

On the Spatial Distribution and Abundance of Herbaceous Plants in Aspen and Conifer
Forest Understories: A Test of Neutral/Statistical Versus Niche-Based Theories?

Jacob Godinez
John Harte, Mentor
REU Program 2008
Rocky Mountain Biological Laboratory

Abstract

The purpose of this study is to test several macroecological statistical theories in the sub-alpine environment. Statistical theories are based on purely statistical assumptions, and these predictions were tested against measured values for macroecological metrics including the species-abundance distribution, the species-area relationship, and the collector's curve. In order to test the theories, it was necessary to calculate these macroecological metrics from raw plant data. The plant data consisted of mostly plant census, along with slope, aspect, canopy, soil temperature, and soil moisture. From the plant census data, the graphs of collector's curve, species-area relationship and species-abundance distribution were analyzed using these measured values and the predicted values from the maximum entropy theory. With almost exact certainty, the maximum entropy theory predictions held true for what was observed in nature.

Keywords: Maximum Entropy Theory, Statistical Models, Species-Area Relationship, Species-Abundance Distribution, Collector's Curve

Introduction

Ecology is defined as a branch of science concerned with the interrelationship of organisms and their environments ("Ecology" 720). A major focus in ecology is macroecology, the study of patterns in the distribution and abundance of multiple species across spatial scales within ecological communities. At the frontier of macroecology is the question of whether plant traits influence these patterns; if the answer is yes, then because plants with different traits live in different local environments, we would expect to see different macroecological patterns in different environments. Theories based on this concept are called "niche-based" theories, where the term niche refers to the abiotic conditions in the local environment that determine the traits of the species that can live there. Contrasting with niche-based or mechanistic theories are statistical theories (or "neutral" theories) that predict macroecological patterns regardless of the collection of traits possessed by the organisms that live in a particular community. One goal of my work is to determine if climate and abiotic factors not only influence the traits and niches

of plants, but also influence the distribution of abundances across species, the species-area relationship, and the spatial distribution of individuals within species across intraplot spatial scales. A second and related goal is to test several statistical theories that predict the form of these macroecological patterns independent of the abiotic conditions in the environment and the set of traits that characterize the species living there.

Because species with different traits live in different abiotic environments, to test whether traits influence patterns, I will study such patterns in both aspen and conifer understory plots with differing aspect and slope because I expect these habitats will possess different abiotic conditions. A set of plant traits and abiotic environmental conditions will be measured to quantify the extent to which this expectation is correct. With the presumption in hand, I will test the macroecological predictions of several statistical theories. The math- and physics-based statistical theories that I will test are the random placement model or RPM (Coleman et. al., 1982), fractal theory or FT (Halley et. al., 2004), and the recently advanced maximum entropy theory or MaxEnt (Harte et. al., 2008).

The Coleman random placement model is based on the assumption that individuals within each species are randomly scattered over the plot or landscape. The fractal theory is based on the assumption that the distribution of individuals is “scale free”, which, figuratively speaking means that whether one looks through a microscope or a telescope, they will see the same geometric patterns. The MaxEnt theory is a mathematically proven method of inferring unique, least-biased probability distributions, if all of one’s prior knowledge of those distributions is encapsulated by a set of specified, empirical constraints. The method entails maximizing information entropy under those

constraints using the method of Lagrange multipliers. In ecological applications, the constraints arise from ratios of four measured state variables: A (the area of an ecosystem), S (the total number of species in the ecosystem), N (the total number of individuals in the ecosystem), and E (the total metabolic rate of all the individuals located within the ecosystem).

To test the theory, it will be necessary to calculate macroecological metrics, which are quantitative measurements of a pattern, from raw plant data. Such metrics include the Species-Abundance Distribution (SAD), the Species-Area Relationship (SAR), and the collector's curve. The SAD will be analyzed as a rank-abundance distribution rather than as a probability density to avoid ambiguities associated with arbitrary binning procedures (Green et. al., 2003).

Among the statistical theories, only MaxEnt makes a prediction for the SAD: a Fisher log-series distribution obeying the equation $\Phi(n) \sim (c/n)\exp(-\lambda n)$, where $\Phi(n)$ is the probability a species has n individuals and c is the normalization constant approximately given by $c=1/\ln(1/\lambda)$. For a system with a large number of individuals, λ depends solely on the ratio S/N ; on the other hand, for a system with a small number of individuals, λ also depends weakly on S and N separately. For the SAR, which is plotted as $\log(\text{\# of species})$ versus $\log(\text{area})$, the RPM predicts curvature, FT predicts a straight line that follows the Power Law behavior (Borda-de-Agua et. al., 2002), and MaxEnt predicts that the slope of a straight line fit will decrease and the goodness of a straight line fit will increase as the ratio of S/N decreases. Lastly, for the collector's curve, MaxEnt predicts that the number of species increases approximately as $\ln(N)$, where N represents the number of individuals in an ecosystem, with the approximation improving

as S/N decreases. To repeat the major issue that I wish to resolve: while it is obvious that abiotic conditions constrain the species list located in an environment, do they influence the patterns characterizing the local distribution and abundance of those species within that environment? By looking at multiple environments from multiple perspectives I will be contributing to our understanding of the degree to which “universal” rather than particular and contingent factors dominate macroecology.

Methods and Materials

This project was centered in the surrounding Aspen and Conifer forests of Gothic, Colorado. There were six square-meter observational plots in each of the two forests types and one plot in the open area between the two forests, for a total of thirteen square-meter research plots. These plots were assigned with respect to slope and aspect across both the Aspen and Conifer Forests. A square-meter frame was used not only to designate the research plot with flags, but was also necessary when conducting a census of the plot. The square-meter frame was divided into a 16 by 16 grid, totaling 256 cells. These small increments made it easier to differentiate what species belonged to which cell, which helped to attain a more accurate census.

The plant census, which counted the number of species and individuals within a research plot, was performed twice this summer to take into account early- and late-blooming species, and was done by placing the square-meter frame onto the plot, and carefully transcribing the number of species in the plot and individuals within each species onto paper. This was easily accomplished by printing a table on paper modeling the square-meter grid and writing the number of species and individuals in the box that

correspond to the correct cell. The census was done only on the herbaceous plants in the forest understory. The plant was traced back to the roots, and multiple plants stemming from the same root were counted as one plant.

The soil temperature was measured in the early morning by inserting a thermometer probe about ten centimeters into the ground, and taking the reading from the probe. In order to keep relative data, I collected the temperature samples of all the plots within an hour timeframe. In addition to the time period imposed on this sampling, there was also a second reading during midday to be able to compare the soil temperatures. When collecting the soil temperature for each plot, the air temperature was measured with the same temperature probe held about one meter above the ground.

The soil moisture was recorded by taking core samples of about ten-to-twelve centimeters deep with a diameter of two-and-a-half centimeters. These core samples were taken to the laboratory, where the wet weight (W) was recorded and then the samples were put into the drying oven at 110°C for twenty-four hours. Then the samples were weighed for the dry weight (D), sieved to 2 mm, and then the rocks were weighed separately (R). Percent soil moisture was calculated as $100 \times (W - D) / (D - R)$.

Aspect was determined for the plots by taking a compass and facing the direction of the slope, and taking the reading with respect to degrees and orientation. The slope for each plot was determined using a 1-3 grading system, where 1 was a mild slope, 2 was a moderate slope, and 3 was a steep slope. The canopy cover (shade) of a plot was determined in a similar manner, where 1 was an open canopy, 2 was a moderate canopy, and 3 was a dense canopy.

Results

To test the MaxEnt prediction for the SAD, the theoretical value of λ was first calculated from S and N for each plot (Harte et. al., 2008). Then, by sampling from the predicted Fisher-log series distribution, $\Phi(n) \sim (c/n)e^{(-\lambda n)}$, the predicted values of abundance could be computed and compared to the observed values. Plots of predicted abundance versus measured abundance were constructed and a straight line fit was assigned to look at the slope (b) and goodness of the straight line fit (R^2). Perfect agreement would correspond to b and R^2 values of 1. The results from all thirteen plots had an average b value of 0.938 and an average R^2 value of 0.944 (see Table 1). An R^2 value of 0.944 translates into 94% of the variance in the data being explained by the MaxEnt prediction.

The Species-Area Relationship (SAR) was analyzed by plotting $\ln(A)$ versus $\ln(S)$, where A represents the area of subplots within each plot and S represents the number of species in the subplots (He and Legendre 1996). The shapes of all the resulting graphs had slight curvature, but not enough to follow the prediction of the random placement model. The equation $S=CA^Z$ is a power function that is often used to describe the equation for an SAR graph (Williams 1996); it corresponds to the fractal model and would result in zero curvature on a log-log plot. Clearly observed, and systematically concave (i.e., negative second derivative) curvature ruled out the fractal theory. To further test the MaxEnt theory, the slope of the straight fit line was examined. The predicted slope values from the MaxEnt theory come from the S/N ratio and the measured slope values were taken from the straight line fit of all the SAR graphs. These

slope values were graphed on a predicted slope versus measured slope plot and the slope and goodness of the straight line fit were measured, expecting to see a Z and R^2 value of 1 for a perfect match. Excluding one outlier (the only plot that had *Artemesia Tridentata* present), 94% of the variance in the data is explained by the prediction ($R^2=0.936$) and the slope was equal to 0.9991 (see Figure 1).

The collector's curve is given by plotting the number of individuals selected against the number of species collected (He and Legendre 2002). In other words, as an individual is identified from an ecosystem, it will correspond to a certain species, but the number of species does not increase exactly with the number of individuals unless there is an ecosystem of unique individuals with no repetition. To extract a collector's curve from our data, each plot was treated as a random draw from a larger pool of individuals. The MaxEnt theory makes a prediction that the number of species, S, in a plot will increase as the log of the number of individuals, $\ln(N)$, in the plot. When the graph of S versus $\ln(N)$ was examined, an overall upward trend was observed but it was not a straight line, rather a line with a positive slope and plenty of scatter. The slope was equal to 2.60 and the R^2 value was equal to 0.62. The scatter around the logarithmic prediction is a measure of the role of environmental variability, which was tested below.

In order to test how environmental variability affected this study, a backwards stepwise multiple linear regression of S (total number of species), N (total number of individuals), and the ratio S/N with respect to environmental factors such as AM and PM soil temperature, air temperature, soil moisture, aspect, slope, and canopy. The statistical program "Sigma Plot" was used in order to conduct this test. The procedure eliminates un-explanatory variables one at a time until it arrives at a model that meets the criterion

of only containing significant variables. When strongly co-dependent variables, such as AM and PM temperature, occur, the procedure eliminates whichever of them is least explanatory. The results for the most successful multiple regression models are as follows: $S = -14.21(4.68) + 1.50(0.29)*AM \text{ soil temp} + 0.107(.033)*\text{soil moisture}$, where AM soil temperature has a P-value < 0.001 and soil moisture has a P-value = 0.009, and the model has an adjusted R^2 value of 0.73; $N = -310.6(189.9) + 39.8(15.1)*AM \text{ soil temp}$, where AM soil temperature has a P-value = 0.023 and the model has an adjusted R^2 value of 0.39; $S/N = 0.045(.013) + 0.0195(.0081)*\text{canopy}$, where canopy has a P-value of 0.035 and the model has an adjusted R^2 value of 0.28. The numbers in parenthesis are standard errors for the coefficients.

Discussion:

One purpose of this study was to calculate macroecological metrics from raw plant data and use that information to test the predictions of several statistical models of macroecology. This included using the plant census information to construct graphs for the species-area curve and species-abundance distribution. The data for the species-area relationship (slope of the straight fit line) and the species-abundance distribution (the predicted and measured abundance values) both appear to fit the Maximum Entropy Theory (Harte et. al., 2008) almost perfectly. This means that 94% of the measured data with respect to the species-abundance distribution and species-area relationship support the Maximum Entropy theory predictions.

Environmental variability was measured in order to account for the one outlier on the predicted slope versus measured slope graph to further test the SAR and the scatter

around the collector's curve. The collector's curve was used to plot the number of individuals versus the species of that individual. This helped to estimate species richness by constructing a collector's curve from the individuals found within randomly selected plots (Ugland and Grey 2003) throughout the Aspen and Conifer forests understory. The environmental factors that were tested included soil temperature, air temperature, soil moisture, slope, aspect, and canopy. From the test, only soil temperature, soil moisture, and canopy had significant effects on the values of S (total number of species), N (total number of individuals), and the ratio S/N. The most significant was with respect to S, where there is only a 0.009 chance that soil moisture does not influence S.

An underlying question when looking at the role of the environment and plant traits was what explains the value of the ratio of S/N with respect to environmental factors? Since the Maximum Entropy Theory makes all of its predictions from the values S, N, and mostly S/N, it is interesting to look at what, if any, significant effects the environmental factors have on each of these values. One explanation for why a statistical theory works well might be that environmental factors and species traits do not matter. But this overlooks the possibility, suggested by our results, that the S, N, and S/N values, which are the governing variables in Maxent, are themselves directly determined by environmental factors. It is noteworthy that the value of S/N is less influenced by environmental variables than is S and N alone, given that the predictions of MaxEnt primarily depend on S/N alone. The implications of this remain to be explored.

It appears from this study that the other two statistical theories, random placement model and fractal theory, fail to explain what is observed in nature. On the other hand, the recently advanced maximum entropy theory describes the plant abundance and

distribution patterns to almost exact certainty. The significance of the environmental factors helped to show that macroecological statistical theories are not only dependent on statistical assumptions, but also depend on the abiotic biological factors of an environment. With great accuracy, the maximum entropy model exhibits early signs of being an essential predictor of the species-area relationship and species-abundance distribution macroecological metrics. It is no wonder why this theory is being considered a unifying theory of ecology.

Works Cited

- Borda-de-Agua, Luis, S.P. Hubbell, and M McAllister. 2002. Species-Area Curves, Diversity Indices, and Species Abundance Distributions: A Multifractal Analysis. *The American Naturalist*. 159(2): 138-155.
- Coleman B.D., M. A. Mares, M. R. Willig and Y.H. Hsieh. 1982. Randomness, Area, and Species Richness. *Ecology*. 63(4): 1121-1133.
- “Ecology.” Webster’s Third New International Dictionary. Springfield: G. & C. Merriam Company, 1976. 720.
- Green J.L., J. Harte, and A. Ostling. 2003. Species richness, endemism, and abundance patterns: tests of two fractal models in a serpentine grassland. *Ecology Letters*. 6(10): 919-928.
- Halley J. M., S. Hartley, A. S. Kallimanis, W. E. Kunin, J. J. Lennon, and S. P. Sgardelis. 2004. Uses and abuses of fractal methodology in ecology. *Ecology Letters*. 7(3): 254-271.
- Harte J., T. Zillio, E. Conlisk, and A. B. Smith. 2008. Maximum Entropy and the State Variable Approach to Macroecology. *Ecology*. In Press.
- He, Fangliang, and Pierre Legendre. 1996. On Species-Area Relations. *The American Naturalist*. 148(4): 719-737.
- He, Fangliang, and Pierre Legendre. 2002. Species Diversity Patterns Derived from Species-Area Models. *Ecology*. 83(5): 1185-1198.
- Ugland, K., J. Gray, K Ellingsen. 2003. The Species-Accumulation Curve and Estimation of Species Richness. *The Journal of Animal Ecology*. 72(5): 888-897.
- Williams, M.R. 1996. Species-Area Curves: The Need to Include Zeroes. *Global Ecology and Biogeography Letters*. 5(2): 91-93.

Table 1: R^2 and b values from the thirteen graphs of predicted abundance versus measured abundance, where the measured abundance comes from the species-abundance distribution and the predicted abundance comes from the maximum entropy model λ value.

Plot #	R^2 values	Slope (b) values
1	0.9751	1.1664
2	0.9828	1.0219
3	0.9885	0.9364
4	0.9351	0.8021
5	0.9842	0.9952
6	0.9351	1.1162
7	0.9485	0.8429
8	0.8697	0.6342
9	0.8715	0.8231
10	0.955	1.1869
11	0.8398	0.6102
12	0.9885	1.0862
13	0.9979	0.9712
Average values	0.944	0.938

Figure 1: The Z value corresponds to the slope of the straight fit line from the species-area relationship (SAR). The predicted Z value is from the Maximum Entropy Theory, and the measured Z value is from the SAR graphs.

