-0.20	1.00
var.tot	0.12	0.28	0.21	0.53	0.15	0.56	0.06
onv.med	-0.06	0.26	0.02	0.59	0.88	-0.26	0.80
onv.var	-0.02	0.45	-0.12	0.64	0.66	0.17	0.65
med.sdn	0.11	0.53	-0.42	0.60	0.91	-0.08	0.91
var.sdn	-0.16	0.24	-0.07	0.69	0.39	0.52	0.40
med.range	0.25	0.60	-0.50	0.56	0.90	-0.10	0.92
var.range	0.04	0.34	-0.05	0.64	0.16	0.72	0.14

Appendix 2b. Spearmean rank correlation coefficients for explanatory variables summarized to the hexagon level.

	var.tot	onv.med	onv.var	med.sdn	var.sdn	med.range	var.range
med.pc2	0.12	-0.06	-0.02	0.11	-0.16	0.25	0.04
var.pc2	0.28	0.26	0.45	0.53	0.24	0.60	0.34
med.pc3	0.21	0.02	-0.12	-0.42	-0.07	-0.50	-0.05
var.pc3	0.53	0.59	0.64	0.60	0.69	0.56	0.64
maxv.med	0.15	0.88	0.66	0.91	0.39	0.90	0.16
maxv.var	0.56	-0.26	0.17	-0.08	0.52	-0.10	0.72
med.tot	0.06	0.80	0.65	0.91	0.40	0.92	0.14
var.tot	1.00	0.13	0.31	0.17	0.54	0.14	0.71
onv.med	0.13	1.00	0.70	0.72	0.42	0.66	0.10
onv.var	0.31	0.70	1.00	0.62	0.54	0.57	0.29
med.sdn	0.17	0.72	0.62	1.00	0.48	0.96	0.26
var.sdn	0.54	0.42	0.54	0.48	1.00	0.37	0.73
med.range	0.14	0.66	0.57	0.96	0.37	1.00	0.22
var.range	0.71	0.10	0.29	0.26	0.73	0.22	1.00

Appendix 2c. Spearmean rank correlation coefficients for explanatory variables summarized to the hexagon level.

	med.sup	var.sup	elev.med	elev.var	elev.div	slope.med	slope.var
med.pc2	0.62	0.05	-0.56	-0.24	0.23	-0.15	-0.21
var.pc2	0.61	0.51	-0.49	0.15	0.44	0.08	0.12
med.pc3	-0.72	-0.31	0.18	0.03	-0.06	0.01	-0.02
var.pc3	0.15	0.60	-0.32	0.30	0.39	0.27	0.32
maxv.med	0.38	0.33	-0.69	0.22	0.69	0.39	0.30
maxv.var	0.11	0.43	0.07	-0.10	-0.22	-0.28	-0.11
med.tot	0.40	0.33	-0.52	0.27	0.56	0.39	0.34
var.tot	-0.02	0.40	-0.33	0.08	0.22	0.05	0.09
onv.med	-0.03	0.18	-0.46	0.30	0.60	0.45	0.36
onv.var	0.11	0.43	-0.34	0.26	0.42	0.29	0.25
med.sdn	0.46	0.41	-0.58	0.21	0.57	0.32	0.29
var.sdn	-0.01	0.57	-0.20	0.08	0.09	0.05	0.13
med.range	0.61	0.41	-0.63	0.18	0.58	0.30	0.25
var.range	0.12	0.63	-0.15	0.09	0.07	-0.05	0.12

Appendix 2d. Spearmean rank correlation coefficients for explanatory variables summarized to the hexagon level.

	slope.div	asp.var	asp.div	precip.mn	seas.mn	seas.var
med.pc2	-0.27	0.17	-0.14	-0.06	0.07	0.00
var.pc2	0.10	0.11	0.15	0.29	-0.22	0.06
med.pc3	-0.01	-0.02	0.17	-0.15	-0.08	-0.09
var.pc3	0.29	0.01	0.02	0.44	-0.41	0.21
maxv.med	0.24	0.09	0.11	0.81	-0.72	0.36
maxv.var	-0.09	0.02	-0.25	-0.31	0.20	-0.10
med.tot	0.30	0.03	0.09	0.76	-0.59	0.31
var.tot	0.08	0.15	-0.02	-0.02	-0.18	0.01
onv.med	0.32	0.02	0.24	0.82	-0.81	0.35
onv.var	0.22	-0.04	0.00	0.55	-0.58	0.24
med.sdn	0.24	0.11	-0.01	0.71	-0.59	0.32
var.sdn	0.11	0.00	-0.11	0.30	-0.34	0.21
med.range	0.20	0.11	-0.01	0.64	-0.50	0.28
var.range	0.12	0.07	-0.14	0.01	-0.08	0.04

Appendix 2e. Spearmean rank correlation coefficients for explanatory variables summarized to the hexagon level.

	med.pc2	var.pc2	med.pc3	var.pc3	maxv.med	maxv.var	med.tot
med.sup	0.62	0.61	-0.72	0.15	0.38	0.11	0.40
var.sup	0.05	0.51	-0.31	0.60	0.33	0.43	0.33
elev.med	-0.56	-0.49	0.18	-0.32	-0.69	0.07	-0.52
elev.var	-0.24	0.15	0.03	0.30	0.22	-0.10	0.27
elev.div	0.23	0.44	-0.06	0.39	0.69	-0.22	0.56
slope.med	-0.15	0.08	0.01	0.27	0.39	-0.28	0.39
slope.var	-0.21	0.12	-0.02	0.32	0.30	-0.11	0.34
slope.div	-0.27	0.10	-0.01	0.29	0.24	-0.09	0.30
asp.var	0.17	0.11	-0.02	0.01	0.09	0.02	0.03
asp.div	-0.14	-0.15	0.17	0.02	0.11	-0.25	0.09
precip.mn	-0.06	0.29	-0.15	0.44	0.81	-0.31	0.76
seas.mn	0.07	-0.22	-0.08	-0.41	-0.72	0.20	-0.59
seas.var	0.00	0.06	-0.09	0.21	0.36	-0.10	0.31

Appendix 2f. Spearmean rank correlation coefficients for explanatory variables summarized to the hexagon level.

	var.tot	onv.med	onv.var	med.sdn	var.sdn	med.range	var.range
med.sup	-0.02	-0.03	0.11	0.46	-0.01	0.61	0.12
var.sup	0.40	0.18	0.43	0.41	0.57	0.41	0.63
elev.med	-0.33	-0.46	-0.34	-0.58	-0.20	-0.63	-0.15
elev.var	0.08	0.30	0.26	0.21	0.08	0.18	0.09
elev.div	0.22	0.60	0.42	0.57	0.09	0.58	0.07
slope.med	0.05	0.45	0.29	0.32	0.05	0.30	-0.05
slope.var	0.09	0.36	0.25	0.29	0.13	0.25	0.12
slope.div	0.08	0.32	0.22	0.24	0.11	0.20	0.12
asp.var	0.15	0.02	-0.04	0.11	0.00	0.11	0.07
asp.div	-0.02	0.24	0.00	-0.01	-0.11	-0.01	-0.14
precip.mn	-0.02	0.82	0.55	0.71	0.30	0.64	0.01
seas.mn	-0.18	-0.81	-0.58	-0.59	-0.34	-0.50	-0.08
seas.var	0.01	0.35	0.24	0.32	0.21	0.28	0.04

Appendix 2g. Spearmean rank correlation coefficients for explanatory variables summarized to the hexagon level.

	med.sup	var.sup	elev.med	elev.var	elev.div	slope.med	slope.var
med.sup	1.00	0.40	-0.44	-0.03	0.24	0.03	0.00
var.sup	0.40	1.00	-0.18	0.15	0.14	0.08	0.18
elev.med	-0.44	-0.18	1.00	0.14	-0.54	-0.04	0.07
elev.var	-0.03	0.15	0.14	1.00	0.55	0.82	0.86
elev.div	0.24	0.14	-0.54	0.55	1.00	0.59	0.50
slope.med	0.03	0.08	-0.04	0.82	0.59	1.00	0.82
slope.var	0.00	0.18	0.07	0.86	0.50	0.82	1.00
slope.div	-0.05	0.13	0.13	0.86	0.49	0.78	0.95
asp.var	0.10	-0.03	-0.19	0.01	0.22	-0.01	0.00
asp.div	-0.20	-0.23	0.06	0.19	0.31	0.30	0.25
precip.mn	0.13	0.18	-0.39	0.38	0.60	0.53	0.43
seas.mn	0.16	-0.07	0.53	-0.06	-0.50	-0.19	-0.13
seas.var	0.08	0.14	-0.16	0.26	0.32	0.37	0.32

Appendix 2h. Spearmean rank correlation coefficients for explanatory variables summarized to the hexagon level.

	slope.div	asp.var	asp.div	precip.mn	seas.mn	seas.var
med.sup	-0.05	0.10	-0.20	0.13	0.16	0.08
var.sup	0.13	-0.03	-0.23	0.18	-0.07	0.14
elev.med	0.13	-0.19	0.06	-0.39	0.53	-0.16
elev.var	0.86	0.01	0.19	0.38	-0.06	0.26
elev.div	0.49	0.22	0.31	0.60	-0.50	0.32
slope.med	0.78	-0.01	0.30	0.53	-0.19	0.37
slope.var	0.95	0.00	0.25	0.43	-0.13	0.32
slope.div	1.00	0.03	0.30	0.40	-0.11	0.28
asp.var	0.03	1.00	0.01	0.06	-0.07	-0.13
asp.div	0.30	0.01	1.00	0.16	-0.18	0.16
precip.mn	0.40	0.06	0.16	1.00	-0.71	0.37
seas.mn	-0.11	-0.07	-0.18	-0.71	1.00	-0.21
seas.var	0.28	-0.13	0.16	0.37	-0.21	1.00

Appendix 3. The number of terminal nodes suggested from 3 methods of cross-validation for each taxa with 10 iterations per method.

	Prune Method	One Standard Error Rule	Adjusted Minimum Deviation
Mammals	11,11,10,10,12,22,12, 11,12,11	8,3,5,7,8,3,11,8,8,5	11,11,10,8,12,3,12,11,12, 11
Birds	5,5,6,5,5,5,25,6,5,21	5,5,5,5,5,5,5,5,5,5	5,5,6,5,5,5,5,5,5,5
Reptiles	16,5,31,16,29,10,25,22, 37,18	9,3,5,8,5,5,8,4,8,4	13,5,8,13,8,6,9,8,9,5
Amphibians	10,10,10,7,6,10,10,6,7, 6	5,5,5,5,3,5,5,5,5,5	10,6,10,6,6,6,6,6,6,6
All Vertebrates	20,16,16,16,24,20,7,26, 28,15	8,8,6,8,13,12,6,12,8,8	8,10,7,16,8,15,7,6,11,8