4 General Community Indices for the Analysis of Nematode Assemblages

Deborah A. Neher[1]* and Brian J. Darby[2†]
[1]*Department of Agriculture, Landscape and Environment, University of Vermont, Vermont, USA;* [2]*Formerly Department of Biology, University of North Dakota, North Dakota, USA*

Abstract
This chapter reviews classical community indices that condense community data into one or a few meaningful variables to simplify analysis and interpretation. Given that no indices can conclusively reveal all ecological processes, the recommendation is to complement univariate approaches that disregard taxon identity with multivariate approaches that preserve taxon identity to improve one's understanding of both the autecology of individual community members and synecology of the community. Common univariate indices include index families such as diversity or maturity. Multivariate approaches include clustering and ordination. Recommendations for computer software and R scripts are included.

4.1 Introduction

To be successful as an indicator, a single index must 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. Limitations of community indices are that they are density-independent and rely on *pattern* to reflect process, and often several processes can result in similar patterns. Productivity, resilience and stability are some of the ecological characteristics relevant to ecosystem management, and some early successful attempts to link diversity with function include Rosenberg (1976) and Schafer (1973) and continue to be investigated (e.g. Lazarova *et al.*, 2021). However, the link between ecosystem processes 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; Brussaard *et al.*, 2004). Appropriate sampling and statistical techniques are critical to valid interpretation of diversity indices. 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). Systematic sampling with equal sampling effort (volume, area) is necessary for comparison among samples (Neher and Campbell, 1996). Although nematode communities vary by season (Neher *et al.*, 2005), a general recommendation to sample at the end of a growing season prevails. Furthermore,

*Email: Deborah.Neher@uvm.edu

 Nematodes as Environmental Indicators: From Theory to Practice, 2nd Edition (T. Kakouli-Duarte *et al*. eds)
DOI: 10.1079/9781800624221.0004

one must consider the time lag that exists between nematode community composition and evidence of ecosystem function (Neher *et al.*, 2012).

Like other microscopic groups, nematodes can be tedious and laborious to extract from their environment, count, preserve and identify. Most data sets that characterize nematode communities are either: (i) quantitative, with several replicates from several sites, but performed at coarse taxonomic resolution (e.g. family or genus); or (ii) have a fine level of taxonomic resolution (e.g. species) but either lack replication or are not quantitative. Admittedly, although molecular tools are not to be used exclusively but rather as additional characters to identify nematodes, the advent of high-throughput amplicon sequencing creates the potential to characterize nematode communities, as well as associated bacteria, fungi and protozoans, for many sites with a level of taxonomic precision that was not feasible previously. Likely, as high-throughput amplicon sequencing becomes more popular with nematode community characterizations, data sets with many sites and identifications with species-level precision will become more popular and may incite the use of a variety of tests and indices that are otherwise commonplace in general ecological literature but scarcely used in the context of nematode ecology.

4.2 Univariate Identity-Independent Indices

Depending on the context, the term 'diversity' is sometimes used to simply describe the number of taxa. In the context of quantitative community ecology, the term 'diversity' is more commonly, and appropriately, used to describe an integration of both numbers of taxa (*species richness*) and equitability among taxa (*species evenness*) (Hurlbert, 1971). Nematode diversity can be computed at the species, genus or trophic level. Most nematode communities are enumerated at coarser resolutions because species identification based on morphology is difficult. 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). At the trophic level, diversity is a measure of food chain length and food web complexity (Neher and Campbell, 1994). A trophic diversity index assumes that greater diversity of trophic groups in soil food webs (i.e. complexity) and longer food chains reflect improved ecosystem function (Moore, 1993; Neher *et al.*, 2019). Furthermore, the appropriate spatial resolution of diversity should be consistent with the objectives of the study. *Alpha* diversity reflects the species diversity of individual localities, *beta* diversity reflects the difference of communities across landscapes, and *gamma* diversity reflects the differences between landscapes across regions (e.g. Kerfahi *et al.*, 2016). In the context of soils, the concept of a landscape could have a variety of interpretations but would generally be operationally defined and would likely involve a treatment or comparison of research interest, potentially even as close as two experimental plots.

4.2.1 Identity-independent indices and their calculation

A variety of identity-independent indices is available to serve different purposes in different circumstances (Table 4.1). Each diversity index weights richness and evenness uniquely, but all diversity indices generally function so that an increase in either richness or evenness will always increase diversity. In some reports, the term diversity continues to refer simply to the total number of species; it is preferable, however, to restrict the use of 'diversity' to incorporate both the number of species and evenness. Formulae for calculating several common indices are summarized (Table 4.1) and accompanied by a customized R-studio script written to compute all indices (Table 4.2).

Determining numbers of species (richness) requires standardization and clear reporting for each experiment or sampling regime to prevent artefacts of sampling effort when comparing richness and diversity indices. Nematode density generally varies widely from sample to sample, so the number of nematodes enumerated is a representative subset of the total number extracted, i.e. an unknown number at the time of sampling. Therefore, species *richness* is the appropriate term to refer to the total number of species found when enumerating a uniform *number* of extracted individuals (e.g. 200 from each sample) from samples of a uniform initial mass or volume. Species *density* differs by referring to the total number of species expressed as a

Table 4.1. Selected richness, diversity and evenness indices that can be calculated for nematode communities. (Author's own table.)

Name	Equation[a]	Application	Reference
Margalef's richness	$D_{Marg} = \frac{(S-1)}{\ln(N)}$	Its use should be restricted to comparing species richness among large communities	Margalef (1958)
Shannon's diversity	$H' = -\Sigma(p_i \ln p_i)$	Sensitive to rare taxa. 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 the Simpson index	Shannon (1948)
Hill's N1	$N1 = \exp[-\Sigma(p_i \ln p_i)] = \exp(H')$	The value of this index can be interpretated as the number of abundant taxa	Hill (1973)
Simpson's dominance (infinite community)	$D = \Sigma p_i^2$	Weights common taxa. Probability that two randomly chosen individuals of an infinite community belong to the same class, thus inversely related to diversity	Simpson (1949)
Simpson's dominance (finite community)	$\lambda = \frac{\sum n_i (n_i - 1)}{N(N-1)}$	Like 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	Simpson (1949)
Hill's N2	$N2 = (\Sigma p_i^2)^{-1} = 1/D$	The value of this index can be interpreted as the number of very abundant taxa	Hill (1973)
Brillouin's diversity	$H = \frac{1}{N} \log \frac{N!}{\prod N_i!}$	Use only on fully censused communities because it is a true statistic and, thus, free from statistical error	Brillouin (1962); Pielou (1975)
Brillouin's maximum diversity	$H_{max} = \frac{1}{N} \ln \frac{N!}{(X!)^{S-r}(Y!)^r}$	Represents maximum possible evenness of a sample of *N* individuals and *S* species	Brillouin (1962); Pielou (1975)
Brillouin's minimum diversity	$H_{min} = \frac{1}{N} \ln \frac{N!}{(N-S+1)!}$	Represents minimum possible evenness of a sample of *N* individuals and *S* species	Brillouin (1962); Pielou (1975)
Brillouin's evenness	$J = \frac{H}{H_{max}}$ or $J' = \frac{H'}{\ln S}$	Evenness represents equality of abundances in a community. Use J for samples (and J′ for collections) to determine the evenness portion of diversity; J or J′ represent observed and maximum diversity, respectively	Brillouin (1962); Pielou (1975)

Continued

Table 4.1. Continued.

Name	Equation[a]	Application	Reference
Brillouin's relative evenness	$V = \frac{H - H_{min}}{Hmin_{max}}$	Unlike J and J′, V is not influenced by species richness (*S*)	Hurlbert (1971); Pielou (1975)
Hill's evenness	$E_{2,1} = \frac{(N_2)}{(N_1)}$	Ratio of very abundant taxa to rare taxa. Approaches value of 1 as a single species become more dominant in a community	Hill (1973)
Heip's evenness	$E_{Heip} = \frac{(e^{H'} - 1)}{(S - 1)}$	More sensitive to variations in rare species richness and/or abundance	Heip (1974)

[a]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 Brillouin's maximum diversity) is the integer portion of (*N*/*S*), *Y* = *X* + 1, and *r* = the remainder of *X*.

Table 4.2. R code to compute the various diversity indices from Table 4.1. (Author's own table.)

```
library(vegan)
library(ggplot2)

## Diversity Indices

# Read in data and separate labels from read counts
speciesraw = read.csv('sequences.csv')

spdata = speciesraw[,-c(1:5)]
labels = speciesraw[,c(1:5)]

# Prepare commonly used calculations

S = ncol(spdata)
N = rowSums(spdata)
pind = spdata/N

# Calculate indices

DMarg = (S - 1) / log(N)
ShannonH = -(rowSums(pind*log(pind), na.rm = TRUE))
N1 = exp(ShannonH)
SimpsonD = rowSums(pind**2)
N2 = 1 / SimpsonD
BrillouinJ = ShannonH / log(S)
HillE = N2 / N1
HeipE = (N1 - 1) / (S - 1)

## Using vegan package

veganShannonH = diversity(spdata, index="shannon")
veganN2 = diversity(spdata, index="invsimpson")

## Graph selected indices

df = cbind(labels, ShannonH)
ggplot(df, aes(x=SEASON, y=ShannonH, fill=SEASON))+
  geom_boxplot()
```

uniform *portion* of all extracted individuals (e.g. 20% of the individuals from each sample). This distinction is important because species richness and density are not necessarily linear in relationship. For example, 20 species found among 200 individuals does not necessarily mean that one will find 40 species from 400 individuals. This type of extrapolation requires rarefaction of original data to estimate the number of species collected from a hypothetical number of individuals or samples. Ecologists use this approach to generate *species area curves* which plot the number of unique species accumulated for each sample as they are sampled (Gotelli and Colwell, 2001). A rarefaction curve used to define uniform portions of extracted individuals acts as an accumulation curve, and is performed retrospectively, after samples have been collected (Fig. 4.1A). Curves that do not fully approach a hypothetical asymptote suggest that additional

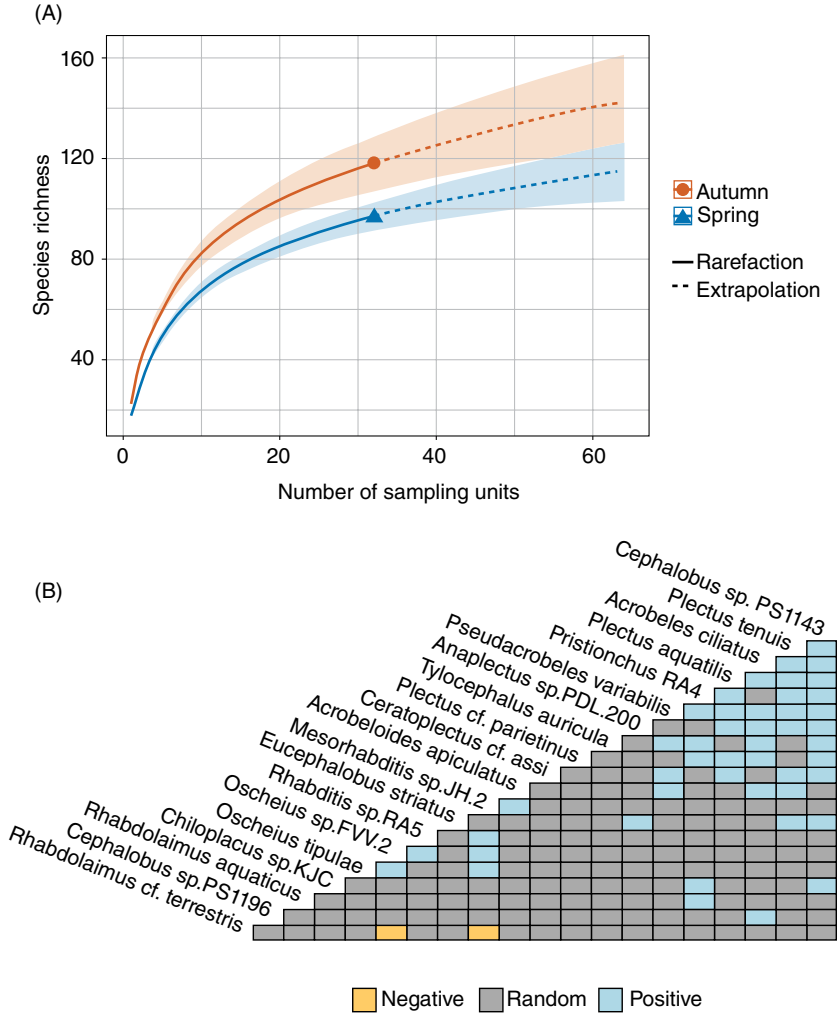

Fig. 4.1. Species rarefaction curves and co-occurrence analysis. (A) Species rarefaction curves for nematode species sampled in the spring and autumn. Solid lines indicate the rarefaction curves, dotted lines indicate extrapolated predictions, and shaded fill indicates bootstrapped confidence intervals. (B) Visualization of the results of a species-based co-occurrence analysis using the R-language package *cooccur* (Griffith *et al*., 2016). Pairwise comparisons with a blue box indicate a positive co-occurrence pattern, while yellow boxes indicate negative co-occurrence patterns. (Data for both (A) and (B) come from nematodes sampled at the Konza tallgrass prairie; Darby *et al*., 2013.)

sampling is likely to yield additional species. The Margalef index (Table 4.1) is useful to adjust the number of species (*S*) for the number of individuals enumerated (*N*).

Evenness represents the relative uniformity in abundance of each taxon within a sample. Heip proposed an evenness index (Table 4.1) to standardize the Shannon's diversity index (H′) by total number of species (*S*). Alternatively, Brillouin developed a series of statistics for censused communities that are computationally complex (Table 4.1). For example, Brillouin's maximum theoretical diversity (= H_{max}) is computed with the assumption that all individuals are distributed as uniformly as possible, and minimum theoretical diversity is computed assuming all individuals are distributed as asymmetrically as possible. Two forms of evenness can be computed, the first as diversity relative to maximum diversity (= J) and the second ('relative evenness') as diversity relative to maximum diversity but scaled to minimum diversity (= V). The first type (not scaled to minimum diversity) can be based on either of two estimates of diversity depending on whether the user wishes to assume a finite or infinite community enumeration. We recommend using Brillouin's sample diversity relative to Brillouin's maximum diversity (= J) when assuming a finite community enumeration, or Shannon's population diversity relative to the natural logarithm of richness (= J′) when assuming infinite community enumeration. The second type (= V, scaled to minimum diversity) uses Brillouin's calculation of diversity from a censused community. Although nematode communities are rarely, if ever, fully censused in nature, the assumption of complete enumeration may be appropriate in some unique applications; for example, small, isolated habitats or virtual individuals in a computationally simulated model community.

There are a variety of diversity indices that incorporate both richness and evenness, and they differ mostly as to the degree to which they are influenced by dominant and rare species. Shannon's diversity (Table 4.1) is a popular diversity index. The exponent of Shannon's index (Hill's N1) can be interpreted as the number of uniformly distributed species that would produce an identical Shannon's index as the non-uniformly distributed community. For example, consider a community with 20 non-uniformly distributed species and a Shannon's index of 2.3. The exponent of 2.3 (Hill's N1) equals 9.97, so, intuitively, approximately 10 uniformly distributed species would be needed to produce a Shannon's index like the community of 20 non-uniformly distributed species. Furthermore, Heip's evenness index = [(9.97 – 1)/(20 – 1)] = 0.47, indicating that about half of the observed species would be necessary to produce a similar Shannon's index if they were distributed uniformly. Simpson's D index (Table 4.1) is considered a dominance index because it increases as species are distributed more unevenly (increasing dominance) and can be interpreted intuitively as the probability that two randomly selected individuals from an infinite community will be the same. The reciprocal of Simpson's index (Hill's N2) is often reported as a diversity index, and like Hill's N1, Hill's N2 can be interpreted as the number of uniformly distributed species that would produce a Simpson's index identical to that of the non-uniform community. Notice that the minimum Simpson's D possible (i.e. least dominance by any taxa) is 1/*S* and the maximum Hill's N2 possible (greatest equitability) is *S*, so we could compute an evenness index similar to Heip's approach as N2/*S*. See Neher and Darby (2006) for a deeper explanation of interpretation of various diversity indices.

4.3 Community Assemblage Models

4.3.1 Species co-occurrence patterns

As high-throughput molecular sequencing gains popularity in nematode ecology, allowing greater taxonomic resolution and facilitating characterization of nematode species for a greater number of samples, we think that analysing co-occurrence patterns among nematode species may become a more fruitful way to understand nematode community assemblages. For example, a co-occurrence matrix from data at the species level using presence–absence data obtained from amplicon sequencing is helpful to define species associates (Fig. 4.1B). Species co-occurrence patterns have an important historical place in the field of ecology and represent an attempt to understand how communities form and are assembled. Species co-occurrence patterns can be linked back to the earliest concepts of community formation, including the competing Gleason

(1926) and Clements (1936) concepts of community assemblage. Co-occurrence patterns were again made the subject of ecologists' attention by Diamond (1975) who suggested that communities are structured with patterns of positive or negative co-occurrences akin to a 'checkerboard' pattern, reflecting cooperative or competitive relationships, respectively. This theory, and its associated models, have been debated extensively (Simberloff, 1978) and generated several studies on the proper use of null models against which to test empirical presence–absence data (Gotelli and Graves, 1996; Weiher and Keddy, 1999). Veech (2013) developed a probabilistic model for analysing co-occurrence patterns and provided software for the R environment (Griffith *et al.*, 2016). This package computes rates of all occurrences between each pairwise species combination from a site by species presence–absence matrix, plus their probability. In this case, pairwise probabilities are obtained from a hypergeometric distribution, much like performing a Fisher's exact test. Other software programs designed for more speciose communities, like bacteria or fungi, compute co-occurrence as correlation. However, read counts from high-throughput amplicon sequencing are poorly correlated with individual abundance, making correlations between nematode species only possible with specimen counts. Species differ in their number of genomic rRNA copy numbers, which is the locus typically used for identification by amplicon sequencing. Therefore, it would take species-specific 'copy-number corrections' to convert high-throughput sequencing reads to specimen counts (Darby *et al.*, 2013). Unfortunately, the accuracy of such copy-number correction factors will vary by season if the species itself varies seasonally in its number of somatic cells per individual (i.e. differences in the number of juveniles) or in the number of reproductive cells (i.e. season reproductive patterns). Ultimately, the most precise community identifications will likely come from a combination of specimen counts by morphology but using sequencing data to assist the identifications.

4.3.2 Ecological succession

Ecological succession refers to a relatively predicable or directional sequence of spatio-temporal patterns of ecological interactions within a community. As species composition changes, it alters the abiotic environment, which in turn selects against the existing community favouring a community composition that performs better under the newly created abiotic environment. The concept originated in plant ecology (Whittaker, 1975) but also applies to invertebrate communities in soil and sediment. Succession usually progresses directionally unless set back by an environmental disturbance such as cultivation, pollution or nutrient enrichment (Neher, 1999). Therefore, quantitative measures of ecological succession can serve as indicators of disturbance. With improved knowledge of synecology of nematode communities, one could identify the type and intensity of disturbance based on an index of succession. Bongers (1990) proposed an index of ecological succession for application to nematodes and Ruf (1998) applied a similar approach to mesostigmatid mites. 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. Successional indices are described in greater detail in Chapter 5 of this volume.

4.3.4 Beta diversity

Beta diversity is another way to examine patterns of ecological succession in nematodes. Whereas alpha diversity reflects the communities at individual localities, beta diversity describes the comparison between communities across a landscape. For example, Podani and Schmera (2011) developed indices that separated the various components of beta diversity into species similarity, richness differences and species replacements (Fig. 4.2). These three indices describe different aspects of beta diversity, or the potential differences between communities at different localities. For a community ecologist, calculating these indices across different communities may help to identify mechanisms of community assemblage or ecological succession. For example, as a community naturally shifts in time, these indices reflect whether taxa maintain similar richness and progress by

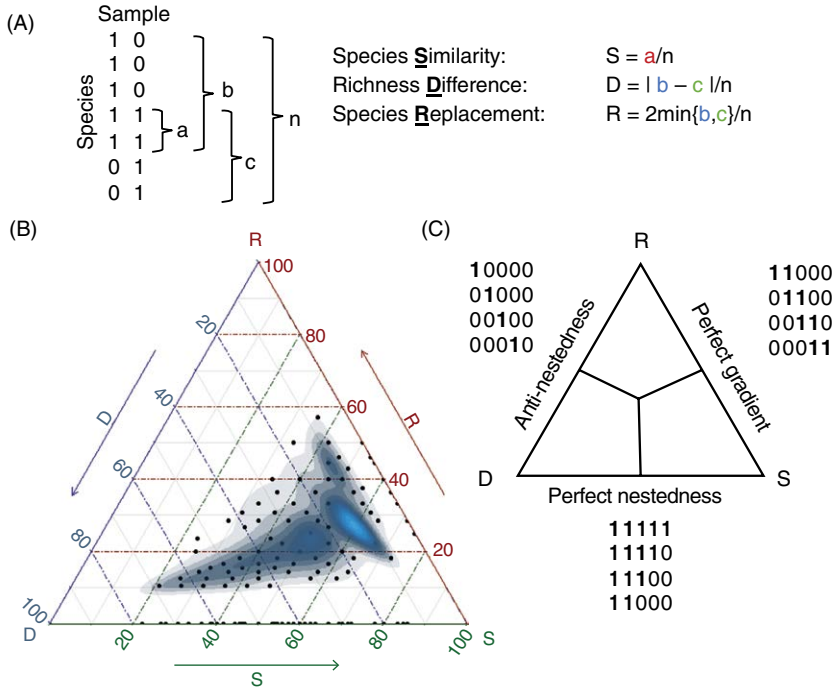

Fig. 4.2. Illustrating the use of species similarity, richness difference and species replacement indices of beta diversity. (A) Indices are calculated for each pairwise sample in a site-by-species presence–absence matrix. (B) Index values for all pairwise comparisons can be plotted on a ternary diagram. (C) The aggregate distribution of indices can reflect beta diversity patterns of anti-nestedness, perfect gradient or perfect nestedness between communities. (The *ggtern* package of R software was used to generate the graphs from unpublished data of co-author Brian Darby.)

species replacement, or whether they change by the addition or subtraction of individual species. One helpful feature of these indices is that the three values sum to 1.0, which allows one to plot the values on a ternary plot to illustrate the dominant component of beta diversity between each pairwise comparison (Fig. 4.2).

4.4 Multivariate Techniques

Multivariate analysis offers both descriptive and inferential procedures to analyse multiple variables simultaneously to reveal the collective interactions of all variables and the effect each variable has on the others. *Descriptive* procedures help to illustrate the overall structure of a data set while *inferential* procedures help to test hypotheses of interactions. Therefore, multivariate analysis has two complementary applications, *exploratory hypothesis-generating* and *inferential hypothesis-testing*, that can be combined into a two-phase approach that might begin with an exploratory phase that seeks patterns in nature by asking 'To what can I ascribe the variation in my data?' The second phase, then, tests the hypotheses that were generated by asking 'Can I reject the null hypothesis that species are unrelated to each other or postulated environmental factor(s)?' In this way, multivariate analysis is useful in evaluating nematode community structure as a biological indicator by keeping the identity of individual taxa explicit throughout the analysis. Below, we discuss two types of multivariate analysis commonly applied to nematode communities: cluster analysis and ordination.

4.4.1 Cluster analysis

Cluster analysis treats each multivariate observation (sample) as a vector and attempts to

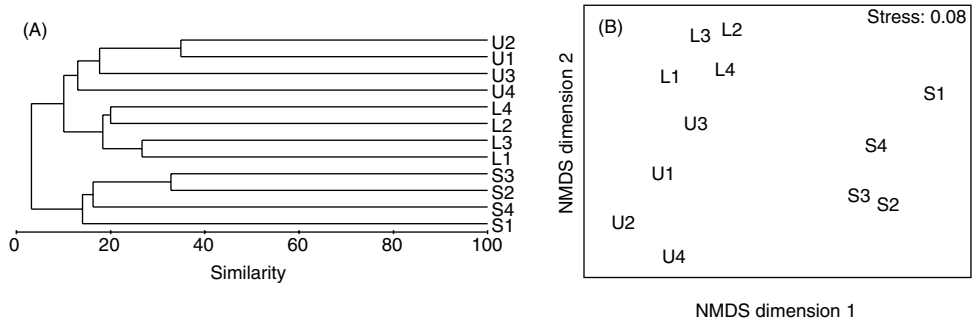

Fig. 4.3. Graphic representations among samples of nematode communities illustrated either as (A) similarities in a dendrogram or (B) dissimilarities on a multidimensional scaling (MDS) biplot. Bray–Curtis similarity was computed on non-transformed abundance data for each pairwise combination of samples. A dendrogram and biplot show the same data in two formats of duplicative information. (Illustrations were created using ANOSIM and CLUSTER modules of Primer-E Version 5.2.9 software; Clarke and Gorley, 2001.)

group vectors that are like each other into clusters (Fig. 4.3). Cluster analysis begins with a (dis)similarity matrix, often computed as the Euclidean distance or Bray–Curtis similarity among all pairs of vectors. *Hierarchical clustering* algorithms are either agglomerative or divisive. *Agglomerative clustering* begins with each vector representing a unique cluster and sequentially combining the two nearest clusters into one until an optimal number of clusters is attained. *Divisive clustering* begins with one cluster containing all vectors and sequentially divides the cluster into two until an optimal number of clusters has been obtained. Agglomerative clustering is most common and there are several methods of determining the distance of vector clusters from each other. The *single linkage (or nearest neighbour) method* determines the distance between two clusters as the minimum distance (e.g. Euclidean) between the two most similar vectors of each cluster, while the *complete linkage (e.g. furthest neighbour) method* determines the distance between two clusters as the maximum distance (e.g. Euclidean) between the two most dissimilar vectors of each cluster. The *average linkage method* defines the distance between two clusters as the average distance of all elements from each cluster, while the *centroid method* defines the distance between two clusters as the distance between the two mean (or median) vectors of a cluster, called the centroids. Finally, *Ward's method* joins clusters to minimize the increase in sum of squares within and between clusters. The result of hierarchical cluster analysis is a dendrogram (i.e. tree diagram) that shows each step of the clustering procedure and the distance at which the clusters merge (e.g. Fig. 4.3A).

Discriminant analysis is a related approach based on an a priori expectation of group members whereas cluster analysis has no preconceived expectation of group members and therefore conducts a posteriori aggregation. With discriminant analysis, one hypothesizes that there are two or more distinct groups and then determines whether the observations divide significantly among those two predicted groups (Afifi *et al.*, 2020).

4.4.2 Ordination

Ordination techniques are popular in community analysis due to their ability to visualize multidimensional data in two-dimensional space (Afifi *et al.*, 2020). There are two main classes of ordination techniques: direct and indirect gradient analysis. *Indirect gradient analysis*, also called *unconstrained*, seeks to interpret patterns from within a data set. *Direct gradient analysis* seeks to extract patterns from known gradients and is therefore *constrained* by the environmental variables supplied. Indirect gradient analysis is divided into distance-based and eigenanalysis-based methods, whereas all direct gradient analyses are eigenanalysis-based methods. Examples of

distance-based indirect gradient ordination include polar ordination (PO), principal coordinates analysis (PCoA) and non-metric multidimensional scaling (NMDS). In PO, two samples most different from each other based on their species composition serve as end points and all other samples are plotted relative to them. In this way, new samples can be added to polar ordination without changing the structure of the ordination diagram. PCoA simply maximizes linear distance measures of the ordination in metric space (using a Euclidean or Bray–Curtis distance matrix), while NMDS is analogous to a non-parametric variant of PCoA by maximizing *rank* distance measures of the ordination in non-metric space (e.g. Fig. 4.3B).

Redundancy analysis (RDA) and canonical correspondence analysis (CCA) are two types of direct gradient analysis that constrains the distribution of taxa by environmental variables (Fig. 4.4). They vary by whether there is a linear (RDA) or unimodal (CCA) link between suites of taxon data with suites of environmental variables. Environmental variables can include treatment classes (coded as nominal 0 or 1 variables) or chemical or physical properties (such as pollutants or temperature) as continuous variables. All these procedures can be performed in R, either through Base R software, or through packages such as *labdsv* or *vegan* (Oksanen *et al.*, 2022). Alternatively, Canoco (ter Braak and Šmilauer, 2012) and Primer-E (Clarke and Gorley, 2001) software packages are simple tools to perform these procedures. In Canoco, abundances are transformed as $\log(x + 1)$ prior to analysis, which is a historically used transformation in nematology. Transformations are unnecessary in Primer-E because the scaling is non-metric multidimensional. CCA results are displayed graphically with biplots. In CCA biplots, each vector for an environmental variable defines an axis, and site or taxa scores can be projected on to that axis. 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. The first axis represents the greatest explained variation, and subsequent axes represent progressively less

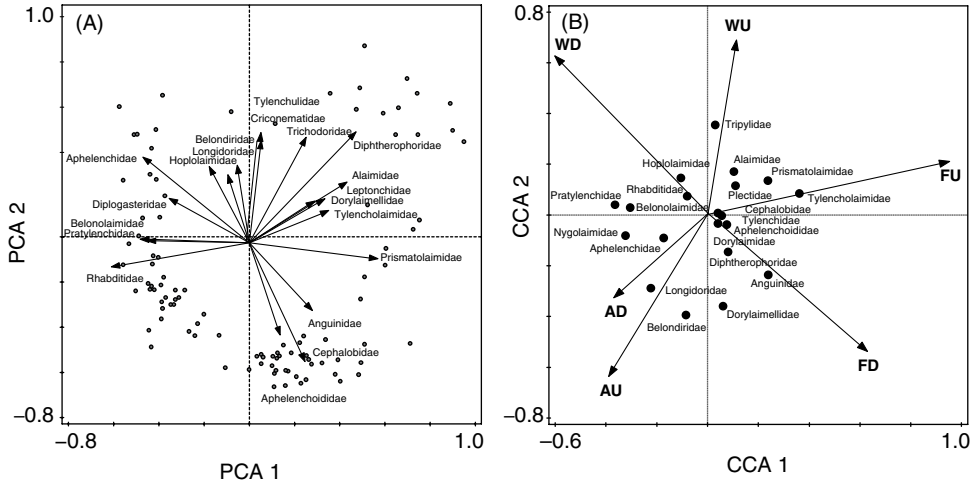

Fig. 4.4. Nematodes in 20 families (out of 39) are illustrated as either (A) principal components analysis (PCA) or (B) canonical correspondence analysis (CCA). PCA is unconstrained, illustrating taxa (arrows) versus samples (grey dots). CCA is constrained by environmental variables representing factorial combinations of three ecosystems (A = agriculture, F = forest, W = wetland) and two disturbance levels (D = disturbed, U = undisturbed) sampled 12 times over two years in North Carolina. The longer vectors explain more variation than shorter vectors. Vectors with acute angles are correlated positively and those in opposite quadrants are correlated negatively. Right angles are orthogonal or independent of each other. (Biplots were generated using Canoco Version 5 software; ter Braak and Šmilauer, 2012. Data from Neher *et al.*, 2005.)

variance. Unfortunately, CCA analyses are restricted to illustrating one instance in time. Therefore, repeated measures of communities through time are consolidated into a single biplot that loses information about temporal patterns.

Principal response curves (PRC) are a multivariate method for the analysis of repeated-measurement designs (van den Brink and ter Braak, 1998, 1999). PRC is based on 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 (Fig. 4.5). 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.

The type of analysis chosen depends on the research question but is critical in terms of output and interpretation. Nematode community

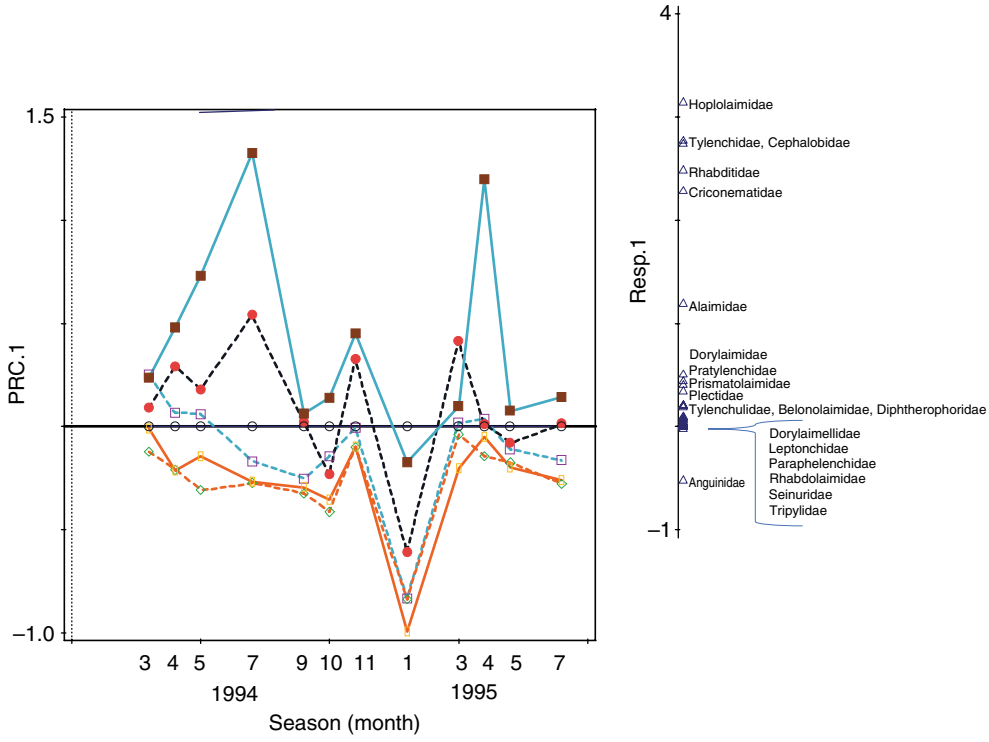

Fig. 4.5. Nematodes in 20 families (out of 39) are illustrated that were sampled repeatedly for 12 times over two years in factorial combinations of three ecosystems (A = agriculture, F = forest, W = wetland) and two disturbance levels (D = disturbed, U = undisturbed). The *x*-axis (black, solid line) represents undisturbed agriculture as a reference baseline for comparison of relative differences for disturbed agriculture (black, dashed), disturbed (dashed) and undisturbed (solid) forests (orange), and disturbed (dashed) and undisturbed (solid) wetlands (blue). On the right are the species scores (Resp. 1) for taxa that explain the fluctuations and contrasts of the ecosystem communities through time. For example, the cluster of taxa at the top (Hoplolaimidae, Tylenchidae, Cephalobidae, Rhabditidae and Criconematidae) distinguish UW from others; the Alaimidae through Diphtherophoridae are associated with the baseline reference; and Dorylamellidae through Tripylidae and Anguinidae follow dips in AD, WD and forests. (Principal response curve (PRC) plot was generated using Canoco Version 5 software; ter Braak and Šmilauer, 2012. Data from Neher *et al*., 2005.)

data collected 12 times over a period of two years was analysed three different ways (Figs 4.4 and 4.5). For example, there is only 50% of taxa in common when unconstrained (Fig. 4.4A) or constrained (Fig. 4.4B). By constraining, one gains information about which taxa are associated with environmental variables. For example, Anguinidae, Longidoridae and Hoplolaimidae are relatively abundant in disturbed forest, agriculture and wetland sites, respectively (Fig. 4.4). PRC plots separate times sampled to further refine which taxa explain differences among treatments and how those varied through time (Fig. 4.5). This information is lost in a CCA (Fig. 4.4B) and there is a 20% difference in taxa explaining variation. From the PCA, we learn that Hoplolaimidae and Anguinidae are associated with fluctuating communities in wetlands and forests, respectively.

4.5 Conclusion

Classical community composition can be analysed using metrics that either disregard or preserve the identity of taxon within the community. Identity-independent methods such as diversity and evenness indices are relatively simple to compute and analyse statistically. However, the user must exercise caution by selecting the form of index most appropriate to the goals of the study and resisting the temptation to singularly extrapolate to a greater ecological meaning without substantial supplementary evidence. Alternatively, indices that incorporate and/or maintain taxon identity can more convincingly be linked to ecological process and function. Measures of ecological succession and species assemblage are univariate forms that can be analysed using traditional statistical tools such as regression and analysis of variance. A variety of multivariate methods are accessible through commercial software packages. Many multivariate approaches capture a one-time snapshot of community composition. However, repeated-measures approaches are becoming available to evaluate changes in community composition through time. Practitioners should be aware of the many limitations, assumptions and caveats of community assemblage and multivariate techniques by consulting with expert statisticians.

References

Afifi, A., May, S., Donatello, R.A. and Clark, V.A. (2020) *Practical Multivariate Analysis*. CRC Press, Boca Raton, Florida.

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.

Brussaard, L., Kuyper, T.W., Didden, W.A.M., de Goede, R.G.M. and Bloem, J. (2004) Biological soil quality from biomass to biodiversity – importance and resilience to management stress and disturbance. In: Schjoenning, P., Elmholt, S. and Christensen, B.T. (eds) *Managing Soil Quality: Challenges in Modern Agriculture*. CAB International, Wallingford, UK, pp. 139–171.

Chapin, F.S., Schultz, E.D. and Mooney, H.A. (1992) Biodiversity and ecosystem processes. *Trends in Ecology and Evolution* 7, 107–108.

Clarke, K.R. and Gorley, R.N. (2001) *Plymouth Routines in Multivariate Ecological Research (Primer) v5: User Manual/Tutorial*. Primer-E Ltd, Plymouth, UK.

Clements, F.E. (1936) Nature and structure of the climax. *Journal of Ecology* 24, 252–284. DOI: 10.2307/2256278.

Darby, B.J., Todd, T.C. and Herman, M.A. (2013) High-throughput amplicon sequencing of rRNA genes requires a copy number correction to accurately reflect the effects of management practices on soil nematode community structure. *Molecular Ecology* 22, 5456–5471. DOI: 10.1111/mec.12480.

Diamond, J.M. (1975) *Ecology and Evolution of Communities*. Harvard University Press, Cambridge, Massachusetts.

Ettema, C.H. (1998) Soil nematode diversity: species coexistence and ecosystem function. *Journal of Nematology* 30, 159–169.

Gimaret-Carpentier, C., Pélissier, R., Pascal, J.-P. and Houllier, F. (1998) Sampling strategies for the assessment of tree species diversity. *Journal of Vegetation Science* 9, 161–172. DOI: 10.2307/3237115.

Gleason, H.A. (1926) The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club* 53, 7–26. DOI: 10.2307/2479933.
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. DOI: 10.1046/j.1461-0248.2001.00230.x.
Gotelli, N.J. and Graves, G.R. (1996) *Null Models in Ecology*. Smithsonian Institution Press, Washington, DC.
Griffith, D.M., Veech, J.A. and Marsh, C.J. (2016) Cooccur: probabilistic species co-occurrence analysis in R. *Journal of Statistical Software* 69, 1–17. DOI: 10.18637/jss.v069.c02.
Heip, C. (1974) A new index measuring evenness. *Journal of the Marine Biological Association* 54, 555–557. DOI: 10.1017/S0025315400022736.
Hill, M.O. (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology* 54, 427–432. DOI: 10.2307/1934352.
Hurlbert, S.H. (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52, 577–586. DOI: 10.2307/1934145.
Kerfahi, D., Tripathi, B.M., Porazinska, D.L., Park, J., Go, R. and Adams, J.M. (2016) Do tropical rain forest soils have greater nematode diversity than High Arctic tundra? A metagenetic comparison of Malaysia and Svalbard. *Global Ecology and Biogeography* 25, 716–728. DOI: 10.1111/geb.12448.
Lazarova, S., Coyne, D., Rodriguez, M.G., Peteira, B. and Ciancio, A. (2021) Functional diversity of soil nematodes in relation to the impact of agriculture – a review. *Diversity* 13, 64. DOI: 10.3390/d13020064.
Margalef, R. (1958) Information theory in ecology. *General Systems* 3, 36–71.
Moore, J.C. (1993) Impact of agricultural practices on soil food web structure: theory and application. *Agriculture, Ecosystems and Environment* 51, 239–247. DOI: 10.1016/0167-8809(94)90047-7.
Neher, D.A. (1999) Soil community composition and ecosystem processes: comparing agricultural ecosystems with natural ecosystems. *Agroforestry Systems* 45, 159–185. DOI: 10.1023/A:1006299100678.
Neher, D.A. and Campbell, C.L. (1994) Nematode communities and microbial biomass in soils with annual and perennial crops. *Applied Soil Ecology* 1, 17–28. DOI: 10.1016/0929-1393(94)90020-5.
Neher, D.A. and Campbell, C.L. (1996) Sampling for regional monitoring of nematode communities in agricultural soils. *Journal of Nematology* 28, 196–208.
Neher, D.A. and Darby, B.J. (2006) Computation and application of nematode community indices: general guidelines. In: Eyualem, A., Traunspurger, W. and Andrássy, I. (eds) *Freshwater Nematodes: Ecology and Taxonomy.* CAB International, Wallingford, UK, pp. 211–222.
Neher, D.A., Wu, J., Barbercheck, M.E. and Anas, O. (2005) Ecosystem type affects interpretation of soil nematode community measures. *Applied Soil Ecology* 30, 47–64. DOI: 10.1016/j.apsoil.2005.01.002.
Neher, D.A., Weicht, T.R. and Barbercheck, M.E. (2012) Linking invertebrate communities to decomposition rate and nitrogen availability in pine forest soils. *Applied Soil Ecology* 54, 14–23.
Neher, D.A., Nishanthan, T., Grabau, Z.J. and Chen, S.Y. (2019) Crop rotation and tillage affect nematode communities more than biocides in monoculture soybean. *Applied Soil Ecology* 140, 89–97. DOI: 10.1016/j.apsoil.2019.03.016.
Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P. *et al.* (2022) Package 'Vegan'. Available at: https://cran.r-project.org/web/packages/vegan/vegan.pdf (accessed 24 October 2023).
Pielou, E.C. (1975) *Ecological Diversity*. Wiley, New York.
Podani, J. and Schmera, D. (2011) A new conceptual and methodological framework for exploring and explaining pattern in presence–absence data. *Oikos* 120, 1625–1638. DOI: 10.1111/j.1600-0706.2011.19451.x.
Rosenberg, R. (1976) Benthic faunal dynamics during succession following pollution abatement in a Swedish estuary. *Oikos* 27, 414–427. DOI: 10.2307/3543460.
Ruf, A. (1998) A maturity index for predatory soil mites (Mesostigmata: Gamasina) as an indicator of environmental impacts of pollution on forest soils. *Applied Soil Ecology* 9, 447–452. DOI: 10.1016/S0929-1393(98)00103-6.
Schafer, C.T. (1973) Distribution of foraminifera near pollution sources in Chaleur Bay. *Water, Air, and Soil Pollution* 2, 219–233. DOI: 10.1007/BF00655698.
Shannon, C.E. (1948) A mathematical theory of communication. *The Bell System Technical Journal* 27, 379–423.
Simberloff, D. (1978) Using island biogeographic distributions to determine if colonization is stochastic. *The American Naturalist* 112, 713–726.
Simpson, E.H. (1949) Measurement of diversity. *Nature* 163, 688. DOI: 10.1038/163688a0.
ter Braak, C.J.F. and Šmilauer, P. (2012) *Canoco Reference Manual and User's Guide: Software for Ordination (version 5.0)*. Biometris, Wageningen, The Netherlands.

van den Brink, P.J. and ter Braak, C.J.F. (1998) Multivariate analysis of stress in experimental ecosystems by principal response curves and similarity analysis. *Aquatic Ecology* 32, 163–178. DOI: 10.1023/A:1009944004756.

van den Brink, P.J. and ter Braak, C.J.F. (1999) Principal response curves: analysis of time-dependent multivariate responses of biological community to stress. *Environmental Toxicology and Chemistry* 18, 138–148. DOI: 10.1002/etc.5620180207.

Veech, J.A. (2013) A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography* 22, 252–260. DOI: 10.1111/j.1466-8238.2012.00789.x.

Weiher, E. and Keddy, P. (1999) Assembly rules as general constraints on community composition. In: Weiher, E. and Keddy, E.C.P. (eds) *Ecological Assembly Rules: Perspectives, Advances, Retreats*. Cambridge University Press, Cambridge, pp. 251–271.

Whittaker, R.H. (1975) *Communities and Ecosystems*, 2nd edn. Macmillan, New York.