Computation and application of nematode community indices: general guidelines

DEBORAH A. NEHER AND BRIAN J. DARBY

Community indices, such as maturity or diversity, condense information regarding the structure and composition of communities into a single metric. Soil health and quality can be inferred from such indices by assuming that communities with different structure and composition function differently. Thus, these indices can be instrumental in monitoring soil and sediment quality as well as assessing ecosystem sustainability and biodiversity.

1 Historical perspective

In the 1980s, interest increased for using nematode communities as indicators for environmental monitoring of terrestrial communities (Freckman, 1988; Bongers, 1990). Initially, simple indices of abundance, proportions, or ratios of nematodes by trophic group were proposed. Subsequently, diversity indices were employed and a Maturity Index (MI) was developed for terrestrial nematodes (Yeates, 1970, 1984; Bongers, 1990). Later, the application of the MI was extended successfully to marine and brackish sediments (Bongers et al., 1991). Since then, others have applied the concept to freshwater systems (Beier and Traunspurger, 2003; Höss et al., 2004).

2 Maturity indices

Maturity indices are used as a measure of the ecological successional status of a soil community. They are based on the principle that different taxa have contrasting sensitivities to stress or disruption of the successional sequence because of their life-history characteristics. Bongers’ original MI proposal had separate indices for free-living (MI) and plant-parasitic nematodes (PPI). The index is represented by a colonizer-persister (c-p) value that ranges from a colonizer (c-p = 1) to a persister (c-p = 5) with the index values representing life-history characteristics associated with r- and K-selection, respectively. Those with a c-p = 1 are r-selected or colonizers, with short generation times, large population fluctuations, and high fecundity. Those with a c-p = 5 are K-selected or persisters, produce few offspring, and generally appear later in succession (Bongers and Bongers, 1998; Bongers and Ferris, 1999). Small and large c-p weights correspond with taxa relatively tolerant and sensitive to ecological disturbance, respectively (Table 1).

| TABLE 1 |
| Maturity Indices are computed as a weighted mean frequency, X = \( \sum \frac{v_i \times f_i}{n} \) where \( v_i = c-p \) value assigned to family, \( f_i \) - frequency of family \( i \) in sample, \( n = \) total number of individuals in a sample. MI may be viewed as a measure of disturbance, with smaller values being indicative of a more disturbed environment and larger values characteristic of a less disturbed environment. The MI decreases with increasing microbial activity and pollution induced stress, when opportunistic nematodes rapidly become dominant. Bongers (1990) proposed separate index calculations for... |
free-living (MI) and plant-parasitic (PPI) nematodes. The PPI may (Neher and Campbell, 1994) or may not (Bongers et al., 1997) correlate positively with MI.

Both stress and enrichment result in MI decrease; stress decreases numbers of sensitive species and enrichment increases abundance of opportunists. To differentiate MI-decreases caused by enrichment and pollution, it is better to omit cp-1 if the aim is to use MI to measure stress, resulting in MI25 (Bongers and Korthals, 1993; Bongers et al., 1995; Korthals et al., 1996). Opportunistic taxa (c-p = 1) are re-evaluated because they are considered enrichment opportunists and their population densities increase rapidly in response to additives of nutrients to soil and may not necessarily reflect long-term changes in soil ecological condition. Those with c-p values between 2 and 5 are more stable temporally and may provide relatively long-term information about environmental conditions. Therefore, it is recommended that marine Oncholaimidae be omitted because they accumulate to generate mono-populations under anoxic enriched conditions known to result in fish-dying. Furthermore, dauer larvae of Rhabditidae, Diplogasteridae, Panagrolaimidae are not included because they are non-feeding or inactive stages at the time of sampling.

Yeates (1994) and Wasilewska (1994) proposed a modification of the index based on merging free-living and plant-parasitic nematodes in a soil community (ΣMI). Neher and Campbell (1996) suggested a ΣMI25 which represents a combination of the concepts proposed by Bongers et al. (1995) and Yeates (1994). In each case, the same basic calculation is performed except different combinations of feeding groups (Yeates et al., 1993) and c-p groups are included. The difference in these two more comprehensive indices is that they include and exclude c-p 1 nematodes, respectively. Because no distinction between free-living and plant-parasitic nematodes is necessary, these indices solve a disagreement on whether to include (Bongers, 1990) or exclude (Neher and Campbell, 1996) Tylenchidae in the PPI. The controversy is based on whether Tylenchidae are considered fungal- or root-feeders.

3 Diversity

Diversity has been equated solely with numbers of taxa, and the popular press has perpetuated this misconception. Rather, a more appropriate indication of diversity integrates numbers of taxa (species richness) and equitability among taxa (species evenness) (Hurlbert, 1971). Two assumptions of most diversity indices include 1) an index limited to one taxonomic group, and 2) all species are equal (Cousins 1991). Good (1953) outlined a generalized diversity index that incorporates richness and evenness into a single value that generally increases with both richness and evenness;

\[ H(\alpha, \beta) = \sum_{i=1}^{S} p_i^\alpha \left(-\ln(p_i)\right)^\beta \]

where \( p_i \) is the relative abundance of taxon \( i \), \( S \) is the total number of species present, and \( \alpha \) and \( \beta \) define structural attributes of the algorithm. Good’s generalized diversity index demonstrates the mathematical relation of several commonly used diversity, dominance, and evenness indices (Table 2). Namely, Shannon’s diversity index can be interpreted as a variant of Good’s diversity index using values of 1 and 1 for \( \alpha \) and \( \beta \), respectively (or H(1,1)). Simpson’s dominance index can be interpreted as H(2,0). Rényi (1961) defined a notation of this generalized equation, which allows a user to put in different coefficients for species richness and dominance concentration (followed by Hill 1973), and Baczkowski et al. (1997) discuss optimal bounds of \( \alpha \) and \( \beta \) for ecological applications. Most other indices constitute logarithmic, exponential, reciprocal,
complementary, or relative transformations of the Shannon or Simpson indices. In addition to the main indices listed in Table 2, numerous alterations have been reported as evenness indices, including \((1-D)/(1-1/S)\), \((1-D)/S\), and \((-\ln D)/(\ln S)\) by Smith and Wilson (1996), and \((1/D-1)/(e^{\text{H'}}-1)\) by Alatalo (1981).

TABLE 2

These variations on a theme demonstrate unique sensitivities to changes in various structural attributes of an abundance distribution within a community. For example, the Camargo diversity index may be more sensitive for assessing structural alterations in aquatic communities than the H’ and MacArthur indices, which are sensitive to the number of taxa present and the whole spectrum of taxon proportions (Camargo, 1992). Secondly, the Camargo index typically increases with the addition of subordinate species (those defined as having a relative abundance less than \(1/S\)) more than most indices, valuing rare species. Beisel et al. (1996) argues that such sensitivity to rare species is an undesirable property of a diversity index and favors the Simpson and McIntosh indices that are more sensitive to changes in dominant taxa. Camargo (1997) rebuts with the argument that the “conviction that rare taxa should not contribute to the response of a dominance index is absurd.” In fact, the Camargo index is one of few diversity indices developed from an \textit{a priori} ecological definition of dominance, that dominance is the appropriation of potential niche space of certain subordinate species by other dominant species (McNaughton and Wolf, 1970).

The debate defining ‘dominance’ and ‘diversity’ raises a second question, that of which taxa to include in an index. Ideally a summary statistic such as a diversity index could relate the abundance structure of an entire community. For a freshwater nematode community, that would include all the bacterivores, algivores, herbivores, predators, and omnivores. However, if the McNaughton and Wolf (1970) definition of dominance is accepted as a complement to diversity, then logically only one trophic group should be included in any diversity or dominance index. This creates a practical problem because the specific feeding habits of most nematode species are diverse, changing, or unknown. In fact, the effect of each species on ecosystem processes has not been determined (Chapin et al., 1992). Secondly, the reduction of diversity indices to specific functional groups entirely contradicts the intended use of indices, that is, to summarize complex and varied community data into a single useful datum (Beisel, 1997).

Not all indices are contested so vigorously. Although the sensitivities of respective indices are still unclear, some generalization can be made from the literature. Typically, Shannon’s H’ index is sensitive to rare taxa, and Simpson’s \(\lambda\) index weights common taxa (Boyle et al., 1990). Hill’s family of diversity numbers are easy to interpret ecologically because the indices define units as taxa (Peet, 1974) but they are not necessarily superior from a statistical perspective (Heip et al., 1988; Ludwig and Reynolds, 1988). Hill’s diversity numbers \(N_0, N_1,\) and \(N_2\) are defined as numbers of all taxa, abundant taxa, and very abundant taxa, respectively (Ludwig and Reynolds, 1988). \(N_1\) equates with an antilog of a Shannon index \((e^{\text{H'}})\), and \(N_2\) equals the reciprocal of a Simpson index \((1/\lambda)\).

Diversity is linked artificially to the taxonomic resolution an investigator employs. Even though diversity is most often equated with species, it can be applied at various taxonomic levels of resolution, such as genotype, genus, family, and trophic group. For free-living nematodes, it is more common to apply diversity indices to taxonomic levels above species because species identifications based on morphology are difficult. Appropriate caution must be taken when
applying indices at the family or trophic group levels. Unfortunately, ambiguity in trophic
classification of nematodes usually occurs because it is inferred by morphology rather than
actual experiments on feeding preferences (Yeates et al., 1993). Furthermore, feeding-habit
groupings may be ambiguous and (or) not mutually exclusive in some cases. For example,
abundant populations of *Aphelenchoides*, *Tylenchus*, *Tylencholaimus*, and *Ditylenchus* can be
classified as ‘‘plant/fungal feeding’’ nematodes (Sohlenius et al., 1977), or some ‘‘predaceous’’
*Mesodorylaimus* sp. can grow and reproduce by feeding on bacteria (Russell, 1986). *Tylenchus*
spp. are often considered fungivores in ecological studies, but the basis of the judgment is
dubious because several species feed and reproduce on roots. In other cases, some species may
always be placed in one category and may have developmental stages or generations that fit in
another category.

Because diversity indices are based on relative abundances of community species they are
insensitive to taxonomic differences between species. Alternatively, biodiversity indices are
based on taxonomic relations among species and ignore species abundances. However, in many
environmental planning and protection programs, there is interest in both species composition
and relative distribution. Therefore, new combined indices have been proposed such as the
quadratic entropy index (Q) incorporates both species relative abundances and a measure for the
pairwise taxonomic differences between species (Izsak and Papp, 2000; Ricotta, 2002). Unluckily, these new combined indices violate part of the mathematical properties of an
ecological diversity index, so they are dubbed ‘weak diversity indices’.

Critical to valid interpretation of diversity indices are appropriate sampling and statistical
techniques. Generally, stratified- or simple-stage cluster sampling are touted as generating less
bias in diversity estimates than simple random sampling (Gimaret-Carpentier et al., 1998;
Mayoral, 1998; di Battista, 2002). Commonly, diversity indices are analyzed statistically with
traditional ANOVA procedures. However, care must be taken to insure that assumptions of
normality and equal variances are met, especially for small sample sizes. Sometimes,
distributions can be transformed to resemble a Gaussian distribution by application of log and
log-normal scales (Hill et al., 2003). Rogers and Hsu (2001) propose an asymptotically correct
method for diversity indices with unequal variances, when sample sizes are equal, and
transformations cannot remedy the situation.

4 Trophic group ratios

Ratios of trophic groups have been proposed to describe relative balance of positive to
negative impacts of nematodes on primary productivity or stage of decomposition. Wasilewska
(1989) proposed a ratio that computes the sum of fungivores and bacterivores divided by plant-
parasites. A ratio greater than one suggests the positive impacts of nematodes outweigh the
negative impacts on plant productivity. Two forms of decomposition pathway indices have been
proposed, differing in the denominator. Statistically, the form fungivores divided by the sum of
fungivores and bacterivores \([F/(F+B)]\) is considered mathematically more stable than simply
dividing fungivores by bacterivores \((F/B)\). Because each variation gives a contrasting result, it is
critical that the user define the ratio employed in the results of a report. Typically, individual or
ratios of trophic groups have not withstood the level of statistical rigor that maturity and trophic
diversity indices do in their ability to differentiate the ecological condition of soils on a large
geographic scale (Neher and Campbell, 1996; Neher et al., 1995).

5 Multivariate approaches
Canonical correspondence analysis (CCA) is useful to compare suites of taxon data with suites of environmental variables. Environmental variables can include treatment classes as nominal 0 or 1 variables or chemical properties or pollutants as continuous variables. Canoco (ter Braak and Smilauer, 2002) and Primer-E (http://www.pml.ac.uk/primer/index.htm) software are simple tools to perform these procedures. In Canoco, abundances were transformed as log (x + 1) prior to analysis. Transformations are unnecessary in Primer because the scaling is non-metric multi-dimensional scaling. CCA results are displayed graphically with bi-plot. In CCA biplots, each vector for an environmental variable defines an axis, and site or genera scores can be projected onto that axis (Jongman et al., 1995). An indication of relative importance of a vector is its length; the angle indicates correlation with other vectors and CCA axes. Eigenvalues for CCA axes indicate the importance of the axes in explaining relationships in the genera-environment data matrices. Unfortunately, CCA analyses are restricted to illustrating one instance in time.

Principal response curves is a multivariate method for the analysis of repeated measurement design. PRC is based on Redundancy Analysis (RDA); each experimental unit and sampling times and unit by time interactions are treated as dummy explanatory variables. The result is a diagram showing the sampling periods on the x-axis and the first Principal Component of the variance explained by treatment on the y-axis. For illustrative purposes, undisturbed condition was treated as a ‘control’, representing a zero baseline, and ‘disturbed’ of the same experimental unit as the ‘treatment’ to focus on the differences between the two states of condition through time. Monte Carlo permutation tests permuting whole time series are applied to compute statistical significance. Van den Brink et al. (2003) provide a review of the analytical procedure and detailed instruction is provided in the manual of Canoco Version 4.5 software (ter Braak and Šmilauer, 2002).
References


Table 1. Colonizer-persister (c-p) value assignments for select nematode families (Bongers and Bongers 1998)

<table>
<thead>
<tr>
<th>c-p value</th>
<th>Family</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rhabditidae, Diplogasteridae (s.l.), Panagrolaimidae, Bunonematidae</td>
</tr>
<tr>
<td>2</td>
<td>Cephalobidae, Plectidae, Monhysteridae, Aphelenchoididae</td>
</tr>
<tr>
<td>3</td>
<td>Teratocephalidae, Chromadoridae, Diphtherophoridae, Prismatolaimidae</td>
</tr>
<tr>
<td>4</td>
<td>Alaimidae, Mononchidae, Leptonchidae, Qudsianematidae, Dorylaimidae</td>
</tr>
<tr>
<td>5</td>
<td>Aporcelaimidae, Actinolaimidae, Thornenematidae, Belondiridae</td>
</tr>
</tbody>
</table>
Table 2. Common indices used to characterize the distribution of abundance within a community.

<table>
<thead>
<tr>
<th>Name</th>
<th>Equation*</th>
<th>Application</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shannon’s diversity</td>
<td>$H' = -\sum (p_i \ln p_i)$</td>
<td>This widely used and versatile index can be applied for both large and small sample sizes. The Shannon index is generally more influenced by rare species than Simpson’s index.</td>
<td>Shannon and Weaver (1949)</td>
</tr>
<tr>
<td>Hill’s diversity</td>
<td>$N_1 = \exp \left[ -\sum (p_i \ln p_i) \right] = \exp (H')$</td>
<td>An exponential form of Shannon’s $H'$, the value of this index can be interpreted as the number of abundant taxa (Ludwig and Reynolds 1988).</td>
<td>Hill (1973)</td>
</tr>
<tr>
<td>Brillouin’s diversity</td>
<td>$H = \frac{1}{N} \log \frac{N!}{\prod N_i !}$</td>
<td>Use only on fully censused communities because it is a true statistic and, thus, free from statistical error.</td>
<td>Brillouin (1962)</td>
</tr>
<tr>
<td>Camargo’s diversity</td>
<td>$d = \sum \left[ p_i - \left( \frac{1}{S} \right) \right]$</td>
<td>Estimates the (structural) asymmetry in relative abundance between dominant and subordinate species, not necessarily differences between dominant species or between subordinate species.</td>
<td>Camargo (1992)</td>
</tr>
<tr>
<td>Margalef’s diversity</td>
<td>$DMg = \frac{(S - 1)}{\ln(N)}$</td>
<td>Though simple to calculate, this index is unaffected by evenness or dominance and is sensitive only to species richness and sample size. Thus, its use should be restricted to comparing species richness among large communities.</td>
<td>Margalef (1958)</td>
</tr>
</tbody>
</table>

* $p_i$ represents the proportion of the $i$-th taxa in a sample, or $n_i$ the number, with $N$ individuals and $S$ total species
<table>
<thead>
<tr>
<th>Name</th>
<th>Equation*</th>
<th>Application</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simpson’s dominance (infinite community)</td>
<td>$D = \sum p_i^2$</td>
<td>Probability that two randomly chosen individuals of an infinite community belong to the same class, thus inversely related to diversity. It is often reported as $1 - D$, but see Hurlbert’s PIE.</td>
<td>Simpson (1949)</td>
</tr>
<tr>
<td>Hill’s reciprocal of $D$</td>
<td>$N_2 = (\sum p_i^2)^{-1} = 1/D$</td>
<td>The reciprocal of Simpson’s $D$, the value of this index can be interpreted as the number of very abundance taxa (Ludwig and Reynolds 1988)</td>
<td>Hill (1973)</td>
</tr>
<tr>
<td>Simpson’s dominance (finite community)</td>
<td>$\lambda = \frac{\sum n_i(i-1)}{N(N-1)}$</td>
<td>Similar to Simpson’s $D$, but corrected for finite communities. Mathematically it is usually more appropriate in ecological studies than Simpson’s $D$, but is used less often.</td>
<td>Simpson (1949)</td>
</tr>
<tr>
<td>Probability of interspecific encounter (PIE)</td>
<td>PIE = $\left( \frac{N}{N-1} \right) \left( 1 - \sum p_i^2 \right)$</td>
<td>Simpson’s dominance index $D$ converted to a diversity index and corrected for finite communities</td>
<td>Hurlbert (1971)</td>
</tr>
<tr>
<td>McIntosh dominance</td>
<td>$D1 = \left( \frac{N - \sqrt{\sum n_i^2}}{N - \sqrt{N}} \right) \approx \sqrt{\sum n_i^2}$</td>
<td>Recommended by Beisel et al. (1996) as the most relevant dominance index as most sensitive to variations on dominant taxa and not highly sensitive to variations on rare or medium taxa.</td>
<td>McIntosh (1967)</td>
</tr>
</tbody>
</table>

* $p_i$ represents the proportion of the $i$-th taxa in a sample, or $n_i$ the number, with $N$ individuals and $S$ total species
<table>
<thead>
<tr>
<th>Name</th>
<th>Equation*</th>
<th>Application</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brillouin’s maximum diversity</td>
<td>$H_{\text{max}} = \frac{1}{N} \ln \left( \frac{N!}{(\frac{N}{S})^{S-1} \left[ \left( \frac{N}{S} \right) + 1 \right]} \right)$</td>
<td>Represents maximum possible evenness of a sample of $N$ individuals and $S$ species.</td>
<td>Brillouin (1962)</td>
</tr>
<tr>
<td>Brillouin’s minimum evenness</td>
<td>$H_{\text{min}} = \frac{1}{N} \ln \left( \frac{N!}{(N - S + 1)} \right)$</td>
<td>Represents minimum possible evenness of a sample of $N$ individuals and $S$ species.</td>
<td>Brillouin (1962)</td>
</tr>
<tr>
<td>Brillouin’s evenness</td>
<td>$J = \frac{H}{H_{\text{max}}}$ or $J' = \frac{H'}{\ln S}$</td>
<td>Use $J$ for samples (and $J'$ for collections) to determine the evenness portion of diversity; $J$ or $J'$ represent observed and maximum diversity, respectively.</td>
<td>Pielou (1966)</td>
</tr>
<tr>
<td>Relative evenness ($J'$)</td>
<td>$V = \frac{H - H_{\text{min}}}{H_{\text{max}} - H_{\text{min}}}$</td>
<td>Unlike $J$ and $J'$, $V$ is not influenced by species richness ($S$)</td>
<td>Hurlbert (1971)</td>
</tr>
<tr>
<td>Heip’s evenness</td>
<td>$E_{\text{Heip}} = \frac{(e^{H'} - 1)}{(S - 1)}$</td>
<td>Hypothesized by Beisel et al. (2003) to be more sensitive to variations in rare species richness and/or abundance.</td>
<td>Heip (1974)</td>
</tr>
</tbody>
</table>

* $p_i$ represents the proportion of the $i$-th taxa in a sample, or $n_i$ the number, with $N$ individuals and $S$ total species.
<table>
<thead>
<tr>
<th>Name</th>
<th>Equation</th>
<th>Application</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morisita’s similarity</td>
<td>$C_H = \frac{2\sum_i p_{ij}p_{ik}}{\sum_j p_{ij} + \sum_k p_{ik}^2}$</td>
<td>Quantitative data; represents the degree of overlap between segment $j$ and all segments combined ($k$). It ranges from 0 (no similarity in community structure) to 1 (complete similarity) and is often expressed as a percentage ($C_H \times 100$).</td>
<td>Morisita (1959)</td>
</tr>
<tr>
<td>Bray-Curtis dissimilarity</td>
<td>$BC_{ij} = \sum \left</td>
<td>\frac{n_{ik} - n_{jk}}{n_{ik} + n_{jk}} \right</td>
<td>$</td>
</tr>
<tr>
<td>Jaccard similarity</td>
<td>$S_J = 100 \times \frac{c}{(a + b + c)}$</td>
<td>Binary data (presence/absence); represents the percent of taxa present that are similar to both groups.</td>
<td>Jaccard (1912)</td>
</tr>
<tr>
<td>Sørensen similarity</td>
<td>$CC = 100 \times \frac{2c}{(a + b + 2c)}$</td>
<td>Represents the percent similarity of each group with respect to taxa present.</td>
<td>Sørensen (1948)</td>
</tr>
</tbody>
</table>

*Quantitative data: $p_{ij}$ or $p_{ik}$ represents the proportion of the $i$-th taxa in sample $j$ or $k$, or $n_i$ the number. Binary data: $a$ the number of taxa unique to group A, $b$ the number of taxa unique to group B, and $c$ the number of taxa common to both group A and group B.*