Describing communitites part 2: Using EcoSim to build null models

[much of this text is copied from the EcoSim help manual]

EcoSim is an interactive computer program for null model analysis in community ecology. EcoSim allows you to test for community patterns with non-experimental data. EcoSim performs randomizations to create "pseudo-communities" (Pianka 1986), then statistically compares the patterns in these randomized communities with those in the real data matrix. These null model tests have wide applicability in both applied and basic ecology.

What, precisely, is a null model?
A null model is a pattern-generating model that is based on randomization of ecological data or randomly sampling from a known or imagined distribution. Null models are based on the principal of the null hypothesis –that patterns in the data do not reflect biological forces, but represent chance variation or sampling effects.

For example, Diamond (1975) hypothesized that competitive interactions lead to checkerboard distributions, in which two competing species never occur together on the same island. Connor and Simberloff (1979) tested Diamond´s hypothesis with a null model in which species were distributed randomly and independently of one another. They compared the expected number of checkerboard pairs in these simulated communities with the number observed in the data matrix.

To do this, you would create a large number (typically >1000) of randomly assembled communities. These "pseudo-communities" are generated by randomly reshuffling the observed species occurrences, subject to certain constraints. Next, a histogram would be constructed of the number of co-occurring species pairs in the pseudo-communities. This histogram tells you the range of values you might expect for the number of co-occurring pairs in an assemblage that was not structured by competition. Finally, the number of co-occurring pairs observed in the actual data matrix would be compared to the distribution of co-occurring pairs for the pseudo-communities.

For example, suppose you found in the observed matrix 680 pairs of species that formed checkerboard distributions and never co-occurred. In the simulated matrices, 985 out of 1000 times, the number of exclusive species pairs was less than 680. Therefore, by chance, the probability of obtaining 680 or more exclusive species pairs is 15/1000 = 0.015. This is less than the conventional level of statistical significance of $p = 0.05$, so we would conclude that the observed data matrix contains more exclusive species pairs than expected by chance, at least compared to the particular null model we used.

EcoSim

We must first install EcoSim if it is not already installed. (Check My Documents to see if it is.)

1. Go to http://www.garyentsminger.com/ecosim/index.htm, and click on Download EcoSim v7.72.

2. Confirm that you want to 'Open with WinZip'

3. In WinZip, click extract, and choose 'My Documents'

4. Close WinZip and go to My Documents. You should see several files with EcoSim in the name.

5. Click on the 'EcoSim700' icon. An empty species x site matrix will appear in a new window.

Co-occurrence

The co-occurrence module lets you test for non-random patterns of species co-occurrence in a presence-absence matrix. For example, suppose a pair of species in an archipelago compete with one another and never occur on the same island. Islands support one species, or the other, but not both. Diamond (1975) identified this as a checkerboard distribution, and argued that the presence of many checkerboard pairs in a community is evidence of deterministic assembly rules. Connor and Simberloff (1979) were among the first to rigorously test such patterns against a null hypothesis of random community assembly.

Step 1. Use the file menu to open the file called "West Indies finches.txt" in folder "Tutorial Datasets". This data set is a presence-absence matrix for finches (Fringillidae) of the West Indies (Gotelli and Abele 1982). These islands have been censused for over a century by many ornithologists, so the species list is probably complete. Each row is one of the 17 finch species in the West Indies, and each column is one of the 19 major islands.

Step 2. Now select co-occurrence from the 'Analyze' menu. Immediately switch to the "general" tab and set the random number seed to 10. Normally, you should use the default seed of 0, which instructs EcoSim to get a fresh random number seed from the system clock each time a new analysis is requested. In this case, by choosing a particular random number seed, your results will match up exactly with those in this tutorial.

Step 3. Notice that there are other indices and you can alter the row and column constraints. The default (fixed-fixed) is the most conservative, so we'll use it. Now click run.

Step 4. Now you've run your first null model! It should be a proud moment. The window has five tabs. The first, "Input matrix" is simply the data you entered. The second, "simulation" is one of the 5000 randomly constructed null communities. The third tab "pairwise" shows the number of checkerboard units, calculated

between all possible pairs of species. Notice that for some species pairs, such as the first two in the matrix, no checkerboard units were found, so the entry for those pairs is a zero. The C-score is calculated as the average of all the pairwise values for a matrix. The C-score is calculated for the observed matrix, and then compared statistically to the C-score values calculated from the sample of simulated matrices.

This comparison is shown in the fourth tab index. On the left, you see that the observed C-score for the finch matrix was 3.79412. In contrast, the average of the 1000 simulated matrices was 2.76281, and only one of the simulated matrices had a C-score larger than the observed. So, compared to the simulated universe of random matrices with identical row and column sums, there is much less co-occurrence in the finch matrix than expected by chance (p = 0.0002).

All of this information is shown in the summary window, which is your complete paper trail of the analysis. As always, you can edit this window as a text file so the output can be annotated. The output includes a standardized effect score of 6.21, indicating that the C-score for the observed finch matrix was over 6 standard deviations greater than the mean!

Step 5. On the 'Summary' tab, click 'Save to summary to file', and save it to the Desktop. This file shows the complete results of your null model test. Closk 'Close' to return to the main EcoSim view.

Size overlap

A seminal paper in the history of ecology is Hutchinson (1959): "Homage to Santa Rosalia, or why are there so many kinds of animals?" In this paper, Hutchinson described a visit to the shrine of Santa Rosalia in Palermo, Sicily. In the pool at the shrine, Hutchinson found three species of co-existing corixiid water beetles. He noticed that when the species were ordered from largest to smallest, the ratio of the body size of each species to the next smallest was about 1.3. He speculated that a body size ratio of 1.3 might represent a minimum size difference between animal species that was necessary to ensure coexistence. Species that are "too similar" in body size might not be able to coexist because they overlap too much in the use of shared resources. This modest suggestion spawned a vast amount of ecological research in which ecologists searched for patterns in the body sizes of coexisting species, often without an appropriate statistical analysis (Simberloff and Boecklen 1981; Wiens 1982).

This module of EcoSim allows you to test for unusual patterns in the body sizes of coexisting species, and to compare those patterns to what might be expected in a random assemblage that was not structured by competition.

Step 1. Use the file menu to open the file called "Desert rodents.txt". This data file gives the body size in grams of coexisting rodent species in the Sonoran and Great

Basin deserts. These data come from Brown (1975). Brown and his colleagues have studied competitive interactions among desert rodents (and among rodents and ants) for many years at these sites (see Brown 1998 for a summary of this work). Experimental studies have established that, in some locations, rodents compete for seed resources (Brown et al. 1986). We can now use EcoSim to see if competition is manifest in the body sizes of coexisting species.

Each row is a different rodent species, and each column is a different site (Great Basin and Sonoran deserts). Each entry in the matrix is the average body size of a particular species in a particular site. A blank indicates that a species does not occur in a particular site. EcoSim ignores these blanks. In this module, it also ignores zeros or negative numbers.

As in most analyses of body size variation, these data ignore small scale among and within populations due to factors such as clinal variation, age and size structure of populations, and sexual size dimorphism.The tests in this module only analyze the pattern of spacing of body sizes or peak flowering time. If you have quantitative data on resource use or flowering times, you should use one of Ecosim's other modules for the analysis of niche overlap.

Step 2. Understanding segment lengths - Before we can start analyzing the Brown data set, we need to understand how EcoSim uses size data to calculate patterns. For the Sonoran data set, the body sizes are:

7.2, 11.4, 17.1, 24.3, 45.3, and 120.

These are in order from smallest to largest. You do not have to enter your data this way, because EcoSim will sort them in order before it starts working. The first thing that EcoSim does it to create a set of segments from the set of body sizes. Each segment represents the difference in size between two consecutive species. Thus, for the Sonoran data set, the segments are:

4.2, 5.7, 7.2, 21.0, 74.7

The first segment is calculated as 11.4 - 7.2 = 4.2, and the last segment is calculated as 120 - 45.3 = 74.7. Because the body size data have been ordered from smallest to largest, the segments will always be non-negative numbers, but they need not increase in length. Whether the segments are large or small depends on whether two consecutive species are very similar in size (small segment) or very different in size (large segment). Also, notice that if there are n species in the community, there will be only n - 1 segment lengths created.

It is essential that you grasp the distinction between the original body sizes and these newly created segments. EcoSim will use a variety of randomization methods to create null communities in which body sizes are randomly chosen. However, the calculation of the pattern in body sizes is based on the segments, as we will explain.

Step 3. Now select size overlap from the analysis menu. Immediately switch to the "general" tab and set the random number seed to 10. Keep all of the other default values. In the Preferences tab, set Colomn to Analyze to 'Sonoran'.

The first tab in the output window is labelled "Input column", and just shows you the original data set that you analyzed. The next tab, "Simulation", shows you one of the null assemblages with randomly chosen body sizes. At first glance these numbers appear very different, but remember that the default uses a log transformation of the data. EcoSim always uses a base 10 for this transformation, but the results would be identical with any other log base.

Notice that the smallest body size is 0.85733 and the largest body size is 2.07918. If you take the anti-logs of these numbers, you get 7.2 and 120, which were the largest and smallest species in the Sonoran data set. In this simulation model, the largest and smallest species always form the fixed endpoints of the distribution.

The next tab, "Simulation Segments", shows the corresponding segments for the simulated data. These segments are calculated as the difference between the sizes (log-transformed) of consecutive species. Note that with 6 species in the Sonoran data set, 5 segments are created.

The "Size Histogram" tab shows you the distribution of simulated body sizes, which are again displayed on a log scale. The window below the histogram gives the mean (1.45155) and variance (0.12999) of these values. For each of the 1000 iterations of this model, one of the body sizes that was randomly generated was chosen and used to construct this histogram.

Because the default null model specified a (log) uniform distribution of body sizes, there are approximately equal numbers of species in each of the bins of the size histogram. It is important to appreciate that although the simulation creates this distribution of body sizes, the statistical test is based on the properties of the segment lengths, calculated from all pairs of adjacent species in each simulation.

This statistical test is illustrated in the "Size Overlap" tab. This histogram shows the distribution of the variances of segment lengths for each of the 1000 simulated communities. The variance in segment length for the Sonoran data is 0.01193, shown in the first panel. The second two panels give the low and high cut points for 12 evenly spaced bins. The final column shows the frequency of simulations in each of the 12 bins. These integer values add up to 1000, the number of iterations specified.

The observed variance of 0.01193 was smaller than 944 of the simulated variances, generating a tail probability of 0.056. The observed variance in segment lengths for the Sonoran size data is suspiciously small, suggesting a pattern of constant body size ratios in the Sonoran rodents. As always, a complete paper trail of your analysis can be found in the "Summary" tab, which can be annotated as a text window.

Rarefaction

Use the "Open" command in the "File" menu to load the file "Pitfall carabids". These data represent pitfall trap collections of carabid beetles reported by Niemelä *et al.* (1988). The traps were placed in young (< 20 years) and old (20-60 years) pine plantations in northern Europe. Each row of the data set represents a different beetle species. If you wish, use the mouse to drag on the width of the column labels so you can read the species names.

The first column shows the data for the old plantations and the second column shows the data for the young plantations. Each entry is the number of individuals collected of a particular species in the two communities. For this module, the data must be non-negative integers that represent counts of individuals. Percentages, biomass, or coverage data cannot be analyzed with the algorithms in this module.

In the young plantations, the pitfall traps yielded 243 individuals and 31 species. In the old plantations, the traps yielded only 63 individuals and 9 species. Is species richness (and other measures of diversity) really higher in the young plantations? It is difficult to say from these data.

Almost 4 times as many individuals were collected from the young plantations, so it isn´t surprising that more species were discovered. Moreover, all the species in the old plantation are a subset of the species in the young plantation. This suggests that if the old plantation were sampled more intensively, it might yield the same diversity patterns.

How can we use EcoSim to help us explore this problem?

Step 1. Use the "Species Diversity" option under the "Analyze" menu to compare the two communities.

In the General tab, set the random number seed to the value 10.

In the Preferences tab, Choose "Species richness" as the species diversity index and choose "Young_plantations" as the column to be analyzed. Choose "User-defined" for the abundance level. This will pop up an edit window, in which you should (erase whatever default values are there first) enter the value 63. You are instructing EcoSim to randomly subsample exactly 63 individuals from the young plantation data set.

Step 2. Now run the simulation, which should take only a second for 100 iterations of a single abundance level.

The Input Column tab shows the single column of original data for the young plantations. The Simulation tab shows the results of a single random draw of 63 individuals from the input column.

Although there are 48 individuals of *Calathus micropterus* in the young plantation data (first row), only 13 individuals are present in the random sample. Three

individuals of *Notiophilus biguttatus* were found in the original young plantation data (third row), but only one of these individuals were chosen in this particular random sample.

The Diversity Curve tab summarizes the simulation results. The columns give the abundance, average and median of species richness, the variance and a low and high boundary for a 95% confidence interval.

The first row (shaded in gray) gives these numbers for the entire young plantation data set. This data set had 243 individuals and 31 species. If all these individuals are randomly sampled, the mean and the median will always be the same, and the variance will always be zero.

The next row gives the results for the abundance level that you specified in the edit dialog box. For random samples of 63 individuals, there was an average of 20.03 species represented, with a median of 20 species and a variance of 3.93785.

The last two columns give us a confidence interval that will allow us to answer the question of which of the two assemblages is most diverse. The confidence interval is from 16 to 24 species. In other words, 95% of the time that a random sample of 63 individuals is drawn from the young plantation assemblage, we expect to find between approximately 16 and 24 species.

However, remembering back to the original data, the 63 individuals collected from the old plantation represent only 9 species. We can conclude that species richness is substantially higher for the young plantation, even after adjusting for sampling differences.

The Summary tab displays all the options that you chose for your simulation. You can save your diversity curve and summary results by clicking the "Save summary to file" or "Save diversity curve" buttons.

<u>Step 3.</u> Close the output window and rerun this simulation again. This time, however, use the abundance level = Default, rather than the user-defined levels.

When you look at the "Diversity Curve" tab, you will now see that there are 33 rows of abundance levels, evenly spaced between a minimum of 1 and a maximum of 243 individuals. Notice that at these extremes, the variance of the simulations is zero. Because only 1 individual is drawn, only 1 species will be represented.

Conversely, if all 243 individuals are drawn, exactly 31 species will always be represented. Between those extremes, the number of species will vary from one run to the next for a particular abundance level, and this variability is reflected in the variance and confidence intervals.

This data could be used to plot a rarefaction curve, showing how the species richness in each random sample changes as the number of individuals in the samples increases.

You can also use EcoSim to compare diversity with other indices besides species richness. They can be selected from the "Species diversity index" box when you are selecting simulation options.