

Chapter 4

General Community Indices that can be used for Analysis of Nematode Assemblages

Deborah A. Neher and Brian J. Darby

University of Vermont, Department of Plant and Soil Science, 105 Carrigan Drive, Burlington, Vermont, 05405, USA.

Introduction

The objective of classical community indices is to condense community data into one or a few variables to simplify analysis, interpretation, or review. To be successful as an indicator, a single index must be able to perform one of two functions: either reflect a past ecological process or predict a future ecological process. The success of community indices to reflect ecological processes or predict patterns depends on the relative completeness of ecological knowledge. A limitation of community indices is that they rely on *pattern* to reflect process, and often several processes can result in similar patterns. Productivity, resilience, and stability are some of the ecological functions relevant to ecosystem management, and some early successful attempts to link diversity with function include Rosenberg (1976) and Schafer (1973). However, the link between ecosystem process and diversity is not always clear even for well-studied communities, so it is not surprising that linkages between ecosystem processes and nematode diversity are also unclear (Ettema, 1998).

Two ecological approaches are necessary for community analysis, both **autoecology**, the study of an individual species and their interaction with the environment, and **synecology**, the study of a community of species interacting together in a predictable manner for several groups of organisms, including nematodes. Nematode communities differ in the degree to which their autecology and synecology are understood and, thus also vary in the potential for classical community indices to reflect ecological processes. Often simple univariate indices are more successfully applied to communities in which the autecology of community members and the synecology of the system results in processes that have distinct and well known patterns. This is not always the case for nematode communities, however. When the ecology of the community or system is poorly understood, more complex community assemblage and multivariate community analysis is required to discern patterns. Application of diversity indices to describe nematode communities is insufficient as a stand-alone indicator of ecological processes because the ecology of nematode communities is simply not known well enough for most habitats.

For ecological community assays, a few routine diversity indices can be reported in ecological studies to benefit meta-analyses in linking past, present, and future studies. However, the current state of knowledge does not permit univariate diversity indices to conclusively reveal ecological processes. Therefore, it is imperative to complement univariate

identity-independent approaches with multivariate identity-explicit approaches to improve our understanding of both the autecology of individual community members and synecology of the community as a whole. In this chapter, we first offer recommendations on performing some of the common identity independent ('diversity') indices and, secondly, suggest methods of incorporating community data into identity-explicit analysis with community assemblage and multivariate techniques.

Univariate Identity Independent Indices

In the broadest sense, diversity can refer to the sum of differences in form and function of life, including multiple scales of organization (ranging from the gene to the biome), space (with alpha diversity reflecting localities, beta diversity reflecting landscapes, and gamma diversity reflecting regions), and diversity of habitat and environmental disturbance types. The following section is concerned mostly with the representation of alpha diversity at the biological organization level of species and above. Although general ecological studies apply the following indices in the context of species, most nematode communities are enumerated at coarser resolutions because species identifications based on morphology are difficult (Neher, 2001). Besides, functional groups are a practical necessity because the effect of individual species on ecosystem processes has yet to be determined (Chapin *et al.*, 1992).

Diversity indices have their roots in post-World War II information theory with the goal of optimizing code length for digital communication. Theoretically, alpha diversity, richness, and evenness indices are applicable to any taxonomic level, which is thought to convey information, whether it is species, genus, family, or trophic group. The appropriate resolution should be determined by the objectives of the study. From an information-theory perspective, if information is lost at coarser resolutions then the corresponding index would be unlikely to distinguish among samples and statistical populations. From an ecological perspective, however, if ecological information is lost at fine resolution, the corresponding index may also be unlikely to distinguish among samples and treatments. To illustrate this, we have computed Shannon's diversity index at the species, genus, family, and trophic group level (Table 4.1) from data published by Yeates and Cook (1998). Each soil type exhibits a unique pattern of diversity between management practices when viewed at various levels of taxonomic resolution.

1 **Table 4.1.** Mean (and standard deviation) of Hill's N_1 diversity ($N_1 = \exp[H']$) at the species, genus, family, and trophic level
 2 for nematodes from organic and conventional grassland management regimes of three Welsh soils: Conway fine silt, Moor Gate
 3 coarse loam, Newport sand ($n = 10$ in all cases). Data from Yeates and Cook (1998).

	Silt [†]		Loam [†]			Sand [†]		
	Conventional	Organic	Conventional		Organic	Conventional		Organic
$N_{1\text{species}}$	10.7	13.1	14.0	**	17.4	17.8		15.8
	(3.70)	(3.48)	(3.02)		(2.61)	(3.70)		(2.47)
$N_{1\text{genus}}$	9.8	12.2	12.6	**	15.4	16.1	*	13.9
	(3.38)	(3.10)	(2.57)		(2.38)	(3.40)		(1.96)
$N_{1\text{family}}$	8.2	8.6	9.7	*	11.1	12.7	**	10.8
	(2.74)	(2.02)	(1.74)		(1.63)	(2.42)		(0.95)
$N_{1\text{trophic}}$	2.7	2.8	3.0		3.1	3.1	*	3.5
	(0.57)	(0.41)	(0.41)		(0.30)	(0.47)		(0.43)

4

5 [†]non-adjusted *t*-test between management regimes of similar soil type. * $P < 0.1$, ** $P < 0.05$;

1 One conceptual challenge with applying diversity, richness, or evenness indices to
2 trophic or functional groups is that a 'nematode community' does not represent an entire soil or
3 aquatic community but rather parts of several communities that may or may not interact with
4 each other directly; the remainder of the communities are comprised of organisms that may be
5 considered outside the scope of the study or simply not enumerable quantitatively from a
6 nematode extraction. For example, soil systems include several bacterivorous water-film faunal
7 groups. Only part of this group is represented by bacterivorous nematodes and the rest of the
8 group is composed of amoebae, flagellates, ciliates, rotifers, and other taxa. There is a paucity of
9 data that compare how changes in the composition of nematode trophic groups parallel those of
10 protozoan communities in the sample.

11 A second conceptual challenge with diversity indices is the interpretation of evenness in
12 reflecting community structure; the ecological implication of a uniform distribution of species
13 from multiple trophic groups is unknown. For example, ecologists may agree that a community
14 with low evenness (e.g., 19 species of very low relative abundance dominated by one or two
15 enrichment bacterivorous nematodes of high relative abundance) might represent a disturbed
16 community, relative to a community with intermediate evenness. However, a community with
17 evenly distributed abundance, e.g., where predators are equally abundant as microbivores, might
18 also reflect a recent disturbance. Furthermore, the mechanisms for how predators and
19 competitors (e.g., protozoa, tardigrades, and microarthropods) influence the diversity of
20 nematode species are still unclear.

22 ***Identity independent indices and their calculation***

23 A variety of identity-independent indices is available to serve different purposes in different
24 circumstances (Hill, 1973; Peet, 1974; Pielou, 1975). The total number of species collected from
25 a sample can be referred to as *species richness* (if representative of a known number of
26 individuals) or *species density* (if representative of a known number per mass or volume of soil).
27 *Evenness* is the equitable distribution of proportions or relative abundance. Diversity, then, is a
28 combination of both richness and evenness elements. Each diversity index weights richness and
29 evenness uniquely, but all diversity indices generally function so that an increase in either
30 richness or evenness will always increase diversity. In some reports, the term *diversity* continues
31 to refer simply to the total number of species; it is preferable, however, to restrict the use of
32 'diversity' to incorporate both the number of species and evenness. Formulae for calculating
33 several common indices are summarized in Table 4.2 and accompanied by a customized SAS
34 macro written to compute all indices (Table 4.3).

1 **Table 4.2.** Selected richness, diversity, and evenness indices that can be calculated for nematode
 2 communities.
 3

Name	Equation*	Reference
Margalef's richness	$D_{\text{Marg}} = \frac{(S - 1)}{\ln(N)}$	Margalef (1958)
Shannon's diversity	$H' = -\sum (p_i \ln p_i)$	Shannon (1948)
Hill's diversity	$N1 = \exp [-\sum (p_i \ln p_i)] = \exp (H')$	Hill (1973)
Simpson's dominance (infinite community)	$D = \sum p_i^2$	Simpson (1949)
Simpson's dominance (finite community)	$\lambda = \frac{\sum n_i(n_i - 1)}{N(N - 1)}$	Simpson (1949)
Hill's reciprocal of D	$N_2 = (\sum p_i^2)^{-1} = 1/D$	Hill (1973)
Brillouin's diversity	$H = \frac{1}{N} \log \frac{N!}{\prod N_i!}$	Brillouin (1962) Pielou (1975)
Brillouin's maximum diversity	$H_{\text{max}} = \frac{1}{N} \ln \frac{N!}{(X!)^{S-r} (Y!)^r}$	Brillouin (1962) Pielou (1975)
Brillouin's minimum diversity	$H_{\text{min}} = \frac{1}{N} \ln \frac{N!}{(N - S + 1)!}$	Brillouin (1962) Pielou (1975)
Brillouin's evenness	$J = \frac{H}{H_{\text{max}}} \text{ or } J' = \frac{H'}{\ln S}$	Brillouin (1962) Pielou (1975)
Brillouin's relative evenness	$V = \frac{H - H_{\text{min}}}{H_{\text{max}} - H_{\text{min}}}$	Hurlbert (1971) Pielou (1975)
Hill's evenness	$E_{2,1} = \frac{(N_2)}{(N_1)}$	Hill (1973)
Heip's evenness	$E_{\text{Heip}} = \frac{(e^{H'} - 1)}{(S - 1)}$	Heip (1974)

- 4 • 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. X (in
 5 Brillouin's maximum diversity) is the integer portion of (N/S) , $Y = X+1$, and $r =$ the remainder of X .

6

1

2 The procedure for enumerating nematodes should be standardized with each experiment
3 or sampling regime to prevent artifacts of sampling effort when reporting richness and diversity
4 indices. Nematodes are enumerated differently than, for example, vascular plant surveys, in that
5 nematodes are not enumerated as they appear *in situ*, but rather are extracted from soil, benthos,
6 or water samples. Nematode density generally varies widely from sample to sample, so the
7 number of nematodes enumerated is a representative subset of the total number extracted, i.e., an
8 unknown number at the time of sampling. Therefore, species richness is the appropriate term to
9 refer to the total number of species found when enumerating a uniform *number of extracted*
10 *individuals* (e.g., 200 from each sample) from samples of a uniform initial mass or volume.
11 Species density differs by referring to total number of species expressed as a uniform *portion of*
12 *all extracted individuals* (e.g., 20% of the individuals from each sample). Regardless, for
13 quantitative comparisons, it is preferable to begin with approximately the same initial mass or
14 volume of soil, water, or benthos. As a rough guide, the range of initial mass or volume for all
15 samples of comparison should be within 5% of the mean. Notice that species richness and
16 density are not necessarily linear in relationship. For example, 20 species found among 200
17 individuals may not extrapolate to 40 species from 400 individuals. For this case, *rarefaction* of
18 original data is necessary to estimate the number of species collected from a hypothetical number
19 of individuals from the same population and species abundance curve (Gotelli and Colwell,
20 2001). It is essential to explicitly state the conditions in which richness indices were computed.
21 Sometimes it is not practical or impossible to enumerate a uniform number or portion of
22 individuals, such as nematodes collected from small, isolated habitats such as pitcher plants,
23 epiphytes, or an insect. In such a case, Margalef's index ($= [S - 1] / \ln[N]$) can be used to adjust
24 the number of species (S) for the number of individuals enumerated (N).

25 Evenness indices appear infrequently in the literature. Heip (1974) proposed an evenness
26 index ($= [\exp(H') - 1] / [S - 1]$) which standardizes the Shannon's diversity index (H') by total
27 number of species (S). Alternatively, Brillouin (1962) developed a series of statistics for
28 censused communities that are computationally complex. Brillouin's maximum theoretical
29 diversity is computed with the assumption that all individuals are distributed as uniformly as
30 possible, and minimum theoretical diversity is computed assuming all individuals are distributed
31 as asymmetrically as possible. Two forms of evenness can be computed, the first as diversity
32 relative to maximum diversity and the second ('relative evenness') as diversity relative to
33 maximum diversity but scaled to minimum diversity. The former relative evenness (not scaled to
34 minimum diversity) can be based on two estimates of diversity depending on whether the user
35 wishes to assume a finite or infinite community enumeration. Use Brillouin's sample diversity
36 relative to Brillouin's maximum diversity when assuming a finite community enumeration, or
37 Shannon's population diversity relative to the natural logarithm of richness when assuming
38 infinite community enumeration. The second 'relative' form of evenness (scaled to minimum
39 diversity) uses Brillouin's calculation of diversity from a censused community. Although
40 nematode communities are rarely, if ever, fully censused in nature, the assumption of complete
41 enumeration may be appropriate in some unique applications, e.g., small isolated habitats or
42 virtual individuals in a computationally simulated model community.

43 Ecologists disagree on the best method to incorporate both richness and evenness, as well
44 as the degree to which dominant and rare species, respectively, should influence the index.
45 Therefore, exercise caution in application and interpretation of diversity indices. Shannon's

1 diversity ($= -\sum [p_i \ln p_i]$, Shannon 1948) is a popular diversity index. The exponent of Shannon's
 2 index ($= \exp[H]$, also called Hill's N1) can be interpreted as the number of uniformly distributed
 3 species that would produce an identical Shannon's index as the non-uniformly distributed
 4 community. For example, consider a community with 20 non-uniformly distributed species and a
 5 Shannon's index of 2.3. The exponent of 2.3 (Hill's N1) equals 9.97, so, intuitively,
 6 approximately 10 uniformly distributed species would be needed to produce a Shannon's index
 7 similar to the community of 20 non-uniformly distributed species. Furthermore, Heip's evenness
 8 index (above) $= ([9.97 - 1] / [20 - 1]) = 0.47$, indicating again that roughly half of the observed
 9 species would be necessary to produce a similar Shannon's index if they were uniformly
 10 distributed. Simpson's index (Simpson, 1949), $D (= \sum p_i^2)$, is considered a dominance index
 11 because it increases as species are distributed more unevenly (an increase in dominance) and can
 12 be interpreted intuitively as the probability that two randomly selected individuals from an
 13 infinite community will be the same. The reciprocal of Simpson's index (Hill's N2 $= 1 / D$) is
 14 often reported as a diversity index, and like Hill's N1, Hill's N2 can be interpreted as the number
 15 of uniformly distributed species that would produce a Simpson's index identical to that of the
 16 non-uniform community. Notice that the minimum Simpson's D possible (i.e., least dominance
 17 by any taxa) is $1/S$ and the maximum Hill's N2 possible (greatest equitability) is S , so we could
 18 compute an evenness index similar to Heip's approach as $N2 / S$.
 19

20 **Table 4.3.** Annotated SAS/IML code to illustrate an approach to implementing the univariate
 21 indices of Table 3.2 from within SAS.

```

22
23 %let species = a b c d e f g h;

24 data countdata;
25     input sample &species;
26 cards;
27 1      12      18      17      12      3      4      18      15
28 2      10      11      28      8      26      3      7      8
29 3      10      11      19      22      18      6      3      13
30 4      12      18      9      13      18      15      12      4
31 5      19      5      4      26      4      11      17      14
32 6      15      18      0      11      14      13      14      14
33 7      14      12      19      12      6      9      13      16
34 8      16      15      19      18      5      9      2      16
35 9      20      4      16      7      26      3      3      21
36 10     14      23      27      5      0      12      14      4

37 ;

38 proc IML;
39 use countdata;
40 read all var {&species} into data;
41 /* N = column vector of count sums */;
42 N = data[,+];
43 /* p = matrix of proportions */;
44 p = j(nrow(data),ncol(data),0); *Pre-allocate space;
45 p = data # (1/N); *elementwise division of data by sums;
46 /* richness = column vector of taxa present */;
47 richness = (data>0)[,+]; *only data > 0 are used;

```

```

1  MargalefsD = (richness - 1) # (1 / log(N)); /* Margalef's D index as richness corrected for N */;
2  /* Shannon's H index as the opposite of the sum of all proportions times ln(proportions) */;
3  nonzeros = loc(p>0); *nonzeros = a row vector of elements of p that are present;
4  ShannonH = j(nrow(data),ncol(data),0); *Pre-allocate space to speed up computation;
5  ShannonH[nonzeros] = p[nonzeros] # log(p[nonzeros]); *all absent species remain zero;
6  ShannonH = -ShannonH[,+]; *opposite to the sum of all columns;
7  /* Simpson's D dominance index (community and sample) as the sum of all proportions squared */;
8  SimpsonD = p[,##];
9  Simpsonfinitelambda = ((data # (data - 1))[,+]) / (N # (N - 1));
10 /* Hill's diversity (N1 and N2) and evenness (E21) */;
11 HillsN1 = exp(ShannonH); *exponent of Shannon's H ;
12 HillsN2 = 1/SimpsonD; *inverse of SimpsonD;
13 HillsE21 = HillsN2 / HillsN1; *ratio of N2 to N1;
14 /* Heips alternative evenness */;
15 BrillouinJprime = ShannonH/log(richness);
16 HeipE = (HillsN1 - 1) / (richness - 1);
17 /* Brillouin's indeces */;
18 if any(N>=100) then largeN = loc(N>=100); *largeN = a row vector of locations where N >= 100;
19 if any(N<100) then smallN = loc(N<100); *smallN = a row vector of locations where N < 100;
20 /* SAS fact(n) may not compute factorials for large n (> 100) so it is necessary to run */;
21 /* an alternate module to compute the log of n! by suming a vector of 1:n */;
22 /* IML module to compute natural log of factorial of large (>100) n */;
23 /* ----- */;
24 /* use: factorial(n)    returns: log(n!)    */;
25 start factorial(n);
26 factorial = j(nrow(n),1,0);
27 do k = 1 to nrow(n);
28     a = 1:n[k];
29     temp = log(a);
30     factorial[k] = temp[,+];
31 end;
32 return (factorial);
33 finish;
34 BrillouinH = j(nrow(data),ncol(data),0); *Pre-allocate space;
35 BrillouinH = log(fact(data))[,+]; *Sum of log(Ni!), which equals the logarithm of the products of Ni!;
36 logNfact = factorial(N); *Call IML module factorial(n) for n > 100;
37 BrillouinH = (1/N) # (logNfact - BrillouinH); *final BrillouinH calculation
38 /* Brillouin's Hmax */;
39 intBrillouinHmax = int(N # (1/richness)); *integer portion;
40 r = mod(N,richness); *remainder, or modulus;
41 BrillouinHmax = j(nrow(data),1,0); *Pre-allocate space;
42 do j = 1 to nrow(data); *repeat for each row;
43 BrillouinHmax[j] = log((factorial(intBrillouinHmax[j]) ## (richness[j] - r[j])) # (factorial((intBrillouinHmax[j] + 1)
44 ## r[j]));
45 end;
46 BrillouinHmax = (1/N) # (logNfact - BrillouinHmax); *final computation
47 /* Brillouin's Hmin */;
48 BrillouinHmin = factorial(N - richness + 1);
49 BrillouinHmin = (1/N) # (factorial(N) - BrillouinHmin);
50 /* Brillouin's evenness (J for samples, Jprime for collections) and relative evenness (Vrel) */;
51 BrillouinJ = BrillouinH / BrillouinHmax; * ;
52 BrillouinJprime = ShannonH / log(richness); * ;
53 BrillouinVrel = (BrillouinH - BrillouinHmin) / (BrillouinHmax - BrillouinHmin); * ;
54 print N richness MargalefsD ShannonH SimpsonD Simpsonfinitelambda HillsN1 HillsN2 HillsE21
55     BrillouinH BrillouinHmax BrillouinJ BrillouinJprime BrillouinHmin BrillouinVrel;
56 CREATE indices var {richness MargalefsD ShannonH SimpsonD Simpsonfinitelambda HillsN1 HillsN2 HillsE21

```

```

1         BrillouinH BrillouinHmax BrillouinJ BrillouinJprime BrillouinHmin BrillouinVrel});
2 APPEND;
3 quit;
4 data final;
5         merge countdata(keep=sample) indices(keep = richness MargalefsD ShannonH SimpsonD
6 Simpsonfinitelambda HillsN1 HillsN2 HillsE21
7         BrillouinH BrillouinHmax BrillouinJ BrillouinJprime BrillouinHmin BrillouinVrel);
8 run;
9 proc print data = final;
10 run;
11

```

12

13 **Community Assemblage Models**

14

15 ***Ecological succession***

16

17 Ecological succession refers to a relatively predictable or directional sequence of spatio-temporal
18 patterns of ecological interactions within a community. As species composition changes, it alters
19 the abiotic environment, which in turn selects against the existing community favoring a
20 community composition that performs better under the newly created abiotic environment. The
21 concept originated in plant ecology (Whittaker, 1975), but also applies to invertebrate
22 communities in soil and sediment. Succession usually progresses directionally unless set back
23 by an environmental disturbance such as cultivation, pollution, or nutrient enrichment (Neher,
24 1999). Therefore, quantitative measures of ecological succession can serve as indicators of
25 disturbance. With improved knowledge of synecology of nematode communities, one could
26 identify the type and intensity of disturbance based on an index of succession.

27 Bongers (1990) proposed an index of ecological succession for application to nematodes
28 whereas Ruf (1998) applied a similar approach to mesostigmatid mites. An alternative approach
29 is to quantify species assemblage patterns. This can be achieved by repeated sampling methods
30 or a Mantel test (Manly, 1997). These approaches are computationally intensive but practical
31 given the speed of current computer systems. Repeated sampling methods include techniques
32 referred to as bootstrap, resampling, jackknife, randomization, and Monte Carlo (Manly, 1997).
33 A Mantel test computes a correlation coefficient among matrices. Each matrix can represent an
34 assemblage of species in a community through time. Data can be entered as raw or transformed
35 in variables that are continuous, ordinal or binary (Peres-Neto and Jackson, 2001). One can test
36 hypotheses that concern the (dis)similarity of order and composition between two communities
37 or treatments. A third variable, e.g., spatial pattern can be adjusted by using a partial Mantel test.
38 These approaches are rank or distribution-free which allow them to be applied to small and
39 unbalanced data. Data are reshuffled or resampled repeatedly for 10,000 to 100,000 times to
40 compute a *P*-value and confidence intervals. The level of significance possible is affected by the
41 choice of distance measure (Jackson, 1995). Distance can be quantified in Euclidean and non-
42 Euclidean spaces (e.g., genetic distance, Bray Curtis). The methods vary in their sampling
43 approach (i.e., with or without replacement) and the sample size (i.e., replacing the whole or

1 subset of original sample). In addition to ecological succession, these approaches can also be
2 applied to quantifying other ecological phenomena including intrinsic rate of increase (r),
3 estimate of genetic distance, ecological divergence or phylogenies, microarrays, and
4 biogeography (Felsenstein, 1985; Hillis and Bull, 1993; Jackson, 1995; Efron *et al.*, 1996; Diniz-
5 Filho *et al.*, 1998; Kerr and Churchill, 2001; Rossi, 1996, 2003).

6 An alternative approach is the Procrustean superimposition approach (Gower, 1971). It
7 differs from Mantel by scaling raw data or their ordination solutions to find optimal
8 superimposition rather than transformation. This avoids the problem that the space between
9 distances of transformed variables may not be necessarily equivalent to ones taken from the
10 space of the original data. This approach is more powerful and results in lower type I error rates
11 than the Mantel test (Peres-Neto and Jackson, 2001). Commercial computer software is
12 available to compute most of these indices (Table 5.4).

1 **Table 4.4.** Software packages containing univariate statistical procedures.

2

			EstimateS ¹	Primer-E ²	R ³	Canoco 4.5	PC-ORD 4	PROTEST ⁴	NTSYS-PC ⁵
Univariate	Diversity			x		x	x		
		Shannon	x						
		Simpson	x						
		Margalef's							
		Evenness				x	x		
		Brillouin							
Ecol Succ	Maturity								
	Simple Mantel			x	x		x		x
	Partial Mantel				x				
	Procrustes							x	x
	Jackknife		x				x		x
	Bootstrap		x						x

3

4 ¹: EstimateS, Statistical Estimation of Species Richness & Shared Species from Samples, viceroy.eeb.uconn.edu/estimates5 ²: Primer-E, version 5 available, Plymouth Routines in Multivariate Ecological Research, www.primer-e.com)6 ³: The R Package, (<http://www.bio.umontreal.ca/Casgrain/en/labo/R/v4/index.html>, French version also available)7 ⁴: PROTEST software available on <http://www.zoo.utoronto.ca/jackson/software/>8 ⁵: NTSYS-PC (Rohlf 1994), Version 2.2 was initially released Sept. 2005, <http://www.exetersoftware.com/cat/ntsyspc/ntsyspcfaq.html>, Exter Software Inc., 100

9 North Country Road, Building B, Setauket, NY 11733

1 **Neutral community assemblage models**

2

3 In addition to the niche-based models that are the impetus for the successional, seasonal,
4 disturbance, and habitat-based studies that dominate historical nematode community analyses,
5 non-neutral models present a necessary alternative perspective to spatial-temporal dynamics of
6 communities. Neutral models have been in use for some time but were brought to the forefront of
7 ecology as Hubbell (2001) presented a neutral model of community dynamics whereby
8 individuals immigrate from a metacommunity into a local community at random with
9 ecologically equivalent fitness. The spatio-temporal dynamics of neutral communities resemble
10 neutral drift, analogous to random genetic drift. The surprising result of the neutral community
11 model is that the distribution of species abundance closely resembles natural abundance
12 distributions. This finding is controversial because, although neutral models can predict many
13 ecological patterns, Hubbell's implicit assumption of neutrality (i.e., ecological equivalence)
14 challenges nearly 150 years of niche-based ecology that sought to delineate the niche boundaries
15 of what were believed to be non-neutral species. However, many authors suggest that neutral and
16 non-neutral models of community assemblage may not necessary be contradictory, but rather
17 complementary (Chave, 2004). Neutral dynamics may occur when non-neutral interactions play
18 out on a reciprocal non-uniform fitness landscape, in effect, equalizing fitness. Neutral processes
19 likely occur within nematode communities because many species appear to be functionally
20 redundant. However, nematodes may not be suitable to test neutral theories because most
21 representative neutral models are spatially and temporally explicit (Holyoak and Loreau, 2006)
22 and the destructive nature of nematode enumerations prevent truly repeated samplings of the
23 same volume. It is important to remember that neutral dynamics may occur over the course of an
24 experiment and it may not be advantageous to force niche-based explanations onto what may be
25 neutral dynamics.

26

27 **Multivariate techniques**

28

29 Multivariate analysis offers both descriptive and inferential procedures to analyze multiple
30 variables simultaneously so as to reveal the collective interactions of all variables and the effect
31 each variable has on the others. *Descriptive* procedures help to illustrate the overall structure of a
32 dataset while *inferential* procedures help to test hypotheses of interactions. Therefore,
33 multivariate analysis has two complementary applications, *exploratory hypothesis-generating*
34 and *inferential hypothesis-testing*, that can be combined into a two-phase approach that might
35 begin with an exploratory phase that seeks patterns in nature by asking "to what can I ascribe the
36 variation in my data?" The second phase, then, tests the hypotheses that were generated by
37 asking "can I reject the null hypothesis that species are unrelated to each other or postulated
38 environmental factor(s)?" In this way, multivariate analysis is useful in evaluating nematode
39 community structure as a biological indicator by keeping the identity of individual taxa explicit
40 throughout the analysis. Below, we discuss two types of multivariate analysis commonly applied
41 to nematode communities, cluster analysis and ordination (see also Trett *et al.*, Chapter 12, this
42 volume). Commercial software packages that compute these procedures are summarized in Table
43 4.5.

Table 4.5. Software packages containing multivariate statistical procedures.

Category	Statistic	SAS ¹	SPSS ²	Statistica ³	SYSTAT ⁴	Canoco ⁵	Primer-E ⁶	PC-ORD ⁷	SYNTAX	R
	Cluster	CLUSTER FASTCLUS	CLUSTER	Join	Join		X	X	X	
	Discriminant	DISCRIM STEPDISC	DISCRIM- INANT	Discriminant	MGLH		X			
Direct Gradient	Canonical correspondence analysis (CCA)					X		X	X	
	Nonmetric Multidimensional Scaling (NMDS)						X	x	X	
	Redundancy Analysis (RDA)					X			X	
	Detrended canonical correspondence analysis (DCCA)					X				
	Canonical correlation analysis	CANCORR								
Indirect gradient	Polar (= Bray- Curtis) Ordination(PO)					X		X		
	Principal Coordinates Analysis (PCoA)					X				
	Principal Components Analysis (PCA)	PRINCOMP FACTOR	FACTOR	Factor	Factor	X	X	X	X	
	Correspondence analysis (CA)	CORRESP				X		X	X	
	Detrended correspondence analysis (DCA)					X		X		
	Principal					X				

	Response Curves (PRC)									
--	--------------------------	--	--	--	--	--	--	--	--	--

- ¹: SAS Version 9.1, (<http://www.sas.com/>)
- ²: SPSS Version 15 (<http://www.spss.com/>)
- ³: Statistica Version 8 (<http://www.statsoft.com/>)
- ⁴: SYSTAT Version 12 (<http://www.systat.com/>)
- ⁵: Canoco Version 4.5, (<http://www.microcomputerpower.com/>)
- ⁶: Primer-E: Plymouth Routines in Multivariate Ecological Research, www.primer-e.com)
- ⁷: PC-ORD, Version 4, MjM Software Design

1 **Cluster analysis**

2

3 Cluster analysis treats each multivariate observation (sample) as a vector and attempts to group
 4 vectors that are similar to each other into clusters (see Figure 12.5). Cluster analysis begins with
 5 a (dis)similarity matrix, often computed as the Euclidean distance among all pairs of vectors.
 6 *Hierarchical clustering* algorithms are either agglomerative or divisive. *Agglomerative*
 7 *clustering* begins with each vector representing a unique cluster and sequentially combining the
 8 two nearest clusters into one until an optimal number of clusters have been obtained. Divisive
 9 clustering begins with one cluster containing all vectors and sequentially divides the cluster into
 10 two until an optimal number of clusters have been obtained. Agglomerative clustering is most
 11 common and there are several methods of determining the distance of vector clusters from each
 12 other. The single linkage (or nearest neighbor) method determines the distance between two
 13 clusters as the minimum distance (e.g., Euclidean) between the two most similar vectors of each
 14 cluster, while the complete linkage (e.g., farthest neighbor) method determines the distance
 15 between two clusters as the maximum distance (e.g., Euclidean) between the two most dissimilar
 16 vectors of each cluster. The average linkage method defines the distance between two clusters as
 17 the average distance of all elements from each cluster, while the *centroid method* defines the
 18 distance between two clusters as the distance between the two mean (or median) vectors of a
 19 cluster, called the centroids. Finally, *Ward's method* joins clusters so as to minimize the increase
 20 in sum of squares within and between clusters. The result of hierarchical cluster analysis is a
 21 dendrogram (i.e., tree diagram) that shows each step of the clustering procedure and the distance
 22 at which the clusters merge.

23 Discriminant Analysis is a related approach based on an *a priori* expectation of group
 24 members whereas cluster analysis has no preconceived expectation of group members and
 25 therefore conducts *a posteriori* aggregation. With discriminant analysis, one hypothesizes that
 26 there are two or more distinct groups, and then determines whether the observations divide
 27 significantly among those two predicted groups (Afifi and Clark, 1997; McGarigal *et al.*, 2000;
 28 Gotelli and Ellison, 2004).

29

30 **Ordination**

31

32 Ordination techniques are popular in community analysis due to their ability to visualize data in
 33 two-dimensional space. There are two main classes of ordination techniques, direct and indirect
 34 gradient analysis. *Indirect gradient analysis*, also called unconstrained, seeks to interpret patterns
 35 from within a dataset. *Direct gradient analysis*, also called constrained, seeks to extract patterns
 36 from known gradients and is therefore *constrained* by the environmental variables supplied.
 37 Indirect gradient analysis is divided into distance-based and eigenanalysis-based methods
 38 whereas all direct gradient analyses are eigenanalysis-based methods. Examples of distance-
 39 based indirect gradient ordination include Polar Ordination (PO), Principal Coordinates Analysis
 40 (PCoA) and Nonmetric Multi-Dimensional Scaling (NMDS). In polar ordination, two samples
 41 most different from each other based on their species composition serve as endpoints and all
 42 other samples are plotted relative to them. In this way, new samples can be added to polar
 43 ordination without changing the structure of the ordination diagram. Principal Coordinates

1 Analysis simply maximizes linear distance measures of the ordination in metric space (using a
 2 distance matrix), while NMDS is analogous to a non-parametric variant of PCoA by maximizing
 3 *rank* distance measures of the ordination in non-metric space. Computer software packages are
 4 commercially available to compute any of these methods (Table 4.5).

5 The concept of eigenanalysis, used in the remaining indirect and direct gradient analyses,
 6 is important but somewhat more tedious. Eigenanalysis is a procedure to reduce the
 7 dimensionality of data that also begins with a square distance, similarity, correlation or
 8 covariance matrix. The result of eigenanalysis, an *eigenvalue* and its corresponding *eigenvector*,
 9 describes the data matrix as a multidimensional volume. Consider a cluster of samples with three
 10 species (*x*, *y*, and *z*) plotted in three-dimensional space (i.e., along axes *x*, *y*, and *z*) that take the
 11 shape of a ball. Eigenanalysis of these data circumscribes a volume around the points and the
 12 dominant eigenvalue (one of three) describes the length of the longest dimension and so
 13 describes the greatest amount of variance in the data. If the second eigenvalue, which is the
 14 length of the second longest dimension, is much shorter than the first eigenvalue, the ellipse
 15 around the points in these two dimensions is oblong, and the three-dimensional volume
 16 resembles the shape of a rugby football. Just as eigenvalues describe the shape of the volume of
 17 data points, eigenvectors describe the orientation of the data points, with the first eigenvector
 18 defining the orientation of the first eigenvalue, and so on. In this way, datasets with multiple
 19 variables can be visualized in multi-dimensional space derived from latent gradients.

20 Both direct and indirect eigenanalysis methods are available to model either linear or
 21 unimodal (humped or convex) responses to environmental gradients. Within indirect
 22 eigenanalysis methods, Principal Components Analysis (PCA) is a special eigenanalysis
 23 expression of PCoA using Euclidean distance and models a linear response of variables while
 24 Correspondence Analysis (CA) models a unimodal response of variables. Within direct
 25 (constrained) methods, Redundancy Analysis (RDA) models a linear response of variables while
 26 Canonical Correspondence Analysis (CCA) models a unimodal response of variables.
 27

28 ***Data transformation***

29
 30 The multivariate normal distribution is analogous to the univariate normal distribution. As
 31 multivariate ordination is an extension of multiple regression, the transformation of variables for
 32 multivariate analysis is similar to the problem of transforming variables for regression. Lepš and
 33 Šmilauer (2003) advise against strict adherence to traditional Gaussian (i.e., normal)
 34 distributions, but rather recommend “the semantics of the hypothesis you are testing”. To them,
 35 the effect of violating multivariate normality on the results of multivariate analysis and
 36 ordination is unclear and often considered insignificant. If one wishes to interpret the association
 37 between variables on a scale of *one measurement unit*, then it is acceptable to use non-
 38 transformed variables. However, many animal populations follow alternative scales, such as a
 39 logarithmic or square-root scale. In these cases, it is appropriate to apply a logarithm or square-
 40 root transformation for species data. In the case of log transformations, a natural logarithm (\ln or
 41 \log_e) is used more often than a \log_{10} -transformation, although both give similar results. The only
 42 computational restrictions to transformations are zeros and negative values. Linear multivariate
 43 models (as in PCA and RDA) can employ negative values so that the data can even be centered
 44 and standardized. However, unimodal models (as in CA and CCA) cannot employ negative

1 numbers and so logarithm transformations employ a constant such as $\log(x + b)$ where b is some
2 small constant to accommodate zeros. Again, NMDS is a non-parametric analog to multivariate
3 analysis.
4

5 **Visualization**

6
7 The result of ordination is a biplot or diagram that illustrates the data, either species scores or
8 sample points or both, plotted in multivariate space typically along two axes, but sometimes
9 three. The goal is to explain up to 80% of the variation. If environmental variables are included
10 to constrain the analysis, such as with direct gradient analysis, it is possible to overlay the
11 environmental variables as vectors through the observations. In this case the length of the vector
12 represents its descriptive importance, the direction indicates the vectors correlation with various
13 sites or species, and the angle between vectors reflects the variable's correlation with other
14 variables. These ordination plots are generally used to represent snapshot datasets taken at one
15 point in time. Principal Response Curves (PRC) is a novel approach to representing repeated
16 measures multivariate data, especially in the context of community responses to stress (Van den
17 Brink *et al.*, 2003; Van den Brink and ter Braak, 1998, 1999). The ordination begins with
18 redundancy analysis (the linear model of direct gradient analysis) with time as a covariate. The
19 resulting diagram is a community coefficient along the y -axis plotted against time along the x -
20 axis with the community coefficient representing the relative change in community composition
21 (as indicated by an accompanying list of species scores) of treatment populations against a
22 control population.
23

24 **Conclusion**

25
26 Classical community composition can be analyzed using metrics that either disregard or preserve
27 the identity of taxon within the community. Identity-independent methods such as diversity and
28 evenness indices are relatively simple to compute and analyze statistically. However, the user
29 must exercise caution by selecting the form of index most appropriate to the goals of the study,
30 and resisting the temptation to singularly extrapolate to a greater ecological meaning without
31 substantial supplementary evidence. Alternatively, indices that incorporate and/or maintain taxon
32 identity can more convincingly be linked to ecological process and function. Measures of
33 ecological succession and species assemblage are univariate forms that can be analyzed using
34 traditional statistical tools such as regression and analysis of variance. Given the advances in
35 computer technology, a variety of multivariate methods are accessible through commercial
36 software packages. Many multivariate approaches capture a one-time “snapshot” of community
37 composition. However, repeated measures approaches are becoming available to evaluate
38 changes in community composition through time. Practitioners should be aware of the many
39 limitations, assumptions, and caveats of community assemblage and multivariate techniques by
40 consulting with expert statisticians. We recommend Legendre and Legendre (1998), Afifi and
41 Clark (1997), Lepš and Šmilauer (2003), and Gotelli and Ellison (2004) as helpful texts for
42 further study of these techniques.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

References

- Afifi, A.A. and Clark, V. (1997) *Computer-Aided Multivariate Analysis*. 3rd edn. Chapman and Hall, New York, New York, USA.
- Bongers, T. (1990) The maturity index: An ecological measure of environmental disturbance based on nematode species composition. *Oecologia* 83, 14-19.
- Brillouin, L. (1962) *Science and Information Theory*, 2nd edn. Academic Press, New York.
- Chapin, F.S., Schultz, E.D. and Mooney, H.A. (1992) Biodiversity and ecosystem processes. *Trends in Ecology and Evolution* 7, 107–108.
- Chave, J. (2004) Neutral theory and community ecology. *Ecology Letters* 7, 241-253.
- Diniz-Filho, J.A.F., Oliveira, L.G., and Silva, M.M. (1998) Explaining the beta diversity of aquatic insects in "Cerrado" streams from central Brazil using multiple Mantel test *Revista Brasileira de Biologia* 58, 223-231.
- Efron, B., Halloran, E., and Holmes, S. (1996) Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences USA* 93, 13429-13429.
- Ettema, C.H. (1998) Soil nematode diversity: Species coexistence and ecosystem function. *Journal of Nematology* 30, 159–169.
- Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783-791.
- Gotelli, N.J. and Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4, 379-391.
- Gotelli, N.J. and A.M. Ellison. (2004). *A Primer of Ecological Statistics*. Sinauer Associates, Sunderland, Massachusetts, USA.
- Gower, J.C. (1971) Statistical methods of comparing different multivariate analyses of the same data. In: Hodson, F. R., Kendall, D. G., and Tautu, P. (eds) *Mathematics in the Archaeological and Historical Sciences*. Edinburgh University, Edinburgh, Scotland, pp. 138-149.
- Heip, C. (1974) A new index measuring evenness. *Journal of the Marine Biological Association* 54, 555–557.
- Hill, M.O. (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology* 54, 427-432.
- Hillis, D.M. and Bull, J.J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* 42, 182-192.
- Holyoak, M. and Loreau, M. (2006) Reconciling empirical ecology with neutral community models. *Ecology* 87, 1370-1377.
- Hubbell, S.P. (2001) *The Unified Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, New Jersey, USA.
- Hurlbert, S.H. (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52, 577-586.
- Jackson, D.A. (1995) PROTEST: A PROcrustean randomization TEST of community environment concordance. *Ecoscience* 2, 297-303.

- 1 Kerr, M.K. and Churchill, G.A. (2001) Bootstrapping cluster analysis: Assessing the reliability
2 of conclusions from microarray experiments. *Proceedings of the National Academy of*
3 *Sciences USA*. 98, 8961-8965.
- 4 Legendre, P. and Legendre, L. (1998) Numerical Ecology. 2nd English edition. Elsevier Science,
5 Amsterdam. The Netherlands.
- 6 Lepš, J. and Šmilauer, P. (2003) *Multivariate Analysis of Ecological Data using CANOCO*.
7 Cambridge University Press, Cambridge, United Kingdom.
- 8 Manly, B.F.J. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd
9 edition. Chapman and Hall, New York, New York, USA.
- 10 Margalef, R. (1958) Information theory in ecology. *General Systems* 3, 36-71.
- 11 McGarigal, K., Chushman, S. and Stafford, S. (2000) *Multivariate Statistics for Wildlife and*
12 *Ecology Research*. Springer-Verlag, New York, New York, USA.
- 13 Neher, D.A. (1999) Soil community composition and ecosystem processes: Comparing
14 agricultural ecosystems with natural ecosystems. *Agroforestry Systems* 45, 159-185.
- 15 Neher, D.A. (2001) Role of nematodes in soil health and their use as indicators. *Journal of*
16 *Nematology* 33, 161-168.
- 17 Peet, R.K. (1974) The measurement of diversity. *Annual Review of Ecology and Systematics* 5,
18 285-307.
- 19 Pielou, E.C. (1975) *Ecological Diversity*. Wiley, New York.
- 20 Peres-Neto, P.R. and Jackson, D.A. (2001) How well do multivariate data sets match? The
21 advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* 129,
22 169-178.
- 23 Rosenberg, R. (1976) Benthic faunal dynamics during succession following pollution abatement
24 in a Swedish estuary. *Oikos* 27, 414-427.
- 25 Rossi, J.P. (1996) Statistical tool for soil biology. 11. Autocorrelogram and Mantel Test.
26 *European Journal of Soil Biology* 32, 195-203.
- 27 Rossi, J.P. (2003) The spatiotemporal pattern of a tropical earthworm species assemblage and its
28 relationship with soil structure. *Pedobiologia* 47, 497-503.
- 29 Ruf, A. (1998) A Maturity Index for predatory soil mites (Mesostigmata :Gamasina) as an
30 indicator of environmental impacts of pollution on forest soils. *Applied Soil Ecology* 9, 447-
31 452.
- 32 Schafer, C.T. (1973) Distribution of foraminifera near pollution sources in Chaleur Bay. *Water*
33 *Air Soil Pollution* 2, 219-233.
- 34 Shannon, C.E. (1948) A mathematical theory of communication. *The Bell System Technical*
35 *Journal* 27:379-423, 623-656.
- 36 Simpson, E.H. (1949) Measurement of diversity. *Nature* 163, 688.
- 37 ter Braak, C.J.F. (1988) CANOCO – an extension of DECORANA to analyze species-
38 environment relationships. *Vegetatio* 75, 159-160.
- 39 ter Braak, C.J.F. (1990) Update notes: CANOCO version 3.10. Agricultural Mathematics Group,
40 Wageningen, The Netherlands.
- 41 Van den Brink, P.J., Van den Brink, N.W. and ter Braak, C.J.F. (2003) Multivariate analysis of
42 ecotoxicological data using ordination: demonstrations of utility on the basis of various
43 examples. *Australasian Journal of Ecotoxicology* 9, 141–156.
- 44 Van den Brink, P. J. and ter Braak, C. J. (1998) Multivariate analysis of stress in experimental
45 ecosystems by principal response curves and similarity analysis. *Aquatic Ecology* 32, 163-
46 178.

- 1 Van den Brink, PJ and ter Braak, CJF. (1999) Principal response curves: Analysis of time-
2 dependent multivariate responses of biological community to stress. *Environmental*
3 *Toxicology and Chemistry* 18, 138-148.
- 4 Whittaker, R.H. (1975) *Communities and Ecosystems*. 2nd edn. Macmillian, New York.
- 5 Yeates, G.W. and Cook, R. (1998) Nematode fauna of three Welsh soils under conventional and
6 organic grassland farm managements. In: de Goede, R. G. M. de and Bongers, T. (eds).
7 *Nematode Communities of Northern Temperate Grassland Ecosystems*. Focus Verlag,
8 Giessenm, pp. 305-313.
- 9