INVESTIGATING EASTERN NORTH AMERICAN TREE SPECIES RICHNESS AND DIVERSITY USING RELEVÉ DATA

by

JOEL SEAN SABIN

(Under the Direction of ELGENE O. BOX)

ABSTRACT

Species richness and diversity are products of interacting ecological, evolutionary, historical, and geographic processes. Processes varying at broader scales may influence data collected over finer scales, although factors unique to the fine scale of the data may obscure the link to broader-scale processes. Braun-Blanquet relevé sampling offers a fast and efficient method for collecting vegetation data. Sampling should be designed in accordance with the specific needs of investigation. However, caution must be applied when submitting relevé data to analysis beyond the original intentions of those who collected the data. No broad relationship was demonstrated between species evenness and environmental data. Mantel tests suggest relationships similar to regression results between species richness and environmental variables. The results of this study are not discordant with other similar research, although the relationships demonstrated are not as strong or significant, possibly because of scale differences among data.

INDEX WORDS: Braun-Blanquet, Relevé, Eastern deciduous forests, Species richness, Species evenness, Topographic heterogeneity, Spatial autocorrelation

INVESTIGATING EASTERN NORTH AMERICAN TREE SPECIES RICHNESS AND DIVERSITY USING RELEVÉ DATA

by

JOEL SEAN SABIN

B. S., Northern Michigan University, 2001

A Thesis Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2005

INVESTIGATING EASTERN NORTH AMERICAN TREE SPECIES RICHNESS AND DIVERSITY USING RELEVÉ DATA

by

JOEL SEAN SABIN

Major Professor: Elgene O. Box

Committee: E. Lynn Usery
Marguerite Madden

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2005

DEDICATION

To my parents John and Jackie Sabin, and to Gary Videto, for introducing me to geography, from my backyard to the entire world. Their encouragement has guided me in the exploration of the human and physical dimensions of the natural sciences.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank Dr. Elgene Box for the generous contribution of data used in this thesis, as well as his support and instruction. I would also like to thank Drs. Marguerite Madden and E. Lynn Usery for their input, advice and instruction with this thesis.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

Biological diversity has been a subject of formal geographic study since Alexander von Humboldt, Charles Darwin, Alfred Russell Wallace and others first theorized the mechanisms behind the global distribution of plant and animal species. The distribution of biological diversity is the product of a combination of processes involving ecology, evolution, history, and geography (Schluter and Ricklefs 1993). The term biodiversity has included biological organisms classified at different taxonomic levels, as well as features of the physical environment (Magurran 2004). A simple definition adopted by Magurran (2004, p. 8) will be used for this research: "the variety and abundance of species in a defined unit of study," with species restricted to plants, mainly trees. The number of currently competing theories to explain species diversity "reflects the prominence of diversity as a bellwether of our understanding of ecological pattern" (Ricklefs, 2004, p.1).

Although no general relationship has been empirically proven between higher species diversity and greater ecological stability (Kent and Coker 1992), biological diversity has long been regarded as an indicator of ecosystem health (Magurran 1988). Moreover, while an ecosystem composed of only a few species may appear stable from the relatively short-term perspective of a human researcher, it may not necessarily be the most desired from a human perspective (noxious non-native plants) or even the most stable when measured over longer terms. Greater species diversity provides biological resilience against future threats of climate change, disease and other causes of ecological collapse. The increasing pressures of anthropogenic selection limit habitat availability and species responses to environmental changes. Over geologic time, plant species have adjusted habitat range in response to climate changes. Now, potential plant species migration is greatly limited due to the broad-scale and rapid rate of human expansion.

Regardless of whether ecosystem changes are attributed to human actions or natural causes, we need to expand our knowledge of the processes influencing species diversity in order to preserve biotic integrity (Tilman and Pacala 1993). With improved understanding of these processes we can more effectively apply human effort and resources. Current conservation measures often focus scarce resources on “popular” species, many that are possibly beyond saving, while we continue to destroy other important ecosystems whose functions we do not clearly understand (Meyer 2004).

There is a need for broad-scale studies in order to understand the ecological processes behind biological diversity and to help inform future management and land-use decision priorities. Methods to estimate species richness are vital, as there is not enough time or resources available to devote towards a complete species inventory. It is estimated that at current discovery rates, a global inventory of plants, animal and fungi species could take 500 years (Magurran 2004, see May 1999). Clearly, research in biological diversity can benefit by using existing data collections.

1.1 Purpose of Research

This research investigates species richness and diversity of forest and tall woodland communities in eastern North America. Species data for trees, shrubs, herbs, and other plant types come from a forest and tall-woodland subset of communities sampled during the Eastern North American Vegetation Survey (ENAVS) (Miyawaki *et al.* 1994). The data were originally collected for a comparative study of the natural vegetation in eastern North America and that in Japan, funded by the Japanese Ministry of Education, Science and Culture (Monbushō). Data were collected in 1988, 1989, and 1990 using the Braun-Blanquet relevé method of survey.

Species richness and diversity information from the ENAVS dataset were examined for relationships among factors within the data, such as species distribution by vertical vegetation layer, site elevation and topographic position, as well as independently collected climate and topographic data.

The primary research objective was to explore any significant relationships between tree species richness and diversity from the ENAVS data and environmental data collected from other sources. A secondary objective was to analyze observed data using spatial analysis and geographic information system (GIS) techniques. This was done in order (1) to explore the utility of secondary data from private and public sources for biodiversity research, (2) to avoid the costs associated with collecting primary data, and (3) to assess the utility of GIS in analysis of relevé data.

The primary methods of analysis included linear regression models, using ordinary least-squares methods. Correlograms and Mantel tests were used to investigate spatial autocorrelation in a data subset. Mantel analysis is based on correlations between resemblance matrices created from data distances and geographic distances (Legendre and Legendre 1998).

Chapter two will present background information from the biogeographical and ecological literature and research related to broad-scale patterns of species richness, highlighting which variables should be analyzed. Chapter three describes data details and methods of analysis. A comparison is made between the Braun-Blanquet and North American quadrat methods of vegetation sampling. Chapter four presents the results from the methods of analysis, and is followed by a discussion of the results and the conclusion in Chapter five.

CHAPTER 2. RESEARCH IN SPECIES RICHNESS AND BIODIVERSITY

2.1 Background

One of the most noted trends in global species richness is the latitudinal gradient, with fewer species being observed as sites are compared among distances further from the equator (Rosenzweig 1995). According to Hawkins *et al.* (2003), there are presently over 30 hypotheses to explain the latitudinal gradient in broad-scale species richness. Testing these hypotheses is difficult due to a lack of high quality environmental and species-diversity data for large latitudinal extents (Huston 1999).

Likewise, longitude has a strong correlation with species richness as an expression of moisture gradients (Currie and Paquin 1987). Species richness in Australia and southern Africa follows a longitudinal trend more closely than latitude (Hawkins and Diniz-Filho 2004, O'Brien 1993). Longitudinal species richness gradients of non-native plant species have been linked to settlement dates, which strongly correlate with longitude (Clemants and Moore 2003).

Latitude and longitude should not be considered causal factors in species richness. Earth-sun relationships cause latitudinal first-order climate variation (O'Brien 1993). Other factors such as topography or large water bodies modify the distribution of water and energy in relationship to species richness. The spatial distribution of data collection locations can influence the contribution from these factors defined in the data. The ENAVS data were gathered in a pattern generally parallel to the East Coast of North America between the Atlantic Ocean and the Appalachian Mountains and diagonal to the axes of latitude and longitude. Some locations extend along the North Gulf Coast and into South Florida.

2.2 Spatial Scale and Area

Species richness estimation is complicated by the effect of spatial scale and area. Biodiversity is inherently scale dependent. Research analysis must be compatible with the spatial scale of the raw data (Meentemeyer and Box 1987). The relationship of species richness to some factor measured at one scale may differ from the same relationship examined at another scale. Many plant species have clumped distributions; therefore, variation will be greater when comparing smaller areas than when comparing larger areas (Gauch 1982).

Larger areas have consistently been shown to contain more species than smaller areas, assuming that climates are similar (Hawkins *et al.* 2003). Species with larger ranges typically have lower extinction rates and greater abundances, both factors linking species richness to area. Larger areas experience greater variation in climate and in environmental heterogeneity, providing a greater variety of potential species niches (Rosenzweig 1995). In homogenous environments, species richness may be reduced through competitive exclusion if smaller areas confine species range and resource availability.

Many important ecological factors can covary with area, while also being independent from it. Controlling for area when using different sized samples may obscure these covarying factors (Whittaker *et al.* 2001). Ratios of species richness to area (species density) are misleading because the relationship of species to area is nonlinear (Gotelli and Colwell 2001). Area is often used as a proxy for plant species abundance, measured by percent species cover, as vegetative reproduction makes it difficult to identify and count individual species members (Pielou 1977). This method removes the dimension of area; however, it also reduces the information content of the abundance data, as with area, the relationship of species cover to individuals is nonlinear.

2.3 Energy and Productivity

Most plants (and animals consuming plants) depend on solar energy and precipitation to sustain biological functions. Water and energy are each required for photosynthesis. The amount of water returning to the atmosphere from evaporation and plant transpiration under an ideal soil-water balance is called potential evapotranspiration (PET). Actual evapotranspiration (AET) accounts for local soil-water balance conditions. Many studies have shown species diversity to correlate with productivity, although explanations are difficult as productivity is related to numerous ecological variables (Rosenzweig 1995). Measures of productivity combine energy and moisture, and two areas with different climates can produce the same AET values; therefore, it may be best to keep energy and moisture as separate variables (O'Brien 1998).

At broad spatial scales, the general trend between species richness and energy is a parabolic curve, with a positive relationship up to a peak, after which the relationship becomes negative with further increases of energy (Rosenzweig 1995, Huston 1999). Huston (1999) relates this to two factors that result from local processes: low diversity (1) as a result of increased mortality due to low productivity, or (2) as a result of competitive exclusion due to high population growth rates associated with high productivity. A simple explanation from Denslow (1980) relates the parabolic relationship between species richness and productivity to the distribution of extremes in high and low productivity; fewer species can take advantage of extreme conditions because they occur less frequently (Tilman and Pacala 1993).

O'Brien (1993) developed the Interim General Model to explain the parabolic relationship between woody plant species richness and climate: Species richness = f [Water + (Energy – Energy^2)]. This model accounted for 78.8% of woody plant variation examined at regional scales (25,000 km^2 grid cells) in southern Africa. It assumes that trees and shrubs maximize water (annual precipitation) consumption and optimize energy (minimum monthly potential evapotranspiration {PET}) consumption (O'Brien 1993).

Stephenson (1998) found that the annual climatic water deficit correlated with whether or not forests could exist, and that annual actual evapotranspiration (AET) (see Thornthwaite and Mather 1955) correlated with the variation between deciduous and coniferous forest types in the western U.S. Currie and Paquin (1987) found AET to account for over 75% of the variation in species richness of North American trees at continental scale; adding the influence of topography increased explanatory power to 86% of variation in richness.

In a comparison of 82 studies, Hawkins *et al.* (2003) found moisture and energy variables to explain over 60% of plant and animal species richness. They emphasize that while moisture and energy variables are not the only factor in species richness, and may not apply to every species, they are essential considerations for biodiversity research. They also note that there appear to be transitions from moisture to energy as the primary limiting factor as observations are made from the equator to the poles.

2.4 Topographic Heterogeneity

Topographic heterogeneity is used as a general term for abiotic influences of landscape structure on species richness. Patterns of species richness correlate at local scale with the diffusion of moisture and energy from topographic heterogeneity (Whittaker *et al.* 2001). Competitive exclusion is reduced in a heterogeneous environmental setting relative to a homogenous one (Huston 1999). Two sites with similar areas and climates may have differences in the way sunlight and precipitation are received and dispersed, creating microclimates that favor one species over another.

Burnett *et al.* (1997) concluded that the species richness of eastern North American deciduous forests is a product of high geomorphological heterogeneity patterns occurring at very local (patch) scale. In a related study, Nichols *et al.* (1998) found geomorphological heterogeneity to account for over 65% of the variation in total plant species at broader landscape scale. By adding topographic relief to a moisture/energy model of tree and shrub species diversity in southern Africa, O'Brien (2000) increased the model coefficient of determination (R^2) from 78.8% to 85.6%, in a pattern geographically unrelated to latitude or longitude.

Rosenzweig (1995) warns that to assume that heterogeneous environments cause species richness is a tautology. Natural selection suggests that species evolve to exploit local habitat conditions (niches) for competitive advantage; therefore the species must come before the niche. However, this logic applies to many environmental factors suggested as causal mechanisms for species richness. A counter argument could be made that heterogeneous environments contribute to species richness by facilitating competitive advantage differently than homogenous environments. Regardless, topographic heterogeneity serves as an indicator of potential species richness, particularly in response to environmental changes.

2.5 Summary

Clearly, there are many factors influencing species richness, each with different scales of influence. Area is a fundamental factor, although its significance is similar to latitude and longitude, mostly in confounding the measurement of other factors (Whittaker *et al.* 2001). Generally, if latitude is held constant, species richness varies with area, while species richness varies with latitude if area is held constant (Rosenzweig 1995, Hawkins *et. al* 2003). Moisture and energy influence species richness over larger extents due to geographic variation in their intensity and duration. Over smaller extents, the variation in moisture and energy appears consistent when measured as climate, and moisture and energy distribution patterns are diffused by topographic variation (Whittaker 1999). Complex ecological and biological factors such as competition and disturbance also influence species richness over many spatial scales. This research focuses on geographic and environmental factors, although ecological processes such as competition and disturbance not directly examined here are acknowledged as important and interrelated influences on species richness.

The relationship between species richness and area demonstrates the importance of considering spatial scale. Many of the studies discussed (Hawkins *et al.* 2003, O'Brien 2000, Nichols *et al.* 1998, Stephenson 1998, Currie and Paquin 1987) used data sampled at coarser scale over large spatial extents. Although the ENAVS data were collected at a sub-continental extent, the minimum scale resolution, or 'grain', of the individual sample areas represents local-scale samples (Palmer and White 1994). This suggests that species richness variation contained in the ENAVS data should be related more to factors whose variation is observed at finer scales, such as topography, soil texture, and soil quality (Whittaker *et al.* 2001). Climate variation is observed over larger extents and may relate more to physiognomic diversity, as measured by plant growth forms and vegetation structure (Neilson *et. al* 1992).

CHAPTER 3. DATA AND METHODS

3.1 The ENAVS Survey

The data for this research were originally collected for a comparative study of the natural vegetation in eastern North America and that in Japan. Both areas are similar in latitudinal positions and geographic extents. Areas examined in the eastern North America survey have been altered over the span of geologic time by global climate variation and the accompanying glacial activity and sea-level changes. Native Americans modified the landscape to some extent, although recent European settlement of the region was the most intensive and extensive (Box 1994). Most of the original vegetation has been altered and most forests are second growth (Miyawaki 1994). Consequently, preferential sampling was necessary for the survey to represent potential natural vegetation. The data are primarily from deciduous forests, indicating stand ages that are generally mid-succession or later.

3.2 The Braun-Blanquet Method and Relevés

Field collecting of data is typically the most expensive and time-consuming activity in conducting research. The Braun-Blanquet relevé method was used to gather the ENAVS data because it provided an efficient way to perform a geographically extensive vegetation survey (Box 1994). In this sampling method, a uniform area representing a plant community is selected and described in terms of vegetation and site characteristics. All species are inventoried in a sample plot called a relevé. Relevés allow large areas to be studied fast, allowing broader comparison of vegetation communities than possible when considering time and funding limitations of other sampling methods (Box 1994).

Species abundance was expressed using a six-class (Braun-Blanquet 1964) cover-abundance scale (Nakamura *et al.*1994). Visual estimates of the percentage of area covered by each species were made for the following vertical vegetation layers: 1) tree canopy, 2) tree understory, 3) shrub, and 4) herb. The same species may be present in several layers. For example, a tree species can be recorded as a seedling in the herb layer, as a young tree in the shrub layer, and as a mature tree in the canopy layer.

The cover-abundance scale is a dimensionless, non-linear scale that permits scale-independent comparisons of vegetation to be sampled from different sized areas. Relevé locations are selectively chosen to contain homogenous vegetation. Area is adjusted based on species composition to ensure adequate sampling of all species (Mueller-Dombois and Ellenberg 1974). Area is increased until all species are presumed to have been identified. The cover-abundance scale begins with the symbol "+" for individual species occurring with less than 1% cover, and increases by integer values from 1 to 5 as species cover increases. A cover-abundance scale value of 5 represents species with greater than 75% cover. The complete scale range is shown in Table 3.1.

Table 3.1: Cover-abundance scale and transformation to percentage.

Braun-Blanquet cover-abundance	Species cover	Middle percent value
+	Less than 1%	0.5 %
1	1 to 5%	3 %
2	5 to 25%	15 %
3	25 to 50%	37.5 %
4	50 to 75%	62.5 %
5	75 to 100%	87.5 %

3.3 Braun-Blanquet Sampling vs. North American Quadrats

The Braun-Blanquet method comes from the European scientific discipline of phytosociology. A prominent feature of the method is that the relevé sampling area can vary in order to contain the necessary homogenous vegetation used to classify a plant community. This differs from the customary North American practice, where predefined-sized quadrats are used to sample vegetation. Quadrats use estimates of species basal area by counting stems and by measuring diameter at breast height, in contrast to the cover-abundance estimates used in relevés.

North American vegetation studies have typically studied forests and grasslands, where identification of individual species members is reasonably straightforward. Phytosociology developed around the study of plant communities, where it is often difficult to differentiate between distinct groups of individual plants and clonal species with connected shoots, etc., belonging to a single plant. Counting individual stems can be very time-consuming in species-rich communities (Knapp 1984). In comparison, relevés can be done much faster, as no measurements are made in the field.

Mueller-Dombois and Ellenberg (1974) relate the different sampling strategies to compositional differences between North American and European vegetation. European forests typically have fewer indigenous tree species. Consequently, European vegetation ecologists focus on understory species. They also note a difference in scale of study extent: vegetation is usually studied over larger areas in North America than in Europe (Mueller-Dombois and Ellenberg 1974).

3.4 Relevé Data Organization

A subset of 331 relevés representing forest and tall woodland communities was selected from the complete ENAVS dataset. Selection was done by computer based on the occurrence of major deciduous tree species. This was to avoid including dissimilar community types in the ENAVS data such as grasslands. From this subset 106 relevés were excluded based on: (1) dominance by evergreen species, including gymnosperms and warm-temperate angiosperms, (2) dominance by wetland or swamp species, (3) urban forests and disturbed areas in early succession, and (4) areas less than 200 m^2 (Elgene Box, personal communication, November 2004). This created a final set of 225 relevés referred to as the deciduous forest and tall woodland data, abbreviated and referred to hereafter as DFW.

Data attributes for vegetation include estimated height, percentage of cover, and number of species, all recorded for each herb, shrub, tree-understory, and canopy layer. Other site-related data include relevé area, total number of species, elevation, and slope. Data were arranged in a matrix with attributes by column and relevés by rows using the program RELSPECS (Box, unpublished). These data will be further referred to as the DFW summary file.

A second file called the "raw table" was produced containing individual species cover-abundance data. The name comes from the tables used in phytosociology for classifying relevé data from its raw form into an ordered form that reveals vegetation associations. The raw table is the starting point for standard phytosociological data analysis.

Cover-abundance scale values were transformed by taking the mid-value of the percent cover range for each of the scale increments; in effect converting the abundance data from a non-linear ordinal scale into a continuous ratio scale (Table 3.1). Species data were stratified into the categories 'tree,' and 'non-tree,' by adding a column of growth form codes to the raw file (Table

3.2). The tree category includes growth forms labeled ‘tree’ and ‘palm.’ The remaining growth forms are simply called ‘non-trees.’ Although further division of the growth forms into more categories was not explored for this research, but it could prove interesting.

Table 3.2: Species growth forms.

	Incidence	% of Species	Code	Type
	191	10.7%	T	Tree
	4	0.2%	P	Palm (tuft tree)
	223	12.5%	S	Shrub
	5	0.3%	D	Dwarf-shrub
	1	0.1%	r	Rosette-shrub
	311	17.4%	G	Graminoid
	819	45.9%	F	Forb
	1	0.1%	f	Fern
	107	6.0%	V	Vine
	5	0.3%	E	Epiphyte
	13	0.7%	c	Cryptogam
	18	1.0%	a	Aquatic
	85	4.8%	(blank)	(unidentified)
Total	1783			

3.5 Study Area

More than 1200 relevés were collected for ENAVS, over an extent of approximately 3000 kilometers, north to south. Few field vegetation studies have collected data over such a broad scale. To ensure that sampling represented species diversity, sites were chosen to include a range of topographic units, including mountains, floodplains, and coastal regions. The entire set of ENAVS relevés represents most of the forest types of North America east of and including the Appalachians. Zonal vegetation types included alpine, boreal/montane, northern temperate, southern temperate, warm-temperate, and subtropical. Azonal types included coastal, wetlands, floodplains, edaphic, and urban. Most relevés are in the northern temperate and southern temperate zones and represent deciduous forest vegetation (Box 1994).

The DFW subset of relevés ranges from 47.65°N (Lac des Cygnes, Quebec, Canada) to 29.55°S latitude (Interlachen, Florida, U.S.A.), and from 90.5°W (McComb, Mississippi, U.S.A.) to 68.22 °W longitude (Franklin, Maine, U.S.A.). Most relevés are located in the Coastal Plain and Piedmont physiographic provinces. The remaining relevés by proportion are within the Appalachian Plateau, New England, Ridge and Valley, Blue Ridge and Saint Lawrence Valley provinces (Figure 3.1).

Elevations and slopes reflect the high proportion of Coastal Plain and Piedmont relevés. Median relevé elevation is 160 meters, with the majority between sea level and 300 meters. The maximum elevation (1600 meters) is located in the Smoky Mountains. Overall, slope ranges from flat to 45% gradient. Median slope is 2%, although most relevés have no appreciable slope (Table 3.3).

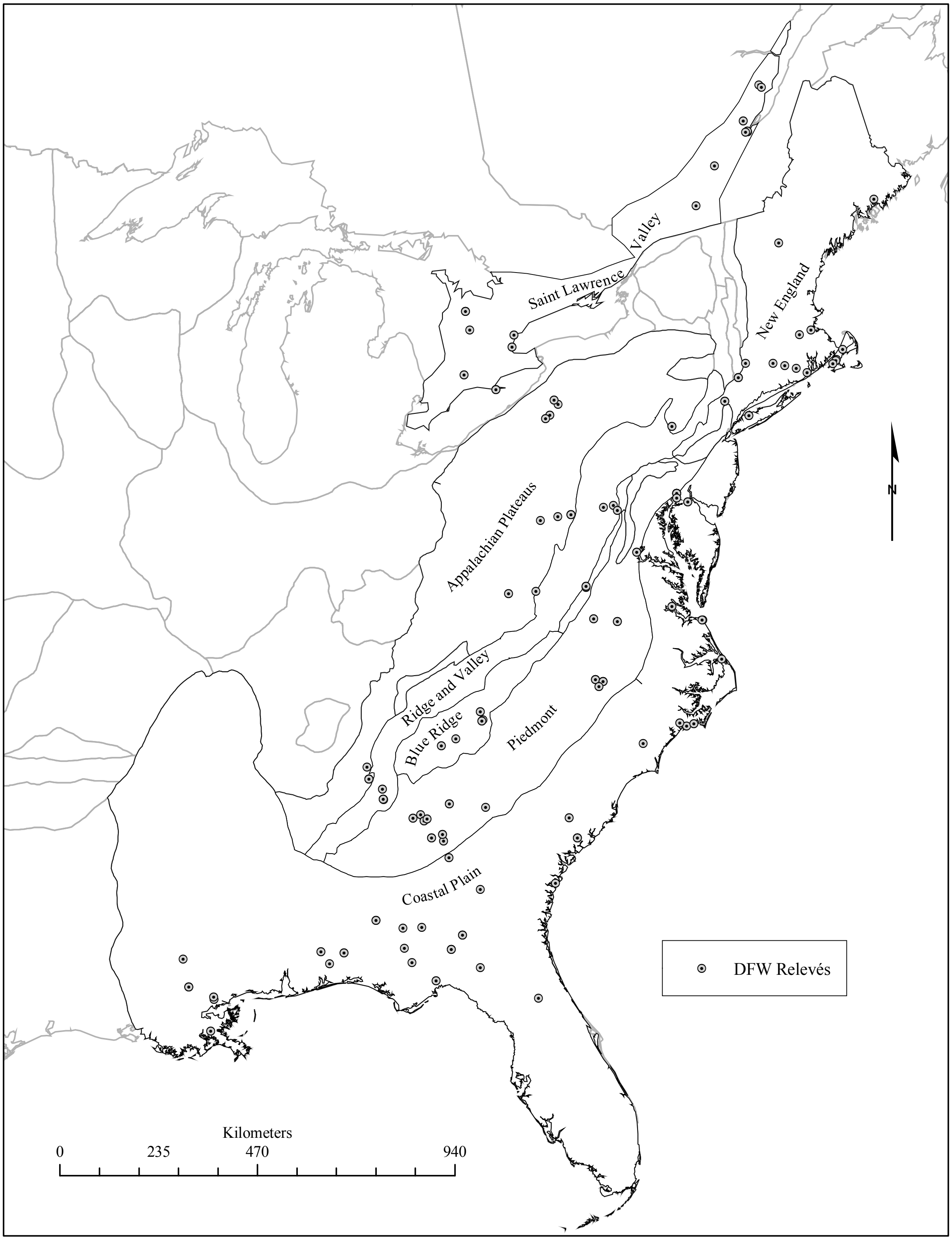


Figure 3.1: Location of the 226 DFW relevés relative to physiographic provinces.

Table 3.3: Descriptive statistics for relevé data.

225 Observations	Min.	Max.	Extent (km.)
Latitude	29.55 N	47.65 N	2248
Longitude	-90.50 W	-68.22 W	1644

	Min.	Max.	Mean	Median	Mode	Std. Dev.
Area (sq. m)	200	2,500	937.3	750	600	647.3
Elevation (m)	0	1,600	258	160	40	305.3
Slope (%)	0	45	5.8	2	0	8.4

Number of species	Min.	Max.	Mean	Median	Mode	Std. Dev.
Trees	3	29	14	14	13	4.9
Non-trees	7	65	25	23	16	10.5
All layers	15	75	39	37	32	12.5
Layer 1 - Tree canopy	1	12	5.3	5	4	2.1
Layer 2 - Tree understory	1	15	5.4	5	4	2.6
Layer 3 - Shrub	0	37	13.8	13	11	6.6
Layer 4 - Herb	4	58	23.7	22	18	10.6

The DFW relevés cover a range of climatic zones, extending from the middle of the warm-temperate zone in Florida to the southern limit of the boreal zone in southeastern Canada. The majority of the relevés (over 60%) are within the typical-temperate climate zone (*sensu* Walter 1977), with about 30% in the warm-temperate zone. Annual average temperature ranges from 1.4 to 22.5°C. Monthly high and low temperatures range from an average of –17.9°C in the coldest month (Quebec), to 28°C in the warmest month (Florida). Temperatures generally follow the latitudinal gradient of the site locations. Precipitation also follows a latitudinal gradient, although not as closely as temperature. Annual average precipitation ranges from 750 to 1936 mm. Monthly high and low precipitation ranges from an average of 43 mm. in the driest month to 273 mm. in the wettest month.

3.6 Climate Data

Species richness and diversity are fundamentally related to the distribution of usable water and energy (Whittaker 1999). Climate conditions were estimated to analyze water and energy relationships with species richness and diversity patterns. Climate data are long-term records from meteorological stations nearest to vegetation sampling locations. Data were mostly compiled from the National Climatic Data Center, although other sources are included (Elgene Box, personal communication). Two computer programs, POLATE (Box, unpublished) and SOLWAT (Box 1982), were used to add climate variables to the DFW relevé statistics file.

Data were used directly from meteorological stations that were reasonably near each relevé. When a station was not near a relevé, combinations of stations were used based on location and data quality. The POLATE program estimates site climate by weighted triangulation, using up to five of the nearest meteorological stations. Averaged values were used when relevés were located between two stations; otherwise values were extrapolated by averaging data included from a third station. Stations with records collected over longer periods were used when two stations were approximately the same distance to a relevé.

Climate data added to the summary file include: mean annual temperature (TY), mean temperature of the warmest month (TMAX), mean temperature of the coldest month (TMIN), mean annual precipitation (PY), mean precipitation of the wettest month (PMAX), and mean precipitation of the driest month (PMIN). Annual potential evapotranspiration (PET) and biotemperature (BT)[1] were calculated, and annual actual evapotranspiration (AET) was calculated from standard water budget procedures (*e.g.*, Mather 1979) as embodied in SOLWAT. A list of descriptive statistics for climate variables is in Table 3.4.

---

[1] Defined as the mean annual temperature, with periods below 0° C considered as 0° C when averaged (Holdridge 1967).

Table 3.4: Descriptive statistics for climate data.

225 Observations	Min.	Max.	Std. Dev.
Average annual temperature (C)	-1.4	22.5	5.28
Average temp. of coldest month (C)	-17.9	16.1	7.27
Average temp. of warmest month (C)	13.5	27.8	3.37
Average annual precipitation (mm)	750	1936	195.88
Avg. precip. of driest month (mm)	43	125	13.32
Avg. precip. of wettest month (mm)	76	273	34.07
PET (mm)	406	1326	245.93
AET (mm)	406	1241	241.85
Biotemperature (C)	4.2	22.5	4.63
Moisture Index (PY/PET)	.96	3.54	.50

3.7 Species Richness and Diversity

Species diversity is measured by two primary components: species richness and species evenness. Species richness is the total number of species in a sample. The second component, evenness, has been referred to by a variety of other names including equitability, heterogeneity, diversity, and dominance. Evenness measures the relative abundances of individuals, for each species. Other units of abundance may be used, such as biomass or cover.

Species richness is insensitive to changes in species abundance, unlike measures of evenness. Species richness is biased by rare species; total species richness increases if a new species has one or a thousand members – this is why diversity indices combine richness and evenness. Species richness and evenness are sensitive to sample size; as sample size increases species richness generally increases while evenness generally decreases (Hurlbert 1971).

Diversity indices can be parametric or non-parametric, depending on the assumptions made about the species distribution. One example of a parametric diversity index is α (Fisher *et al.* 1943), which is a parameter of the log series model. Although it is a relatively robust measure, α requires that individual units are used to measure abundance and therefore not applicable to this research (Magurran 2004). Non-parametric diversity indices have been used frequently to summarize ecological communities, since they were introduced in the late 1950's, although they have been criticized over the past 30 years (Hurlbert 1971, May 1975, Magurran 1988 and 2004, Rosenzweig 1994).

Because values from diversity indices are related to the numbers of species and abundances per species, the same value can apply to any number of entirely different vegetation patterns. Species richness and evenness can be positively or negatively correlated, depending on a community's composition. For this reason, they should not be used to compare drastically differing communities (Magurran 1988). The contribution an abundant species makes to a diversity index reflects the structure of the community in which it occurs, not the species' ecological importance (Hurlbert 1971). High species diversity can result from an equal distribution of many common or undesirable species, while low species diversity may be due to unique or rare species.

Combined measures of diversity such as the Shannon-Weaver (1949) or Simpson (1949) index were not used in this research, as the vegetation patterns are too complex to be meaningfully summarized with one measure, especially over broad variety of habitat types. Additionally, the dimensionless nature of the abundance data limits the application of these indices. Most indices require integer values representing individual organisms. Converting 30% to 30 plants or trees is unjustified. The relationship between cover-abundance and number of individuals is not likely to be linear because different plant species are of different sizes and shapes. Species richness and evenness were examined separately rather than using diversity indices combining these measures.

Methods for measuring species evenness are typically based on counts of individual species members. The evenness index E_{var} uses proportional differences of species abundance variance to provide values that are unit-free (Magurran 2004, Smith and Wilson 1996). Maximum evenness equals one and minimum evenness equals zero. E_{var} was decided as the best method to make use of the original species cover-abundance scale, using the transformed percent cover values from the raw table file. A column of E_{var} values was added to the DFW summary file for further analysis.

A resemblance matrix was created using tree species data from the raw table. The resemblance matrix represents beta-diversity, a measure of species turnover between sites. These data were used in partial Mantel analysis to compare matrices of tree species beta-diversity and environmental variables while controlling for geographic distance and spatial structure.

3.8 Topographic Heterogeneity Indices from DEM Data

Landform characteristics influence local distribution patterns of moisture and energy (Bailey 1998, McNab 1993, Miller 1996, O'Brien 1998, 2000). Variation of moisture and energy specific to landform shape creates local climate conditions called topoclimates or microclimates (Bailey 1998, see Thornthwaite 1954). Topographic heterogeneity also influences variation in soil drainage and composition. Microclimate and soil factors each correlate with species richness patterns measured at local scale.

Topographic heterogeneity has been examined through use of non-parametric diversity indices. Miller (1986) describes a method using the Shannon index applied to elevation diversity, in a study of rare plants in the Appalachians. However, this study compared topography and species richness over greater scale resolution, using point-data samples occurring within classes of topographic relief. Classes were determined from elevation grids of areas ranging from 180 to 4,000 km^2. The influence of topographic heterogeneity and climate together were not included in their study.

Burnett *et al.* (1998) and Nichols *et al.* (1998) also used the Shannon index for comparing point-data for species richness within large-area grids, which were classified by topographic relief. Each study combined soil survey data with DEM data to create classes of geomorphological heterogeneity based upon geological differences (as opposed to biologically related soil differences). Burnett *et al.* (1998) sampled vegetation in plots classified as having high or low geomorphological heterogeneity and compared differences, while Nichols *et al.* (1998) used existing species data for sites within their classes of geomorphological heterogeneity.

Similar ideas were explored using DEM data (United States Geological Survey, National Elevation Dataset {USGS-NED} 2004) with the ENAVS data, although there are significant differences between the data used as well as the research designs of each of the cited studies. Burnett *et al.* (1998) collected vegetation data after classifying geomorphological heterogeneity and did not consider topography over a gradient or climate, which were both objectives in this research. Miller (1986) used elevation data covering a much larger extent (180 to 4,000 km^2) than the DEM data in this research (.04 km^2). DEM data coverage of a smaller extent was used with the DFW relevés because topographic heterogeneity is theorized to be a local-scale influence and there is no clear reason why topography hundreds of kilometers away from a site should be expected to influence the vegetation there. Additionally, larger DEM extents would have required many raster data to be collected and processed.

Topographic index values were created using one-third arc-second DEM data (10-meter horizontal resolution), downloaded from the National Map Seamless Data Distribution System (USGS, National Elevation Dataset 2004). At the date of collection, coverage of 10-meter DEM's was not available for the entire extent of the relevés (approximately 2250 x 1650 km.). Consequently, data were limited to 144 of the 331 ENAVS relevés. Although 30-meter resolution data were available for the entire ENAVS extent, the 10-meter data were chosen to match to the scale of individual relevés better. Some of these relevés are not in the 225 DFW relevés, as the decision to reduce the dataset to 225 relevés had not been made at the time that the DEM data were collected. None of the 144 relevés fall outside the DFW sampling extent (Figure 3.2).

A collection of 20 x 20 cell grids of DEM data was downloaded, with each representing approximately .04 km^2 of ground area. A 3 x 3 window of pixels centered on the nearest relevé location was selected from this larger sample to create the topographic indices, approximately equal to an area of 900 m^2. Because the USGS-NED uses the geographic (latitude/longitude) coordinate system, DEM data were projected to the Universal Transverse Mercator (UTM) coordinate system, to assign similar unit values (meters) to the horizontal and vertical coordinates (x, y, and z). Nearest-neighbor resampling was used for the projection transformation. Elevation and percent-slope surface values created from these data were exported to text format for calculating topographic indices using Excel (Microsoft 2000). ArcGIS (version 9) was used for reprojection and creation of slope surfaces.

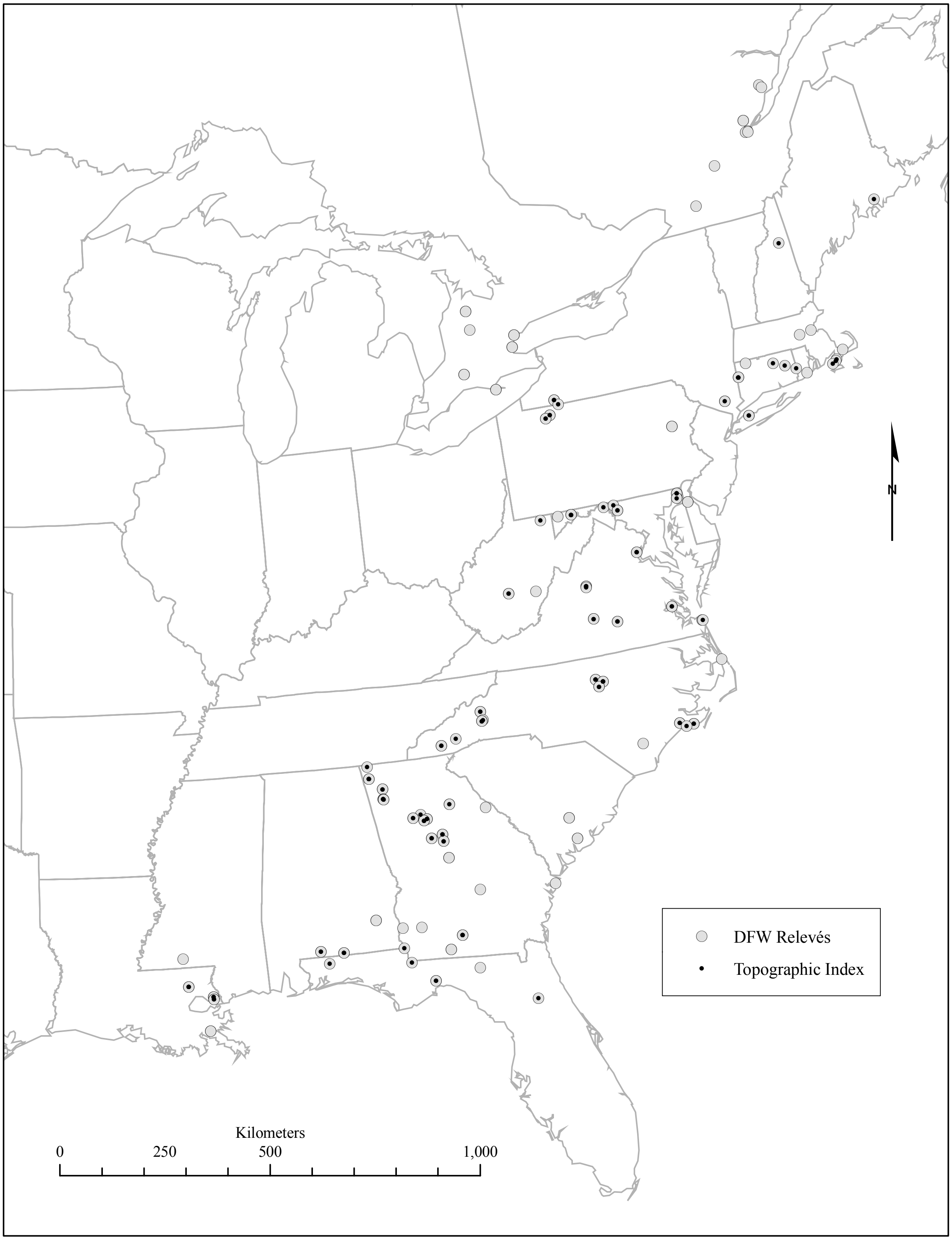


Figure 3.2: Locations of relevés with topographic index, relative to all 225 DFW relevés.

Non-parametric diversity indices were tested with slope data from the NED-DEM coverage of the ENAVS relevés (Table 3.4). The small 3 x 3-cell window of slope values meant that topographic heterogeneity indices were based on only nine numbers. Using the Shannon-Weaver index produced a limited range of index values because it is constrained by sample size (DeBenedictis 1973, Magurran 2004). The Simpson index is less sensitive to sample size; however, slope values less than 1% produced negative index values. Rescaling the slope data produced positive Simpson index values, but with a limited range.

Smith and Wilson's (1996) evenness index E_{var} works with any unit of measurement (Magurran 2004). A single E_{var} value was calculated for each 3 x 3-cell slope sample, and then multiplied by the number of different slope values in each sample to produce a diversity index. This provided a greater overall range of index values by weighting topographically complex areas more than flat areas. Using an index combining the number of different values (richness) and evenness in one measure in this application is not as problematic as it is with species data. There are fewer possible slope values from each DEM sample than possible species in a vegetation sample, and the number of values is fixed by the size of the pixel window. The E_{var} based diversity indices (EVARDIV) were added to the DFW summary file.

Table 3.4: Comparison of indices tested as topographic heterogeneity index.

Index		Range	Comments
Shannon & Weaver (1949)	H	1.77 - 2.19	Limited range of values
Simpson (1948)	D	.79 - .99	Negative values/ limited range
Smith and Wilson (1996)	E_{var}	.23 - .99	Limited range of values
Modified version	*EVARDIV*	2.12 – 8.99	Greatest range, strongest correlation

3.9 Data Considerations

An objective of this research was to investigate the merits of subjecting data collected for one specific purpose to analysis intended for a different purpose. The ENAVS data are attractive for biogeographic research due to the large spatial extent of the vegetation samples. At the same time, the relevés represent a collection of non-randomly clustered, fine-scale observations, and substantial portions of the overall sampling extent are unaccounted by the data. The objective of the original survey was to represent a variety of natural vegetation, which often meant going to areas where a vegetation type still existed without substantial influence from human "development." This introduces sampling bias, although such bias exists in much field-recorded sampling and is often practically unavoidable, and even was desired in the case of ENAVS (Box, personal communication).

The overall shape of the sampling extent further confounds analysis. The overall pattern of relevé distribution has a shape that is roughly linear, generally along a northeast-to-southwest major axis. This is also a practically unavoidable consequence, related mostly to the shape of North America. As a result, analysis is complicated because environmental factors that vary along the direction of the major axis of the sampling extent are weighted by more observations than factors orthogonal to this axis. To complicate matters further, two water-energy patterns exist over the sampling extent. One is influenced by the angle of the sun at different latitudes, while the other is from elevation change beginning at the seacoast and crossing over the Appalachian Mountains. Each of these patterns affects both temperature and precipitation levels.

3.10 Issues Related to Braun-Blanquet Sampling Methods

The Braun-Blanquet relevé sampling method has been criticized by those in favor of more “quantitative” methods for many reasons, including: (1) errors in species abundance estimation, (2) the nonlinear scaling of cover-abundance values, and (3) the use of sample areas with varying sizes. The first two issues could have affected the measurement of species evenness but this is difficult to prove, as no relationships were suggested between species evenness and other variables. The third problem deserves more concern, as species richness estimates are scale dependent; discussion follows.

3.11 Species Abundance Estimation

Cover-abundance scales can introduce bias through overestimation of prominent species and underestimation of less prominent species (Kent and Coker 1995). For instance, a person with little vegetation sampling experience can perceive a higher abundance of noticeable flowering plants while missing less conspicuous species, or an expert may be biased toward species related to his/her domain of study. The degree of potential bias varies according to the experience of the data collector. The ENAVS survey used experienced workers to collect data (Box 1994). Gauch (1982) notes that the natural variation among sample abundances limits the practical value in making measurements with accuracy much better than 5% of mean sample abundance, and that natural variation is typically larger than error introduced by visual estimation (see Orlóci 1978).

Estimation bias can also result from the effects of seasonality, and comparisons between sites must be made under similar seasonal conditions (Poore 1955). This condition was partly met in the case of the ENAVS study; surveys for 1989 and 1990 were made during

September/October, but the 1988 survey was made in June. After the data have been sampled, nothing can be done concerning species estimation errors. Errors in cover-abundance estimation may obscure any relationships between species evenness and environmental factors. Although the impact may be negligible, there is no way to estimate it.

3.12 Nonlinear Scaling

Table 3.1 shows that the first three Braun-Blanquet cover-abundance scale values do not cover equally spaced intervals. This is because the scale combines cover and abundance values. Values 2 through 5 designate the portion of a species crown contributing to cover, while the lower values usually designate individual species members. The scale intervals reflect the general structure of a natural plant community such as a forest. Species that are sparsely distributed contribute the least to the total biomass, while species contributing the most biomass usually have the greatest cover. Using this system, several plant communities can be analyzed in the same time that it would take for precise measuring of one community. The method is very useful to community classification, a main object of phytosociology (Mueller-Dombois and Ellenberg 1974).

Magurran (2004, p.140) is critical of using cover-abundance scales to measure species diversity, stating that they “provide the most resolution at maximum and minimum coverage,” and that “the nonlinear nature of the data they generate impedes interpretation.” The potential for diversity estimates from cover-abundance data to be biased towards the upper and lower cover values should not obscure general trends in species evenness. Relationships between abundance estimates and other factors may still be informing, if limited in statistical strength.

Note that issues related to diversity estimates from cover-abundance scales concern the evenness of species distribution, not species richness. Species richness is not affected by scaling since it is based on the presence or absence of a species. Species richness can be biased by area, which may be of concern considering that relevés typically vary in size.

3.13 Variation in Relevé Area

Whittaker *et al.* (2001) state that the simplest and best method to control for the effect of area on species richness is to use equal-sized sample areas, and that if a factor does not vary over fixed areas it cannot be a cause for non-random patterns of species richness. Since the primary use of the relevé method is for plant community description and classification, relevés are not usually chosen with the goal of estimating species richness patterns or other parameters (Chytrý 2001). Therefore, relevé sizes often vary to aid in classification.

Braun-Blanquet cover-abundance is considered an absolute scale, where cover percent refers to relevé area, and permits comparison of vegetation sampled from different sized areas (Mueller-Dombois and Ellenberg 1974). Relevé locations are selectively chosen to contain homogenous vegetation. Relevé area sizes are adjusted based on species composition, with minimal areas chosen that ensure complete sampling of all species for different vegetation communities (Mueller-Dombois and Ellenberg 1974). However, there is theoretical and empirical support that area is highly correlated with species richness (Chytrý 2001, Whittaker *et al.* 2001, and Rosenzweig 1995). This introduces two problems: (1) controlling for area could obscure environmental variables that covary with area (Whittaker *et al.* 2001), and (2) species richness could be inflated in species-poor areas. Based on comparison of 12,975 relevés from the Czech Republic, Chytrý (2001) has demonstrated that larger areas are likely to be sampled where vegetation is species-poor, in order to aid phytosociological classification. The effect of sampling area could introduce bias in the DFW data by inflating estimates from species-poor areas.

The DFW relevés range in area from 200 to 2,500 square meters. Mean area is approximately 937 m^2, with a median area of 750 m^2 (Table 3.3). Species-area curves were produced (Figure 3.2) in semi-logarithmic space using local linear regression to compare with the trends reported by Chytrý (2001); it appears that there is some bias toward larger relevé areas where tree species richness is lower. The largest relevés are at the northern limit of the sampling extent where lower species richness would be expected. Expanding the area of a relevé to include all species in a community type changes the scale relationship between the sample and factors affecting species richness, including environmental inputs such as solar energy and water.

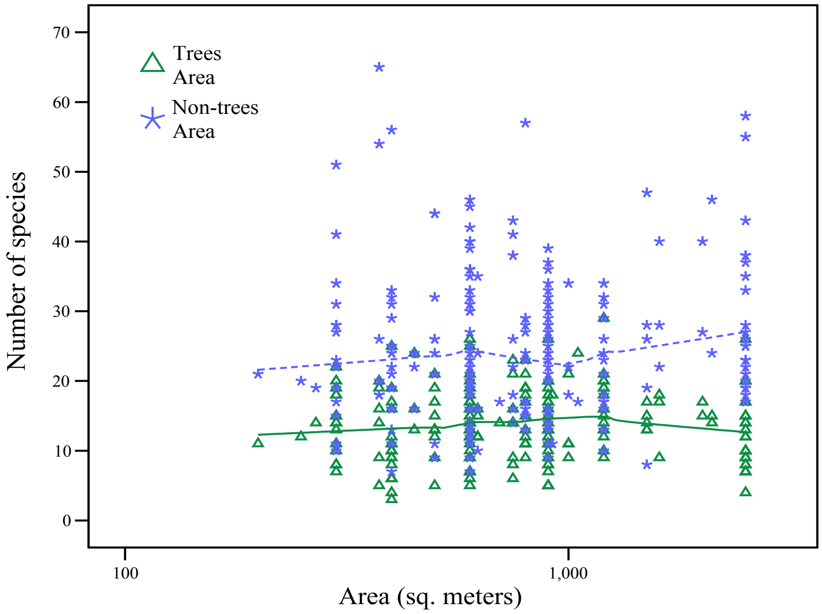


Figure 3.2: Species/area relationships for tree and non-tree growth forms.

3.14 Spatial Autocorrelation

Statistical analysis of the DFW data was complicated by several factors. Sampling sites were non-randomly selected and have a clustered dispersal. In addition, many sites represent more than one relevé. This was not a concern when the objective was a large-area descriptive survey (Box 1994). However, it violates the classical statistical assumption of independent data, and biases parameter estimates by unequally weighting different locations (Griffiths 1999). Information content is reduced when data are non-independent because nearby observations are likely to carry similar information, due to the phenomenon of spatial autocorrelation. (Cliff and Ord 1981). Spatial autocorrelation, the "first law of geography" (Tobler 1970), affects most spatially distributed data, including factors influencing ecology, especially with large-scale observational data (Lennon 2000).

Spatial autocorrelation refers to the similarity of two observations as a function of their distance in space, where observations are more alike (positive autocorrelation) or different (negative autocorrelation) based on their spatial proximity. When autocorrelation is present in both the dependent variable and independent variable, the basic regression assumption of independent errors is violated. The significance of the relationship between the dependent and independent variables is biased toward those that have the highest autocorrelation. Spatial autocorrelation is essentially a problem related to missing data, and is apparent in the error term due to information accounted for by the dependent variables.

It has been demonstrated that such bias is not present if the dependent variable does not exhibit autocorrelation, even with autocorrelated independent variables (Lennon 2000, Legendre *et. al* 2003). Moran's *I* coefficient is commonly used to measure spatial autocorrelation (Cliff and Ord 1981, Legendre and Legendre 1989). Values range from 1 to –1, representing positive to negative autocorrelation. A correlogram graphically displays spatial autocorrelation by plotting values of Moran's *I* coefficient between pairs of observations over different distance classes.

Significance testing of Moran's *I* coefficient requires that the assumption of second-order stationarity is satisfied, where the mean expected value and spatial covariance is equal over the sampling extent, with finite variance (Legendre and Legendre 1989). A less strict assumption of intrinsic stationarity requires only that the variance of the increments of a process is defined, and is dependent upon the distance between observations. Legendre and Legendre (1989) state that the intrinsic assumption is not met when measures of autocorrelation do not adequately describe the surface being investigated. As an example, they describe a region including both mountains and plains, and strongly advise against using correlograms in such cases unless the region is subdivided into like units.

The full extent of the DFW data is similar to the example described above. Therefore, the DFW data were subdivided into groups of relevés based upon physiographic provinces in order to examine for spatial autocorrelation (Table 3.4). This allowed areas with different topographic characteristics to be independently analyzed. A somewhat subjective process was used to create the relevé groups. In some instances physiographic provinces were combined, to create four relatively similar sized groups of relevés (Figure 3.3). Names and abbreviations are as follows: 1) New England and Saint Lawrence Valley (NE_SL), 2) Appalachian Highlands (AH), 3) Piedmont (P) and 4) Coastal Plain (CP).

Data used in this research are subject to four sources of spatial autocorrelation: (1) spurious, from processes related to the spatial pattern of the data observations; (2) interpolative, from climate data interpolation, (3) true autocorrelation, from interactions among observations, and (4) induced, from relationships with other autocorrelated variables (Fortin *et al.* 2002). The spatial characteristics for a subset of the DFW data were examined using Pattern Analysis, Spatial Statistics and Geographic Exegesis (PASSAGE) software developed by Michael Rosenburg (2001). Relationships between variables used in regression analysis were examined with respect to distance and direction. Information based on the geographic distances between sites was used to calculate correlation coefficients adjusted for spatial autocorrelation.

Table 3.4: Subset groups.

Group	Abbreviation	Observations
Appalachian Highlands	AH	58
Coastal Plain	CP	71
New England & Saint Lawrence Valley	NE_SL	45
Piedmont	P	51

Regression analyses were performed on these groups using the same independent variables (PET, PY and ELEV) to predict tree species richness as were used with the full set of DFW relevés. Only the model residuals from the Piedmont data subset displayed significant spatial autocorrelation and were examined further with the use of trend surface analysis (Legendre and Legendre 1998). Analysis based on the Piedmont subset data is used as a rough proxy for the full dataset. Although the Piedmont subset data do not represent the identical information, and introduce the modifiable areal units problem (MAUP) (Openshaw 1984) as a result of different data aggregation, they may present some insight to the effect of spatial autocorrelation on the DFW relevés as a whole, as an attempt to address a problem often overlooked in biogeographical studies (Lennon 2000).

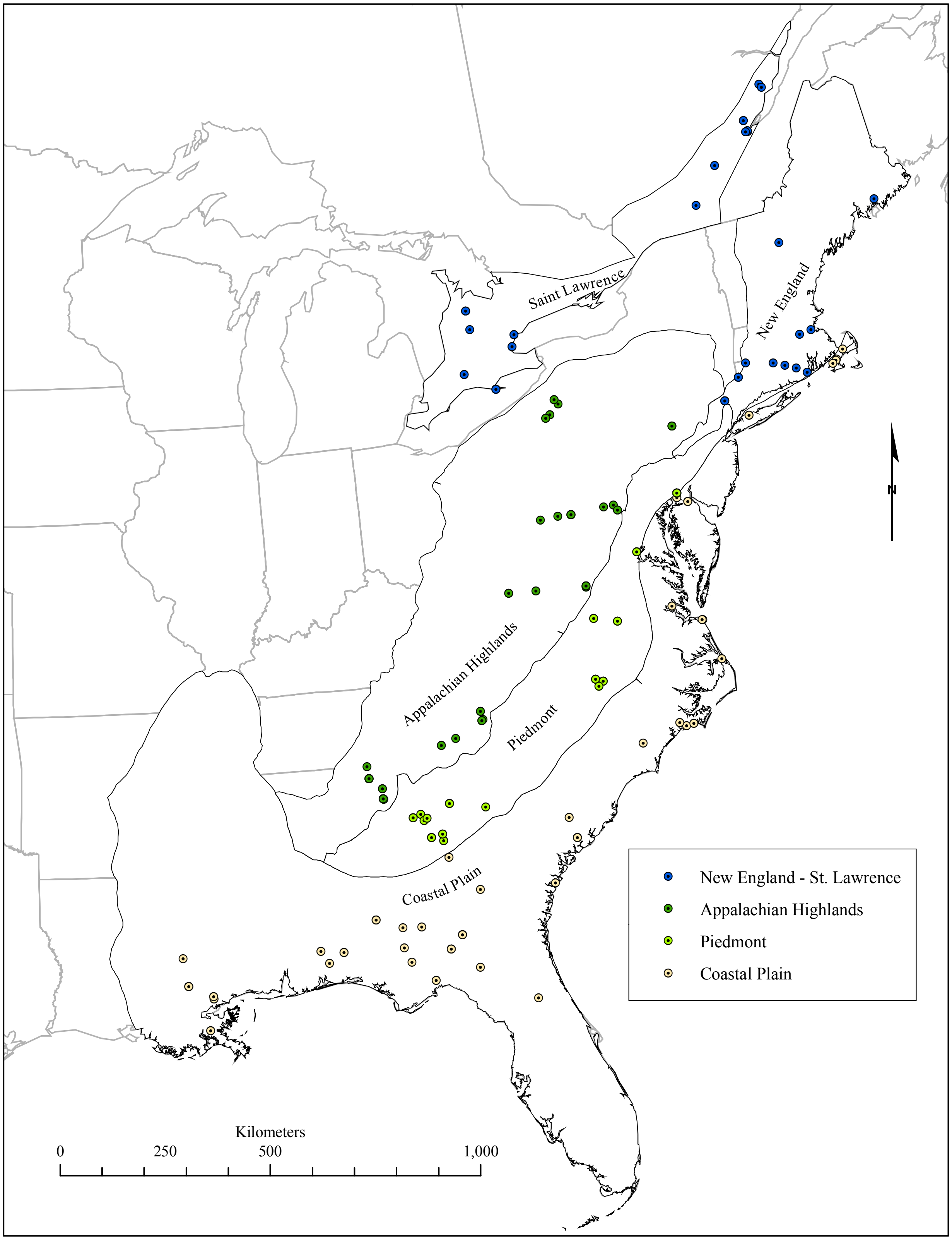


Figure 3.3: DFW relevé subsets based on physiographic regions.

3.15 Correlogram Analysis

Correlograms were created for the variables in the Piedmont subset using PASSAGE. Relevé locations were reprojected from their original geographic (latitude/longitude) coordinates to the Albers USA Contiguous Equidistant Conic projection, so that distance was measured in meters. A distance matrix was created using PASSAGE, based on the absolute Euclidean geographic distance between relevé locations. Lag classes were created for 16 distance intervals using "Sturge's rule" to determine the number of intervals (Legendre and Legendre 1998, p. 717). The distance matrix, distance classes and regression model variables were used to calculate Moran's *I* coefficient for model variables and residuals at different lag distances over the spatial extent of the relevés.

3.16 Mantel Analysis

The Mantel statistic is a measure of correlation between data distances. Two matrices are compared based on the similarity or dissimilarity among data observations in each matrix. The Mantel statistic *Z* was calculated for combinations of tree species, environmental and geographic data from the Piedmont data subset. The statistic *Z* is standardized to provide a correlation coefficient with a range of –1 to 1.

A variety of distance metrics can be used to measure data resemblance. Data distances for the variable NSPP_T were created using absolute Euclidean distance. Squared Euclidean distance was used for matrices based on more than one variable, such as geographic distances using XY coordinate based monomial terms, or environmental distances using the variables PET, PY and ELEV. The Morisita-Horn index was used to create a metric of tree species similarity (beta diversity) from raw table species data.

A partial Mantel test involves the comparison of two matrices while a third is held constant. With this method, spatial autocorrelation can be examined by controlling for broad-scale spatial structure with a matrix created from geographic data distances (Legendre and Legendre 1998). Mantel and partial Mantel tests were performed using PASSAGE.

Significance testing of the Mantel statistic is done using randomization methods since the distance matrix data violate independence assumptions (Urban 2003, Manly 1997). Because Mantel test significance is based on probabilities from observed data rather than a theoretical population distribution, generalizations should not be applied to populations beyond the observed sample (Manly 1997). The assumption of knowing an actual population distribution is somewhat disputable, especially with field-observed biological data (Manly 1997).

3.17 Bearing Analysis

A correlogram assumes that spatial autocorrelation occurs equally in all directions. However, spatial autocorrelation can vary along directional trends (anisotropy). Typical examples include temperature and moisture gradients, and sediment deposition patterns. Phenomena can be observed as a function of direction, through correlations between the geographic and angular distance correlations of data variables in a process called bearing analysis (Falsetti and Sokal 1993).

Bearing analysis was performed on variables in the Piedmont subset group using PASSAGE. Correlation strength was determined by a Mantel test comparing matrices of data distances with geographic distances weighted by angular distances (Rosenburg 2001). Bearing plots were created to graphically display the bearing directions where the distances between observations for data values and geographic locations have the greatest positive or negative correlations. This is a directional measure of spatial autocorrelation. The bearing plots were used along with the correlograms to examine data trends as expressed by observations the in the Piedmont relevé group. Bearing analysis indicates how the shape and pattern of the sampling extent influences the spatial autocorrelation structure and the significance of different environmental variables.

CHAPTER 4. RESULTS

4.1 Regression Analysis

Regression analyses were carried out to examine how species richness and diversity relate to climate and topographic data. Total species richness, total species evenness, and species richness by vertical layer were regressed against the climate variables and other data from the 225 DFW relevés. For the subset of 144 relevés where DEM data coverage was available, topographic indices were also included as independent variables in regression models.

The strongest and most significant bivariate regression relationship among vegetation layers for the 225 DFW relevés was between shrub-layer species richness and PET (Table A.1). In this model, annual PET accounted for approximately 27% of the variation ($r^2 = .272$) in shrub-layer species richness. With a 10-millimeter increase in PET, an increase of approximately one species in the shrub-layer would be expected. No significant relationship was found between total species evenness (E_{var}) and any of the other variables.

It has been theorized that the effects of climate variation relate to physiognomic diversity, as measured by plant growth forms, more closely than to individual species richness (Neilson *et. al* 1992). Trees have conservative growth strategies and respond to long-term conditions more closely than other vegetation (Loehle 1988). Their physical characteristics and longevity leave them more exposed to climate conditions. Tree species in the raw table were coded according to growth form, as assigned by Box (personal communication, November 2004) (Table 3.2). This also made it possible to examine how tree species richness relates to species richness by vegetation layer, and how species evenness compared between trees and non-trees (Figure 4.1).

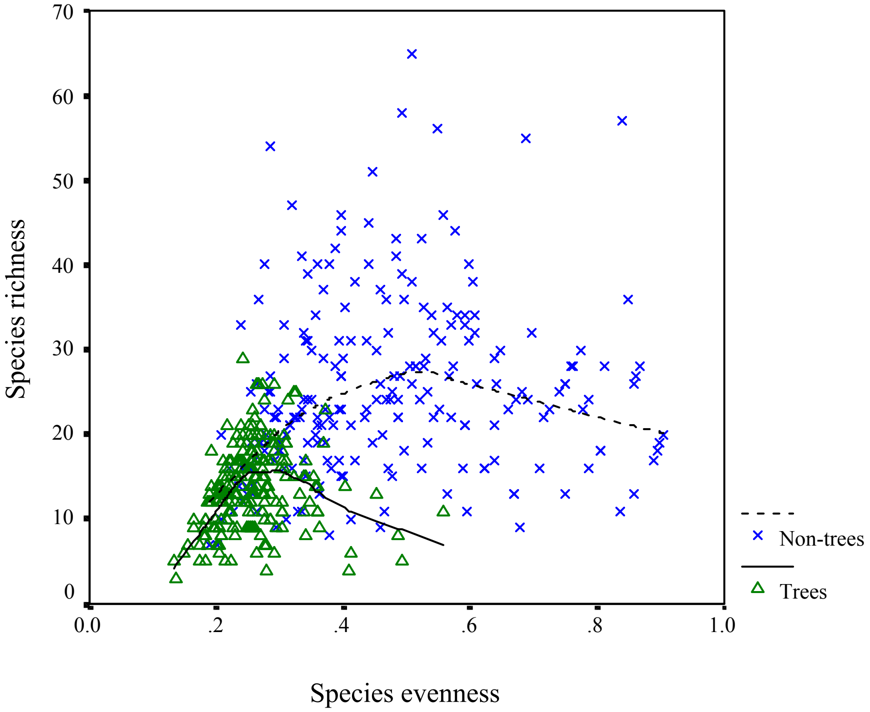


Figure 4.1: Species richness/evenness relationships for tree and non-tree growth forms. Trend lines fitted using locally weighted regression.

Tree species richness showed the strongest correlation to shrub-layer species richness (Table 4.1). This may relate to competition for light, with fewer tree species dominating the canopy layer. Note that this is comparing the correlation between the numbers of species in each relevé, classified by growth form as trees, with the total number of species (of any growth form) in each vertical layer. It should not be confused as a direct percentage of tree species per vertical layer; it is an indirect indication of tree species distribution by layer, inferred through correlation.

Table 4.1 indicates that it is probable that the shrub layer contains more tree species than the tree canopy or tree understory layers. This explains why the relationship of PET was stronger with shrub-layer species richness than with the tree canopy layer. Evenness of non-tree species moderately correlates with tree species richness and shrub-layer species richness (Table 4.1). Species richness in the tree understory and tree canopy layers correlates weakly with tree species evenness. Species richness of trees and non-trees appears to follow a slightly skewed positive unimodal distribution when plotted against species evenness, with each growth form peaking near the middle of its species evenness range (Figure 4.1).

Table 4.1: Correlations between species richness and species evenness (E_{var}) by vertical layer.

Pearson's r correlations	Species Richness		E_{var}	
	Trees	Non-trees	Trees	Non-trees
Tree species	1	0.2155	0.167	0.351
Sig. (1-tailed)	-	0.0006	0.0060	0.0000
Non-tree species	0.2155	1	0.064	0.151
Sig. (1-tailed)	0.0006	-	0.1703	0.0120
Tree Canopy (NSPP1)	0.4987	0.1452	0.214	0.147
Sig. (1-tailed)	0.0000	0.0147	0.0006	0.0138
Tree Understory (NSPP2)	0.4492	0.1674	0.261	0.100
Sig. (1-tailed)	0.0000	0.0060	0.0000	0.0675
Shrub (NSPP3)	0.7468	0.3166	0.081	0.232
Sig. (1-tailed)	0.0000	0.0000	0.1137	0.0002
Herb (NSPP4)	0.2339	0.9050	0.018	0.185
Sig. (1-tailed)	0.0002	0.0000	0.3969	0.0027

Tree species richness was used as the dependent variable in several bivariate regression models (Table 4.2, Models 2-10). A number of independent variables, each one typically varying with latitude, explained roughly 30% of the variation in tree species richness. Average temperature of the warmest month had the strongest coefficient among temperature variables. Longitude was included with these variables and has almost as much explanatory power, reflecting the northeast-to-southwest orientation of the data extent.

Table 4.2: Comparison of bivariate models (Models 2-10).

Tree species richness = *f* (x)

Model	Independent variable:	r^2	Coefficient
2	Latitude	0.292	-0.589
3	Longitude	0.208	-0.455
4	Average annual temp.	0.323	0.596
5	Average temp. of warmest month	0.329	0.914
6	Average temp. of coldest month	0.303	0.425
7	Biotemperature	0.313	0.669
8	PET	0.296	0.012
9	AET	0.286	0.012
10	NPP	0.292	0.010

P-value = 0.000, for all models.

4.2 Multiple Regression Models

Multiple regression analysis was used to investigate the combined contributions of different variables. A model was created based on O'Brien's (1993) model mentioned in Section 2.3, with the exception that mean annual PET was used in place of minimum monthly PET. She found that minimum monthly PET was more strongly correlated with species richness in Africa, and using it instead of annual PET avoided problems related to differences in summer versus winter rainfall. The quadratic term PET^2 was included to describe the relationship between NSPP_T and PET better.

The fit-line in a scatterplot of tree species richness versus PET shows this relationship (Figure 4.2); the optimum level of PET appears to be around 1000 millimeters of evapotranspiration. Tree species richness is limited at lower PET levels primarily with shorter growing seasons associated with northern latitudes. Once past the optimum range, tree species richness is presumed to be limited by a lack of moisture; however, PET does not exceed precipitation in any of the climate observations, suggesting some other mechanism as a limiting factor. This will be discussed further in the next chapter.

The following regression model: tree species richness = *f* (PY + (PET + PET^2)) (Table A.2, Model 11), explained more variation in tree species richness (r^2 = .389) than any of the models in Table 4.4 However, the partial coefficient for precipitation (PY) was weak and not statistically significant. Adding elevation (ELEV) to this model improved the explained variation and increased the strength and significance of PY (Table A.3, Model 12).

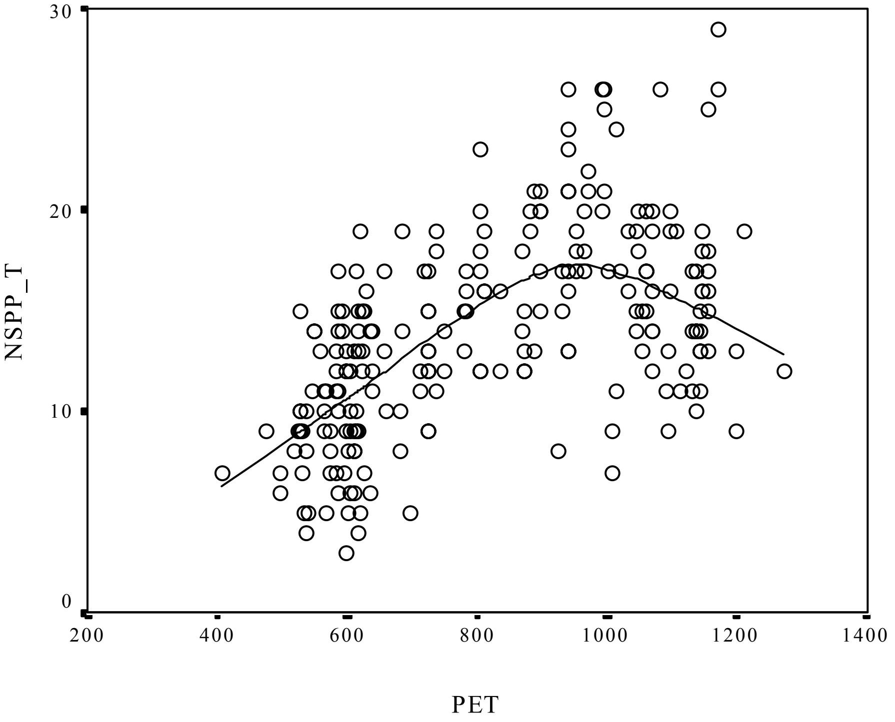


Figure 4.2: Tree species richness versus PET. Line fitted by locally weighted regression.

Although the significance of PY improved after ELEV was included, the partial slope coefficient remained weak and still had a negative sign. Experiments using different measures of temperature/energy as independent variables generally lowered the strength and significance of the partial slope coefficient for PY, although it was always negative. The negative sign is counterintuitive, given that a positive relationship between precipitation and species richness is supported by many researchers (Hawkins *et. al* 2003, Whittaker *et al.* 2001, O'Brien *et al.* 2000, O'Brien 1988, 1993 and 1998, Currie and Paquin 1987, etc.) and considering that the positive bivariate correlation between PY and independent variable NSPP_T ($r = .134$, P-value = .044).

There are many potential reasons for a slope coefficient sign to be the opposite of that expected, including high variance due to multicollinearity or low variation in the dependent variables, selection bias, outliers and interaction terms (Kennedy 2005). High multicollinearity was not indicated for PY using common diagnostic methods. The condition index number for the overall model is high due to the quadratic term of PET. Variance inflation factor (VIF) scores for PY and ELEV are below usual rule-of-thumb threshold levels (Table A.3). There is a strong correlation between the coefficients for PY and ELEV, but that did not change the sign of PY.

Some outlier observations were identified as having high influential leverage. The strongest outliers were three relevés located in the Appalachian Mountains of North Carolina, each with a west-facing aspect and elevations from 1500 to 1600 meters above sea level. Adiabatic cooling produced by orographic uplifting causes these relevés to have the highest precipitation values in the dataset, along with nearly the lowest values of PET. However, removing these relevés reduced the strength and significance of the precipitation coefficient, and hurt the overall model fit, while the sign of the precipitation coefficient remained negative.

The contribution of each variable was estimated by partial regressions to partition their explained variation (Legendre and Legendre 1998). The combined contributions of PET, PET^2, PY and ELEV account for 3 % of the variation explained, while PET, PET^2 and PY together account for 35 % of the variation explained and ELEV alone accounts for 2 %. The 60% residual variation that remains is unexplained model error.

Partial regression plots help to clarify how the interactions of PET, PET^2, PY and ELEV contribute to the variation explained (Figures 4.2 and 4.3). The nonlinear nature of NSPP_T versus PET shown in Figure 4.1 is seen again in Figure 4.2, as expressed by the linear and quadratic PET terms. The scatterplot on the left side of Figure 4.3 shows that once the covariance of PET, PET^2 and PY is accounted for, the variance attributed solely to PY has virtually no linear form. The outliers mentioned above can be seen on the right side of the plot, where they add negligible leverage influence, considering that they are relatively large outliers.

The scatterplot on the right side of Figure 4.3 shows the relationship between NSPP_T and PY after the covariance of ELEV has been accounted for in addition to the covariance of PET, PET^2 and PY. The relationship is still relatively weak, but the cloud of points is slightly more linear and indicates a more significant relationship between NSPP_T and PY, when ELEV is in the model. Before elevation was considered, the limiting effect of high PET was the strongest negative influence on NSPP_T. It appears that for these models precipitation may also be a limiting factor in tree species richness, especially after elevation is accounted for.

However, with the partial coefficients for PY and ELEV being so close to zero, interpretation of the precise contribution of each variable to the model is not generally useful information. The standardized partial coefficient in Model 12 (Figure 4.6) indicates that ELEV contributes slightly more to the explained variation in tree species richness than PY. The Akaike information criterion (AIC) supports the inclusion of ELEV (Table 4.3). An AIC decrease of 3 units is considered a sufficient improvement to support one set of variables over another (Burnham and Anderson 1998).

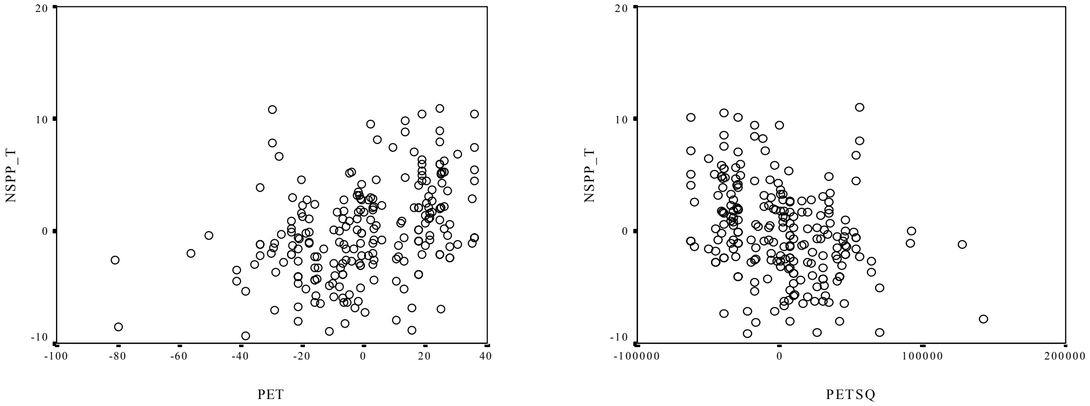


Figure 4.2: Partial regression plots.
Left: NSPP_T vs. PET, net of PET^2 & PY.

Right: NSPP_T vs. PET^2, net of PET & PY.
(Both from Model 11, Table A.2)

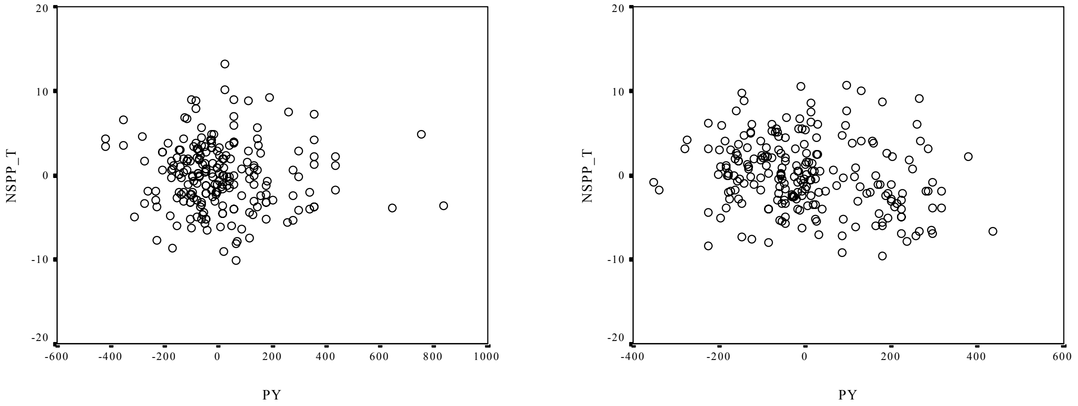


Figure 4.3: Partial regression plots.
Left: NSPP_T vs. PY, net of PET & PET^2.
(from Model 11, Table A.2)

Right: NSPP_T vs. PY, net of PET, PET^2 & ELEV.
(from Model 12, Table A.3)

Individual interpretation of the coefficients for PET and PET^2 is not possible, since any change in PET always means a change in PET^2. The expected change in tree species richness will vary depending on where an observation falls in the overall range of PET values. In general, interpretation of individual coefficients is difficult in many cases involving observation data; none of the variables in these models are likely to change independently of each other.

Table 4.3: Change in Akaike Information Criterion

Model	Independent variables	AIC
11	PY + PET + PET^2	1256.52
12	PY + PET + PET^2 + ELEV	1250.03
	Difference	6.49

4.3 Models Using the Topographic Index

Two multiple regression models were created using the set of 144 relevés that included topographic index values. The variables PET, PET^2, PY and ELEV explained 36% of the statistical variation in tree species richness ($r^2 = .361$) (Table A.4, Model 13). Adding the topographic diversity index (EVARDIV) increased the coefficient of determination to $r^2 = .417$ (Table A.5, Model 14). Interpretation of the partial slope coefficient for EVARDIV is difficult. Higher values of EVARDIV signify sites with greater variation in slope. An increase of approximately five tree species is expected for each one unit increase of slope aspect diversity, or more generally, that tree species increase as slope aspect diversity increases.

Comparison between models using the complete DFW dataset and models using the dataset with the topographic index are not equal, due to the inclusion and exclusion of different data among models. This limits the generalization of results only to models using the same observations. Model 14 can be compared to Model 13, but not to Model 11, as differences among models using different observations may simply reflect less variation among variables to explain (Burt and Barber 1996).

The addition of EVARDIV increased the explained variation of tree species richness by roughly eight-percent. The AIC value for Model 14 is significantly less than for Model 13 (Table 4.4). Adding the topographic heterogeneity index appears to have produced a model that explains a greater amount of variation in tree species richness. However, the EVARDIV information comes at the expense of less information from the other variables due to the limited DEM data.

Table 4.4: Change in Akaike Information Criterion

Model	Independent variables	AIC
13	PY + PET + PET^2 + ELEV	813.14
14	PY + PET + PET^2 + ELEV + EVARDIV	801.88
	Difference	11.26

4.4 Correlogram Analysis

Correlograms and other methods for examining spatial autocorrelation must be used under the assumption of intrinsic stationarity in order to have significant meanings (Legendre and Legendre 1998). For these reasons, spatial autocorrelation analyses were only performed on a subset of the DFW relevés. Multiple regression analysis was performed on the subset groups listed in Table 3.4 and shown in Figure 3.3. The quadratic energy function (PET^2) was not used in these models. Instead, a linear function was used to keep the models simpler, since the smaller sampling extents represented a limited range of PET values. The results show differences in each model due to different relevés being included in each group (Table A.6, Models 15 – 18).

Correlograms were created using the dependent and independent variables, and residuals from the models in Table A.6. The correlograms for the model residuals from the relevé groups

AH, CP and NE_SL (Models 15-17) do not indicate significant spatial autocorrelation. The PY coefficient remains negative and the coefficients for PY and ELEV are still weak and have low significance. The r^2 is still a valid estimate of the NSPP_T variation explained by PET, PY and ELEV. However, the individual contributions of each variable cannot be confidently estimated.

In the group AH (Model 15, top of Figure 4.4), adding elevation appears to have accounted for the autocorrelation structure in the independent variable NSPP_T and the dependent variable PET. Correlograms for this group show patchy distributions influenced by the topography of the Appalachians. Spatial autocorrelation ($I = .17$) is present in the first lag distance for NSPP_T, but overall it is insignificant. Low autocorrelation for NSPP_T, and the inclusion of ELEV appear to have made autocorrelation insignificant in the residual correlogram, although all the dependent variables exhibit significant autocorrelation at some lags.

The correlogram for the group CP (Model 16, bottom of Figure 4.4) indicates very little autocorrelation for the dependent variable NSPP_T. As a result the residuals for this model do not exhibit significant autocorrelation, although it is present in each of the dependent variables. The CP group is the largest subset of relevés, and covers the greatest longitudinal extent as well as a considerable latitudinal extent. This may influence the similarity between observed pairs.

The correlogram for the group NE_SL (Model 17, top of Figure 4.5) has a combination of low autocorrelation in NSPP_T and an effect of spatial structuring on PET similar to Model 15. A sharp reversal in trends for PY and ELEV occurring between the fifth and sixth lags coincides with the Saint Lawrence River valley. At this distance the sites compared include those within the valley, the Ontario peninsula and east of the Saint Lawrence River valley.

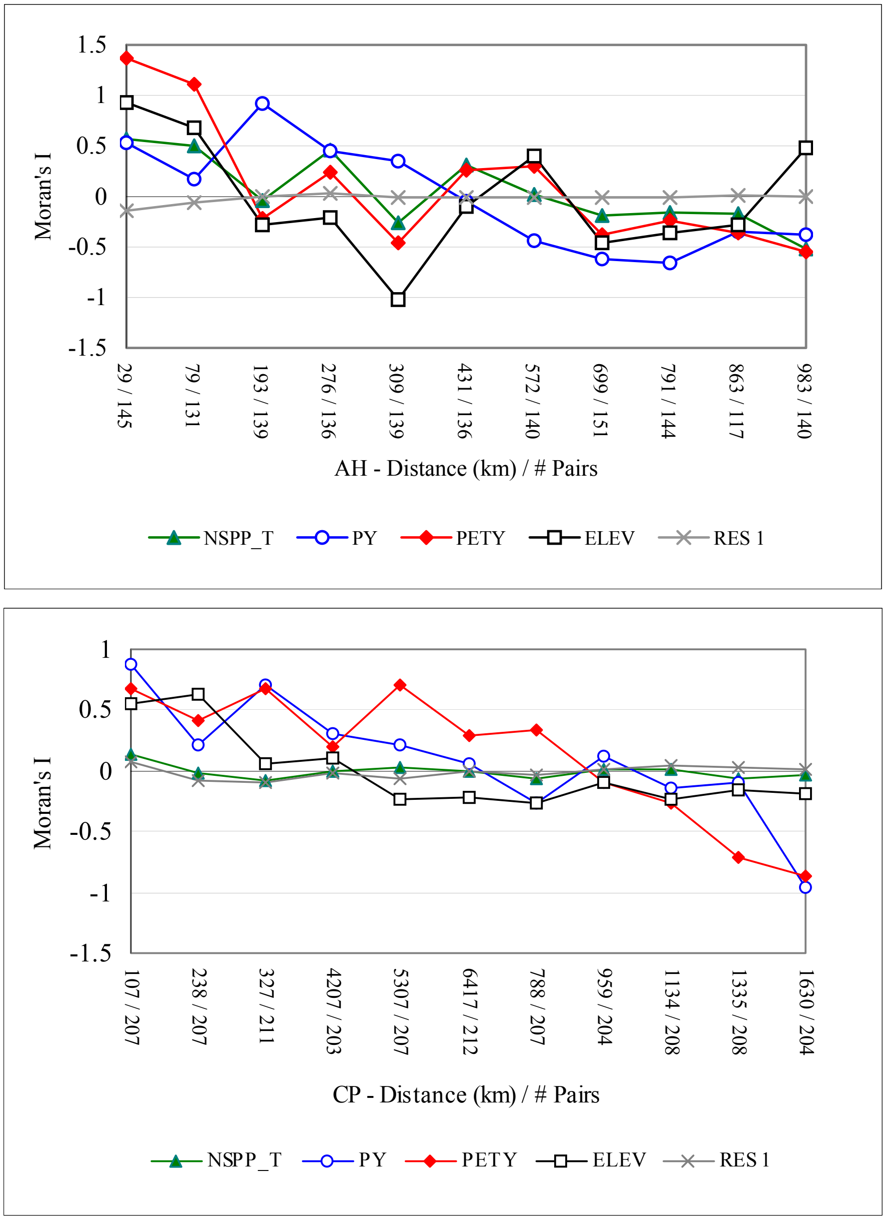


Figure 4.4: Correlograms of the variables and residuals in the multiple regression model: NSPP_T = *f* (PET + PY + ELEV).

For the relevé groups AH (Model 15, top) and CP (Model 16, bottom).

Moran's *I* measures positive/negative autocorrelation based on similarity/dissimilarity between sample pairs as a function of distance. Values are displayed over 12 distance lags (12th lag not shown).

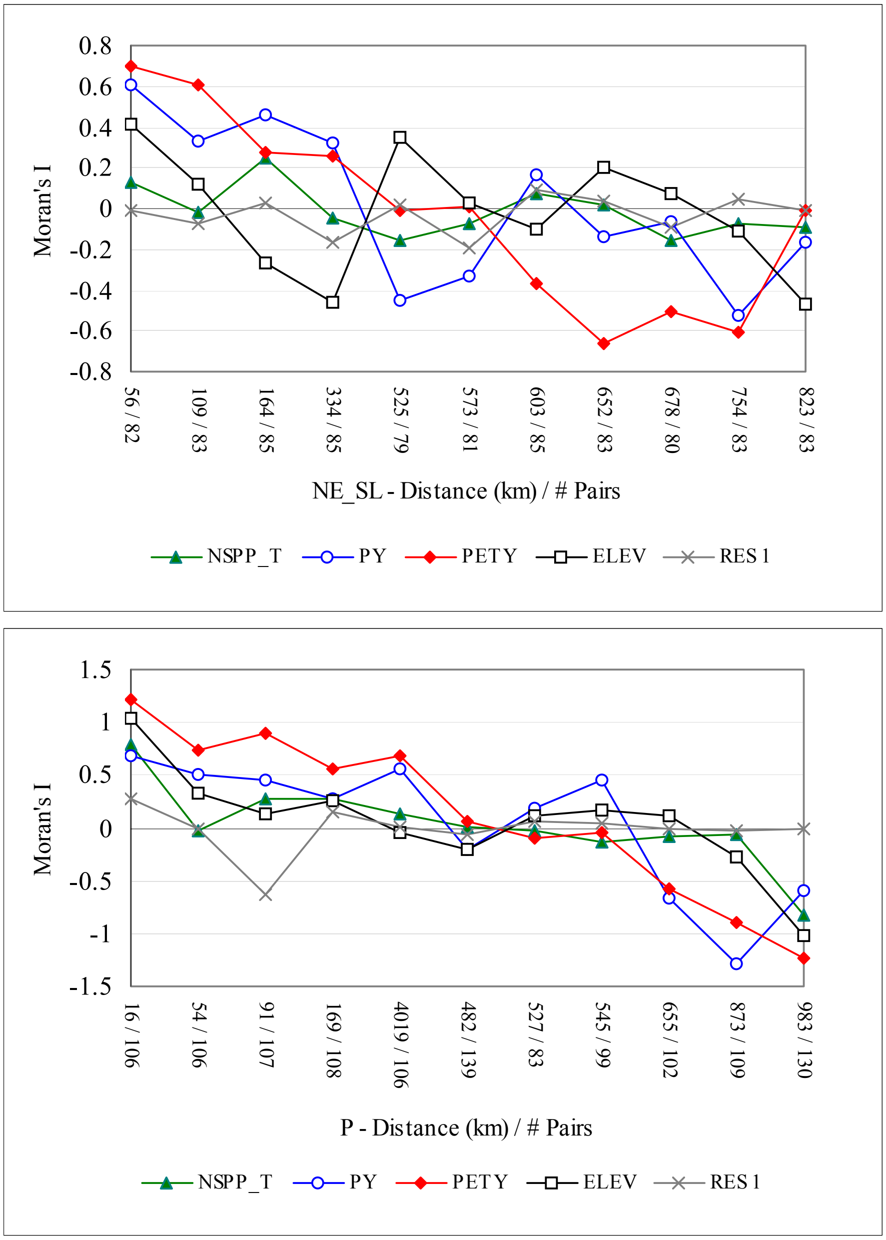


Figure 4.5: Correlograms of the variables and residuals in the multiple regression model: NSPP_T = *f* (PET + PY + ELEV).

For the relevé groups NE_SL (Model 17, top) and P (Model 18, bottom).

Moran's *I* measures positive/negative autocorrelation based on similarity/dissimilarity between sample pairs as a function of distance. Values are displayed over 12 distance lags (12th lag not shown).

4.5 Spatial Autocorrelation and the Piedmont Group

The correlogram for the Piedmont group variables (Figure 4.5, bottom) indicates significant spatial autocorrelation present in the model residuals. All of the variables have high values of Moran's *I* at the first lag distance. Values of *I* for PET and PY follow a decreasing trend up to the fifth lag distance, where climate observations between pairs begin to increase in similarity as a function of lag distance. This is due to greater variation in elevation among pairs observed at the fifth lag distance. The correlogram for PY indicates a strong positive increase at the eighth lag, compared to the other correlograms. Significant negative autocorrelation is present in at the third lag distance of the residual correlogram (Figure 4.6). The reasons for this, and the positive autocorrelation for PY at the eighth lag are not clear.

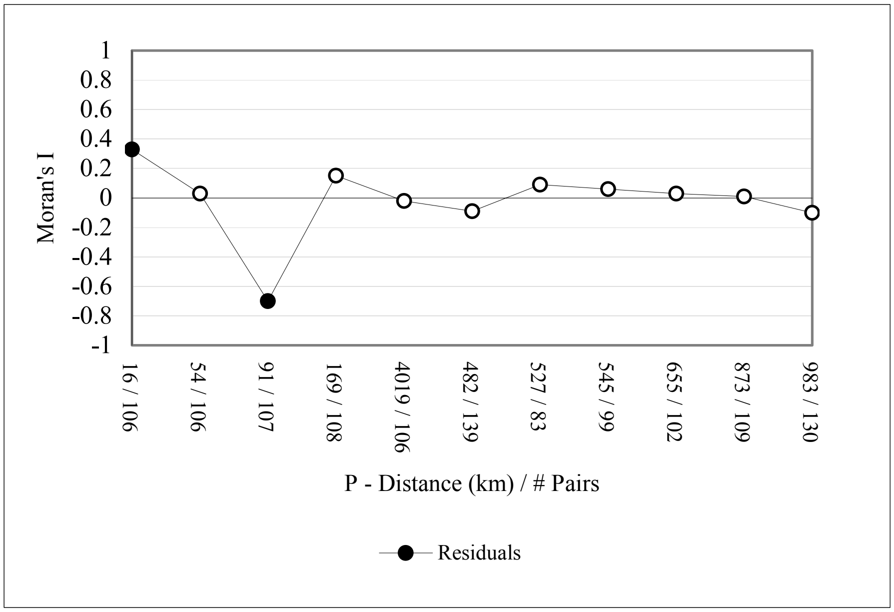


Figure 4.6: Correlogram of residuals from Model 18. Filled symbols represent significant spatial autocorrelation.

The affect of spatial autocorrelation on the significance of correlations between variables was tested using a modified t-test function in PASSAGE (Clifford, Richardson, and Hémon 1989). After the degrees of freedom were adjusted to account for autocorrelation, the correlation coefficients for PET, PY and ELEV lose significance, with PET being affected the most (Table 4.12). The information content of these variables is reduced due to the lack of independence among observations. The higher effective sample size for PY may indicate that variation in PY occurs over a finer scale grain than PET and ELEV.

Table 4.5: CRH corrected correlations.

NSPP_T vs.	PET	PY	ELEV
Pearson's r correlation	0.5828	0.3745	0.4589
Sig. (2-tailed)	0.0000	0.0068	0.0007
CRH corrected sig.	0.3932	0.3855	0.3574
Effective sample size	4.2	7.5	6

Actual sample size = 51

4.6 Trend Surface Analysis

Legendre and Legendre (1998) describe methods to account for broad-scale autocorrelation structure in regression model error terms through the use of polynomial trend surfaces. A polynomial expression created from a set of XY spatial coordinates can account for spatial structure that is often a source of autocorrelation over broad scales. Borcard and Legendre (2004) developed the program SpaceMaker2, to construct spatial matrices based on this idea.

A trend surface was created using SpaceMaker2 to obtain a third-order polynomial function of the coordinates for the Piedmont relevé group. The program centers the coordinates and computes a set of monomial terms. Following Legendre and Legendre (1998), all monomial terms were regressed against tree species richness, and non-significant terms were removed to provide the most parsimonious model with the highest coefficient of variation. The most significant terms were latitude*longitude (MON 4), longitude^2 (MON 5), and longitude^3 (MON 9) (Table A.7, Model 19). The variable ELEV became insignificant after the addition of the monomials, and was dropped from the model. The resulting model explains approximately 10% more of the variation in tree species richness after the monomial terms were included.

No significant autocorrelation is present in the correlogram for the model residuals (Figure 4.7), indicating that the monomial terms account for some spatially related element missing from the model (Diniz-Filho *et. al* 2003). Apparently, the spatial structure, most likely related to the mountains, is better represented by the monomials than by a single elevation value. Strong multicollinearity is present after adding the monomial terms (Table A.7). PET and the monomials have the highest VIF levels. This is expected considering that the monomials were created from the site coordinates, and the latitudinal temperature trend influencing PET.

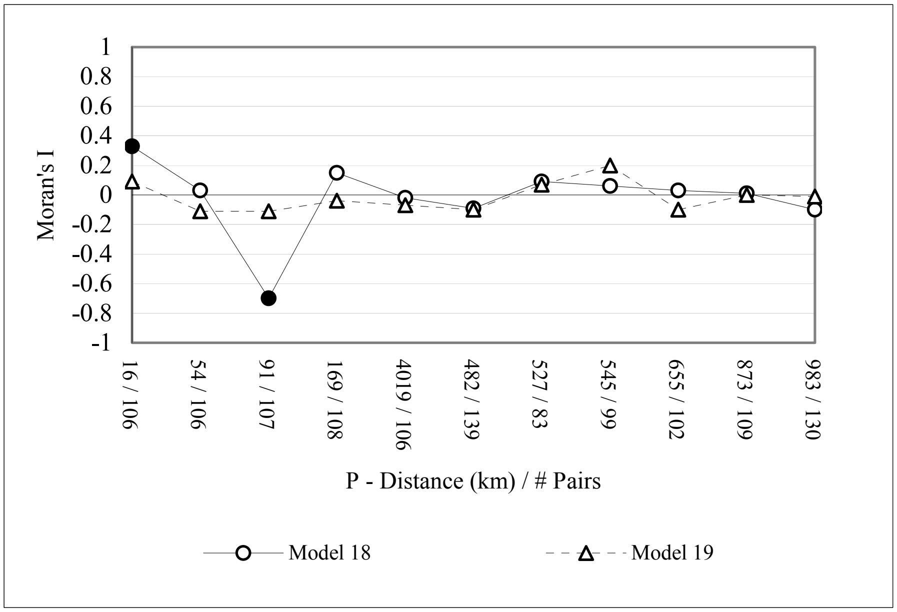


Figure 4.7: Comparison of correlograms for residuals from Model 18 & 19. Filled symbols indicate significant autocorrelation.

The effect of the monomial terms on PY is apparent when comparing partial regression plots (Figure 4.8). The right plot in Figure 4.8 shows that PY has a stronger linear form once the covariance related to the spatial structure represented by MON 4, MON 5 and MON9 is accounted for along with the covariance of PET and ELEV. While the strength and significance of PY have improved, the sign of the coefficient has remained negative. The relationship between tree species richness and precipitation may express more than random patterns within the data. This model-specific result appears to have come from the choice of sites included in the Piedmont group. The smaller dataset increases the relative influence of ELEV on PY.

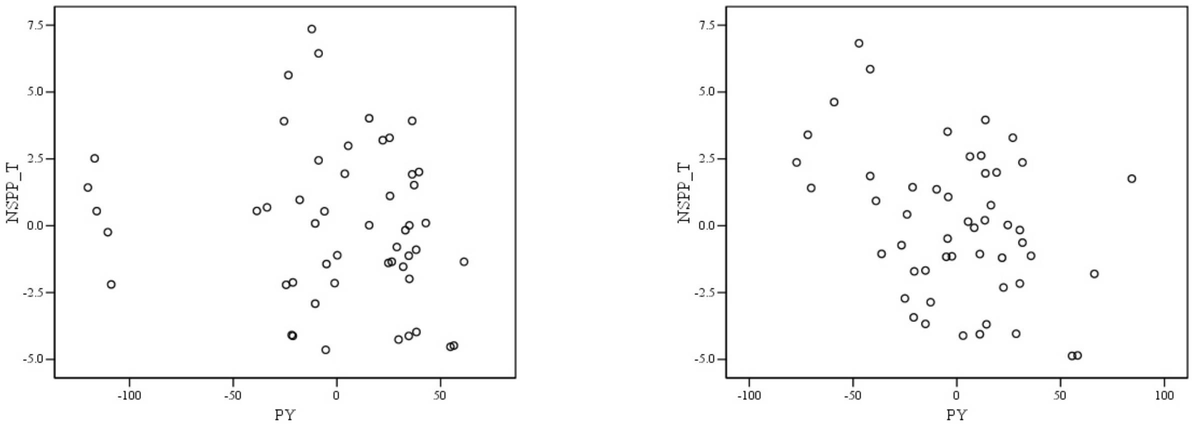


Figure 4.8: Partial regression plots.
Left: NSPP_T vs. PY, net of PET & ELEV.
(from Model 18, Table A.6)

Right: NSPP_T vs. PY, net of PET, ELEV, MON 4, MON 5 & MON 9
(from Model 19, Table A.7)

4.7 Mantel Tests with Tree Species Richness

Mantel tests and partial Mantel tests support the multiple-regression results from Model 19. The variables PET, PY and ELEV were used in explanatory matrices, individually (Table A.8, top) and in combinations (Table A.9, top). The resulting correlations between tree species richness and environmental matrices are similar to those from regression analysis, although correlations between distance matrices and correlations between raw data observations do not have the same meanings (Legendre and Legendre 1998). The Mantel coefficient indicates how differences or similarities in environmental conditions among sites correlate with differences or similarities in tree-species richness patterns among sites.

A Mantel test between NSPP_T and PY resulted in a weak negative correlation, similar to the partial slope coefficient for PY from the regression models. The correlation is not significant however after 4999 randomizations (equivalent to a 95% confidence interval). A partial Mantel test in which a distance matrix was held constant produced a similar result, with even less significance. The bottom sections of Tables A.8 and A.9 include a spatial distance matrix created from the monomial terms. This is similar to using trend surfaces in partial regression (Legendre and Legendre 1998).

The distance matrices for NSPP_T and PET had the strongest and most significant correlation in the Mantel and partial Mantel tests. The relationship between NSPP_T and PET appears stronger after accounting for the monomial matrix. The addition of PY and ELEV to the environmental matrix increased the coefficient, demonstrating an interaction contributes to the overall correlation. The inclusion of the monomial terms reduces the strength and significance in each test, indicating reduced spatial autocorrelation. Once the spatial variation is accounted for, less variation is explained by the environmental matrix variables, especially PY and ELEV.

4.8 Mantel Tests with Tree Species Beta-diversity

The matrix design of the Mantel test allows the species composition of each relevé to be considered. A matrix was created using the Morisita-Horn index as a measure of similarity/dissimilarity based on species common or uncommon to each relevé. This distance metric represents Whittaker's (1960) concept of species replacement or turnover, called β diversity (Magurran 2004). Morisita-Horn index values were calculated using EstimateS software version 7.5 (Colwell 2005) and tree species data from the raw table. Because this software requires integer data, the percent-cover values were converted to presence/absence data before calculating the index. This represents a loss of information, as species are not weighted by abundance and therefore rare species may be over-represented. With further investigation a method might be devised to avoid this obstacle and use the species cover-abundance data.

Mantel tests were performed using matrices for tree species β diversity (β_T) and the environmental variables PET, PY and ELEV (Table A.10). In the previous Mantel tests and regression models, PET had the strongest relationship with the tree species richness, while precipitation did not display a strong or significant relationship. Essentially the same relationship is shown by the Mantel tests using β diversity. In the Mantel tests for either measure of tree species distance, PY and ELEV are not significant as individual variables. For β diversity, the coefficient indicates how differences or similarities in environmental conditions among sites correlate with differences or similarities in tree species composition among relevés. When the monomial matrix (MON 4, 5 and 9) is held constant in partial Mantel tests of PY, ELEV and PET versus ß_T (Table A.11), the correlation coefficient strength is reduced in each test, and only PET versus β_T is significant after 5000 randomized iterations.

Tables A.10 and A.11 suggest that PET correlates more with which species are found at a given site, rather than to how many different species are found there. This supports the concept that broad-scale climate variation exhibits a stronger relationship to physiognomic growth form than to species richness (Neilson *et. al* 1992). However, the relationship displayed by PY may indicate that it varies over a scale grain somewhere between PET and the tree species data.

The same relationship may exist for the entire DFW dataset (Table A.12, top). The ß_T distance matrix has a stronger and more significant correlation with the environmental distance matrix composed of PET, PY and ELEV, for the group of 225 relevés. The actual statistical significance may be lower due to the effects of spatial autocorrelation, but this cannot be examined without violating the assumption of intrinsic stationarity.

A beta-diversity distance matrix was created to compare with the topographic index data. Matching ID codes were chosen between sites from the raw table file corresponding to the 144 relevés with topographic index data. This created a group of only 65 relevés; raw-table data for the remaining relevés with topographic index data exist but would have required extensive time to reorganize. The Mantel test results for these 65 relevés are similar to the other Mantel tests (Table A.12, bottom), although the magnitude is not as great as for the test using all 225 DFW relevés. This is a MAUP effect caused by using a different aggregation of relevés, which weakened the strength and significance of the correlation.

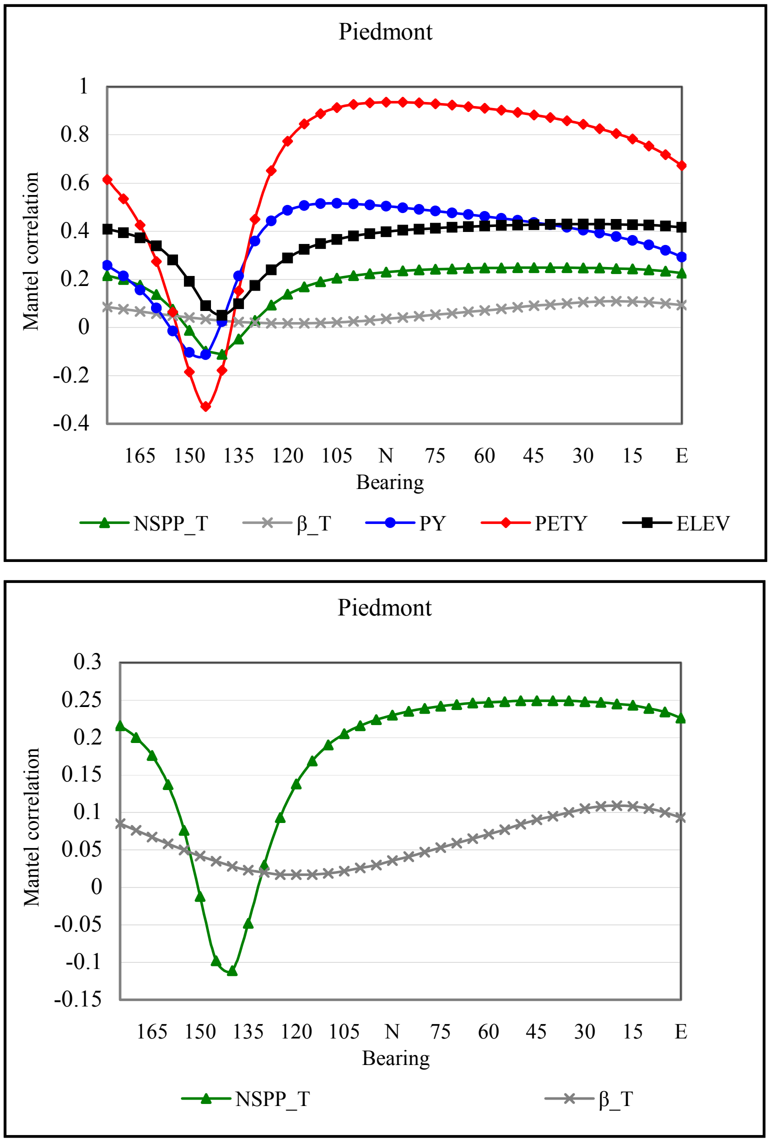
4.9 Bearing Analysis Results

The correlograms in subsection 4.4 assume that autocorrelation is the same in all directions (isotropic). Ecological data are typically anisotropic, with directional trends in the similarity or dissimilarity of neighboring observations. Bearing plots were created using PASSAGE to examine how autocorrelation varies directionally in the Piedmont relevé group.

Figure 4.9 shows that correlations between data and geographic distances are relatively constant for the variables NSPP_T, β_T, PY, PET and ELEV, except for a negative correlation on a north-northwest bearing. Sites in close proximity have less similar values than sites further apart along a north-northwest direction that crosses the Appalachians, where topography affects climate and substrate characteristics, and ultimately influences tree species. Beta-diversity is relatively constant in all directions. It is also more resistant to spatial autocorrelation, in part because matrix-based values are often less autocorrelated than raw data values (Fortin 2002).

The lines overlaid on the map in Figure 4.10 display information from the bearing plots in Figure 4.9. They show the bearings of the greatest positive and negative distance correlations in reference to the spatial pattern of the relevé locations. The lengths of the lines are scaled relative to bearing correlation values, and angles represent the direction of each trend. PET is clearly the most spatially autocorrelated, followed by PY and ELEV, and then NSPP_T and β_T.

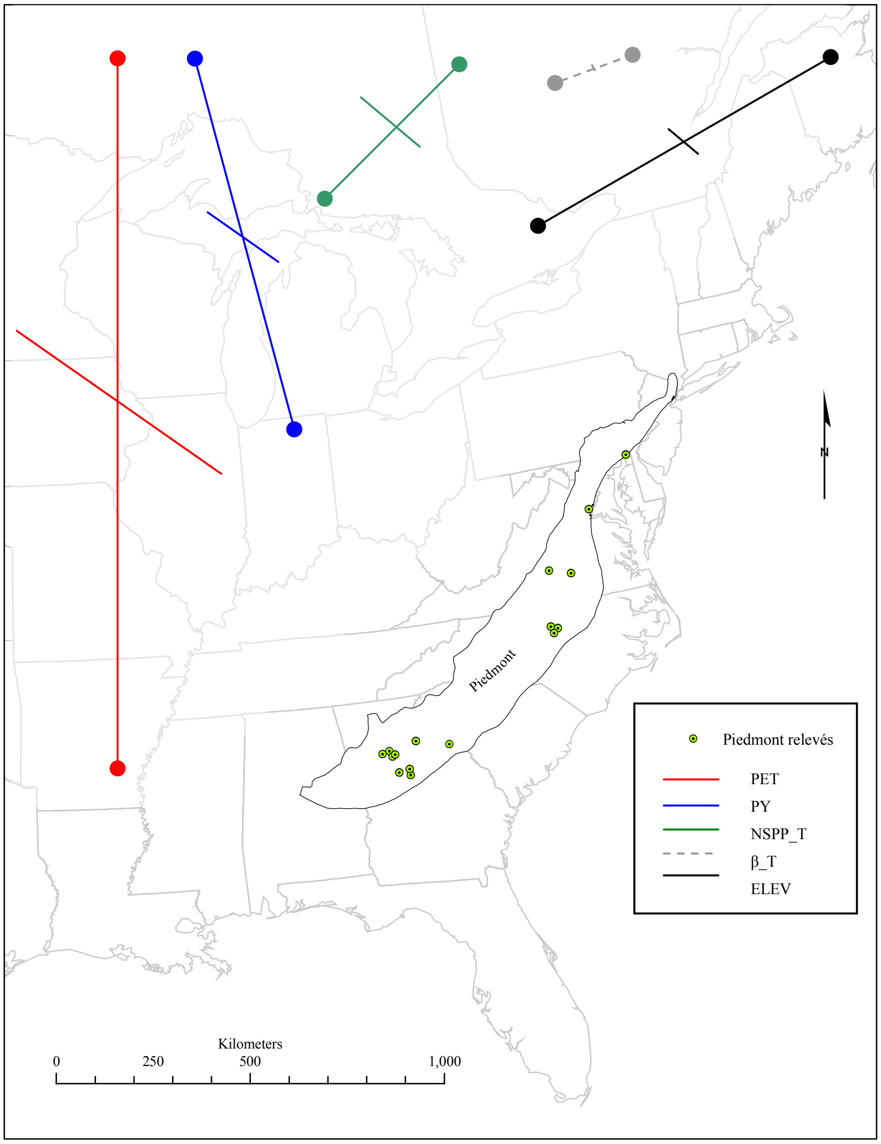
Note that the lines on the map do not indicate a gradient of low to high values; they are measures of positive and negative spatial autocorrelation. In each direction, values are more similar (or dissimilar) among sites in close proximity, and less similar (or dissimilar) among sites that are farther apart. The bearing of positive correlation is different for each variable, while all negative correlations are along a bearing that crosses the mountains.



Top: Tree species richness, β diversity and environmental variables.
Bottom: Tree species richness and β diversity, using different Y-axis scale.

Positive values indicate bearing of highest correlation between data distances and geographic distances; negative values indicate the bearing of least correlation. The effect of the Appalachian Mts is shown by negative values coinciding with a NNW bearing. The strong gradient in PET appears to be a combination of latitudinal extent and elevation. Precipitation is less correlated with bearing direction than PET.

Figure 4.9: Bearing Plots for relevés in the Piedmont physiographic group.



Trends for the following variables, from left to right: PET, PY, NSPP_T, β_T and ELEV. Line lengths are scaled relative to correlation strength between distance and observed value.

●—● = Bearing of greatest positive correlation.

—— = Bearing of greatest negative correlation.

Figure 4.10: Data trends for the Piedmont group, based on bearing plots in Figure 4.9.

CHAPTER 5. DISCUSSION

The results from the regression models and the Mantel tests support the theory of water and energy as factors correlating with the distribution of tree species richness. None of the regression models, however, explains more than 30% of the statistical variation in tree species richness. Clearly, there are other factors missing from the models. Energy as measured by PET was a significant, although relatively weak, explanatory variable. Precipitation and elevation had less strength and significance as explanatory variables.

The relationship between tree species richness and the variables PET and PY is complicated by two conflicting patterns. A parabolic pattern is related to the latitude of each relevé. Temperature changes as a result of the angle of solar insolation, which also influences precipitation because of the higher water holding capacity of warmer air. Tree species richness indicates an optimized relationship with PET, while the relationship with PY ranges from being ambiguous to negative, depending on how the data are aggregated.

Within the trend of NSPP_T versus PET and PY is scatter related to fine-scale site-specific variation, as well as a negative trend related to elevation. Temperature and precipitation are affected by elevation in opposite directions. If the data were sampled over a broader and more uniform extent, this trend might simply represent the “width” of the scatterplot points in a linear or parabolic relationship.

5.1 Interpreting the Negative Precipitation Coefficient

In general, the consistently negative relationship between tree species richness and precipitation appears to be based on spurious random variation, although there may be an explanation in the context of the Piedmont group of relevés. The PY coefficient would be interpreted to mean that a decrease of approximately two species would be expected with each 100-millimeter increase in PY, holding PET and the other variables constant. While this goes against most broad-scale species richness theories, it may be an accurate reflection of the interaction among the model variables.

A 100-millimeter increase in PY, holding PET constant, would not necessarily be expected to increase species richness. A latitudinal temperature gradient is present and generally PET and PY increase from north to south. However, sites at higher elevations experience a loss of temperature and a gain of precipitation, as elevation increases. Observed tree species richness is lower at higher-elevation sites, while observed precipitation is relatively high. Elevation is significant to the variation explained because it accounts for instances where the variables PET and PY act independently. The PY partial slope coefficient was stronger for the Piedmont group of relevés due to the MAUP effect, as sites with higher elevation were emphasized, in contrast to the full DFW dataset.

Sites with poor drainage may have fewer species if PY were increased while PET was held constant. Anaerobic conditions limit species richness, resulting in fewer tree species in locations with excess PY, such as swamps or flood-prone locations. Although relevés clearly fitting this description were removed when the DFW file was created, inspection of the original data indicates that some relevés with overly wet conditions remained in the dataset. The PY partial slope coefficient remained negative in models created without these relevés, although the strength and significance of the coefficient was reduced.

The negative PY coefficient can also be explained from the perspective of PET. An increase of approximately four tree species is expected if PET were increased by 100 millimeters while PY was held constant. The models based on 225 DFW relevés show similar relationships between NSPP_T versus PET. This relationship may hold true under certain conditions, although higher temperatures and lower precipitation would generally produce dry conditions and fewer tree species.

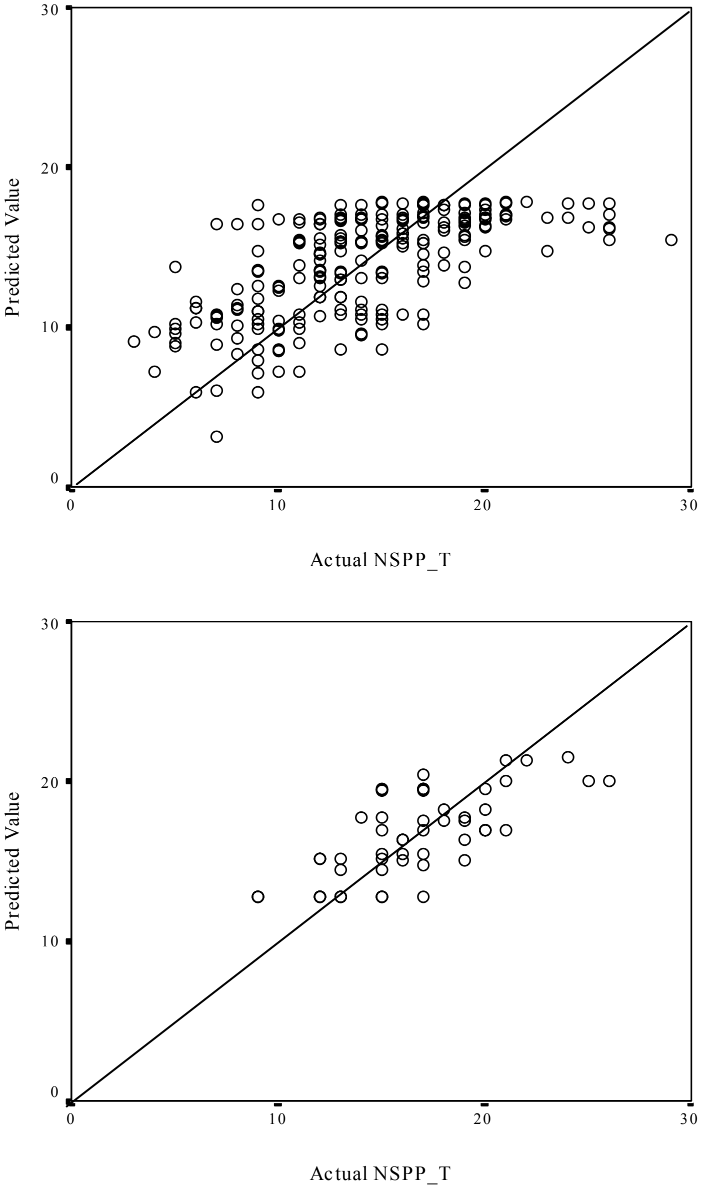
In sites where soils have greater water holding capacity, the tree species richness may not be affected as much by an increase in PET, with a constant amount of PY. Where soils do not have good water holding capacity, such as in sandy coastal-plain soils, tree species richness may suffer from drought stress by increased PET, if this is not accompanied by increased PY. When tree species richness is generalized over broad scales, situations such as these may represent mild scatter in an otherwise strong trend. When tree species richness is examined at finer scales, with fewer observations, small variations have a stronger affect on what relationships are expressed through the data.

The choice of which sites are represented in the dataset could be a factor influencing the relationship between tree species richness and precipitation. The Coastal Plain is represented by the most DFW relevés when classified by physiographic province. This combined with relevés that are “too wet” may be a reason for the negative relationship between NSPP_T and PY.

However, the negative precipitation coefficient was consistent in various data aggregations, including experiments involving the removal of coastal-plain sites as well as random group assignment. This suggests a generally weak relationship, manipulated by different data aggregations. The most likely explanation for the negative slope coefficient is that the precipitation data do not exhibit a strong linear relationship to tree species richness. This was shown by the standardized coefficients in all of the regression models, and in the Mantel tests.

The main reason for this weak relationship could be related to scale. O'Brien (1998) suggests that climate must be measured at distances over 100 km apart, or between areas of 10,000 km^2, to exhibit sufficient variation in relation to species richness, although it is not clear how she arrives at this conclusion. Whittaker and Field (2000) note that for some variables, relationships to species richness can be observed at several scales. PET may significantly contribute to the explained variation of tree species richness at several scales, while PY does not. It is also possible that PY varies at a finer scale than PET, and the vegetation data do not adequately match this scale for a relationship to be expressed.

O'Brien (1998) and Whittaker and Field (2000) point out that the Interim General Model tends to over-predict species richness in areas with high precipitation and low PET. This appears to be the case when examining predicted versus actual tree species richness scatterplots (Figure 5.1), as well as the mapped residuals (Figure 5.2). O'Brien (1998) gives two reasons for this, specifically when her model is applied to data from North America; the model may be sensitive to differences in re-speciation rates between Africa and North America following the last glaciation period, or perhaps more interesting, there may be a limit in the relationship between water and species richness rather than a maximizing function.



Top: 225 releves (Model 12).

Bottom: 51 Piedmont releves (Model 19).

Overpredicted observations left of the diagonal are mostly sites with wet conditions. Observations in the upper right represent underprediction in species-rich sites, most likely due to karst influenced substrates.

Figure 5.1: Regression model fit; actual vs. predicted values of tree species richness.

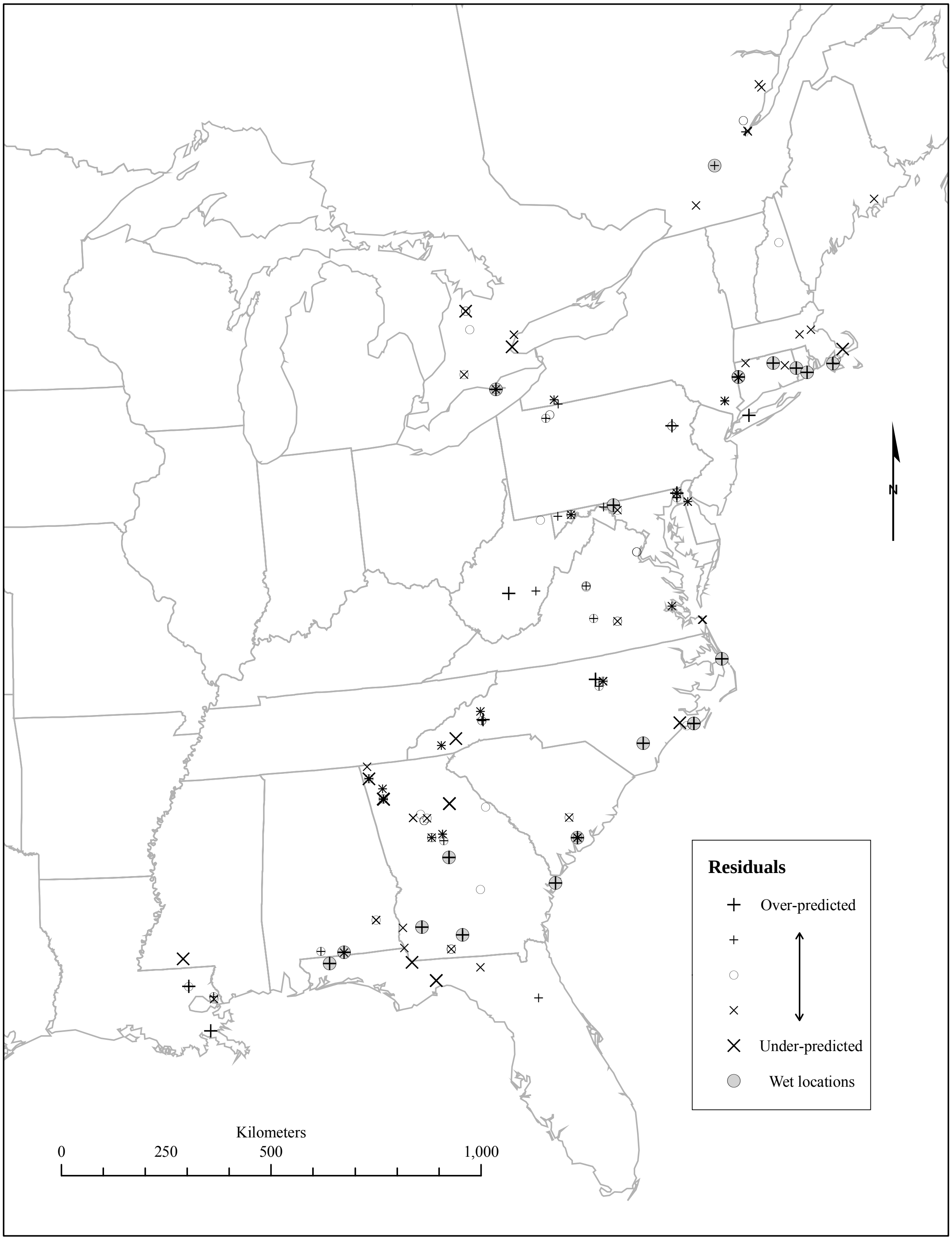


Figure 5.2: Residuals from regression model. Model over-prediction coincides with wet sites.

Loehle (1998) contends that northern and southern range-limit differences in tree species are the result of tradeoffs between cold hardiness and maximum height growth. The resources applied to cold-weather adaptations are not available for trees to devote toward competition for light. He notes that trees species exhibit parabolic growth responses to temperature in regions where moisture stress is not present, such as eastern North America. Indeed, PET does not exceed precipitation in any of the 226 climate observations used in this research, indicating that water is generally available for transpiration by trees. Although Loehle's (1998) research concerns growth rates and range limits, his concepts should be applicable to O'Brien's (1998) species richness model, given that growth rates represent biological activity. This may partly explain the differences in the relationship between tree species richness and precipitation in this research, in contrast to studies using data from other regions; certainly water is a limiting factor, but only in cases where it is actually in limited supply.

O'Brien's (1998) model is endorsed as a powerful general explanation for global species diversity, because it considers fundamental interactions of energy and water (Hawkins *et al.* 2003, Whittaker *et al.* 2001). The model is acknowledged to provide greater global significance than models based exclusively in temperate or tropical climate zones. This is because the study area of Southern Africa where the model is based extends into the temperate and tropical climate zones (O'Brien 1993). However, the results from this research suggest that a model based on Africa inadequately explains the relationship between species richness and precipitation expressed by the relevé and climate data from Eastern North America.

CHAPTER 6. CONCLUSION

The Braun-Blanquet relevé method of vegetation sampling presents challenges for statistical and spatial analysis. Some relate to unique characteristics of the methodology; some are related to the more general difficulties caused by the naturally random variability inherent in ecological data; and some come from the application of standard statistical methods to data that rarely meet the conditional assumptions.

The greatest concern with the relevé method itself comes from the use of non-standardized plot sizes and the arbitrary, ordinal abundance values which do not conform to statistical assumptions (Robert Colwell, personal communication, November 2004). It is difficult to make use of the dimensionless cover-abundance metric. The evenness index E_{var} (Smith and Wilson 1996) is one diversity metric that accommodates species abundance expressed as a percentage, but the values are constrained by the ordinal nature of the cover scale.

Cover-abundance scale values rely on three-dimensional phenomena being accurately represented in two dimensions, by visual estimation. Confidence can be gained at the expense of information content by ignoring abundance and considering only whether a species is present or absent. This still leaves the possibility that absence may indicate that a species was missed during sampling rather than not being present in the sampled area. This problem exists for any method of sampling, and relevés have an advantage over range maps and data from floristic lists in that they are complete vegetation inventories.

Any confidence in the presence or absence of a species is undermined when the species data come from different-sized samples. Species richness is sensitive to area, as previously noted. Larger relevés were used in sites at the northern range of the sampling extent, and sizes vary over the entire sampling extent. It was noted by Chytrý (2001) that relevé sizes are often enlarged at species-poor sites. Species richness is lower at these sites for a reason; that reason is obscured when area is not held constant.

The accuracy of visual estimation is hard to control and depends on the quality of the estimator. Gauch (1982) points out that this may not be a significant source of error, considering natural variation. Area, however, is a factor that can be controlled in the sampling process. Area can vary if needed for purposes of classification, but a standard area could be used as a starting point, allowing for greater confidence in species richness estimates.

Vegetation sampling strategy decisions are limited by choices between greater accuracy per sample or more samples per study (Gauch 1982). The data used in this research were originally collected with the latter in mind, as this satisfied the original intent of the vegetation survey (Box 1994). Although there are many samples, scale differences between the vegetation data and environmental data, and a lack of local-scale data limit the explanatory power of models used in this research. The water-energy-related variables show correlation with species richness, but other variables are missing from the analysis, most likely substrate and soils data.

A major asset of the ENAVS data was also its main disadvantage; the data cover a broad extent, yet they are essentially fine-scale data. To “fill in the gaps” would require further sampling. Broader sampling of vegetation may help to give a more accurate representation of the relationship between tree species richness and climate. Relevés of 10,000 km^2 are obviously not possible, but relevés of a standard size could be used in combination with other vegetation data to estimate species richness at a sampling grain closer to that of precipitation patterns.

Climate data may have the appearance of being precise quantitative measures, in contrast to relevés made by visually estimating cover. However, the climate data used in this research are interpolated or extrapolated in many cases, and may not precisely reflect local climate. Precipitation is difficult to measure accurately, although long-term averaged records most likely smooth out measurement errors and patchy rainfall patterns. Further exploration of climate data would have been useful to this research. The use of long-term data is limited to what records are available.

The addition of fine-scale soil data would add valuable information missing from this research. An attempt was made to utilize the U.S. Department of Agriculture State Soil Geographic (STATSGO) and Soil Survey Geographic (SSURGO) digital soil databases in a geographic information system, but SSURGO data were not readily available for the majority of sampling locations, and STATSGO data appeared to be too coarse in scale in preliminary analyses. Perhaps further investigation would be productive in this area. Consistent and accurate accounting of local-scale factors such as topography and soil characteristics would provide measures reflecting higher-order factors influencing the distribution of moisture in relation to species richness. Making use of increasingly available portable devices that incorporate GIS and GPS functions would be beneficial toward this purpose.

> “Large-scale biogeography is largely an armchair pursuit; the vast majority of studies are derived from data originally published for other purposes” (Gotelli and Graves 1996, p. 308).

This research was primarily exploratory; the data were collected for one purpose and were examined to see if they could be applied to other purposes. The results suggest that relevé methods of vegetation data collection are useful in some applications, but consideration of specific applications should be made in advance of data collection. Collecting additional data after the initial sampling has taken place is difficult, especially in the case of the ENAVS data, given the sampling extent. Digital data are increasingly available but gaps still exist in coverage, especially for local-scale data.

Consideration should also be given for methods to account for spatial autocorrelation and spatial dependence inherent in ecological data. This is likely to be a reason behind large r^2 values from studies using data aggregated to large-area cell grids; none of the studies cited mentions spatial autocorrelation. Spatial autocorrelation presents problems for classical statistics, but investigation of the sources of these problems may be informing. This research used relatively unrefined methods of spatial statistical analysis and further exploration is warranted. The added complexity of including spatial considerations when designing data sampling methods and analysis would be of benefit to the results.

REFERENCES

Bailey, R. G., 1998. *Ecosystem Geography*. Springer-Verlag, New York.

Boulinier, T. J. Nichols, J. R. Sauer, J. E. Hines, and K. H. Pollock, 1998. Estimating Species Richness: The Importance of Heterogeneity in Species Detectability. *Ecology*, **79** (5): 1018-1029.

Borcard, D. and P. Legendre, 2004. *SpaceMaker2 – Users Guide*. Departément de sciences biologiques, Université de Montréal. Available from: http://www.fas.umontreal.ca/biol/legendre [Accessed March 11, 2005].

Borcard, D., P. Legendre and P. Drapeau, 1992. Partialling out the Spatial Component of Ecological Variation. *Ecology*, **73** (3): 1045-1055.

Box, E. O., 1982. SOLWAT: A Minimal-input Soil Water Simulation System Applicable to a Full Range of Natural Situations. *User's Manual for Version 2.2*, Computer-printable documentation (unpublished) available from the author. (FORTRAN program originally written in 1975 at University of North Carolina; PC version adapted from mainframe version 2.4 in May 1993, at Tokyo University).

Box, E. O., 1994. Eastern North America: Natural Environment and Sampling Strategy, In: Miyawaki, A., K. Iwatsuki, M. M. Grandtner (eds), 1994. *Vegetation in Eastern North America*, University of Tokyo Press, Japan.

Box, E. O., (unpublished). *POLATE: Inter/Extrapolates Climate Data to New Locations Based on Specified Nearby Base Sites.* (FORTRAN program with no printed documentation; PC version adapted from mainframe version in July 2000, at Yokohama National University).

Berry, W. D. and S. Feldman, 1985. Multiple Regression in Practice. *Series: Quantitative Applications in the Social Sciences*, Sage, California.

Braun-Blanquet, J., 1932. *Plant sociology: The Study of Plant Communities; Authorized English Translation of Pflanzensoziologie, by Dr. J. Braun-Blanquet.* Translated, revised and edited by Fuller, G. D. and H. S. Conard, McGraw-Hill, New York.

Burnham, K. P., and D. R. Anderson. 1998. *Model Selection and Inference: A Practical Information-theoretic Approach*. Springer-Verlag, New York.

Burnett, M. R., P. V. August, J. H. Brown Jr., and K. T. Killingbeck, 1998. The Influence of Geomorphological Heterogeneity on Biodiversity. *Conservation Biology*, **12** (2): 363-370.

Burrough, P. A., and R. A. McDonnell, 1998. *Principles of Geographical Information Systems.* Oxford University Press, New York.

Burt, J. E., and G. M. Barber, 1996. *Elementary Statistics for Geographers*. 2nd edition, Guilford Press, New York.

Chytrý, M., 2001. Phytosociological Data Give Biased Estimates of Species Richness, *Journal of Vegetation Science*, **12**: 439-444.

Colwell, R. K. 2005. *EstimateS: Statistical Estimation of Species Richness and Shared Species from Samples.* Version 7.5. Available from: http://purl.oclc.org/estimates [Accessed February 8, 2005].

Clemants, S. and G. Moore, 2003. *Patterns of Species Richness in Eight Northeastern United States Cities.* Urban Floras, **1** (1). Available from: http://www.urbanhabitats.org/v01n01/speciesdiversity_abs.html [Accessed December 8, 2003].

Cliff, A. D. and J. K Ord, 1981. *Spatial processes -Models and Applications*. Pion, London.

Clifford, P., S. Richardson, and D. Hémon, 1989. Assessing the Significance of the Correlation Between Two Spatial Processes. *Biometrics*, **45**: 123-134.

Cornell, H. V. and J. H. Lawton, 1992. Species Interactions, Local and Regional Processes, and Limits to the Richness of Ecological Communities: A Theoretical Perspective. *Journal of Animal Ecology,* **61**, pp. 1-12.

Currie, D. J. 1991. Energy and Large-Scale of Animal and Plant-Species Richness, *The American Naturalist*, **137**, (1): 27-49.

Currie, D. J. and V. Paquin, 1987. Large-scale Biogeographical Patterns of Species Richness in Trees, *Nature*, **329**, (6137): 326-327.

DeBenedictis, P. A., 1973. On the Correlation Between Certain Diversity Indices. *The American Naturalist*, **107**, (954): 295-302.

Diniz-Filho, J. A. F., and L. M. Bini, 2005. Modelling Geographical Patterns in Species Richness Using Eigenvector-based Spatial Filters. *Global Ecology and Biogeography*, **14**: 177-185.

Diniz-Filho, J. A. F., L. M. Bini and B. A. Hawkins, 2003. Spatial Autocorrelation and Red Herrings in Geographical Ecology. *Global Ecology and Biogeography*, **12**: 53-64.

Elton, C., 1958. *The Ecology of Invasions by Plants and Animals.* Methuen, London, U.K.

Falsetti, A. B., and R. R. Sokal, 1993. Genetic structure of human populations in the British Isles. *Annals of Human Biology*, **20**: 215-229.

Fisher, A.G., 1960. Latitudinal Variation in Organic Diversity. *Evolution*, **14**: 64-81.

Fortin, M-J., M. R. T. Dale, and J. ver Hoef, 2002. Spatial Analysis in Ecology. In: A. H. El-Shaarawi and W. W. Piegorsch (eds), 2002. *Encyclopedia of Environmetrics*, **4**: 2051-2058

Gotelli, N. J. and G. R. Graves, 1996. *Null Models in Ecology*, Smithsonian Institution Press, Washington D. C.

Gotelli, N. J. and R. K. Colwell, 2001. Quantifying Biodiversity: Procedures and Pitfalls in the Measurement and Comparison of Species Richness. *Ecology Letters*, **4**: 379-391.

Guo, Q. and W. L. Berry, 1998. Species Richness and Biomass: Dissection of the Hump-Shaped Relationship. *Ecology*, **79** (7): 2555-2560.

Griffiths, D., 1999. On Investigating Local-Regional Species Richness Relationships. *Journal of Animal Ecology*, **68** (5): 1051-1055.

Hawkins, B. A., R. Field, H. V. Cornell, D. J. Currie, J-F. Guégan, D. M. Kaufman, J. T. Kerr, G. G. Mittlebach, T. Oberdorff, E. M. O’Brien, E. E. Porter and J. R. G. Turner, 2003. Energy, Water, and Broad-Scale Geographic Patterns of Species Richness. *Ecology*, **84** (12): 3105-3117.

H-Acevedo, D. and D. J. Currie, 2003. Does Climate Determine Broad-Scale Patterns of Species Richness? A Test of the Causal Link by Natural Experiment, *Global Ecology and Biogeography*, **12**: 461-473.

Hengeveld, R., 1996. Measuring Ecological Biodiversity. *Biodiversity Letters*, **3** (2): 58-65.

Hulrbert, S. H., 1971. The Nonconcept of Species Diversity: A Critique and Alternative Parameters. *Ecology*, **52** (4): 577-586.

Huston, M., 1979. A General Hypothesis of Species Diversity. *American Naturalist*, **113**: 81-101.

Huston, M., 1997. Hidden Treatments in Ecological Experiments: Re-evaluating the Ecosystem Function of Biodiversity, *Oecologia*, **110**: 449-460.

Huston, M., 1999. Local Processes and Regional Patterns: Appropriate Scales for Understanding Variation in the Diversity of Plants and Animals. *Oikos*, **86** (3): 393-401.

Huston, M., 2002. Introductory Essay: Critical Issues for Improving Predictions. In: Scott, J. M., P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall and F.B. Samson, 2002 (eds), *Predicting Species Occurrences: Issues of Accuracy and Scale*, Island Press, Washington.

Kennedy, P., 2005. Oh No! I Got the Wrong Sign! What Should I Do? *Journal of Economic Education*, **36** (1): 3-28.

Knapp, R., 1984. Considerations on Quantitative Parameters and Qualitative Attributes in Vegetation Analysis and in Phytosociological Relevés, In: Knapp, R., 1984. *Sampling Methods and Taxon Analysis in Vegetation Science*, Junk, Boston.

Kenkel, N. C., P. Juhász-Nagy and J. Podani, 1989. On Sampling Procedures in Population and Community Ecology, *Vegetatio*, **83**: 195-207.

Latham, R. E., and R. E. Ricklefs, 1993. Continental Comparisons of Temperate-Zone Tree Species Diversity, In: Ricklefs, R. E. and D. Schluter, 1993 (eds). *Species Diversity in Ecological Communities*, The University of Chicago Press, Chicago.

Legendre, P., 1993. Spatial Autocorrelation: Trouble or New Paradigm? *Ecology*, **76** (4): 1659-1673.

Legendre, P., and L. Legendre, 1998. *Numerical ecology*: *Second English edition*, Elsvier, New York.

Lennon, J. J., 2000. Red-shifts and Red Herrings in Geographical Ecology. *Ecography*, **23**: 101-113.

Liebhold, A. M., and J. Gurevitch. 2002. Integrating the Statistical Analysis of Spatial Data in Ecology. *Ecography,* **25**: 553-557.

Loehle, C., 1998. Height Growth Rate Tradeoffs Determine Northern and Southern Range Limits of Trees. *Journal of Biogeography*, **25**: 735-742.

Lyons, S. K., and M. R. Willig, 1999. A Hemispheric Assesment of Scale Dependence in Latitudinal Gradients of Species Richness. *Ecology*, **80** (8): 2483-2491.

Magurran, A. E., 1988. *Ecological Diversity and Its Measurement*, Princeton University Press, Princeton, New Jersey.

Magurran, A. E., 2003. *Measuring Biological Diversity.* Blackwell Publishing, Malden, Maine.

Manly, B. F. J., 1997. *Randomization, Bootstrapping and Monte Carlo Methods in Biology*. 2nd edition. Chapman and Hall, New York.

Margalef, R., 1972. Homage to Evelyn Hutchinson, or Why is There an Upper Limit to Diversity? *Trans. Connect. Acad. Arts. Sci.*, **44**: 211-235.

May, R. M., 1973. *Stability and Complexity in Model Ecosytems*. Princeton University Press, Princeton, New Jersey.

May, R.. M., 1975. Patterns of Species Abundance and Diversity. In: M. L. Cody and J. M. Diamond (eds). *Ecology and Evolution of Communites,* 81-120. Harvard University Press, Cambridge.

May, R. M., 1988. How Many Species are There on Earth? *Science*, **241** (4872): 1441-1449.

McNab, W. H., 1993. A Topographic Index to Quantify the Effect of Mesocale Landform on Site Productivity. *Canadian Journal of Forest Resources*, **23**: 1100-1107.

Meentemeyer, V. and E. O. Box, 1987. Scale Effects in Lanscape Studies. In: Turner, M. G. (ed.) *Landscape Heterogeneity and Disturbance,* 15-34. Springer-Verlag, New York.

Meyer, S. M., 2004. End of the Wild. *Boston Review*, April/May 2004. Available at: http://bostonreview.net/BR29.2/meyer.html [Accesed June 28, 2004].

Miller, R. I., 1986. Predicting Rare Plant Distribution Patterns in the Southern Appalachians of the South-Eastern U.S.A. *Journal of Biogeography*, **13** (4): 293-311.

Miyawaki, A., 1994. Introduction and Synopsis. In: Miyawaki, A., K. Iwatsuki, M. M. Grandtner (eds). *Vegetation in Eastern North America*, University of Tokyo Press, Japan.

Monk, C. D., 1967. Tree Species Diversity in the Eastern Deciduos Forest with Particular Reference to North Florida. *American Naturalist*, **101** (918): 173-187.

Moore, J. J., 1962. The Braun-Blanquet System: A Reassessment. *Journal of Ecology*, **50** (3): 761-769.

Mueller-Dombois, D. and H. Ellenberg, 1974. *Aims and Methods of Vegetation Ecology*. John Wiley and Sons, New York.

Nakamura, Y., M. M. Grandtner and N. Villeneuve, 1994. Boreal and Oroboreal Forests of Eastern North America and Japan, In: Miyawaki, A., K. Iwatsuki, M. M. Grandtner (eds). *Vegetation in Eastern North America*. University of Tokyo Press, Japan.

Neilson, R. P., G. A. King, R. L. DeVelice, and J. M. Lenihan, 1992. Regional and local vegetation patterns: The responses of vegetation diversity to subcontinental air masses, In: Hansen, A. J. and F. di Castri (eds). *Landscape boundaries: Consequences for biotic diversity and ecological flows*. Springer-Verlag, New York.

Nichols. W. F., K. T. Killingbeck, and P. V. August, 1998. The Influence of Geomorphological Heterogeneity on Biodiversity. *Conservation Biology*, **12** (2): 363-370.

O'Brien, E. M., 1988. Climatic Correlates of Species Richness for Woody Edible Plants Across Southern Africa. *Monogr. Syst. Bot. Missouri Botanical Garden*, **25**: 385-401.

O’Brien, E. M., 1993. Climatic Gradients in Woody Plant Species Richness: Towards an Explanation Based on Analysis of Southern Africa’s Woody Flora. *Journal of Biogeography*, **20** (2): 181-198.

O’Brien, E. M., 1998. Water-Energy Dynamics, Climate, and Prediction of Woody Plant Species Richness: An Interim Model. *Journal of Biogeography*, **25**, (2): 379-398.

O'Brien, E. M., R. Field, and R. J. Whittaker, 2000. Climatic Gradients in Woody-Plant (Tree and Shrub) Diversity: Water-energy Dynamics, Residual Variation, and Topography. *Oikos*, **89** (3): 588-600.

Openshaw, S., 1984. The Modifiable Areal Unit Problem. *Concepts and Techniques in Modern Geography* (CATMOG) 38: 41.

O’Sullivan, D., and D. Unwin, 2003. *Geographic Information Analysis*. John Wiley and Sons, Hoboken, New Jersey.

Orlóci, L., 1978. *Multivariate Analysis in Vegetation Research*. 2nd edition. Junk, The Hague.

Palmer, M. W., 1994. Variation in Species Richness: Towards a Unification of Hypotheses. *Folia Geobot. Phytotax.* **29**: 511-530.

Palmer, M. W., and P. S. White, 1994. Scale Dependence and the Species-Area Relationship. *American Naturalist*, **144**: 717-740.

Pianka, E. R., 1966. Latitudinal Gradients in Species Diversity: A Review of Concepts, *The American Naturalist*, Vol. 100, No. 910, pp.33-46.

Pielou, E. C., 1975. *Ecological Diversity*. Wiley-Interscience, New York.

Pielou, E. C., 1977. *Mathematical Ecology*. 2nd edition. Wiley-Interscience, New York.

Poore, M. E. D., 1955. The Use of Phytosociological Methods in Ecological Investigations: I. The Braun-Blanquet System. *Journal of Ecology*, **43** (1): 226-244.

Poore, M. E. D., 1955. The Use of Phytosociological Methods in Ecological Investigations: II. Practical Issues Involved in an Attempt to Apply the Braun-Blanquet System. *Journal of Ecology*, **43** (1): 254-269.

Poore, M. E. D., 1955. The Use of Phytosociological Methods in Ecological Investigations: III. Practical Application. *Journal of Ecology*, **43** (2): 606-651.

Poore, M. E. D., 1955. The Use of Phytosociological Methods in Ecological Investigations: IV. General Discussion of Phytosociological Problems, *Journal of Ecology*, **44** (1): 28-50.

Rhode, K., 1992. Latitudinal Gradients in Species Diversity: The Search for the Primary Cause, *Oikos*, **65**: 514-527.

Ricklefs, R. E., 2004. A Comprehensive Framework for Global Patterns in Biodiversity. *Ecology Letters*, **7**: 1-15.

Rogerson, P., 2001. *Statistical Methods for Geography*. Sage, New York.

Rosenberg, M. S. 2001. *PASSAGE. Pattern Analysis, Spatial Statistics, and Geographic Exegesis*. Version 1.1, release 3.4. Department of Biology, Arizona State University, Tempe, Arizona.

Rosenzweig, M L., 1995. *Species Diversity in Time and Space*. Cambridge University Press, New York.

Rosenzweig, M. L. and Z. Abramsky, 1993. How are Diversity and Productivity Related? In: Ricklefs, R. E. and D. Schluter (eds). *Species Diversity in Ecological Communities*. The University of Chicago Press, Chicago.

Schluter, D. and R. E. Ricklefs, 1993. Species Diversity: An Introduction to the Problem In: Ricklefs, R. E. and D. Schluter (eds). *Species Diversity in Ecological Communities*. The University of Chicago Press, Chicago.

Shannon, C. E. and W. Weaver. 1949. *The Mathematical Theory of Communication*. University of Illinois, Urbana, Illinois

Simpson, E. H. 1949. Measurement of Diversity. *Nature*, **163**: 688.

Smouse, P. E., J. C. Long and R. R. Sokal. 1986. Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Systematic Zoology,* **35** (4): 627-632.

Southwood, T. R. E. and P. A. Henderson. 2000. *Ecological Methods*. Blackwell, Oxford.

Stephenson, N.L., 1998. Actual Evapotranspiration and Deficit: Biologically Meaningful Correlates of Vegetation Distribution across Spatial Scales, *Journal of Biogeography,* **25** (5): 855-870.

Taylor, P. H. and S. D. Gaines, 1999. Can Rapoports Rule be Rescued? Modeling Causes of the Latitudinal Gradient in Species Richness. *Ecology*, **80** (8): 2474-2483.

Thornthwaite, C. W, 1954. Topoclimatology. In: *Proceedings of the Toronto Meteorological Conference, September 9-15, 1953.* Toronto: Royal Meteorological Society. 227-232

Thornthwaite, C. W. and J.R. Mather, 1957. *Instructions and Tables for Computing Potential Evapotranspiration and the Water Balance*. Laboratory of Climatology. Centerton, N.J.

Tilman, D. 1999. The Ecological Consequences of Changes in Biodiversity: A Search for General Principles. *Ecology*, **80** (5): 1455-1475.

Tilman, D. and S. Pacala, 1993. The Maintenance of Species Richness in Plant Communities In: Ricklefs, R. E. and D. Schluter (eds). *Species Diversity in Ecological Communities*. The University of Chicago Press, Chicago.

United Sates Geological Survey, 2003. *Water Resources of the United States: Water Resouces NSDI Node*. Physiographic divisions of the conterminous U. S., Available from: ftp://water.usgs.gov/pub/dsdl/physio.e00.gz [Accessed December 8, 2003].

United Sates Geological Survey, 2004. *The National Map Seamless Data Distribution System Viewer: The National Elevation Dataset*. 1/3" NED.
Available from: http://seamless.usgs.gov/viewer.htm [Accessed September 29, 2004].

Urban, D. 2001. *Spatial analysis in ecology: Mantel's test*. Available from: http://www.env.duke.edu/landscape/classes/env352/mantel.pdf. [Accessed March 29, 2005].

van Rensburg, B. J., 2002. Species Richness, Environmental Correlates, and Spatial Scale: A Test Using South African Bird. *The American Naturalist*, **159** (5): 566-576.

Weiher, E. 1999. The Combined Effects of Scale and Productivity on Species Richness, *Journal of Ecology*, **87**: 1005-1011.

Whittaker, R. H., 1963. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs*, **30**: 279-338.

Whittaker, R. J., 1999. Scaling, Energetics and Diversity. *Nature*, **401**: 865-866.

Whittaker, R. J., K. J. Willis, and R. Field, 2001. Scale and Species Richness: Towards a General Hierarchical Theory of Species Diversity. *Journal of Biogeography*, **28**: 453-470.

White, P. S. and R. Miller, 1988. Topographic Models of Vascular Plant Richness in the Southern Appalachian High Peaks. *Journal of Ecology*, **76** (1): 192-199.

Willig, M. R., 2003. Challenges to Understanding Dynamics of Biodiversity in Time and Space. *Paleobiology*, **29** (1): 30-33.

Zevenbergen, L. W., and C. R. Thorne, 1987. Quantitative Analysis of Land Surface Topography. *Earth Surface Processes and Landforms*, **12**: 47-56.

APPENDIX

Table A.1: Model 1, Shrub layer species richness = *f* (PET).

r	r^2	Adj. r^2	Std. Error
0.522	0.272	0.269	5.675

225 relevés	Coefficients	Std. Error	t	Sig.
(Constant)	0.8890	0.0000	0.6070	0.5445
PET	0.0157	0.0020	9.1320	0.0000

Table A.2: Model 11, Tree species richness = *f* (PY + (PET + PET^2)).

r	r^2	Adj. r^2	Std. Error
0.623	0.389	0.380	3.914

225 relevés	Coefficients	Std. Error	Std. Coef.	t	Sig.
(Constant)	-21.12305	5.44595		-3.87867	0.00014
PET	0.07920	0.01229	3.50593	6.44368	0.00000
PET^2	-0.00004	0.00001	-2.95779	-5.37447	0.00000
PY	-0.00106	0.00149	-0.04307	-0.71294	0.47664

Multicollinearity diagnostics

Condition index 98.7

	VIF
PET	107.0
PET^2	109.5
PY	1.3

Coefficient correlations	PET	PET^2	PY
PET	1	0.209	-0.255
PET^2	0.209	1	-0.994
PY	-0.255	-0.994	1

Table A.3: Model 12, Tree species richness = *f* (PY + (PET + PET^2) + ELEV).

r	r^2	Adj. r^2	Std. Error
0.641	0.411	0.401	3.849

225 relevés	Coefficients	Std. Error	Std. Coef.	t	Sig.
(Constant)	-21.51316	5.35801		-4.01514	0.00008
PET	0.08313	0.01216	3.67968	6.83373	0.00000
PET^2	-0.00004	0.00001	-2.96846	-5.48399	0.00000
PY	-0.00397	0.00177	-0.16140	-2.24135	0.02600
ELEV	0.00351	0.00121	0.21569	2.90820	0.00401

Multicollinearity diagnostics

Condition index 103.8

	VIF
PET	108.3
PET^2	109.5
PY	1.9
ELEV	2.1

Coefficient correlations	PET	PET^2	PY	ELEV
PET	1	-0.565	0.111	-0.007
PET^2	-0.565	1	0.108	-0.207
PY	0.111	0.108	1	-0.989
ELEV	-0.007	-0.207	-0.989	1

Table A.4: Model 13, Tree species richness = *f* (PY + (PET + PET^2) + ELEV).

r	r^2	Adj. r^2	Std. Error
0.601	0.361	0.343	4.005

144 relevés	Coefficients	Std. Error	Std. Coef.	t	Sig.
(Constant)	-9.05573	4.83992		-1.87105	0.06344
PET	0.05097	0.01006	2.09164	5.06828	0.00000
PET^2	-0.00002	0.00001	-1.30965	-3.34012	0.00108
PY	-0.00575	0.00213	-0.24193	-2.70117	0.00777
ELEV	0.00487	0.00155	0.33506	3.13089	0.00212

Multicollinearity diagnostics

Condition index 65.8

	VIF
PET	37.0
PET^2	33.4
PY	1.7
ELEV	2.5

Coefficient correlations	PET	PET^2	PY	ELEV
PET	1	-0.576	-0.199	0.385
PET^2	-0.576	1	-0.012	-0.153
PY	-0.199	-0.012	1	-0.965
ELEV	0.385	-0.153	-0.965	1

Table A.5: Model 14, Tree species richness = *f* (PY + (PET + PET^2) + ELEV + EVARDIV).

r	r^2	Adj. r^2	Std. Error
0.646	0.417	0.396	3.838

144 relevés	Coefficients	Std. Error	Std. Coef.	t	Sig.
(Constant)	-13.09663	4.76912		-2.74613	0.00683
PET	0.04994	0.00964	2.04949	5.17929	0.00000
PET^2	-0.00002	0.00001	-1.22243	-3.24631	0.00147
PY	-0.00585	0.00204	-0.24600	-2.86557	0.00481
ELEV	0.00411	0.00150	0.28321	2.73506	0.00706
EVARDIV	0.56822	0.15572	0.25186	3.64909	0.00037

Multicollinearity diagnostics

Condition index 72.5

	VIF
PET	37.1
PET^2	33.6
PY	1.7
ELEV	2.5
EVARDIV	1.1

Coefficient correlations	PET	PET^2	PY	ELEV	EVARDIV
PET	1	-0.013	-0.137	0.063	-0.029
PET^2	-0.013	1	-0.569	-0.013	-0.152
PY	-0.137	-0.569	1	-0.206	0.385
ELEV	0.063	-0.013	-0.206	1	-0.964
EVARDIV	-0.029	-0.152	0.385	-0.964	1

Table A.6: Models 15 – 18, Tree species richness = *f* (PET + PY + ELEV).

Model (Group)	r	r^2	Adj. r^2	Std. Error
15 (AH)	0.709	0.502	0.475	3.589
16 (CP)	0.281	0.079	0.038	4.704
17 (NE_SL)	0.414	0.172	0.111	3.508
18 (P)	0.622	0.386	0.347	2.979

Model (Group)		Coefficients	Std. Error	Std. Coef.	t	Sig.
15 (AH)	(Constant)	-7.864	4.311		-1.824	0.074
	PETY	0.031	0.006	0.856	5.140	0.000
	PY	-0.001	0.003	-0.062	-0.479	0.634
	ELEV	0.003	0.002	0.217	1.172	0.246
16 (CP)	(Constant)	10.125	4.304		2.353	0.022
	PETY	0.010	0.005	0.362	2.203	0.031
	PY	-0.004	0.004	-0.156	-0.976	0.333
	ELEV	0.001	0.013	0.005	0.042	0.966
17 (NE_SL)	(Constant)	8.658	8.573		1.010	0.318
	PETY	0.017	0.013	0.219	1.328	0.192
	PY	-0.009	0.003	-0.376	-2.549	0.015
	ELEV	0.004	0.004	0.195	1.155	0.255
18 (P)	(Constant)	10.775	8.648		1.246	0.219
	PETY	0.018	0.005	0.563	3.561	0.001
	PY	-0.010	0.009	-0.193	-1.100	0.277
	ELEV	0.012	0.007	0.285	1.875	0.067

Table A.7: Model 19, tree species richness = *f* (PET + PY + MON4 + MON5 + MON9).

r	r^2	Adj. r^2	Std. Error
0.726	0.527	0.474	2.673

51 relevés	Coefficients	Std. Error	Std. Coef.	t	Sig.
(Constant)	6.364	9.960		0.639	0.526
PET	0.045	0.013	1.422	3.438	0.001
PY	-0.025	0.009	-0.479	-2.678	0.010
MON4	153.596	38.612	3.046	3.978	0.000
MON5	-123.288	29.986	-3.748	-4.112	0.000
MON9	52.474	20.319	1.303	2.583	0.013

Multicollinearity diagnostics

Condition index 111.55

VIF	
PET	16.28
PY	3.04
MON4	55.80
MON5	79.07
MON9	24.21

Coefficient correlations	PY	PET	MON4	MON5	MON9
PET	1	-0.592	0.464	-0.571	0.930
PY	-0.592	1	-0.601	0.615	-0.485
MON4	0.464	-0.601	1	-0.980	0.585
MON5	-0.571	0.615	-0.980	1	-0.703
MON9	0.930	-0.485	0.585	-0.703	1

Table A.8: Mantel and partial Mantel test results for matrices based on tree species richness and single environmental variables.

Mantel – 51 Piedmont Relevés

Matrix 1	NSPP_T	NSPP_T	NSPP_T
Matrix 2	PET	PY	ELEV
Observed *Z*	1613170.2	809530	1086088
Correlation	0.2334	-0.0142	0.1002
t	5.2821	-0.1996	1.5013
Left-tailed p	1	0.42089	0.93336
Right-tailed p	0	0.57911	0.06664
Randomization			
Total Iterations	4999	4999	4999
> Observed *Z*	4999	2207	4561
< Observed *Z*	0	2792	438
= Observed *Z*	1	1	1
Left-tailed p	1.0000	0.4416	0.9124
Right-tailed p	0.0002	0.5586	0.0878
Two-tailed p	0.0002	0.5586	0.0878

Partial Mantel – 51 Piedmont Relevés

Matrix 1	NSPP_T	NSPP_T	NSPP_T
Matrix 2	PET	PY	ELEV
Constant	MON 4, 5 & 9	MON 4, 5 & 9	MON 4, 5 & 9
Observed *Z*	96688.6578	-23059.9169	22233.4137
Correlation	0.1591	-0.0469	0.0412
t	3.3718	-0.628	0.5737
Left-tailed p	0.99963	0.265	0.71693
Right-tailed p	0.00037	0.735	0.28307
Randomization			
Total Iterations	4999	4999	4999
> Observed *Z*	4995	1386	3496
< Observed *Z*	4	3613	1503
= Observed *Z*	1	1	1
Left-tailed p	0.9992	0.2774	0.6994
Right-tailed p	0.0010	0.7228	0.3008
Two-tailed p	0.0014	0.5278	0.5830

Table A.9: Mantel and partial Mantel test results for matrices based on tree species richness and multiple environmental variables.

Mantel – 51 Piedmont Relevés

Matrix 1	NSPP_T	NSPP_T
Matrix 2	PET & PY	PET, PY & ELEV
Observed *Z*	1880582.547	2271831.023
Correlation	0.1764	0.163
t	4.0955	3.623
Left-tailed p	0.99998	0.99985
Right-tailed p	0.00002	0.00015
Randomization		
Total Iterations	4999	4999
> Observed *Z*	4996	4994
< Observed *Z*	3	5
= Observed *Z*	1	1
Left-tailed p	0.9994	0.9990
Right-tailed p	0.0008	0.0012
Two-tailed p	0.0008	0.0012

Partial Mantel – 51 Piedmont Relevés

Matrix 1	NSPP_T	NSPP_T
Matrix 2	PET& PY	PET, PY & ELEV
Constant	MON 4, 5 & 9	MON 4, 5 & 9
Observed *Z*	64091.6258	54989.6799
Correlation	0.0939	0.0772
t	1.864	1.4104
Left-tailed p	0.96884	0.92078
Right-tailed p	0.03116	0.07922
Randomization		
Total Iterations	4999	4999
> Observed *Z*	4819	4581
< Observed *Z*	180	418
= Observed *Z*	1	1
Left-tailed p	0.9640	0.9164
Right-tailed p	0.0362	0.0838
Two-tailed p	0.0642	0.1598

Table A.10: Mantel and partial Mantel test results for matrices based on tree species turnover (beta diversity) and single environmental variables.

Mantel – 51 Piedmont Relevés

Matrix 1	Beta_T	Beta_T	Beta_T
Matrix 2	PET	PY	ELEV
Observed *Z*	212527.578	117119.12	149481.72
Correlation	0.2910	0.0838	0.1684
t	6.6522	1.1951	2.5577
Left-tailed p	1	0.88398	0.99473
Right-tailed p	0	0.11602	0.00527
Randomization			
Total Iterations	4999	4999	4999
> Observed *Z*	4999	4367	4961
< Observed *Z*	0	632	38
= Observed *Z*	1	1	1
Left-tailed p	1.0000	0.8736	0.9924
Right-tailed p	0.0002	0.1266	0.0078
Two-tailed p	0.0002	0.1266	0.0078

Partial Mantel – 51 Piedmont Relevés

Matrix 1	Beta_T	Beta_T	Beta_T
Matrix 2	PET	PY	ELEV
Constant	MON 4, 5 & 9	MON 4, 5 & 9	MON 4, 5 & 9
Observed *Z*	5190.2118	982.8507	2295.5239
Correlation	0.1816	0.0425	0.0903
t	3.7472	0.5495	1.2167
Left-tailed p	0.99991	0.70867	0.88814
Right-tailed p	0.00009	0.29133	0.11186
Randomization			
Total Iterations	4999	4999	4999
> Observed *Z*	4997	3632	4477
< Observed *Z*	2	1367	522
= Observed *Z*	1	1	1
Left-tailed p	0.9996	0.7266	0.8956
Right-tailed p	0.0006	0.2736	0.1046
Two-tailed p	0.0006	0.5860	0.2208

Table A.11: Mantel and partial Mantel test results for matrices based on tree species turnover (beta diversity) and multiple environmental variables.

Mantel – 51 Piedmont Relevés

Matrix 1	Beta_T	Beta_T
Matrix 2	PET & PY	PET, PY & ELEV
Observed *Z*	254795.6994	312097.6253
Correlation	0.2642	0.2879
t	6.1936	6.4661
Left-tailed p	1	1
Right-tailed p	0	0
Randomization		
Total Iterations	4999	4999
> Observed *Z*	4999	4999
< Observed *Z*	0	0
= Observed *Z*	1	1
Left-tailed p	1.0000	1.0000
Right-tailed p	0.0002	0.0002
Two-tailed p	0.0002	0.0002

Partial Mantel – 51 Piedmont Relevés

Matrix 1	Beta_T	Beta_T
Matrix 2	PET& PY	PET, PY & ELEV
Constant	MON 4, 5 & 9	MON 4, 5 & 9
Observed *Z*	5025.1347	6218.5176
Correlation	0.1566	0.1857
t	3.0204	3.2901
Left-tailed p	0.99874	0.9995
Right-tailed p	0.00126	0.0005
Randomization		
Total Iterations	4999	4999
> Observed *Z*	4993	4996
< Observed *Z*	6	3
= Observed *Z*	1	1
Left-tailed p	0.9988	0.9994
Right-tailed p	0.0014	0.0008
Two-tailed p	0.0014	0.0008

Table A.12: Mantel tests comparing tree species richness and beta diversity to multiple environmental variables. Top: 225 DFW relevés. Bottom: 65 relevés with topographic index. Correlation strength & significance are likely to be inflated by spatial autocorrelation.

Mantel - 225 DFW relevés

Matrix 1	NSPP_T	Beta_T
Matrix 2	PET, PET^2, PY & ELEV	PET, PET^2, PY & ELEV
Observed *Z*	132236396872.53	17202523455.83
Correlation	0.1965	0.4607
t	9.9832	25.6516
Left-tailed p	1	1
Right-tailed p	0	0
Randomization		
Total Iterations	4999	4999
> Observed *Z*	4999	4999
< Observed *Z*	0	0
= Observed *Z*	1	1
Left-tailed p	1.0000	1.0000
Right-tailed p	0.0002	0.0002
Two-tailed p	0.0002	0.0002

Mantel - 65 relevés w/ topographic index

Matrix 1	NSPP_T	Beta_T
Matrix 2	PET, PET^2, PY & EVARDIV	PET, PET^2, PY & EVARDIV
Observed *Z*	12382781982.26	1385667001.25
Correlation	0.2325	0.4449
t	5.179	9.9783
Left-tailed p	1	1
Right-tailed p	0	0
Randomization		
Total Iterations	4999	4999
> Observed *Z*	4999	4999
< Observed *Z*	0	0
= Observed *Z*	1	1
Left-tailed p	1.0000	1.0000
Right-tailed p	0.0002	0.0002
Two-tailed p	0.0002	0.0002