A Comparison of Three Diversity Indices Based on Their Components of Richness and Evenness

Author(s): T. M. DeJong

Source: *Oikos*, Vol. 26, No. 2 (1975), pp. 222-227
Published by: Wiley on behalf of Nordic Society Oikos
Stable URL: http://www.jstor.org/stable/3543712
Accessed: 01-05-2017 17:02 UTC

OIKOS 26: 222–227. Copenhagen 1975

A comparison of three diversity indices based on their components of richness and evenness

T. M. DeJong

Department of Biological Sciences, California State University, Fullerton

DeJong, T. M. 1975. A comparison of three diversity indices based on their components of richness and evenness. – Oikos 26: 222–227.

Species diversity was calculated using three different indices on sets of real and artificial data. Each index was analyzed to determine its relationship to the two component parts of diversity, richness and evenness. Shannon's information formula, $H' = C\Sigma p_i \log 2\, p_i$, is found to be linearly related to evenness and to the $\log_2$ of the number of species. Simpson's Index, $D = 1 - \sum \frac{n(n-1)}{N(N-1)}$ and McIntosh's Index, $D' = \frac{N - \sqrt{\Sigma n_i^2}}{N - \sqrt{N}}$, when expressed in probits are found to be linearly related to evenness and to the $\log_2$ of the number of species. Relationships between, and usefulness of, the indices are discussed.

T. M. DeJong, Botany Dept, Univ. of California, Davis, California 95616, USA.

Рассчитено видовое разнообразие с помощью трех различных индексов по сериям реальных и моделированных данных. Каждый индекс проанализирован для определения его зависимости от двух компонент разнообразия, богатства и выровненности. Инфорационная формула Шэннона $H' = C\Sigma p_i \log 2$ p_i находится в линейной зависимости от выровненности и $\log_2$ от общего количества видов. Индекс Симпсона $D = 1 - \sum \frac{n(n-1)}{N(N-1)}$ и индекс Макинтоша $D' = \frac{N - \sqrt{\Sigma n_i^2}}{N - \sqrt{N}}$ выраженныеı в битах, находятся в линейной зависимости от выровненности и $\log_2$ от числа видов. Обсуждаются отношения между индексами и возможности их применения.

Manuscript accepted 12 November 1974

1. Introduction

Species diversity is one of the basic concepts of ecology that has been used to characterize communities and ecosystems. At first glance the concept appears to be rather simple but ecologists and mathematicians have been searching for ways to express the various aspects of diversity since 1922 (Gleason 1922) even though the term did not appear in the literature until 1943 (Fisher et al. 1943). Consequently, the concept of species diversity has been defined in many ways, and several different indices have been developed to express it.

The objective of this paper is to compare three diversity indices in their sensitivity to the two aspects of diversity: species richness and species evenness. Species richness is usually thought of as the number of species per sample. Species evenness (equitability) is a parameter which indicates relative abundances of the various species in a sample. Species evenness increases as species are more evenly distributed in a sample such that maximum evenness is obtained when all the species are equally abundant. Species diversity increases as the number of species per sample increases and as the abundances of species within a sample become more even (Pielou 1969, Kricher 1972). The diversity indices chosen are those of Simpson, Shannon, and McIntosh.

2. Diversity indices

Simpson (1949) introduced an index of diversity which is a measure of concentration of species. Its numerical values increase as diversity decreases (Risser and Rice 1971). A common variation of Simpson's index yields values on a probability scale from 0.0 to 1.0 in ascending order with increased diversity

$$D = 1 - \frac{\sum_{i=1}^{S} n_i(n_i - 1)}{N(N-1)} \qquad (1)$$

(Pielou 1969). In this equation S = the number of species, n_i = the number of individuals belonging to the ith species, and N = the total number of individuals in the sample. Eqn (1) represents the probability that two individuals, picked independently and at random from a population, will belong to different species. Hurlbert (1971) renamed this index the "probability of interspecific encounter".

The most common remark made about the Simpson index as a measure of diversity is that it is too strongly affected by the abundance of the two or three most abundant species in a community (Williams 1964, Sanders 1968, Risser and Rice 1971, Whittaker 1972). Loya (1972) plotted Simpson's diversity values against the number of species (S) and observed that Simpson's index reached its maximum values after the first 10 to 12 species were encountered. Loya said that the index is "... insensitive to the relative contribution of the rare species encountered along a transect". He concluded that it was mainly an evenness index like Shannon's (H'/H'_{max}) evenness index.

The most frequently used index of species diversity is a derivative of Shannon's Information Theory of Communication (Shannon and Weaver 1949). In this index

$$H' = C \sum_{i=1}^{S} p_i \log p_i \qquad (2)$$

C is a constant and p_i can be estimated by n_i/N. Logarithms to the base 2, e, and 10 can be used in the equation and the information units obtained are called "binary digits" or "bits"; "natural bels" or "nats"; and "bels", "decimal digits" or "decits", respectively. H'is dimensionless and is a measure of uncertainty. If an individual is picked at random from an infinite population, H' is a measure of how uncertain one is that the individual picked will be of a particular species. H' is therefore thought to be an intuitive measure of diversity since uncertainty will increase as species diversity in a population increases (Pielou 1969).

Loya (1972) discovered a plateau-effect similar to that with Simpson's index when he plotted diversity according to Shannon's index against S. He explained the plateau by saying that even though additional species should tend to increase the diversity values, the addition of these species changes the relative abundances of the species. Thus, the added species would tend to decrease the index of diversity. Monk (1967) and Kricher (1972) came to essentially the same conclusion as Loya when they found a curvilinear correlation between H' and S. They thought that the plateau-effect was a result of decreased evenness since added species tend to be rare.

Loya and Monk both postulated that the richness component of Shannon's diversity formula was the "most significant" factor in determining the diversity values when a relatively small number of species is involved. Loya contended that beyond this point, "the relative significance of the evenness component (H'/H'_{max}) increases to counteract the positive effect of additional species until a plateau is observed...".

The third diversity index is that of McIntosh (1967)

$$D' = \frac{N - \sqrt{\sum_{i=1}^{S} n_i^2}}{N - \sqrt{N}} \qquad (3)$$

which is independent of sample size and yields values which are a percentage of the maximum possible diversity for a sample of the same size.

3. Evenness indices

The indices of eveness used in this study are based on a general formula suggested for all diversity-related evenness indices by Fager (1972). In this formula

$$\text{Evenness} = \frac{\text{calculated diversity} - \text{minimum diversity}}{\text{maximum diversity} - \text{minimum diversity}} \quad (4)$$

minimum and maximum diversity are calculated for a sample of the same number of species (S) and individuals (N).

Following Fager's general formula, the eveness "corollary" for Simpson's diversity index is

$$SE = \frac{\left[1 - \frac{\sum_{i=1}^{S} n_i(n_i - 1)}{N(N-1)}\right] - \left[\frac{(S-1)(2N-S)}{N(N-1)}\right]}{\left[\frac{N(S-1)}{S(N-1)}\right] - \left[\frac{(S-1)(2N-S)}{N(N-1}\right]} \quad (5)$$

(Fager 1972). In a similar manner, the eveness corollary for Shannon's formula is

$$J^* = \frac{\left[C \sum_{i=1}^{S} p_i \log p_i\right] - \left[\log N - \frac{N-(S+1)}{N} \log (N-(S+1))\right]}{\log S - \left[\log N - \frac{N-(S+1)}{N} \log (N-(S+1))\right]} \quad (6)$$

Fager 1972). This evenness index is slightly different from the index most commonly used in connection with Shannon's diversity index. Note the difference between Eqn (7)

$$J' = \frac{H'}{\log S} \quad (7)$$

(Pielou 1969), and Eqn (6). Eqn (6) was used in this study so that it would be comparable to the Simpson corollary of evenness.

Eqn (8) is the evenness corollary to McIntosh's diversity index

$$ME = \frac{\frac{N - \sqrt{\sum_{i=1}^{S} n_i^2}}{N - \sqrt{N}} - \frac{N - \sqrt{(S-1) + (N-(S-1))^2}}{N - \sqrt{N}}}{\frac{N - \sqrt{S\left[\frac{N}{S}\right]^2}}{N - \sqrt{N}} - \frac{N - \sqrt{(S-1) + (N-(S-1))^2}}{N - \sqrt{N}}} \quad (8)$$

This index differs from Eqn (9) which was suggested by Pielou (1969)

$$\frac{\triangle}{\max(\triangle/N, S)} = \frac{N - \sqrt{\sum_{i=1}^{S} n_i^2}}{N - N/\sqrt{S}} \quad (9)$$

but Eqn (8) was used so that it would be comparable to the Simpson corollary of evenness.

4. Methods and results

To analyze Simpson's, Shannon's, and McIntosh's diversity indices and their relationship to richness and evenness, it was necessary to use community data based on large sample sizes and having a broad range of species richness. The primary data used in this analysis were taken from the data of Whittaker (1960), collected on six moisture gradient transects at six different elevational belts in the Siskiyou Mountains. Species richness

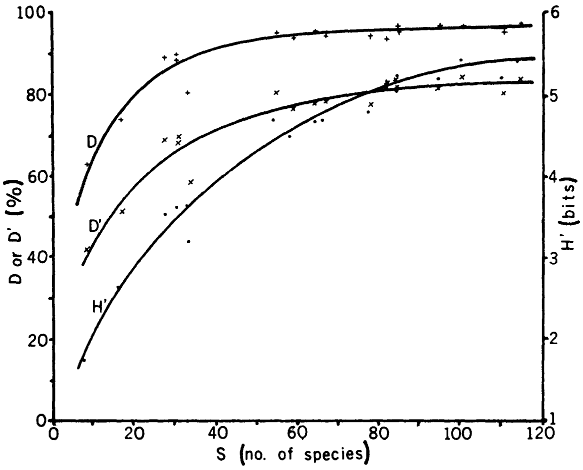

Fig. 1. Relationship between three indices of diversity (Simpson's D, McIntosh's D′, Shannon's H′) and species richness (S) in vegetation of the Siskiyou Mountains. (Data from Whittaker 1960.)

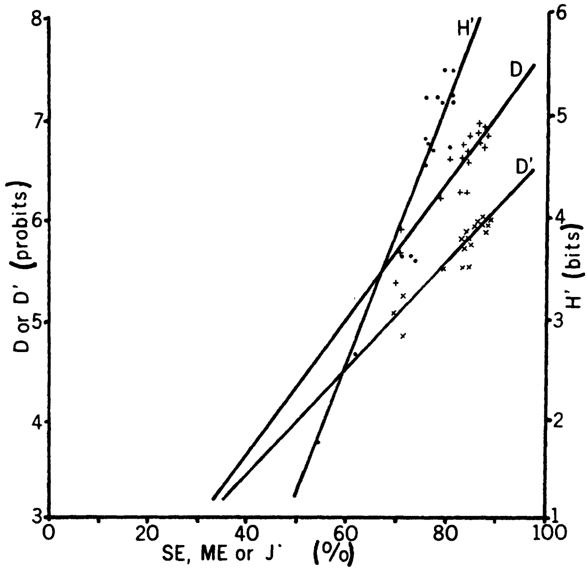

Fig. 3. Linear relationships between three indices of diversity (D, D′, H′) and their respective evenness corollaries (SE, ME, J*) in the vegetation of the Siskiyou Mountains. (Data from Whittaker 1960.)
Simpson's (D) regression: $y = 6.598x + 0.972$, $r = 0.929$
McIntosh's (D′) regression: $y = 5.235x + 1.283$, $r = 0.937$
Shannon's (H′) regression: $y = 13.039x - 5.316$, $r = 0.937$

ranged from 8 species of shrubs and seedlings at one elevation to 114 shrubs, seedlings, and herbs at another elevation. The one drawback in using Whittaker's data was that some species were noted as present but not sampled. These species were treated as if they represented one occurrance of the species on the transect.

Each of the three diversity indices was used to calculate species diversity for (1) shrubs and seedlings, (2) herbs, and (3) shrubs, seedlings, and herbs, found in each of the six elevational transects. The results of these diversity calculations were plotted against species richness, (S), (Fig. 1).

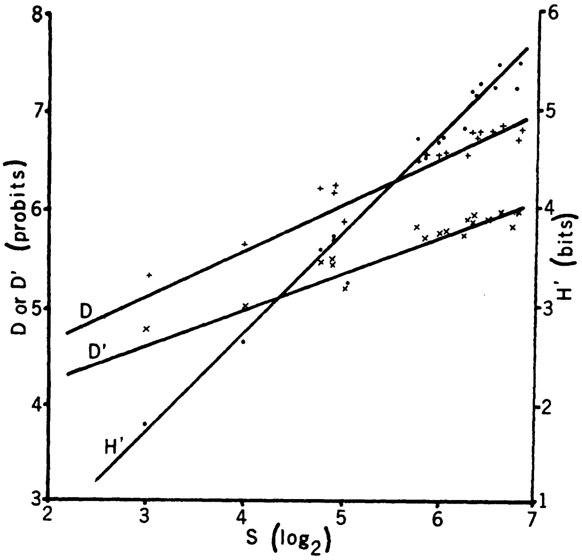

Fig. 2. Linear relationships between three indices of diversity (D, D′, H′) and species richness ($\log_2 S$) in vegetation of the Siskiyou Mountains. (Data from Whittaker 1960.)
Simpson's (D) regression: $y = 0.407x + 4.124$, $r = 0.960$
McIntosh's (D′) regression: $y = 0.314x + 3.857$, $r = 0.956$
Shannon's (H′) regression: $y = 1.001x - 1.313$, $r = 0.986$

Note the curvilinear correlation between the number of species and each diversity index. Transformations of these same data showed that if Shannon's diversity values were plotted against $\log_2$ of species richness, the plot was linear (Fig. 2).

Simpson's index expresses the probability of interspecific encounter and its values are on a non-linear probability scale. Using a table provided by Fisher and Yates (1963), Simpson's diversity values were transformed into probits and plotted against the $\log_2$ of species richness. Probits are normal equivalent deviates coded by the addition of 5.0. Normal equivalent deviates are based on standard deviations which correspond to a cumulative percentage (Sokal and Rohlf 1969). The resulting plot showed a linear correlation (Fig. 2).

McIntosh's index expresses diversity as a percentage of the maximum possible diversity for samples of the same size. Its diversity values were also converted to probits and plotted against the $\log_2$ of species richness. A linear relationship resulted (Fig. 2).

Thus, all three indices can be linearly related to species richness, providing they are expressed in the proper units. The plateau effect discussed by others (e.g. Loya 1972), therefore, is a mathematical, rather than ecological, phenomenon.

Linear plots of index value vs. species evenness can also be obtained (Fig. 3). Species evenness was calculated using the corollary evenness indices that were formulated for each of the three diversity indices, Eqns (5), (6) and (8). However, when using the Simpson and McIntosh evenness corollaries, the respective diversity

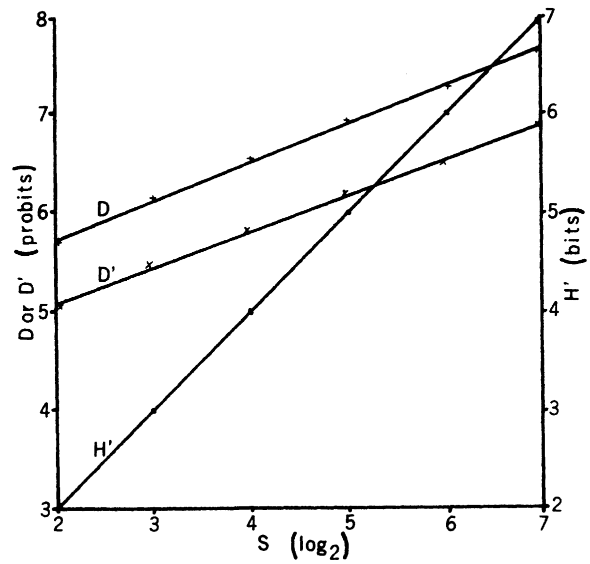

Fig. 4. Relationship between each of the three indices of diversity and species richness expressed as $\log_2 S$ in contrived test data which had maximum evenness in all samples. Simpson's (D) regression: $y = 0.386x + 4.971$, $r = 0.998$ McIntosh's (D′) regression: $y = 0.358x + 4.392$, $r = 0.999$ Shannon's (H′) regression: $y = x$, $r = 1.000$

values were converted to probits, since taking a percent of a percent probability is misleading and yields values which are skewed toward maximum values. Species diversity expressed by H′, D in probits, and D′ in probits was plotted against species evenness as expressed by their respective evenness indices.

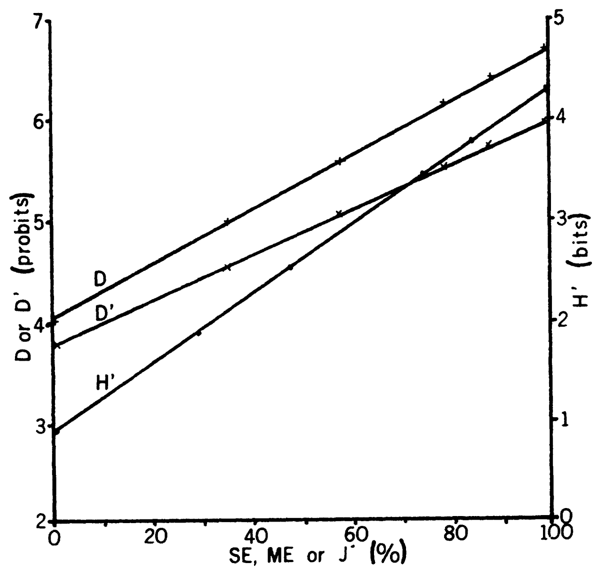

Fig. 5. Relationship between each of the three diversity indices and eveness in data contrived such that richness was the same in all of the samples.
Simpson's (D) regression: $y = 2.603x + 4.090$, $r = 1.000$
McIntosh's (D′) regression: $y = 2.247x + 3.730$, $r = 1.000$
Shannon's (H′) regression: $y = 3.465x + 0.857$, $r = 1.000$

Thus, linear relationships exist between the diversity indices and species richness as well as species evenness. However, the data used were taken from natural samples of vegetation in which richness and evenness are not independent; therefore, the relative contributions of species richness or evenness to each of the index values remains to be determined.

Sets of artificial data were used to test the hypothesis that the diversity indices are in fact linearly related to species richness and evenness in the manner shown above. The first set of test data was designed to show the effect of species richness on each of the diversity indices. The test was based on six samples each having 200 individuals and maximum evenness throughout. Within a sample, all the species were represented by the same number of individuals.

Fig. 4 shows a perfect correlation between Shannon's index of diversity (H′) and the $\log_2 S$. Even though not plotted on the same scale, the diversity values for Shannon's index increase much more rapidly with increased richness than they do for Simpson's or McIntosh's index. The slopes of the lines indicate that Simpson diversity values increase slightly more than McIntosh diversity values with an increase in species richness.

A second set of data was designed to test the individual diversity indices' responses to species evenness, independent of any changes in species richness. These data consisted of six samples, all having 20 species and 200 individuals. The first sample represented maximum evenness, and the last sample represented minimum evenness possible in such a sample. The intermediate samples were arbitrarily chosen to give a range of intermediate values for evenness. Species diversity was again calculated using H′, D and D′ and plotted against species evenness as it was calculated by Eqns (5), (6) and (8). Fig. 5 shows that there is a perfect correlation between species diversity and species evenness in all three indices.

Fig. 2 shows that Shannon's index of diversity was the most strongly correlated with species richness and Simpson's index was the least. This suggests that Shannon's index is more strongly influenced by its richness component than its evenness component when compared to Simpson's index. This can be shown by a comparison of the ratios of the regression coefficients for each index as it is plotted against evenness and richness in Figs 4 and 5 (Tab. 1). The evenness-richness

Tab. 1. Relationships of the regression coefficients of evenness and richness on the three diversity indices.

Index	Regression coefficient with evenness	Regression coefficient with richness	Evenness: richness quotient
Shannon's.....	3.465	1.000	3.465
Simpson's	2.603	0.386	6.743
McIntosh's	2.247	0.358	6.276

quotients (Tab. 1) represent the effect of the evenness component over the effect of the richness component on each of the three diversity indices. These quotients are valid for comparison between diversity indices because the diversity values are all expressed on linear scales. Identical scales were used for evenness in each of the diversity vs. evenness graphs, as well as for richness in each of the diversity vs. richness graphs. A comparison of these quotients shows that Shannon's index places almost twice as much weight on the richness component as does either McIntosh's or Simpson's index. Similarly, Simpson's index is influenced by evenness slightly more than McIntosh's and much more than Shannon's.

To determine if all three indices were measuring similar differences in evenness, linear regressions were calculated between the values of J* and SE, J* and ME, and SE and ME, (Eqns (5), (6) and (8)) from the Siskiyou Mountain data. All three regressions had correlation coefficients greater than 0.96 showing that all three indices measured similar evenness differences.

5. Conclusions

This study has shown the relationship of the two components of species diversity to three indices of diversity. Since species diversity has never been defined specifically in terms of exact amounts of richness and evenness, no actual attempt has been made to judge which diversity index is best or should be used. However, the results of this study clarify certain aspects of both Simpson's and McIntosh's indices which may make them more useful in measuring diversity than they have been in the past. Both indices were originally expressed on a non-linear probability scale which tends to obscure the actual differences in higher diversity values. It is suggested that these indices are much more clearly representative of diversity when expressed in probits or graphed on probability paper. It has also been shown that both McIntosh's and Simpson's indices are influenced considerably more by species evenness and less by richness than is Shannon's index. All three indices have been shown to be linearly related to species evenness and species richness expressed in logarithms to the base 2.

Acknowledgements

I wish to express my gratitude to Dr T. L. Hanes and Dr J. Burk for their support and constructive criticism throughout the study, and to Dr R. H. Whittaker for allowing me to use his data for this analysis. I would also like to thank Mr R. Boston for his assistance with computer programming, Dr M. G. Barbour for editing help, and Mrs R. De Jong, my wife, for her help in preparing this manuscript.

References

Fager, E. W. 1972. Diversity: a sampling study. – Am. Nat. 106: 293–309.

Fisher, R. A., Corbet, A. S. and Williams, C. B. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. – J. Anim. Ecol. 12: 42–58.

Fisher, R. A. and Yates, F. 1963. Statistical tables for biological, agricultural and medical research. – Oliver and Boyd, Edinburgh.

Gleason, H. A. 1922. On the relation between species and area. – Ecology 3: 156–162.

Hurlbert, S. H. 1971. The non-concept of species diversity: a critique and alternative parameters. – Ecology 52: 577–586.

Kricher, J. C. 1972. Bird species diversity: the effect of species richness and equitability on the diversity index. – Ecology 53: 278–282.

Loya, Y. 1972. Community structure and species diversity of hermatypic corals at Eilat, Red Sea. – Mar. Biol. 13: 100–123.

McIntosh, R. P. 1967. An index of diversity and the relation of certain concepts to diversity. – Ecology 48: 392–404.

Monk, C. D. 1967. Tree species diversity in the deciduous forest with particular reference to north central Florida. – Am. Nat. 101: 173–187.

Pielou, E. C. 1969. An introduction to mathematical ecology. – Wiley-Interscience, New York.

Risser, P. G. and Rice, E. L. 1971. Diversity in tree species in Oklahoma upland forests. – Ecology 52: 876–880.

Sanders, H. L. 1968. Marine benthic diversity: a comparative study. – Am. Nat. 102: 243–282.

Shannon, C. E. and Weaver, W. 1949. The mathematical theory of communication. – Univ. Illinois Press, Urbana.

Simpson, E. H. 1949. Measurement of diversity. – Nature, Lond. 163: 688.

Sokal, R. R. and Rohlf, F. J. 1969. Biometry. – Freeman, San Francisco.

Whittaker, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. – Ecol. Monogr. 30: 279–338.

– 1972. Evolution and measurement of species diversity. – Taxon. 21(2/3): 213–251.

Williams, C. B. 1964. Patterns in the balance of nature. – Academic Press, New York.