



Hill numbers everywhere. Does it make ecological sense?

Carlo Ricotta^{a,*}, Enrico Feoli^b

^a Department of Environmental Biology, University of Rome La Sapienza, Rome, Italy

^b Department of Life Sciences, University of Trieste, Trieste, Italy

ARTICLE INFO

Keywords:

Diversity measures

Doubling property

Effective numbers of species

Species addition

Uncertainty measures

ABSTRACT

A supposed weakness of most diversity measures is their non-linearity with respect to species addition. Even for a community where all species have equal abundance, each added species usually leads to a smaller increment in the diversity measure than the species added before it. A recent proposal to solve this problem was to transform classical diversity measures to 'effective numbers of species' or 'Hill numbers.' For any community with diversity D , the effective number of species N is the number of equally abundant species that is needed to get a diversity value equal to D . The conversion of classical diversity measures to Hill numbers makes them linear with respect to species addition such that, given two equally large and completely distinct communities, each with diversity D , if these communities are pooled, the diversity of the pooled communities is $2D$. According to this proposal, Hill numbers have been widely adopted in ecological literature as the ultimate solution for diversity analysis regardless of the scientific question at hand. In contrast, we believe that assuming a non-linear response of diversity measures to species addition is more suitable for many ecological questions. Building on this idea, we have introduced a typification of diversity measures based on how quickly diversity increases as species are added.

1. Introduction

From the very beginning of the discipline, community ecologists have used diversity measures to explore the complex mechanisms that drive the spatial and temporal patterns of species coexistence. Since species assemblages are complex multi-dimensional objects, their study and representation is impossible without multivariate methods of exploratory data analysis. The main purpose of such analysis includes the summarization, visualization, and description of biological patterns and their relationships, whereas estimation and statistical inference are in most cases of secondary importance (Podani, 2000). From the point of view of a community ecologist, diversity measures can be thus defined as a set of multivariate summary statistics for quantifying various aspects of community structure in terms of species richness, abundance, phylogeny, functional traits, etc. (Solow and Polasky, 1994; Ricotta, 2005). Accordingly, while some of the most popular diversity measures, such as the Shannon entropy (Shannon, 1948), the Simpson diversity (Simpson, 1949), or their parametric generalizations (e.g., Rényi, 1970; Patil and Taillie, 1982) have different origins, they have all been used in community ecology to summarize the uncertainty in predicting the relative abundance of species in a given assemblage.

Let p_i be the relative abundance of species i ($i = 1, 2, \dots, S$) with $0 \leq p_i \leq 1$ and $\sum_{i=1}^S p_i = 1$. The Simpson diversity $1 - \sum_{i=1}^S p_i^2$ is the probability that two individuals selected at random with replacement from a given community or assemblage do not belong to the same species, while the Shannon entropy $-\sum_{i=1}^S p_i \ln p_i$ quantifies uncertainty in information-theoretical terms (for details, see e.g. Rényi, 1970). High diversity thus implies high unpredictability. In both cases, for non-empty communities, diversity is zero if the community contains only one species and progressively increases for increasing species richness and evenness, such that for a given number of species, diversity is maximal if all S species occur in equal abundance (i.e., $p_i = 1/S$ for all $i = 1, 2, \dots, S$). Therefore, diversity can be seen as a generic notion encompassing many distinct measures that summarize uncertainty from many different viewpoints.

In this context, a supposed drawback already observed by McArthur (1965) and Whittaker (1972) is that, apart from species richness, most diversity measures are non-linear with respect to species addition. Therefore, even for a completely equitable community where all species have equal abundance, each added species leads to a smaller increment in the value of diversity than the species added before it (Jost et al., 2010; Ricotta et al., 2021). To solve this problem, Hill (1973) and Jost

* Corresponding author.

E-mail address: carlo.ricotta@uniroma1.it (C. Ricotta).

(2006, 2007) proposed to transform classical diversity measures to ‘effective numbers of species’ (also called species equivalents or Hill numbers). For a given diversity measure D , the effective number of species N is the number of equally abundant species (i.e., all with abundance $p_i = 1/N$) that is needed in order that its diversity be D (Patil and Taillie, 1982). Jost (2006, Appendix 1) also demonstrated that the Hill numbers of all measures of diversity that can be expressed as monotonic functions of $\sum_{i=1}^S p_i^q$ (with $q \geq 0$) or limits of such functions as q approaches unity, are given by the formula:

$$N^q = \left(\sum_{i=1}^S p_i^q \right)^{1/(1-q)} \quad (1)$$

According to Eq. (1), for $q = 0$, $N^0 = S$ (i.e., the effective number of species of species richness is species richness itself). For $q = 2$, we obtain the effective number of species of the Simpson diversity $N^2 = 1/\sum_{i=1}^S p_i^2$, while for $q = 1$ Eq. (1) is not defined, but its limit for $q \rightarrow 1$ gives the effective number of species of the Shannon entropy: $N^1 = \exp\left(-\sum_{i=1}^S p_i \ln p_i\right)$.

A relevant property of Hill numbers is their linearity with respect to pooling, such that given two equally large and completely distinct species assemblages, each with diversity D , if these assemblages are combined, the diversity of the combined assemblages is $2D$. This ‘doubling property’ is at the heart of the independence between alpha and beta diversity in multiplicative diversity decomposition demonstrated by Jost (2007). Hill numbers thus represent a general recipe for transforming a large number of classical diversity measures to the common currency of a species richness scale: “the conversion of properly defined frequency-based measures to their number equivalents makes them linear with respect to our intuitive ideas of diversity. They are now almost as easy to interpret as species richness and much more reliable and informative” (Jost, 2007).

2. Are Hill numbers always the optimal solution?

Since the seminal work of Jost (2006, 2007), Hill numbers have been extensively used in community ecology for summarizing different facets of diversity, such as functional and phylogenetic diversity (Leinster and Cobbold, 2012; Chao et al., 2014), acoustic and soundscape diversity (Luybaert et al., 2022), or functional redundancy (Dick, 2023), to mention just a few. However, in many papers Hill numbers are presented as if they were the optimal solution for any kind of diversity analysis regardless of the scientific question at hand. Jost himself indirectly contributed to support this idea by naming Hill numbers ‘true diversities’ (Jost, 2006).

On the contrary, by condensing a complex multi-dimensional object, such as the abundance structure of a species assemblage into a single measure, information is inevitably lost, and there is no true or magic diversity that is able of portraying all aspects of community composition in a satisfactory way (Ricotta, 2010). In this view, a subtle, albeit relevant weakness of Hill numbers is specifically related to their linear behavior with respect to species addition: while it is reasonable to assume that for conservation issues, the diversity of a community with one hundred equally abundant species should be twice the diversity of a community with fifty equally abundant species, there are many other fields of ecology, such as species interaction networks (Delmas et al., 2019; Momal et al., 2020), community stability and responses to disturbance (Tu et al., 2019; Arese Lucini et al., 2020), or the impact of environmental changes on ecosystem functions and the associated services (Hou et al., 2023), where processes and relationships are intrinsically non-linear.

Such non-linear processes challenge the necessity of relying on diversity measures that are linear with respect to species addition. In those cases, it is convenient to imagine that, even when all species are equally

common (a very unlikely condition in nature), the addition of one species to a hypothetical community composed of, say, five species increases diversity more than the addition of one species to a community composed of five hundred species. In other words, diversity measures that increase linearly with the number of species are not always ecologically justified. In some cases, diversity measures that assume a nonlinear rate of change in diversity with the addition of species may be more useful in accounting for the fact that the increase in diversity is higher when a new species is found in a species-poor community compared to a species-rich one, both in absolute and relative terms. This approach usually provides a more comprehensive and realistic picture of ecological complexity. In this context, it is worth noting that while the properties of Hill numbers are widely accepted by ecologists, the properties of evenness measures, which represent another significant aspect of community structure, have received considerably less attention and are still a subject of debate (Jost, 2010; Chao and Ricotta, 2019; Ricotta et al., 2022).

In his seminal paper, Jost (2006) made a distinction between (non-linear) uncertainty and (linear) diversity. We believe instead that all diversity measures are essentially measures of uncertainty and that Hill numbers are simply a class of such measures with a linear response to species addition. Based on this definition, the various measures of diversity/uncertainty can be differentiated looking at the speed with which the uncertainty associated with species addition increases with increasing species richness.

For example, for a completely even community with species abundances $p_i = 1/S$ for all $i = 1, 2, \dots, S$, the first derivative of the Shannon diversity is equal to $\frac{d}{dS} \ln S = \frac{1}{S}$, whereas for the Simpson diversity it is $\frac{d}{dS} \left(1 - \frac{1}{S}\right) = \frac{1}{S^2}$. Accordingly, the uncertainty associated to species addition increases more rapidly for the Shannon diversity compared to the Simpson diversity (Fig. 1).

This effect is even more evident for the parametric diversity measure of Patil and Taillie (1982):

$$D^q = \sum_{i=1}^S p_i \frac{1 - p_i^{q-1}}{q - 1} \quad (2)$$

Unlike the Shannon or the Simpson diversity which are point descriptors of diversity, D^q represents a continuum of diversity measures

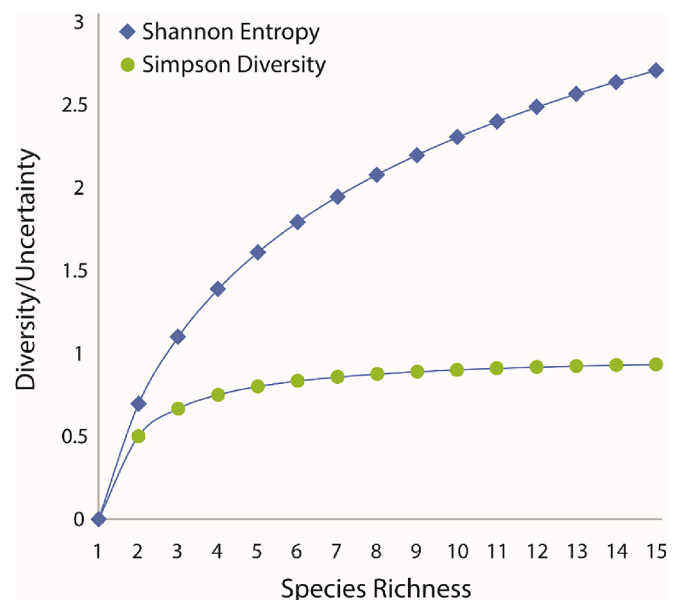


Fig. 1. Plot of the Shannon and the Simpson diversity for a sequence of perfectly even assemblages composed of 1, 2, ..., 15 species. As shown by the figure, the rate of increase of the Simpson diversity/uncertainty with increasing species richness is much lower than that of the Shannon diversity.

which are obtained by varying the parameter q ($q \geq 0$). The different measures obtained by changing the values of the parameter q are thus different moments of the same generalized diversity function (Ricotta et al., 2021). A few characteristic values of the parameter q recover more traditional diversity measures. For example, for $q = 0$, $D^0 = S - 1$ (a linear function of the number of species in the assemblage that assigns zero diversity to single species communities), for $q = 2$, D^2 is the Simpson diversity $1 - \sum_{i=1}^S p_i^2$, while for $q = 1$, D^1 is not defined, but its limit for $q \rightarrow 1$ is equal to the Shannon entropy $-\sum_{i=1}^S p_i \ln p_i$.

From Eq. (2), it is easily shown that for a completely even community, the first derivative of D^q is $\frac{d}{dq} D^q = \frac{1}{3q}$. That is, by increasing the parameter q , the rate of increase of the uncertainty associated to species addition is progressively reduced. By contrast, for the Hill numbers N^q , the first derivative for a complete even community is always equal to $\frac{d}{dq} N^q = 1$. Therefore, irrespective of the value of q , the rate of increase of the uncertainty associated to species addition is a linear function of species richness.

In summary, classical and less classical measures of diversity can be ranked based on their rate of increase of the uncertainty associated to species addition. At one extreme, we have species richness and the Hill numbers with their linear behavior; at the other extreme, we have measures such as the Simpson diversity for which the uncertainty associated with species addition decreases very rapidly with increasing species richness. For practical purposes, one should use the type of diversity which is most adequate to solve the specific problem at hand based on the desired index sensitivity to rare and common species. While the Simpson index responds more strongly to changes in the abundance of common species, becoming progressively less sensitive to changes in abundance as rarer species are considered, the Shannon entropy decreases more slowly with increasing species richness, thus being more sensitive to changes in the abundance of rarer species.

3. Conclusion

The moral of the story is always the same: a diversity measure that is able to summarize all complex and non-linear aspects of diversity/uncertainty cannot exist. Therefore, forcing diversity measures into the Procrustes bed of Hill numbers to get at all costs a linear response to species addition is not always a good idea because it reduces the ecologist toolbox to a single family of measures which is unable to fully capture the complex and often non-linear essence of most ecological processes.

Funding

CR was supported by a research grant from the University of Rome 'La Sapienza' (SP1221845D187EC4).

CRedit authorship contribution statement

Carlo Ricotta: Writing – original draft, Methodology, Conceptualization. **Enrico Feoli:** Writing – original draft, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

- Arese Lucini, F., Morone, F., Tomassone, M.S., Makse, H.A., 2020. Diversity increases the stability of ecosystems. *PLoS One* 15, e0228692.
- Chao, A., Chiu, C.-H., Jost, L., 2014. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers. *Annu. Rev. Ecol. Syst.* 45, 297–324.
- Chao, A., Ricotta, C., 2019. Quantifying evenness and linking it to diversity, beta diversity, and similarity. *Ecology* 100, e02852.
- Delmas, E., Besson, M., Brice, M.H., Burkle, L.A., Dalla Riva, G.V., Fortin, M.J., Gravel, G., Guimarães Jr., P.R., Hembry, D.H., Newman, E.A., Olesen, J.M., Pires, M. M., Yeakel, J.D., Poisot, T., 2019. Analysing ecological networks of species interactions. *Biol. Rev.* 94, 16–36.
- Dick, D.G., 2023. Measuring functional redundancy using generalized hill numbers. *Funct. Ecol.* 37, 1304–1314.
- Hill, M., 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54, 427–432.
- Hou, J., Fanin, N., Zizhao, N., 2023. A network prediction model to quantify relationship between biodiversity and ecosystem functioning (BEF). *Methods Ecol. Evol.* 14, 2907–2916.
- Jost, L., 2006. Entropy and diversity. *Oikos* 113, 363–375.
- Jost, L., 2007. Partitioning diversity into independent alpha and beta components. *Ecology* 88, 2427–2439.
- Jost, L., 2010. The relation between evenness and diversity. *Diversity* 2, 207–232.
- Jost, L., DeVries, P.J., Walla, T., Greeney, H., Chao, A., Ricotta, C., 2010. Partitioning diversity for conservation analyses. *Divers. Distrib.* 16, 65–76.
- Leinster, T., Cobbold, C.A., 2012. Measuring diversity: the importance of species similarity. *Ecology* 93, 477–489.
- Luypaert, T., Bueno, A.S., Masseli, G.S., Kaefer, I.L., Campos-Cerqueira, M., Peres, C.A., Haugaasen, T., 2022. A framework for quantifying soundscape diversity using hill numbers. *Methods Ecol. Evol.* 13, 2262–2274.
- McArthur, R., 1965. Patterns of species diversity. *Biol. Rev.* 40, 510–533.
- Momal, R., Robin, S., Ambroise, C., 2020. Tree-based inference of species interaction networks from abundance data. *Methods Ecol. Evol.* 11, 621–632.
- Patil, G.P., Taillie, C., 1982. Diversity as a concept and its measurement. *J. Am. Stat. Assoc.* 77, 548–561.
- Podani, J., 2000. Introduction to the exploration of Multivariate biological data. Backhuys Publishers, Leiden, NL.
- Rényi, A., 1970. Probability theory. North-Holland Publishing, Amsterdam, NL.
- Ricotta, C., 2005. Through the jungle of biological diversity. *Acta Biotheor.* 53, 29–38.
- Ricotta, C., 2010. On beta diversity decomposition: trouble shared is not trouble halved. *Ecology* 91, 1981–1983.
- Ricotta, C., Szeidl, L., Pavoine, S., 2021. Towards a unifying framework for diversity and dissimilarity coefficients. *Ecol. Ind.* 129, 107971.
- Ricotta, C., Bacaro, G., Maccherini, S., Pavoine, S., 2022. Functional imbalance not functional evenness is the third component of community structure. *Ecol. Ind.* 140, 109035.
- Shannon, C., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423.
- Simpson, E.H., 1949. Measurement of diversity. *Nature* 163, 688.
- Solow, A.R., Polasky, S., 1994. Measuring biological diversity. *Environ. Ecol. Stat.* 1, 95–107.
- Tu, C., Suweis, S., Grilli, J., Formentin, M., Maritan, A., 2019. Reconciling cooperation, biodiversity and stability in complex ecological communities. *Sci. Rep.* 9, 5580.
- Whittaker, R.H., 1972. Evolution and measurement of species diversity. *Taxon* 21, 213–251.