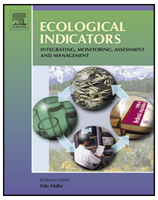
Ecological Indicators 67 (2016) 703–713

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind



Biased richness and evenness relationships within Shannon–Wiener index values

W.L Strong*

Yukon Research Centre, Yukon College, 500 College Drive, P.O. Box 2799, Whitehorse, Yukon Y1A 5K4, Canada

ARTICLE INFO

Article history:
Received 15 October 2015
Received in revised form 21 February 2016
Accepted 21 March 2016
Available online 25 April 2016

Keywords:
Biodiversity
Entropy
Evenness
Plant community
Richness
Shannon–Wiener

ABSTRACT

The purpose of this analysis was to empirically model and graphically illustrate the numerical relationships between richness (S, 4–35 species) and evenness (E) with respect to Shannon–Wiener index (H', $\log_e$-based) values. Thirty-two richness-based third-order polynomial regression models ($R > 0.99$, $P < 0.001$, $n = 28–71$) were constructed to characterize these relationships. A composite diagram showed richness varied curvilinearly, with steepness increasing and the spacing between curves decreasing with greater evenness and H'. Maximum H' values for each richness curve were equal to $\log_e S$ (when $E = 1$), whereas minima were approximated by evenness values of $\sim 1/S$ (when $H' = 0$). It was concluded from multiple and polynomial regression analyses that: (i) evenness contributed more than richness ($E{:}S \geq 3{:}1$) to determining H', based on standardized partial *beta*-coefficients; (ii) the differential in $E{:}S$ ratios increased with greater richness; (iii) the patterns of H' sample variation between maximum unevenness and perfect evenness was convexo-concave shaped; and (iv) richness as an explanatory variable of H' was likely an artifact of evenness (0–1 scale) being rescaled according to individual H' maxima. H' was redefined as a logarithm-weighted measure of evenness at a given level of richness, which means H' is either an imperfect index of diversity or a biased measure of evenness. It was also found that the fundamental components of the Shannon–Wiener index measure dominance concentration rather than evenness, with the reversal in emphasis due to multiplication of the H' equation by -1. H'-derived effective species numbers ($\exp H'$, D) increasingly deviated from those of the diversity model $D = S \times E$ in response to increasing richness (up to 69% for 35 species), particularly when evenness was between 0.15 and 0.40. Of two cross-validated H' prediction methods ($P < 0.001$, $n = 325$), the collective use of individual richness-based polynomial regression equations ($r = 0.954$) was better than a single multiple regression model that incorporated a broad spectrum of richness levels ($r = 0.882$). A simple graphic model was constructed to illustrate patterns of evenness variation as a function of changing richness and H' values. Based on the identified biases, particularly $E{:}S$ ratios, it was recommended that use of H' be discontinued as a basis for assessing diversity in ecological research or, at the very least, accompanied by independent analyzes of richness and evenness.

1. Introduction

The Shannon–Wiener index (Shannon and Weaver, 1949), also known as the Shannon–Weiner [*sic*] diversity, Shannon–Weaver diversity, Shannon (species) diversity, Shannon entropy, Shannon information index, and H', has been used in biological studies as a measure of diversity since the mid-1950s (e.g., MacArthur, 1955; Patten, 1959). As a measure of entropy (Hill, 1973; Jost, 2006; Shannon and Weaver, 1949), or the amount of variation among abundance values, H' is currently defined as the degree of “uncertainty in the . . . identity of an individual that is randomly chosen from . . . [a] dataset” (Tuomisto, 2012, p. 1206; and is similarly defined in ecological reference texts by authors such as Barbour et al., 1999; Kent and Coker, 1992; Krebs, 2009; Magurran, 2003), which is similar to the characterization given by Shannon (1948, pp. 392–393). Others, however, have emphasized the interaction between richness (number of unique taxa or species) and evenness (distribution of abundance, Smith and Wilson, 1996) as the key attribute of H' (e.g., Buzas and Hayek, 1996; Hurlbert, 1971; Lloyd

---

Abbreviations: D, diversity; E, evenness; E_{Dw}, Lorenz curve derived evenness index; E', Camargo evenness index; H', Shannon–Wiener index; $-H'$, an H' value prior to multiplication by -1; IRPR, independent richness-based polynomial regression; J', Pielou evenness index; $\log_e$, natural logarithm; K–S, Kolmogorov–Smirnov one-sample test; H'_{max}, maximum value of H'; p_i, proportional abundance values; P, probability level; r, Pearson product-moment correlation coefficient; R, correlation coefficient with >1 independent-variable; R^2_{adj}, sample size adjusted explained variance; S, species richness; SEE, standard error of estimate.

* Tel.: +1 867 667 2924; fax: +1 867 667 2924.
E-mail address: wstrong@yukoncollege.yk.ca

http://dx.doi.org/10.1016/j.ecolind.2016.03.043
1470-160X/

and Ghelandi, 1964; Molles, 2002; Risser and Rice, 1971; Whittaker, 1972).

Despite a long history of use, uncertainty appears to still exist regarding both the understanding and interpretation of H', although transformation to numbers of equally common species or "effective species" (exp H') has been advocated as a standard for comparing diversity among samples and studies (e.g., Hill, 1973; Jost, 2006; MacArthur, 1965). This numerical transformation, however, does little to advance the ecological understanding of the relationship between richness and evenness. Some studies evade the interaction issue by merely indicating that a statistically significant difference occurred between compared samples without explanation. Opinions on the merits of H', as a measure of diversity, range from being a "dubious index" with "no direct biological interpretation" (Goodman, 1975, p. 244; also Camargo, 2008, p. 282) and an index with "serious conceptual and statistical problems which make comparisons of species richness or species abundance across communities nearly impossible" (Barrantes and Sandoval, 2009, p. 451) to a reluctance to recommend its use (Magurran, 2003, p. 101; Whittaker, 1972, p. 224) to being "the most profound and useful of all diversity indices" (Jost, 2006, p. 364).

Yue et al. (2007) and Barrantes and Sandoval (2009) considered H' to be the best known and most widely used of all diversity indices. A word search of the Thomson Reuters WEB OF SCIENCE™ database found that H' was used in at least 1527 research studies in 2012 through 2014. This number does not include academic theses and dissertations (PROQUEST® Dissertation and Theses Global database), which represented at least an additional 1717 documents, nor technical reports. The number of annually published studies that referenced H' grew from >119 in 1994 to >1073 in 2014, or a nine-fold increase.

A review of 200 arbitrarily selected biological publications from the 2012 to 2014 period (W.L Strong, unpublished data) indicated that ecosystem diversity was the focus of 65% of the studies, with genetics representing an additional 21%. Inferential statistical testing was used to identify differences among H' values in two-thirds of these studies. The remaining studies primarily used regression and correlation to describe relationships between H', and various abiotic (e.g., Çağlar and Albayrak, 2012) and biotic factors (e.g., Con et al., 2013), or made only qualitatively interpretations of the data (e.g., Akhoundi et al., 2013). About 58% of the biological studies used natural logarithms ($\log_e$) as a basis for calculating H', with ~21% not clearly identifying the applied logarithm base, although it affects the numerical outcome of H' and can limit the comparison of results among studies (Magurran, 2003). The use of H'-derived effective species numbers as a basis for diversity comparison was not a widely applied practice among biologists, based on their use in only four of the 200 reviewed studies (Adie et al., 2013; Külköylüoğlu et al., 2012, 2013; Skácelová and Lepš, 2014).

Three issues complicate the interpretation of H' from an ecological perspective: (1) the currently advocated definition emphasizes uncertainty (e.g., Jost, 2006, p. 363–364; Tuomisto, 2012, p. 1206), as if diversity were merely a statistical probability issue rather than an ecological phenomenon; (2) the differential logarithm weighting of proportional abundance values in the equation affects the outcome of H' calculations (Odum, 1969; Peet, 1974, p. 295; Ricotta, 2003, p. 184; Whittaker, 1972, p. 224) and, therefore, its interpretation; and (3) the numerical relationship between richness and evenness, and H' is vague. For example, is or how is richness represented in H'? This question arises because H' is calculated from dimensionless proportional abundance values, without richness being a parameter in the Shannon–Wiener equation. In addition, do richness and evenness have equally weighted roles in the calculation of H'? Based on the conceptual model advocated by Jost (2010, p. 212) and Tuomisto (2012, p. 1205) and indirectly by other such as Hurlbert (1971, p. 577), diversity (D) should be proportional to the relationship between richness (S) and evenness (E) (i.e., D (as represented by exp H') = $S \times E$).

Strong quantitative associations have been reported between H' and richness (DeJong, 1975; Monk, 1967; Risser and Rice, 1971; Tramer, 1969) and H' and (un)evenness (DeJong, 1975; Risser and Rice, 1971), so presumably multivariate regression equations could be constructed to quantify their numerical relationships as predictors of H' (Buzas and Hayek, 1996; Stirling and Wilsey, 2001); but no comprehensive models appear to exist that clearly demonstrate such interactions. With regard to the Stirling and Wilsey (2001) analysis, their multivariate models are of questionable validity, because the Pielou (1966) J' evenness index ($J' = H'/\log_e S$, $\log_e S$ is also known as $H'_{\max}$) was used as an explanatory variable of H', i.e., there is a lack of independence between H' and J'. Their use of J' as an evenness measure probably stemmed from its popularity relative to other indices at the time. Among the 200 reviewed biological publications, J' occurred in 73 (36.5%). The large and growing number of studies that use H' suggest it is becoming an expected standard for assessing biological diversity, and the above indicated issues and questions need to be addressed if H' values are to be understood from a numerical perspective rather than just calculated. From a practical perspective, the use of H' as a resource management tool (e.g., evaluation of herbicide and mechanical treatment effects as forest site management practices – Seiwa et al., 2012; Wu et al., 2013) could have substantial and long-term adverse environmental implications, if it is prone to misinterpretation.

Tuomisto (2010) suggested that the understanding of diversity lay not in the form of the numbers (e.g., entropies, probabilities, effective species), but in the ecological meaning of the variation in the abundance values that are the basis for calculating such indices. Based on this philosophical perspective, the objectives of this analysis were to: (i) graphically model the numerical relationships between evenness and $\log_e$-based H' values, and richness (5–35 species); (ii) estimate the amount of potential variation that occurs among H' values, when calculated for different species abundance combinations at individual levels of richness; (iii) construct multiple regression models to predict H' as a basis for determining if richness and evenness contribute equally to its calculation; (iv) determine if H'-derived "effective species" numbers (D) conform to the diversity model $D = S \times E$; (v) propose a more ecologically-oriented definition of H' than is currently advocated; and (vi) create an easily understood two-dimensional graph to summarize the numerical relationships of evenness to richness and H' with respect to the analyzed data. Although H' is used in various fields of research, this assessment is primarily oriented toward terrestrial vegetation and plant community analyses.

2. Materials and methods

2.1. Synthetic numerical sequences (DATASET A)

Data for modeling the relationship between richness and evenness, and H' were created by compiling a series of synthetic numerical sequences for each richness level for four through 35 species (DATASET A). This range of richness appears to encompass the number of species that typically occur in terrestrial plant community composition samples or relevés (e.g., De Grandpré et al., 2011; Hart and Chen, 2008; Strong, 2015; Wu et al., 2013). To create each numerical series, an initial sequence of abundance values with an indisputable H' and evenness value was established. The initial sequence for each series consisted of equally abundant species that totaled 100 (e.g., first sequence for five species – 20, 20, 20, 20,