# NEMATODES AS ENVIRONMENTAL INDICATORS

Edited by

**Michael J. Wilson**

*Institute of Biological and Environmental Sciences,*
*The University of Aberdeen, Aberdeen, Scotland, UK*

**Thomais Kakouli-Duarte**

*EnviroCORE Department of Science and Health,*
*Institute of Technology,*
*Carlow, Ireland*

www.cabi.org

**CABI is a trading name of CAB International**

CABI Head Office
Nosworthy Way
Wallingford
Oxfordshire OX10 8DE
UK

Tel: +44 (0)1491 832111
Fax: +44 (0)1491 833508
E-mail: cabi@cabi.org
Website: www.cabi.org

CABI North American Office
875 Massachusetts Avenue
7th Floor
Cambridge, MA 02139
USA

Tel: +1 617 395 4056
Fax: +1 617 354 6875
E-mail: cabi-nao@cabi.org



A catalogue record for this book is available from the British Library, London, UK.

**Library of Congress Cataloging-in-Publication Data**

Nematodes as environmental indicators / edited by Michael J. Wilson, Thomais Kakouli-Duarte.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-84593-385-2 (alk. paper)
1. Nematodes–Ecology. 2. Indicators (Biology) I. Wilson, Michael J. (Michael John), 1964- II. Kakouli-Duarte, Thomais. III. Title.

QL391.N4N382 2009
592'.5717--dc22

2008049111

ISBN-13: 978 1 84593 385 2

Typeset by SPi, Pondicherry, India.
Printed and bound in the UK by the MPG Books Group.

The paper used for the text pages in this book is FSC certified. The FSC (Forest Stewardship Council) is an international network to promote responsible management of the world's forests.

# Contents

Contributors vii
Preface xi

**Part I Nematodes and Bioindication – General Considerations**

1 The Role of Nematodes in Ecosystems 1
*G.W. Yeates, H. Ferris, T. Moens and W.H. Van der Putten*

2 Nematode Diversity in Terrestrial, Freshwater Aquatic and Marine Systems 45
*M. Hodda, L. Peters and W. Traunspurger*

3 Molecular Markers, Indicator Taxa, and Community Indices: the Issue of Bioindication Accuracy 94
*K. Ekschmitt and G. Korthals*

**Part II Analysis of Nematode Assemblages in Environmental Samples**

4 General Community Indices that can be used for Analysis of Nematode Assemblages 107
*D.A. Neher and B.J. Darby*

5 Indices Developed Specifically for Analysis of Nematode Assemblages 124
*H. Ferris and T. Bongers*

Villenave, C., Bongers, T., Ekschmitt, K., Djigal, D. and Chotte, J.L. (2001) Changes in nematode communities following cultivation of soils after fallow periods of different length. *Applied Soil Ecology* 17, 43–52.

Wu, J.H., Fu, C.Z., Liang, Y.L. and Chen, J.K. (2004) Distribution of the meiofaunal community in a eutrophic shallow lake of China. *Archiv für Hydrobiologie* 159, 555–575.

Yeates G.W., Bongers T., de Goede R.G.M., Freckman D.W. and Georgieva S.S. (1993) Feeding habits in soil nematode families and genera – an outline for soil ecologists. *Journal of Nematology* 25, 315–331.

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

Deborah A. Neher and Brian J. Darby

*Department of Plant and Soil Science, University of Vermont, 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 its 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, second, 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-Second World War 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.



**Table 4.1.** Mean (and standard deviation) of Hill's N1 diversity (N1 = exp[H']) at the species, genus, family, and trophic level for nematodes from organic and conventional grassland management regimes of three Welsh soils: Conway fine silt, Moor Gate coarse loam and Newport sand ($n$ = 10 in all cases). Data from Yeates and Cook (1998).

| | Silt[a] | | Loam[a] | | Sand[a] | |
|---|---|---|---|---|---|---|
| | Conventional | Organic | Conventional | Organic | Conventional | Organic |
| $N1_{species}$ | 10.7 | 13.1 | 14.0 ** | 17.4 | 17.8 | 15.8 |
| | (3.70) | (3.48) | (3.02) | (2.61) | (3.70) | (2.47) |
| $N1_{genus}$ | 9.8 | 12.2 | 12.6 ** | 15.4 | 16.1 * | 13.9 |
| | (3.38) | (3.10) | (2.57) | (2.38) | (3.40) | (1.96) |
| $N1_{family}$ | 8.2 | 8.6 | 9.7 * | 11.1 | 12.7 ** | 10.8 |
| | (2.74) | (2.02) | (1.74) | (1.63) | (2.42) | (0.95) |
| $N1_{trophic}$ | 2.7 | 2.8 | 3.0 | 3.1 | 3.1 * | 3.5 |
| | (0.57) | (0.41) | (0.41) | (0.30) | (0.47) | (0.43) |

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

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

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

### Identity-independent indices and their calculation

A variety of identity-independent indices is available to serve different purposes in different circumstances (Hill, 1973; Peet, 1974; Pielou, 1975). The total

number of species collected from a sample can be referred to as *species richness* (if representative of a known number of individuals) or *species density* (if representative of a known number per mass or volume of soil). *Evenness* is the equitable distribution of proportions or relative abundance. Diversity, then, is a combination of both richness and evenness elements. 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 in Table 4.2 and accompanied by a customized SAS macro written to compute all indices (Fig. 4.1).

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

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

*$p_i$ 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.



```
%let species = a b c d e f g h;
data countdata;
    input sample &species;
cards;
 1    12    18    17    12    3    4    18    15
 2    10    11    28    8     26   3    7     8
 3    10    11    19    22    18   6    3     13
 4    12    18    9     13    18   15   12    4
 5    19    5     4     26    4    11   17    14
 6    15    18    0     11    14   13   14    14
 7    14    12    19    12    6    9    13    16
 8    16    15    19    18    5    9    2     16
 9    20    4     16    7     26   3    3     21
10    14    23    27    5     0    12   14    4
;
proc IML;
use countdata;
read all var {&species} into data;
/* N = column vector of count sums */;
N = data[,+];
/* p = matrix of proportions */;
p = j(nrow(data),ncol(data),0); *Pre-allocate space;
p = data # (1/N); *elementwise division of data by sums;
/* richness = column vector of taxa present */;
richness = (data>0)[,+]; *only data > 0 are used;
MargalefsD = (richness - 1) # (1 / log(N)); /* Margalef's D index as richness corrected for N */;
/* Shannon's H index as the opposite of the sum of all proportions times ln(proportions) */;
nonzeros = loc(p>0); *nonzeros = a row vector of elements of p that are present;
ShannonH = j(nrow(data),ncol(data),0); *Pre-allocate space to speed up computation;
ShannonH[nonzeros] = p[nonzeros] # log(p[nonzeros]); *all absent species remain zero;
ShannonH = -ShannonH[,+]; *opposite to the sum of all columns;
/* Simpson's D dominance index (community and sample) as the sum of all proportions squared */;
SimpsonD = p[,##];
Simpsonfinitelambda = ((data # (data - 1))[,+]) / (N # (N - 1));
/* Hill's diversity (N1 and N2) and evenness (E21) */;
HillsN1 = exp(ShannonH); *exponent of Shannon's H ;
HillsN2 = 1/SimpsonD; *inverse of SimpsonD;
HillsE21 = HillsN2 / HillsN1; *ratio of N2 to N1;
/* Heips alternative evenness */;
BrillouinJprime = ShannonH/log(richness);
HeipE = (HillsN1 - 1) / (richness - 1);
/* Brillouin's indeces */;
if any(N>=100) then largeN = loc(N>=100); *largeN = a row vector of locations where N >= 100;
if any(N<100) then smallN = loc(N<100); *smallN = a row vector of locations where N < 100;
/* SAS fact(n) may not compute factorials for large n (> 100) so it is necessary to run */;
/* an alternate module to compute the log of n! by summing a vector of 1:n */;
/* IML module to compute natural log of factorial of large (>100) n */;
/* ------------------------------------------------------------- */;
/* use: factorial(n) returns: log(n!) */;
start factorial(n);
factorial = j(nrow(n),1,0);
do k = 1 to nrow(n);
```

**Fig. 4.1.** Annotated SAS/IML code to illustrate an approach to implementing the univariate indices of Table 4.2 from within SAS.

```
    a = 1:n[k];
    temp = log(a);
    factorial[k] = temp[,+];
end;
return (factorial);
finish;
BrillouinH = j(nrow(data),ncol(data),0); *Pre-allocate space;
BrillouinH = log(fact(data))[,+]; *Sum of log(Ni!), which equals the logarithm of the products of Ni!;
logNfact = factorial(N); *Call IML module factorial(n) for n > 100;
BrillouinH = (1/N) # (logNfact - BrillouinH); *final BrillouinH calculation
/* Brillouin's Hmax */;
intBrillouinHmax = int(N # (1/richness)); *integer portion;
r = mod(N,richness); *remainder, or modulus;
BrillouinHmax = j(nrow(data),1,0); *Pre-allocate space;
do j = 1 to nrow(data); *repeat for each row;
BrillouinHmax[j] = log((factorial(intBrillouinHmax[j]) ## (richness[j] - r[j])) # (factorial((intBrillouinH
   max[j] + 1)) ## r[j]));
end;
BrillouinHmax = (1/N) # (logNfact - BrillouinHmax); *final computation
/* Brillouin's Hmin */;
BrillouinHmin = factorial(N - richness + 1);
BrillouinHmin = (1/N) # (factorial(N) - BrillouinHmin);
/* Brillouin's evenness (J for samples, Jprime for collections) and relative evenness (Vrel) */;
BrillouinJ = BrillouinH / BrillouinHmax; * ;
BrillouinJprime = ShannonH / log(richness); * ;
BrillouinVrel = (BrillouinH - BrillouinHmin) / (BrillouinHmax - BrillouinHmin); * ;
print N richness MargalefsD ShannonH SimpsonD Simpsonfinitelambda HillsN1 HillsN2 HillsE21
    BrillouinH BrillouinHmax BrillouinJ BrillouinJprime BrillouinHmin BrillouinVrel;
CREATE indices var {richness MargalefsD ShannonH SimpsonD Simpsonfinitelambda HillsN1
    HillsN2 HillsE21
    BrillouinH BrillouinHmax BrillouinJ BrillouinJprime BrillouinHmin BrillouinVrel};
APPEND;
quit;
data final;
    merge countdata(keep=sample) indices(keep = richness MargalefsD ShannonH SimpsonD
    Simpsonfinitelambda HillsN1 HillsN2 HillsE21
    BrillouinH BrillouinHmax BrillouinJ BrillouinJprime BrillouinHmin BrillouinVrel);
run;
proc print data = final;
run;
```

**Fig. 4.1.** Continued

The procedure for enumerating nematodes should be standardized with each experiment or sampling regime to prevent artifacts of sampling effort when reporting richness and diversity indices. Nematodes are enumerated differently than, for example, vascular plant surveys, in that nematodes are not enumerated as they appear *in situ*, but rather are extracted from soil, benthos, or water samples. 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 total number of species expressed as a uniform *portion of all extracted individuals* (e.g. 20% of the individuals from each sample). Regardless, for quantitative comparisons, it is preferable to begin with approximately the same initial mass or volume of soil, water, or benthos. As a rough guide, the range of initial mass or volume for all samples of comparison should be within 5% of the mean. Notice that species richness and density are not necessarily linear in relationship. For example, 20 species found among 200 individuals may not extrapolate to 40 species from 400 individuals. For this case, *rarefaction* of original data is necessary to estimate the number of species collected from a hypothetical number of individuals from the same population and species abundance curve (Gotelli and Colwell, 2001). It is essential to explicitly state the conditions in which richness indices were computed. Sometimes it is neither practical nor possible to enumerate a uniform number or portion of individuals, such as nematodes collected from small, isolated habitats such as pitcher plants, epiphytes, or an insect. In such a case, Margalef's index (= [S – 1] / ln[N]) can be used to adjust the number of species (S) for the number of individuals enumerated (N).

Evenness indices appear infrequently in the literature. Heip (1974) proposed an evenness index (= [exp(H') –1] / [S – 1]) which standardizes the Shannon's diversity index (H') by total number of species (S). Alternatively, Brillouin (1962) developed a series of statistics for censused communities that are computationally complex. Brillouin's maximum theoretical diversity 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 and the second ('relative evenness') as diversity relative to maximum diversity but scaled to minimum diversity. The former relative evenness (not scaled to minimum diversity) can be based on two estimates of diversity depending on whether the user wishes to assume a finite or infinite community enumeration. Use Brillouin's sample diversity relative to Brillouin's maximum diversity when assuming a finite community enumeration, or Shannon's population diversity relative to the natural logarithm of richness when assuming infinite community enumeration. The second 'relative' form of evenness (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, e.g. small isolated habitats or virtual individuals in a computationally simulated model community.

Ecologists disagree on the best method to incorporate both richness and evenness, as well as the degree to which dominant and rare species, respectively, should influence the index. Therefore, exercise caution in application and interpretation of diversity indices. Shannon's diversity (= $-\Sigma[\, p_i \ln p_i]$, Shannon, 1948) is a popular diversity index. The exponent of Shannon's index (=exp[H'], also called 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 similar to the community of 20 non-uniformly distributed species. Furthermore, Heip's evenness index (above) = ([9.97 – 1] / [20 – 1]) = 0.47, indicating again that roughly half of the observed species would be necessary to produce a similar Shannon's index if they were uniformly distributed. Simpson's index (D (= $\Sigma p_i^2$), Simpson, 1949), is considered a dominance index because it increases as species are distributed more unevenly (an increase in 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 = 1 / D) 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.

## Community Assemblage Models

### 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 favoring 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 whereas Ruf (1998) applied a similar approach to mesostigmatid mites. An alternative approach is to quantify species assemblage patterns. This can be achieved by repeated sampling methods or a Mantel test (Manly, 1997). These approaches are computationally intensive but practical given the speed of current computer systems. Repeated sampling methods include techniques referred to as bootstrap, resampling, jackknife, randomization, and Monte Carlo (Manly, 1997). A Mantel test computes a correlation coefficient among matrices. Each matrix can represent an assemblage of species in a community through time. Data can be entered as raw or transformed in variables that are continuous, ordinal or binary (Peres-Neto and Jackson, 2001). One can test hypotheses that concern the (dis)similarity of order and composition between two communities or treatments. A third variable, e.g. spatial pattern can be adjusted by using a partial Mantel test. These approaches are rank or distribution-free which allow them to be applied to small and unbalanced data. Data are reshuffled or resampled repeatedly for 10,000 to 100,000 times to compute a *P*-value and confidence intervals. The level of significance possible is affected by the choice of distance measure (Jackson, 1995). Distance can be quantified in Euclidean and non-Euclidean spaces (e.g. genetic distance, Bray Curtis). The methods vary in their sampling approach (i.e. with or without replacement) and the sample size (i.e. replacing the whole or subset of original sample). In addition to ecological succession, these approaches can also be applied to quantifying other ecological phenomena including intrinsic rate of increase (*r*), estimate of genetic distance, ecological divergence or phylogenies, microarrays, and biogeography (Felsenstein, 1985; Hillis and Bull, 1993; Jackson, 1995; Efron *et al.*, 1996; Rossi, 1996, 2003; Diniz-Filho *et al.*, 1998; Kerr and Churchill, 2001).

An alternative approach is the Procrustean superimposition approach (Gower, 1971). It differs from Mantel by scaling raw data or their ordination solutions to find optimal superimposition rather than transformation. This avoids the problem that the space between distances of transformed variables may not be necessarily equivalent to ones taken from the space of the original data. This approach is more powerful and results in lower type I error rates than the Mantel test (Peres-Neto and Jackson, 2001). Commercial computer software is available to compute most of these indices (Table 4.3).

### Neutral community assemblage models

In addition to the niche-based models that are the impetus for the successional, seasonal, disturbance, and habitat-based studies that dominate historical nematode community analyses, non-neutral models present a necessary alternative perspective to spatio-temporal dynamics of communities. Neutral models have been in use for some time but were brought to the forefront of ecology as Hubbell (2001) presented a neutral model of community dynamics whereby individuals immigrate from a metacommunity into a local community at random with ecologically equivalent fitness. The spatio-temporal dynamics of neutral communities resemble neutral drift, analogous to random genetic drift. The surprising result of the neutral community model is that the distribution of species abundance closely resembles natural abundance distributions. This finding is controversial because, although neutral models can predict many ecological patterns, Hubbell's implicit assumption of neutrality (i.e. ecological equivalence) challenges nearly 150 years of niche-based ecology that sought to delineate the niche boundaries of what were believed to be non-neutral species. However, many authors suggest that neutral and non-neutral models of community assemblage may not necessary be contradictory, but rather complementary (Chave, 2004). Neutral dynamics may occur when non-neutral interactions play out on a reciprocal

**Table 4.3.** Software packages containing univariate statistical procedures.

| | | EstimateS[a] | Primer-E[b] | R[c] | Canoco 4.5 | PC-ORD 4 | PROTEST[d] | NTSYS-PC[e] |
|---|---|---|---|---|---|---|---|---|
| Univariate | Diversity | | × | | × | × | | |
| | Shannon | × | | | | | | |
| | Simpson | × | | | | | | |
| | Margalef's | | | | | | | |
| | Evenness | | | | × | × | | |
| | Brillouin | | | | | | | |
| Ecol. Succ. | Maturity | | | | | | | |
| | Simple Mantel | | × | × | | × | | × |
| | Partial Mantel | | | × | | | | |
| | Procrustes | | | | | | × | × |
| | Jackknife | × | | | | × | | × |
| | Bootstrap | × | | | | | | × |

[a]EstimateS, Statistical Estimation of Species Richness & Shared Species from Samples (viceroy.eeb.uconn.edu/estimates).
[b]Primer-E, version 5 available, Plymouth Routines in Multivariate Ecological Research (www.primer-e.com).
[c]The R Package, (http://www.bio.umontreal.ca/Casgrain/en/labo/R/v4/index.html), French version also available.
[d]PROTEST software available on http://www.zoo.utoronto.ca/jackson/software/.
[e]NTSYS-PC (Rohlf, 1994), Version 2.2 was initially released Sept. 2005 (http://www.exetersoftware.com/cat/ntyspc/ntsyspcfaq.html), Exeter Software Inc., 100 North Country Road, Building B, Setauket, NY 11733.



non-uniform fitness landscape, in effect, equalizing fitness. Neutral processes likely occur within nematode communities because many species appear to be functionally redundant. However, nematodes may not be suitable to test neutral theories because most representative neutral models are spatially and temporally explicit (Holyoak and Loreau, 2006) and the destructive nature of nematode enumerations prevent truly repeated samplings of the same volume. It is important to remember that neutral dynamics may occur over the course of an experiment and it may not be advantageous to force niche-based explanations onto what may be neutral dynamics.

## Multivariate Techniques

Multivariate analysis offers both descriptive and inferential procedures to analyse multiple variables simultaneously so as 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 dataset 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 (see also Trett *et al.*, Chapter 12, this volume). Commercial software packages that compute these procedures are summarized in Table 4.4.

### Cluster analysis

Cluster analysis treats each multivariate observation (sample) as a vector and attempts to group vectors that are similar to each other into clusters (see Figure 12.5). Cluster analysis begins with a (dis)similarity matrix, often computed as the Euclidean distance 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 have been obtained. Divisive clustering begins with one cluster containing all vectors and sequentially divides the cluster into two until an optimal number of clusters have 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

**Table 4.4.** Software packages containing multivariate statistical procedures.

| Category | Statistic | SAS[a] | SPSS[b] | Statistica[c] | SYSTAT[d] | Canoco[e] | Primer-E[f] | PC-ORD[g] | SYN-TAX |
|---|---|---|---|---|---|---|---|---|---|
| | Cluster | CLUSTER<br>FASTCLUS | CLUSTER | Join | Join | | × | × | × |
| | Discriminant | DISCRIM<br>STEPDISC | DISCRIMINANT | Discriminant | MGLH | | × | | |
| Direct Gradient | Canonical correspondence analysis (CCA) | | | | | × | | × | × |
| | Nonmetric multidimensional scaling (NMDS) | | | | | | × | × | × |
| | Redundancy Analysis (RDA) | | | | | × | | | × |
| | Detrended canonical correspondence analysis (DCCA) | | | | | × | | | |
| | Canonical correlation analysis | CANCORR | | | | | | | |
| Indirect gradient | Polar (= Bray-Curtis) Ordination (PO) | | | | | × | × | | |
| | Principal coordinates analysis (PCoA) | | | | | × | | | |
| | Principal components analysis (PCA) | PRINCOMP<br>FACTOR | FACTOR | Factor | Factor | × | × | × | × |
| | Correspondence analysis (CA) | CORRESP | | | | × | | × | × |
| | Detrended correspondence analysis (DCA) | | | | | × | | × | |
| | Principal response curves (PRC) | | | | | × | | | |

[a]SAS Version 9.1 (http://www.sas.com/).
[b]SPSS Version 15 (http://www.spss.com/).
[c]Statistica Version 8 (http://www.statsoft.com/).
[d]SYSTAT Version 12 (http://www.systat.com/).
[e]Canoco Version 4.5 (http://www.microcomputerpower.com/).
[f]Primer-E: Plymouth Routines in Multivariate Ecological Research (http://www.primer-e.com).
[g]PC-ORD, Version 4, MjM Software Design.



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. farthest 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 so as to minimize the increase in sum of squares within and between clusters. The result of hierarchical cluster analysis is a dendrogram (tree diagram) that shows each step of the clustering procedure and the distance at which the clusters merge.

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 and Clark, 1997; McGarigal *et al.*, 2000; Gotelli and Ellison, 2004).

## Ordination

Ordination techniques are popular in community analysis due to their ability to visualize data in two-dimensional space. 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 dataset. *Direct gradient analysis*, also called constrained, 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 Nonmetric Multi-Dimensional Scaling (NMDS). In polar ordination, two samples most different from each other based on their species composition serve as endpoints 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. Principal Coordinates Analysis simply maximizes linear distance measures of the ordination in metric space (using a 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. Computer software packages are commercially available to compute any of these methods (Table 4.4).

The concept of eigenanalysis, used in the remaining indirect and direct gradient analyses, is important but somewhat more tedious. Eigenanalysis is a procedure to reduce the dimensionality of data that also begins with a square distance, similarity, correlation or covariance matrix. The result of

eigenanalysis, an *eigenvalue* and its corresponding *eigenvector*, describes the data matrix as a multidimensional volume. Consider a cluster of samples with three species ($x$, $y$, and $z$) plotted in three-dimensional space (i.e. along axes $x$, $y$, and z) that take the shape of a ball. Eigenanalysis of these data circumscribes a volume around the points and the dominant eigenvalue (one of three) describes the length of the longest dimension and so describes the greatest amount of variance in the data. If the second eigenvalue, which is the length of the second longest dimension, is much shorter than the first, the ellipse around the points in these two dimensions is oblong, and the three-dimensional volume resembles the shape of a rugby football. Just as eigenvalues describe the shape of the volume of data points, eigenvectors describe the orientation of the data points, with the first eigenvector defining the orientation of the first eigenvalue, and so on. In this way, datasets with multiple variables can be visualized in multi-dimensional space derived from latent gradients.

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

### Data transformation

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



### Visualization

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

## Conclusion

Classical community composition can be analysed using metrics that either disregard or preserve the identity of taxa 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. Given the advances in computer technology, 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. We recommend Afifi and Clark (1997), Legendre and Legendre (1998), Lepš and Šmilauer (2003), and Gotelli and Ellison (2004) as helpful texts for further study of these techniques.

## References


Afifi, A.A. and Clark, V. (1997) *Computer-Aided Multivariate Analysis*. 3rd edn. Chapman and Hall, New York.

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, 13429–13434.

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 Ellison, A.M. (2004). *A Primer of Ecological Statistics*. Sinauer Associates, Sunderland, Massachusetts.

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, 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.

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.

Kerr, M.K. and Churchill, G.A. (2001) Bootstrapping cluster analysis: Assessing the reliability of conclusions from microarray experiments. *Proceedings of the National Academy of Sciences USA* 98, 8961–8965.

Legendre, P. and Legendre, L. (1998) *Numerical Ecology*. 2nd English edition. Elsevier Science, Amsterdam.





Lepš, J. and Šmilauer, P. (2003) *Multivariate Analysis of Ecological Data using CANOCO*. Cambridge University Press, Cambridge.

Manly, B.F.J. (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd edition. Chapman and Hall, New York.

Margalef, R. (1958) Information theory in ecology. *General Systems* 3, 36–71.

McGarigal, K., Cushman, S. and Stafford, S. (2000) *Multivariate Statistics for Wildlife and Ecology Research*. Springer-Verlag, New York.

Neher, D.A. (1999) Soil community composition and ecosystem processes: Comparing agricultural ecosystems with natural ecosystems. *Agroforestry Systems* 45, 159–185.

Neher, D.A. (2001) Role of nematodes in soil health and their use as indicators. *Journal of Nematology* 33, 161–168.

Peet, R.K. (1974) The measurement of diversity. *Annual Review of Ecology and Systematics* 5, 285–307.

Peres-Neto, P.R. and Jackson, D.A. (2001) How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* 129, 169–178.

Pielou, E.C. (1975) *Ecological Diversity*. Wiley, New York.

Rohlf, F.J. (1994) NTSYS-PC: Numerical taxonomy and multivariate analysis system, version 1.80. Exeter Software, Setaukat, NY.

Rosenberg, R. (1976) Benthic faunal dynamics during succession following pollution abatement in a Swedish estuary. *Oikos* 27, 414–427.

Rossi, J.P. (1996) Statistical tool for soil biology. 11. Autocorrelogram and Mantel Test. *European Journal of Soil Biology* 32, 195–203.

Rossi, J.P. (2003) The spatiotemporal pattern of a tropical earthworm species assemblage and its relationship with soil structure. *Pedobiologia* 47, 497–503.

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.

Schafer, C.T. (1973) Distribution of foraminifera near pollution sources in Chaleur Bay. *Water Air Soil Pollution* 2, 219–233.

Shannon, C.E. (1948) A mathematical theory of communication. *Bell System Technical Journal* 27, 379–423, 623–656.

Simpson, E.H. (1949) Measurement of diversity. *Nature* 163, 688.

ter Braak, C.J.F. (1988) CANOCO – an extension of DECORANA to analyze species–environment relationships. *Vegetatio* 75, 159–160.

ter Braak, C.J.F. (1990) Update notes: CANOCO version 3.10. Agricultural Mathematics Group, Wageningen, The Netherlands.

Van den Brink, P.J. and ter Braak, C.J. (1998) Multivariate analysis of stress in experimental ecosystems by principal response curves and similarity analysis. *Aquatic Ecology* 32, 163–178.

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.

Van den Brink, P.J., Van den Brink, N.W. and ter Braak, C.J.F. (2003) Multivariate analysis of ecotoxicological data using ordination: demonstrations of utility on the basis of various examples. *Australasian Journal of Ecotoxicology* 9, 141–156.

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

Yeates, G.W. and Cook, R. (1998) Nematode fauna of three Welsh soils under conventional and organic grassland farm managements. In: de Goede, R.G.M. de and Bongers, T. (eds) *Nematode Communities of Northern Temperate Grassland Ecosystems*. Focus Verlag, Giessen, pp. 305–313.