

Learning To Measure Biodiversity: Two Agent-Based Models That Simulate Sampling Methods & Provide Data for Calculating Diversity Indices

• THOMAS JONES, THOMAS LAUGHLIN



There is certainly no substitute for experiencing biodiversity first-hand. Nothing could be more effective than a wilderness experience to demonstrate the importance of conserving biodiversity. That said, computer models have several features that are helpful in understanding how biodiversity is measured. For one, they are easily used when natural resources, transportation, and time constraints limit the use of field laboratories. Moreover, computer models can generate realistic data, in large quantities, that can be used to calculate different indices of diversity. Here we provide background on concepts and measures of biodiversity, and present two agent-based simulation models that can be accessed from the Internet. One model simulates stream sampling for invertebrate biodiversity, while the other simulates sampling for plant diversity. The stream sampling model is simpler, and effective as a demonstration as well as for generating data for analysis. It is designed to be used at the high school and college levels. The plant sampling model is more advanced, and is useful in teaching experimental design and to generate data for advanced analyses. It is appropriate for the college level, or AP biology.

○ Background on Biodiversity

It is common to hear ecologists talking about biodiversity, particularly in terms of conservation biology. Biodiversity is a frequently used term, and is considered to refer to something important, yet, it is often unclear exactly what is being discussed. In fact, “biodiversity” can refer to many things, and has many specific definitions.

The most common definition of biodiversity, sometimes called species diversity, refers to the number of different species in a given area. The greatest biodiversity by this measure would be the number of different species found in tropical rainforests, and is estimated to be in the millions. Biologists sometimes refer to the diversity within a particular group or taxon. For example, the southern Appalachians are a world “hotspot” for salamander diversity. Several indices of species diversity are discussed below.

Biodiversity, in the broad sense, can also refer to variation within species, or among populations. Many species have populations which can be differentiated by gene frequencies, morphology, or behavior. Typically this occurs within species having large ranges. Consider, for example, the extraordinary differences among human populations across the globe. This variation reflects underlying

differences in allele proportions among populations and is called genetic diversity. Genetic diversity is of special concern regarding endangered species because small populations tend to lose genetic diversity rapidly through random genetic drift or bottlenecks. Without genetic diversity, populations lose their ability to adapt to changing environments. An extreme example of a species with low genetic diversity is the cheetah. Cheetahs are so genetically similar that many can accept skin grafts from unrelated individuals without tissue rejection (O’Brien et al., 1985).

On a larger scale, we can consider ecosystem diversity. In this case, we are not considering individual species, but rather a species assemblage in a particular habitat. Ecosystem diversity is a broad concept, encompassing levels of ecological organization above species (e.g., habitat, community, or ecosystem). An example of a major threat to ecosystem diversity in the U.S. is the loss of wetlands to development. It is difficult to quantify ecosystem diversity, because the edges of habitat patches and communities are difficult to define. However, it can be argued that the most natural way to preserve all levels of biodiversity is to protect as much and as varied habitat as possible, and then let nature take care of the rest.

○ Measures of Biodiversity

Though biodiversity is a conceptually broad subject, there are numerous quantitative indices of biodiversity. Each of these has strengths and limitations, and is used in different situations. Several different indices are typically reported together to give a more complete description of a situation.

- **Abundance.** This is not actually a measure of biodiversity per se, rather it is a count (or estimate) of all individuals present. Usually abundance refers to a certain species, but total abundance refers to all the individuals of all species present.
- **Species richness.** This is the simplest measure of biodiversity, being the tally of all species found in the area or community in question. While potentially easy to calculate, this measure is limited in that it lacks information on the relative abundance of the various species.
- **Community dominance index.** This measure reflects how large a proportion of the total species present (in terms of

Biodiversity is a frequently-used term, and is considered to refer to something important, yet, it is often unclear exactly what is being discussed.

numbers of individuals) is made up of the two most abundant species. It is calculated by the formula:

$$100 \frac{y_1 + y_2}{y}$$

Where y_1 and y_2 are the abundances of the two most common species in the sample and y is the total abundance of the sample.

- **Simpson diversity index.** Simpson (1949) developed a measure which accounts for the proportions of all species in the sample. Unlike the community dominance index, this calculates the value from all the species present, not just the two most abundant. It is calculated by the formula:

$$D = \sum_{i=1}^S p_i^2$$

Where p_i is the proportional abundance of the i th species (abundance of i / total abundance) and S is the species richness. In this case, a high value indicates that there is low diversity in the sample.

- **Shannon-Weiner index.** This index (Shannon, 1948) measures how evenly distributed the numbers of the species present in a sample are on a logarithmic scale. It is calculated by the formula:

$$H' = -\sum_{i=1}^S p_i \ln p_i$$

Where p_i is the proportional abundance of the i th species (abundance of i / total abundance) and S is the species richness. The higher the index, the more evenly distributed the sample is. The index has a maximum value when all species are in equal proportions.

- **Species discovery curve.** When sampling for biodiversity, how does one know when all species have been found? Well, in short, one doesn't. However, one can get an idea of how many more one *could* find by plotting the cumulative number of species found against some standardized measure of sampling effort (Figure 1). The function will be a curve approaching an asymptote of the actual number

of species present (Fisher et al., 1943). The reason it is curved is not (necessarily) because the biologists are getting tired of searching. Rather, it reflects the fact that, as the number of species found accumulates, there is a diminishing probability that the next individual found will be of a new species. The curve is also a function of the fact that the most common species are usually found first, and the rare species are more likely to be missed. From the species discovery curve, one can estimate how much more effort it will take to find new species. It allows one to know when most of the species have been found ... and decide when it is time to give up.

- **Jaccard similarity index.** This measure is used to compare the species assemblages of two areas or habitats (Jaccard, 1901). It is used to define how similar two samples are in terms of species present. An example of when it would be useful is in deciding on which watersheds to focus on for preserving the most biodiversity. If one could only protect two of several, one would choose the two with the lowest similarity index. The index is calculated as the number of

$$J(A, B) = \frac{A \cap B}{A \cup B}$$

species two samples have in common, divided by the total number of species represented in both samples.

○ Using Computer Models To Teach Biodiversity

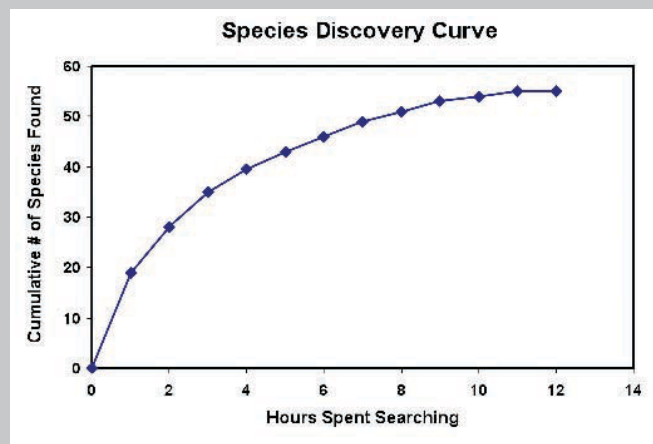
As a rule, it is practically impossible to know how many species, and how many individuals of each, exist in a given area. For this reason, students of biodiversity must understand the concept of sampling. Computer simulations allow different sampling strategies to be tested in class and, thus, facilitate inquiry-based learning (Latham & Scully, 2008). For example, it is easy to compare the results of focusing all of the sampling effort in one area vs. running transects across a large area. These simulations also allow a student to experience the relationship of sampling efforts to cumulative species discovered.

The models presented here were developed in the Netlogo programming environment (Wilensky, 1999), and run as Web-based applets on the Virtual Ecology Web site. The models presented here are similar in general appearance. There is a world-view window that provides a real time look at what is happening in the simulation. There are a variety of sliders, switches, and buttons to set parameters and control the experiment. Finally, there are monitors and graphs that provide results. Students will likely master the interface very quickly, perhaps without even reading instructions. In fact, not giving instructions may be one approach to teaching with these models that would encourage inquiry and exploration.

○ Estimating Stream Diversity: Model Details

This model is loosely based on the water quality monitoring procedure of the Save Our Streams project of the Izaak Walton League of America (2006). The underlying premise is that water quality affects the biodiversity of a stream. Thus, surveying the macroinvertebrates in a stream is a way to indicate water quality. When the model opens, the world view will show a section of stream, with 16 sample buckets below (Figure 2.). Various controls (Table 1) and data monitors (Table 2) surround the world

Figure 1. The species discovery curve illustrates the diminishing returns relationship between searching/sampling effort and the likelihood of new species being found. As more species are discovered, it requires more and more effort to find the remaining few.



view. When “Run” is clicked, the stream will begin to flow, and a variety of invertebrates (Table 3) begin to flow downstream (as if stirred up by an unseen researcher upstream). The appearance of animals in the stream is stochastic both in how many, and in what species are in view at any given time. When “Open Seine” is clicked, a net will appear in the middle of the stream and remain open for the amount of time set on the “Sampling-Time” slider, or until “Close Seine” is clicked. Animals who get caught in the seine are immediately sorted into their appropriate bins which are labeled in their corresponding count monitors below. In actual stream sampling, you would sort through the net after a round of seining, but during the simulation, students can observe the progression of the species discovery curve (Figure 1) shown in the “Cumulative Species vs. Time” plot (Figure 2).

The “Reset” button will clear all monitors and the plot. Students can replicate rounds of seining and overlay curves on the plot by clicking “Release” (which lets all the captured animals go downstream), and then re-opening the seine. Clicking “Clear Plot” will erase all the traces on the plot. Along with the plot, the model continuously monitors the counts of each species caught (individual species abundance), the total number of individuals caught (total abundance), and the number of different species caught (species richness).

The primary variable of which students will explore the effect is the level of pollution and its effect in the stream. This variable can be set to none, moderate, or severe, from the “Pollution” drop-down menu, and changes the probabilities of each animal appearing in the stream based on their sensitivities to pollution (Figure 3). When “Pollution” is set to none, all species have similar probabilities of appearing. In the moderately-polluted stream, the pollution-sensitive species have very low probabilities of appearing, the less-sensitive species have intermediate probabilities, and the tolerant species have higher probabilities. When pollution is severe, sensitive species will not appear, less-sensitive species have a low probability, and most of the species captured will be pollution tolerant.

Figure 2. The Interface of the Estimating Stream Diversity Simulation.

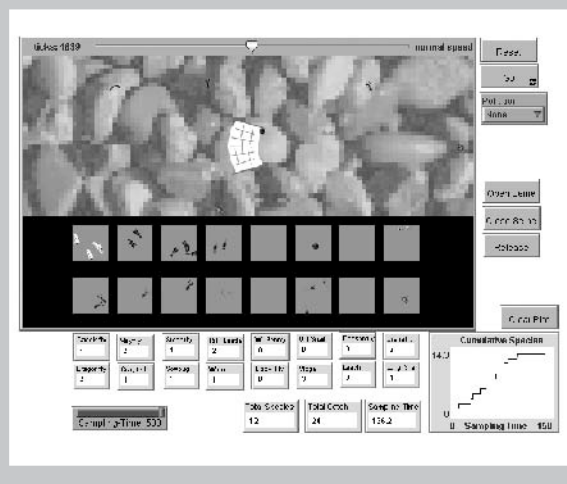


Figure 3. The Interface of the Estimating Plant Diversity Simulation.

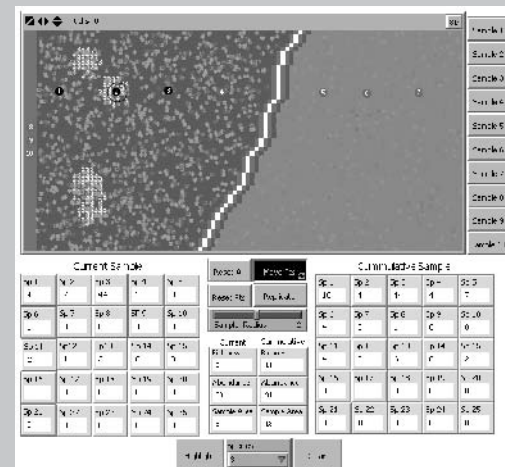


Table 1. Controls for the Estimating Stream Diversity Simulation.

Reset	Clears the graph, sets time to 0, removes the seine, releases animals.
Go	Sets the stream in motion with animals drifting down.
Open Seine	Places the seine in the middle of the stream, starts sampling & plotting data.
Close Seine	Removes the seine, stops plotting data.
Release	Dumps captured animals downstream, clears graph.
Speed	The slider above the world view controls the speed of the simulation.

Table 2. Reporters for the Estimating Stream Diversity Simulation.

Individual Species Counts	The numbers of each species caught (row below world view)
Total Catch	Number of all individuals caught
Total Species	Number of different species represented in sample
Line Graph	Cumulative species vs. Sampling time

○ Estimating Plant Diversity: Model Details

This model simulates plants growing in two habitats separated into a southeast-facing xeric slope from a northwest-facing mesic slope (Figure 3). To the left (northwest) of the ridge, the background is green, symbolizing mesic habitat; and to the right (southeast) of the ridge, it is a yellow-brown, simulating xeric habitat. There are 25 different plant species in the entire simulation that all appear similar, except for subtle differences in color. These species represent wildflowers common to the southern Appalachians (Horn & Cathcart, 2005), with some being exclusive to the mesic habitat, others to the xeric, and some which can be found in both (Table 4). In both habitats, some of the species are common while some are rare. Individual species distributions are either clumped or random. The total abundance of plants in the xeric habitat is about half that of the mesic habitat. There are a variety of controls (Table 5) and banks of monitors (Table 6) beside and below the world view.

When the model is initially opened, or if “Reset All” is clicked, ten numbered red circles appear in a sidebar on the left of the

world view. These are the sampling points. When “Move Pts” is clicked (and until it is un-clicked), the sample points can be dragged around the world view with the computer mouse. If a point is placed in the habitat, it can be sampled, using a column of buttons labeled “Sample (1-10)” to the right of the world view. When a sample button is clicked, the corresponding point will turn black and a black ring will appear around it that roughly corresponds to the sampling radius (set by the “Sample Radius” slider). The counts of each species in the current sample are given in the lefthand bank of monitors. Points which have been sampled will remain black, with the black ring denoting which sample is being displayed as the current sample. The righthand bank of 25 monitors displays the cumulative counts of each species among all the points sampled. Between the two monitor banks, two columns of three monitors give the species richness, total abundance, and sample area for the current and cumulative samples.

Other features of this model are designed to allow students to experiment with various sampling schemes. The “Reset Pts” button puts the sample points back on the sidebar without clearing the cumulative data. This allows further sampling of the current world. The “Replicate” button leaves the sample points in place, but clears all the data, and resets the plants in the world. This allows students to quickly replicate sampling schemes for robust comparisons of different communities. Finally, students can check to see where individuals of a species are by choosing a species from the “Species” drop-down menu and clicking “Highlight.” For example, in the world view of Figure 3, Species 3 is highlighted. The sample point 2 is located in the middle of a clump of Species 3, and the abundance of Species 3 in the “Current Sample” monitor is correspondingly high. The ability to highlight specific species allows students to see how different sampling schemes might miss entire species.

○ Using Diversity Models in the Classroom

Both of these models provide a resource for inquiry-based exploration of biodiversity. In general, they are designed to be run like a real experiment in which results are recorded separately for analysis. Students typically have a spreadsheet program open

Table 3. Cast of characters for Stream Diversity Simulation.

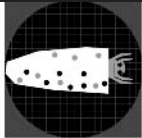

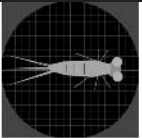


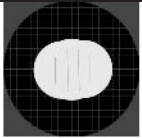
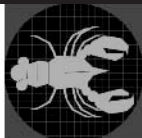
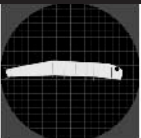
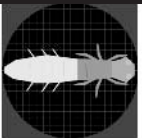
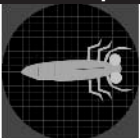
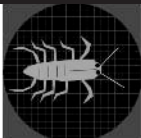
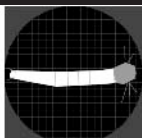


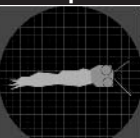
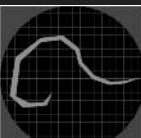
Pollution Sensitive Species					
					
Caddisfly	Gilled Snail	Mayfly	Riffle Beetle	Stonefly	Water Penny
Pollution Less-Sensitive Species					
					
Crayfish	Cranefly	Dobsonfly	Dragonfly	Sowbug	
Pollution Tolerant Species					
					
Blackfly	Leech	Lunged Snail	Midge	Worm	

Table 4. Checklist of Plant Species in the Estimating Plant Diversity Simulation.

Sp 1	Bloodroot	Mesic	Common	Random
Sp 2	Spring Beauty	Mesic	Common	Random
Sp 3	Jack-in-the-Pulpit	Mesic	Common	Clumped
Sp 4	Partridgeberry	Mesic	Common	Random
Sp 5	Dwarf Crested Iris	Mesic	Common	Random
Sp 6	American Gensing	Mesic	Occasional	Random
Sp 7	Showy Orchis	Mesic	Occasional	Clumped
Sp 8	Yellow Trout lilly	Mesic	Occasional	Random
Sp 9	Canada mayflower	Mesic	Rare	Random
Sp 10	Red Trillium	Mesic	Rare	Clumped
Sp 11	White Wood Aster	Xeric	Common	Random
Sp 12	Pipsissewa	Xeric	Common	Random
Sp 13	Fire Pink	Xeric	Common	Random
Sp 14	Common Cinquefoil	Xeric	Common	Random
Sp 15	Roundleaf Ragwort	Xeric	Common	Random
Sp 16	White Milkweed	Xeric	Occasional	Random
Sp 17	Wild Columbine	Xeric	Occasional	Clumped
Sp 18	Woodland Sunflower	Xeric	Occasional	Clumped
Sp 19	Early Buttercup	Xeric	Rare	Random
Sp 20	Whorled Coreopsis	Xeric	Rare	Clumped
Sp 21	Wild Geranium	Both	Common	Random
Sp 22	Rose-pink	Both	Common	Random
Sp 23	Galax	Both	Occasional	Random
Sp 24	Pink Lady's Slipper	Both	Occasional	Clumped
Sp 25	Maryland Golden Aster	Both	Rare	Random

Table 5. Controls for the Estimating Plant Diversity Simulation.

Reset All	Clears all reporters, resets/returns sample points, recreates the population.
Move Pts	Enables the user to move sample points around the habitat.
Reset Pts	Enables points to be re-sampled.
Replicate	Enables re-sampling, recreates the population, clears all reporters.
Sample Radius	Adjusts the size of the area that will be sampled around each point.
Sample (1-10)	Counts the numbers of each species around the identified point.
Highlight	Identifies individual plants with their species number.
Species	Drop-down menu to select which species will be highlighted.
Clear	Removes all of the identifiers.

Table 6. Reporters for the Estimating Plant Diversity Simulation.

Current Sample (Sp1-Sp25)	Counts for each species in the highlighted sample
Cumulative Sample (Sp1-Sp25)	Counts for each species in the total sampled area
Current (3 stacked reporters)	Species richness, total abundance, and sampled area for the current sample
Cumulative (3 stacked reporters)	Species richness, total abundance, and sampled area for the cumulative sample

along with the running model. The models are stochastic in the data they generate, which is biologically realistic. Finding trends among the variation in data is an important component of science (especially biology). With these models, students can generate enough data to see trends, perform statistical analyses, and get significant results. Stochastic models have the added advantage that no two runs are identical, helping ensure that students do their own work.

The Estimating Stream Diversity model is useful as an in-class demonstration or as a virtual laboratory. It generates data that can be used to calculate various diversity indices. It is also useful in illustrating how pollution affects stream diversity, and introduces students to some of the invertebrates they might find in local streams. We use this model as a pre-lab assignment prior to an actual stream sampling field trip. In this way, students are sure to be familiar with the general lab procedure, and we are able to make more efficient use of time in the field.

The Estimating Plant Diversity model is designed for deeper exploration of experimental design and data analysis. While not flashy in its interface, the model generates large quantities of data, which would be impractical to gather in the field with a laboratory class. We use this model as a follow-up lab to the stream sampling. Students are asked to compare the efficacy of various sampling schemes (e.g., focusing intense effort on a localized area or taking small samples across a large area). They are asked to discuss what schemes would be the most useful if time and resources were limited. Along with calculating various diversity indices for each habitat, the students are asked to compare the habitats (abundances, species present, and the Jaccard similarity index).

○ Assessment & Conclusions

We have found that the Estimating Stream Diversity and Estimating Plant Diversity simulations provide students with an easily-accessible environment in which to learn how to sample and ana-

lyze biodiversity. The stream diversity model can serve as a good primer for actual field experiments or to generate simple data. With the plant diversity model, students can devise their own experiments with replication and statistically analyze the results either individually or as group exercises. In conclusion, these agent-based simulations, while no substitute for actual field experience, can provide an opportunity to delve deeper into subject areas and methods of data analysis.

○ Availability

The model can be run from the Virtual Ecology Web site under the Biodiversity menu located at: <http://faculty.etsu.edu/jonestc/Virtualecology.htm>

The model can be run in any browser with the Java 1.4.1 (or higher) plug-in.

○ Acknowledgments

Thanks to U. Wilensky and all who created NetLogo. We also thank Josh Rudd and Arash Mehraban for assisting in classroom implementation of the model. This work was created as part of the SYMBIOSIS Project, funded by a grant from the Howard Hughes Medical Institute (Grant # 52005872).•

References

- Fisher, R.A., Corbet, A.S. & Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, 12, 42-58
- Horn, D. & Cathcart, T. (2005). *Wildflowers of Tennessee the Ohio Valley and The Southern Appalachians*. Auburn, WA: Lone Pine Publishing.
- Izaak Walton League of America. (2006). *A Handbook for Stream Enhancement & Stewardship*. Blacksburg, VA: The McDonald & Woodward Publishing Company.
- Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin del la Société Vaudoise des Sciences Naturelles*, 37, 547-579.
- Latham, L.G. & Scully, E.P. (2008). CRITTERS!: A realistic simulation for teaching evolutionary biology. *The American Biology Teacher*, 70, 30-33.
- O'Brien, S.J., Roelke, M.E., Marker, L., Newman, A., Winkler, C.A., Meltzer, D., Colly, L., Evermann, J.F., Bush, M. & Wildt, D.E. (1985). Genetic basis for vulnerability in the cheetah. *Science*, 227, 1428-1434.
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423.
- Simpson, E.H. (1949). Measurement of diversity. *Nature*, 163, 688.
- Wilensky, U. (1999). NetLogo. <http://ccl.northwestern.edu/netlogo/>. Center for Connected Learning and Computer-Based Modeling. Northwestern University, Evanston, IL.

BIO

THOMAS JONES (jonestc@etsu.edu) and THOMAS LAUGHLIN (Laughlin@etsu.edu) are faculty in the Department of Biological Sciences at East Tennessee State University, Johnson City, TN 37614.