The use of taxonomic diversity indices in the assessment of perturbed community recovery

R. E. Desrochers & M. Anand
Biology Department, Laurentian University, Sudbury, Ontario, Canada

Abstract

The spatial dynamics of ecological diversity are studied in two regions (Sudbury, Canada and Harjavalta, Finland) damaged by air pollution from copper and nickel smelting. Perturbation from pollution is assumed to be most intense near the source and to decrease with distance. Monitoring sites were therefore selected to traverse these pollution gradients. Using Rényi's generalized entropy as a measure of diversity, a monotonic pattern of increasing diversity is discovered at the recovering Canadian sites but not at the Finnish sites. Quadratic entropy and a related information-theoretical measure of taxonomic diversity, taxonomic entropy, were calculated in the hope that these diversity indices, which incorporate taxonomic distances, would provide a better understanding of these unexpected results. Quadratic entropy has the additional advantage of making use of pairwise taxonomic distances between species in a highly intuitive manner. Taxonomic has a clear information-theoretical meaning and can be calculated in the same units of measurement as Rényi's generalized entropy thereby facilitating the comparison of classical diversity to taxonomic diversity. Quadratic entropy was found to contribute little insight as to the state of ecological recovery relative to taxonomic entropy.

1 Introduction

Statistical ecologists have long relied upon information theory and its associated measures of entropy for indices of diversity that include both number of species and their relative proportions [1–3]. The desire to quantify ecological diversity stems from the belief that diversity encapsulates ecological properties that are not only of interest but that should be actively protected since it may be related to stability and productivity [4]. Pielou [1,3] outlined properties that should be included in a meaningful diversity index. The first is that any given number of species, s, the diversity index should be maximal if all species proportions are equal. Secondly if all species proportions are equal the diversity index should increase with increasing s. However, diversity indices that incorporate information about the taxonomic relatedness species may be more valuable. An ecological assemblage composed of distantly related taxa is intuitively believed to be more diverse than an ecological assemblage composed of more closely related taxa, therefore, valuable indices of diversity should account for taxonomic relatedness amongst species [5–8].

2 Quantitative methodology

Possibly the best known diversity index is Shannon entropy (eqn. (1)), which can be calculated using Rényi's generalized entropy (eqn. (2)) of order alpha approaching 1,

$$H = -\sum_{i=1}^{s} p_i \log_2 p_i \tag{1}$$

$$H^{\alpha} = \frac{\log_2 \sum_{i=1}^{s} p_i^{\alpha}}{1-\alpha} \tag{2}$$

where s is the total number of species and p_i is the proportion of the ith species in the sampling unit [9]. Log base 2 renders units of measure for diversity in bits.

Quadratic entropy [10] is a measure of taxonomic diversity that has received much attention recently in ecological literature [5–8, 11]. Rao's quadratic entropy, Q, is calculated here using

$$Q = \sum_{i>j=1}^{s} d_{ij} p_i p_j \tag{3}$$

where $\overline{D} = d_{ij}$ (where $d_{ij} = d_{ji}$ and $d_{ii}=0$) is the taxonomic distance matrix ($\overline{D}$ is symmetric), s is the total number of species, p_i and p_j are the proportions of the ith and the jth species respectively in the sampling unit and d_{ij} is the taxonomic distance between the ith and jth species [6–8, 10]. Similarly to Shimatani [7] and Clarke and Warwick [6], d_{ij} is defined by a taxonomic hierarchical classification where $d_{ij}=1$ if both species belong to the same genus, $d_{ij}=2$ if both species belong to the same family but different genera, $d_{ij}=3$ if both species belong to the same order but different families, $d_{ij}=3.5$ if both species belong to the same subclass but different orders, $d_{ij}=4$ if both species belong to the same class but different subclasses, $d_{ij}=4.5$ if both species are both angiosperms but are from different classes and finally $d_{ij}=5$ otherwise.

As this is measure is not information-theoretical it is not directly comparable to any measures associated with Rényi's generalized entropy. To resolve this problem, Ricotta and Avena [8] have developed an information-theoretical measure of taxonomic diversity termed 'taxonomic entropy' which is analogous to quadratic entropy. Taxonomic entropy, H(K,P), was calculated using

$$H(K,P) = \sum_{i=1}^{s} p_i \log_2 k_i \tag{4}$$

where s is the total number of species, p_i is the proportion of the ith species in the sampling unit and k_i is the taxonomic distinctness of the ith species. The vector $K=(k_1, k_2,\ldots, k_s)$ is the row totals of the distance matrix $\overline{D}$ resulting in a vector $V=(v_1, v_2,\ldots, v_s)$ that is standardized by dividing each element by the sum of the vector, which is equivalent to the sum of all elements of $\overline{D}$ (eqn. 5).

$$k_i = v_i \Big/ \sum d_{ij} \tag{5}$$

H(K,P) is the sum of Shannon diversity (eqn.1) and the information gained with taxonomic information. This information gain can be quantified using

$$H(K\|P) = \sum_{i=1}^{s} k_i \log_2\left(k_i \big/ p_i \right) \tag{6}$$

The quantity $H(K\|P)$ will become large if dominant species (species present in high proportions) are the most taxonomically distinct and minor species are the least taxonomically distinct and vice versa. While the taxonomic distance matrix used here is based upon Linnean classification, distance matrices can also be derived from genetic analysis [6–8].

3 Theoretical analysis

In addition to applying these to phytosociological relevés from recovering plants communities in Canada and Finland, they were first applied to theoretical relevés and taxonomic distance matrices designed to test the effect of community evenness and taxonomic distance on quadratic entropy and taxonomic entropy. Nine 5 species relevés were created, the first with perfect evenness (i.e. 0.2, 0.2, 0.2, 0.2, 0.2), the last with minimal evenness (i.e. 0.9, 0.01, 0.005, 0.01, 0.075) and seven with intermediate levels of evenness. The diversity indices were computed using 4 taxonomic distance matrices. The first assumes maximal taxonomic distance amongst the five species (matrix A), which would occur in the case where all five species belonged to different classes. The second assumes minimal taxonomic distance amongst the five species (matrix B), which would occur in the case where all five species belonged to the same genus. The third and fourth (matrices C and D) assume intermediate levels of taxonomic distinction amongst species.

0	5	5	5	5
5	0	5	5	5
5	5	0	5	5
5	5	5	0	5
5	5	5	5	0

Matrix A

0	1	1	1	1
1	0	1	1	1
1	1	0	1	1
1	1	1	0	1
1	1	1	1	0

Matrix B

0	1	2	3	4
1	0	1	2	3
2	1	0	1	2
3	2	1	0	1
4	3	2	1	0

Matrix C

0	1	5	1	5
1	0	1	5	1
5	1	0	1	5
1	5	1	0	1
5	1	5	1	0

Matrix D

Transactions on Ecology and the Environment vol 63, © 2003 WIT Press, www.witpress.com, ISSN 1743-3541

Shannon entropy increases with increasing evenness and is maximal at perfect evenness (at $H_{max}=\log_2 s$). Because both quadratic entropy and taxonomic entropy incorporate relative taxonomic distances in addition to species proportions neither will be maximal at perfect evenness. In fact depending upon the taxonomic relationship of a given species to all others in the community its deletion from the community may result in an increase in overall taxonomic diversity [7, 8]. Quadratic entropy tends to increase with increasing evenness and with increasing taxonomic distance amongst species (Fig. 1). The distribution of species proportions relative to the distribution of taxonomic distances will allow quadratic entropy to be greater at lower degrees of evenness and will allow shifting rankings of matrices C and D.

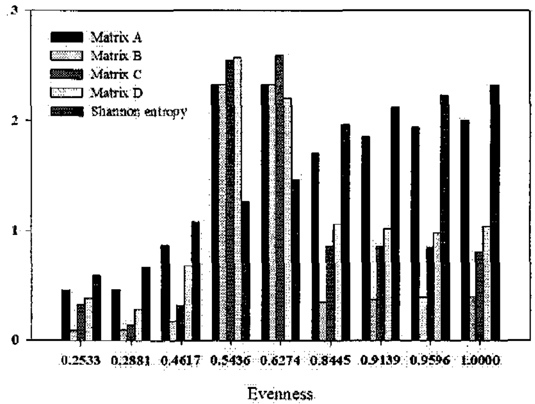

Figure 1: Quadratic entropy in relation to Shannon entropy (bits) at 9 cases of evenness for 4 taxonomic distance matrices (A–D).

Taxonomic entropy and the taxonomic information gain appear to vary little with increasing evenness (Figs. 2 and 3). The distribution of species proportions relative to the distribution of taxonomic distances will again allow shifting rankings of matrices C and D. Due to the manner in which taxonomic distinctness (k_i) is calculated for each species (eqn. (5)), taxonomic entropy is unable to distinguish between the rare cases of maximal taxonomic distinctness (matrix A) and minimal taxonomic distinctness (matrix B). These will equal H_{max} regardless of evenness because the sum of p_i always equals one. While not ecologically meaningful, it is mathematically reasonable as there should be no information gained with the incorporation of equiprobable terms [8].

Quadratic entropy presents the obvious advantage of distinguishing, in most cases, between the rare cases of maximal and minimal taxonomic distances represented by matrices A and B respectively. It should be noted that quadratic entropy is related to Simpson's diversity index (Rényi's generalized entropy of order $\alpha=2$, eqn. (2)) and therefore the relative taxonomic contribution to quadratic entropy can isolated [7]. However, the taxonomic contribution cannot be easily partitioned because the relationship between the Simpson index and the taxonomic contribution is not additive.

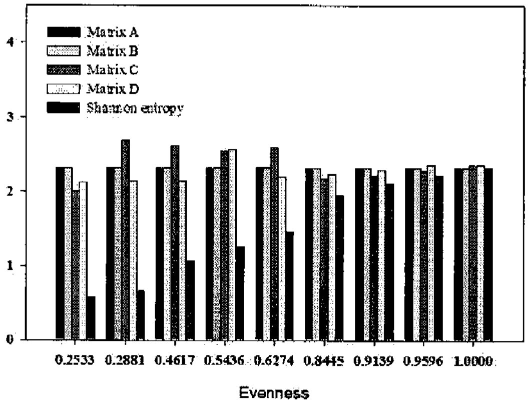

Figure 2: Taxonomic entropy in relation to Shannon entropy (bits) at 9 cases of evenness for 4 taxonomic distance matrices (A–D).

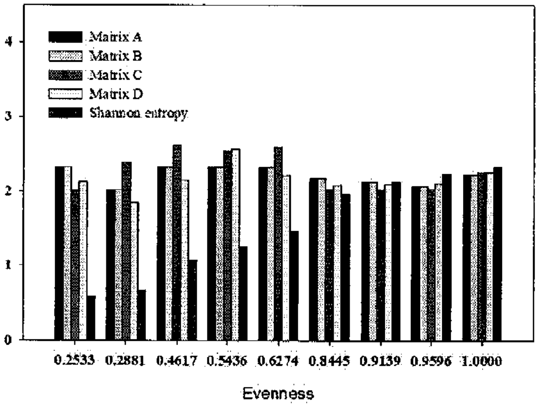

Figure 3: Taxonomic information gain in relation to Shannon entropy (bits) at 9 cases of evenness for 4 taxonomic distance matrices (A–D).

4 Case studies: site and data description

To assess the ability of these measures to provide insightful quantification of ecological community recovery, they were applied to recovering plants communities along pollution gradients in Canada and Finland.

The Canadian data were collected in 2001 from an area perturbed by a now decommissioned smelter complex near Sudbury, Ontario, Canada. While in operation from 1929 to 1972, the Coniston smelter complex subjected the local soil and vegetation to sulphur dioxide emissions and heavy metal fallout consisting primarily of copper and nickel. The vegetation in the area has had a 30 years period of recovery since the termination of smelting activities. The data were collected from six sites extending 36 km in a southerly direction from the smelter complex as contaminants were carried primarily southward by the prevailing winds [12]. All understory vascular plant species were sampled along two 100 m transects at each site. Cover abundance was estimated using the Braun-Blanquet method within each of 100 contiguous 1 m^2 quadrats. The diversity estimates were calculated on a per quadrat basis and averaged [13].

The Finnish data were collected in 1993 and published in 2001 [14] from an area perturbed by the Outokumpu Harjavalta Metals smelter complex near Harjavalta, SW Finland. The Outokumpu Harjavalta Metals smelter complex also processes nickel and copper, and sulphur dioxide, nickel and copper largely contaminate the surrounding vegetation. The data were collected from six sites extending 8 km in a southeasterly direction from the smelter complex. All understory vascular and non-vascular plant species were sampled within three sample plots at each site. Cover abundance was estimated in sixteen 1 m^2 sub-plots per sample plot using a modified point quadrat method. The diversity estimates were calculated on a per sample plot basis and averaged.

5 Results and discussion

As expected there is a significant increasing linear or monotonic trend in the three measures of diversity of the vascular plant community with increasing distance from the source of ecological perturbation in the Sudbury case: Shannon entropy (r^2=0.92), quadratic entropy (r^2=0.91) and taxonomic entropy (r^2=0.75) (Fig. 4). While the dynamics of quadratic entropy closely mimic the dynamics of Shannon entropy, the magnitude of these cannot be compared directly as Shannon entropy is measured in bits whereas quadratic entropy is not.

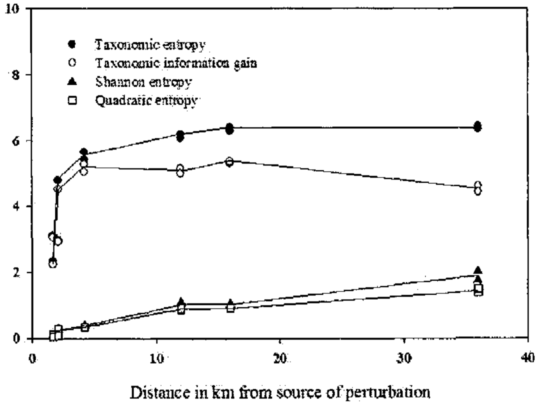

Figure 4: The vascular plant community in Sudbury, Canada where average quadratic entropy, Shannon entropy, taxonomic entropy and taxonomic information gain against distance from the source of perturbation.

Taxonomic entropy and the taxonomic information gain on the other hand can be directly, compared to Shannon entropy as these are measured in bits, where taxonomic entropy is the sum of Shannon entropy and the taxonomic information gain. The comparison of these three allows us to assess the relative contributions of Shannon entropy and taxonomic information to overall diversity. Taxonomic information gain is consistently much greater than Shannon entropy, suggesting that the taxonomic component plays an important role in characterizing these communities (Fig. 4). It is expected that decreasing stress (increasing distance from smelters) should lead to an increase in both classical diversity (Shannon entropy) and taxonomic diversity (taxonomic entropy and quadratic entropy), where specifically the taxonomic information gain should decrease relative to Shannon entropy with increasing stress or perturbation [5, 15]. We found this to be generally true in the Sudbury case, however no simple trend could be detected in the dynamics of the taxonomic information gain (r^2=0.39). While this was not the case for the Canadian vascular plant community, it appears to apply to the Finnish case (Fig. 5).

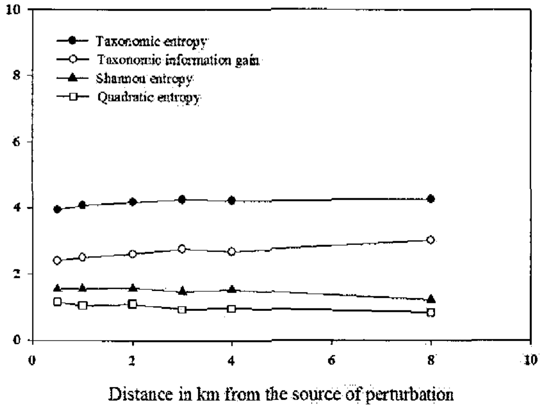

Figure 5: The vascular plant community for Harjavalta, Finland where average quadratic entropy, Shannon entropy, taxonomic entropy and taxonomic information gain against distance from the source of perturbation.

The taxonomic information gain for the vascular plants at the monitoring sites in Harjavalta increases linearly with increasing distance from the smelter complex and thereby decreasing stress (r^2=0.91). Shannon entropy and quadratic entropy and taxonomic entropy do not increase as expected with increasing distance from the source of ecological perturbation (Fig. 5). Both Shannon entropy and quadratic entropy seem to be decreasing linearly with increasing distance (r^2=0.82 and r^2=0.80 respectively). Here again the dynamics of quadratic entropy closely mimic the dynamics of Shannon entropy. Taxonomic entropy appears to increase slightly initially and stabilize with distance, however this trend is not significant (r^2=0.53). Most interesting here is the contribution of taxonomic information to taxonomic entropy relative to Shannon entropy. As would have predicted Warwick and Clarke [5], the relative contribution of taxonomic information gain increases with increasing distance whereas Shannon entropy decreases with increasing distance. Also the taxonomic information gain is the only measure that discerns a trend of ecological improvement with decreasing stress.

The trend in taxonomic entropy for the vascular and non-vascular plant communities at the monitoring sites in Harjavalta appears to be quite similar to that observe for the vascular plant species alone (though the values are greater, as expected, with the inclusion of non-vascular plants), but no significant increasing or decreasing trend could be detected for any of the measures (Fig. 6). However, the interplay between Shannon entropy and the taxonomic information gain is most dramatic here. Shannon entropy is low at proximal sites in response to the small number of species despite relatively high community evenness and later increases with increasing evenness and number of species. At these sites the taxonomic relationship amongst species is such that the information gain is sufficient to compensate for low initial values of Shannon entropy. The taxonomic information gain will become large if dominant species (species present in larger proportions) are the most taxonomically distinct and minor species are the least taxonomically distinct and vice versa [8].

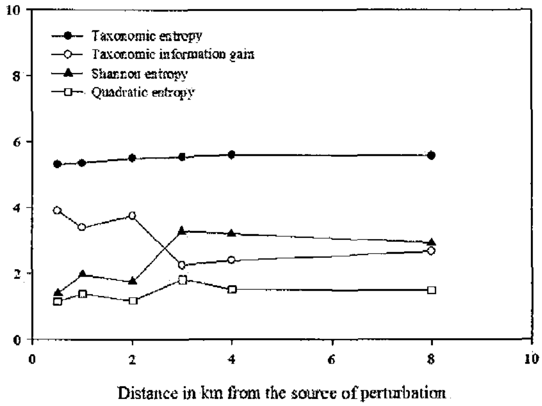

Figure 6: Average quadratic entropy, Shannon entropy, taxonomic entropy and taxonomic information gain against distance in kilometers from the source of perturbation for all vascular and non-vascular plant communities in Harjavalta, Finland.

It can thus be concluded that at proximal sites those species that are present in larger proportions are taxonomically more distinct than those present in smaller proportions. The subsequent decline in the relative contribution of taxonomic information with the addition of species can be related to a shift in dominance of species that are less taxonomically distinct. With distance, the two components eventually contribute approximately half of the overall taxonomic diversity. It is surprising that the inclusion of non-vascular plants decreases the relative contribution of taxonomic information to taxonomic entropy.

Shannon entropy is generally greater in the Finnish cases than in the Canadian case, however, taxonomic entropy is generally greater in Sudbury than in Harjavalta. Taxonomic entropy for the vascular plant community at four kilometers from the smelters (5.6 bits) in Sudbury more closely resembles the taxonomic entropy for the total plant community at four kilometers from the smelters (5.6 bits) in Harjavalta than that of the vascular plant community (4.2 bits). This may have important implications for cross-continental comparisons of ecological recovery.

6 Conclusions

Quadratic entropy and taxonomic entropy violate both of Pielou's axioms of a meaningful diversity index. Neither is necessarily maximal at perfect evenness and can be maximized with the exclusion of a species [7, 8]. We suggest that a new axiom should be developed based on the evenness of taxonomic distances.

In all three cases studied, the dynamics of quadratic entropy greatly resembled the dynamics of Shannon entropy thereby providing no additional information regarding the recovery of the community, which is not entirely unexpected [11]. Despite its ability to better distinguish between rare cases where all species present are taxonomically equidistant, the utility of this measure is questionable as it was unable to provide different information than Shannon entropy. By making use of pairwise distance between species, quadratic entropy incorporates taxonomic relatedness into a diversity index in a highly intuitive manner, however, its use is also cumbered by its intensive computation relative to taxonomic entropy and the relative difficulty of isolating the taxonomic contribution to quadratic entropy.

While taxonomic entropy revealed little more than an increase (however slight) in taxonomic diversity with increasing distance from a source of perturbation, the taxonomic information gain revealed important information as to the changing contributions of taxonomic information and Shannon entropy to overall taxonomic diversity. It is likely this measure that will prove most useful in ecological monitoring at local scales.

Acknowledgments

C. Ricotta is gratefully acknowledged for advance copies of [8] and invaluable discussions. We would also like to thank M. Salemaa for generously providing the Finnish data. Funding was provided by the Natural Sciences and Engineering Research Council of Canada grant to M.A. and an Ontario Graduate Scholarship to R.D.

References

[1] Magurran, A.E. *Ecological Diversity and Its Measurement*, Princeton University Press: Princeton, New Jersey, pp. 1–45, 1988.

[2] Pielou, E. *Ecological Diversity*, John Wiley & Sons: Toronto, Ontario, pp. 1–17, 1975.

[3] Baczkowski, A.J., Joanes, D.N. & Shamia, G.M. Properties of a generalized diversity index. *Journal of Theoretical Biology*, **188**, pp. 207–213, 1997.

[4] Naeem, S. & Li, S. Biodiversity enhances ecosystem reliability. *Nature*, **390**, pp. 507–509, 1997.

[5] Warwick, R.M. & Clarke, K.R. New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. *Marine Ecology Progress Series*, **129**, pp. 301–305, 1995.

[6] Clarke, K.R. & Warwick, R.M. A taxonomic distinctness index and its statistical properties. *Journal of Applied Ecology*, **35**, pp. 523–531, 1998.

[7] Shimatani, K. On the measurement of species diversity incorporating species differences. *Oikos*, **93**, pp. 135–147, 2001.

[8] Ricotta, C. & Avena, G.C. An information-theoretical measure of taxonomic diversity. *Acta Biotheorica*, in press.

[9] He, X.S. & Orlóci, L. Comparative diversity analysis of vegetation. *Abstracta Botanica*, **17**, pp. 79–86, 1993

[10] Rao, C.R. Diversity and Dissimilarity Coefficients: A Unified Approach. *Theoretical Population Biology*, **21**, pp. 24–43, 1982.

[11] Izsák, J. & Papp, L. Application of the quadratic entropy indices for diversity studies of drosophilid assemblages. *Environmental and Ecological Statistics*, **2**, pp. 213–224, 1995.

[12] Gunn, J.M. *Restoration and Recovery of an Industrial Region: Progress in Restoring the Smelter-Damaged Landscape near Sudbury, Canada*, Springer-Verlag: Berlin and New York, 358 p., 1995.

[13] Anand, M., Tucker, B. & Desrochers, R. Ecological monitoring of terrestrial ecosystem recovery from man-made perturbation: assessing community complexity. *Proc. of the 10th Int. Conf. On Modelling, Monitoring and Management of Air Pollution*, eds C.A. Brebbia, J.F. Martin-Duque, WIT Press: Southampton, pp. 341–350, 2002.

[14] Salemaa, M., Vanha-Majamaa, I. & Derome, J. Understorey vegetation along a heavy-metal pollution gradient in SW Finland. *Environmental Pollution*, **112**, pp. 339–350, 2001.

[15] Enquist, B.J., Haskell, J.P. & Tiffney, B.H. General patterns of taxonomic and biomass partitioning in extant and fossil plant communities. *Nature*, **419**, pp. 610–613, 2002.