Detecting trends in species composition

Thomas E. Philippi, University of Georgia
Philip Dixon, University of Georgia
Barbara E. Taylor, University of Georgia

Available at: https://works.bepress.com/philip-dixon/36/
DETECTING TRENDS IN SPECIES COMPOSITION

THOMAS E. PHILIPPI, PHILIP M. DIXON, AND BARBARA E. TAYLOR

Savannah River Ecology Laboratory, Drawer E, Aiken, South Carolina 29802-0005 USA

Abstract. Species composition reflects a combination of environmental and historical events at a site; hence, changes in species composition can provide a sensitive measure of ecologically relevant changes in the environment. Here, we consider the analysis of species composition when multiple sites are followed through time.

Analyses of temporal trends in species composition either summarize species composition into a few metrics (indices or axis scores) or analyze the similarity among sites. We develop and illustrate the similarity approach. Each pair of samples represents a pair of replicates, a pair from the same site at different times, a pair from different sites at the same time, or an unrelated pair. Differences among times can be estimated by comparing average temporal dissimilarity to average replicate dissimilarity. Temporal trends can be described by one of three statistics that measure progressive change, the correlation of temporal dissimilarity with the length of time between samples. These methods are illustrated using data on changes in a South Carolina zooplankton assemblage following disturbance, and changes in bird species composition on Skokholm Island, Wales.

It is difficult to define and interpret temporal trends. Some definitions of interesting trends, like increasing divergence from another set of sample plots, place additional requirements on the sampling design. Including replicate samples or clustering sample plots and including “control” plots for comparison with sentinel sites would contribute to an understanding of changes in species composition.

Key words: community change; dissimilarity; multidimensional scaling; ordination; randomization tests; species composition; temporal differences; trends.

INTRODUCTION

Most discussions of monitoring, including those in this special feature, deal with scalar quantities: environmental conditions such as pH, or abundances of individual species. However, not all quantities that might be useful to monitor are scalars; species composition at monitored sites is a vector, or multivariate quantity. In general, species composition may be analyzed as vectors of 0’s and 1’s representing absence or presence of species, vectors of abundances of each species, or vectors of proportional abundance. Although most of the principles and problems applicable to detecting change or trends in scalars also apply to species composition, the multivariate nature of species composition presents additional problems. Change or differences in composition among samples may be quantified, but may be difficult to interpret. Trends are much more difficult to define, especially if species composition is responding to changes in more than one stressor. Here, we review the forms of questions addressed with species composition data, propose some methods using dissimilarity metrics, and then illustrate the methods using changes in a South Carolina zooplankton assemblage and in the bird community of Skokholm Island, Wales. Similar sorts of questions and analyses are appropriate for many communities with multiple sources of variation, e.g., environmental stress and spatial gradients (Vargo et al. 1982, Fausch et al. 1990, Clarke 1993).

WHY MONITOR SPECIES COMPOSITION?

A need to monitor species composition may arise in at least three ways: species composition may provide a strong signal of the environmental factor of interest; it may be of direct ecological interest; or it may be the only feasible measurement. Because environmental factors differentially affect species, species composition can provide information about environmental factors (Austin and Austin 1980, ter Braak and Prentice 1988). Indirect estimation of environmental characteristics from species composition can be extremely useful if it is difficult to define or directly measure relevant characteristics of the environment. Two examples are paleoecological reconstruction of environmental conditions, such as climate or pH, from historical species composition and modern species–environment relationships (Oksanen et al. 1988, Birks et al. 1990), and using changes in the shrubby and herbaceous vegetation to monitor changes in site capability for timber
production, which might take decades to measure directly (Korstan 1919, Leak 1992). Infrequent impacts (e.g., occasional discharges) may be difficult to detect directly without continuous monitoring, but species composition may effectively and appropriately integrate such acute impacts over time, providing a means of detection and an estimate of severity or importance. Fish species composition has been used as an index of pollution in rivers and streams (Fausch et al. 1990), and changes in marine benthic invertebrates have been used as measures of environmental impact (Warwick and Clarke 1991).

Species composition itself may be the relevant environmental end point, especially in an environmental impact context where the concern is that there be no change in species composition due to operation of some facility (e.g., Gray et al. 1990, Environment Canada 1995). Finally, percent species composition may be the only feasible measurement (e.g., in palynological studies). Such values are not independent among species; a species can have a low percent abundance in a particular sample, either because it has low absolute abundance in that sample, or because another species has extraordinarily high abundance in that sample (e.g., Hellawell 1977). Therefore, it can be quite misleading to treat such data as reflecting patterns in individual species. For some taxa (e.g., soil microorganisms), it is only possible to detect presence or absence; thus, an individual species provides little information.

Changes in species composition may be of interest for both retrospective and prospective monitoring. Although compositional change may be an “effect” relative to some other primary stressor, it may, at the same time, be a “stressor” predictive of future changes in ecosystem properties such as biomass and nutrient cycling, and thus may allow earlier detection and intervention. Underwood (1994) suggested that, although univariate chemical or indicator species measures may respond rapidly with strong signals to other stressors and, thus, may be preferable to assemblage data, analysis of species or assemblage composition generally is required if the goal is to predict community or ecosystem consequences of a change.

In a typical monitoring design, samples are collected or measured at multiple sites monitored over time. Ideally, replicate samples are collected, but this is often not done. Although many different quantitative techniques have been used to display and summarize patterns in species composition (e.g., Legendre and Legendre 1983, Washington 1984, Digby and Kempton 1987, Magurran 1988, and ter Braak and Prentice 1988), we will focus on the analysis of dissimilarity among samples, because it most directly focuses on questions about temporal change. Each pair of samples is either a replicate pair from the same site and time, a temporal pair from the same site at different times, a spatial pair from different sites at the same time, or a mixed pair from different sites at different times.

We divide questions addressed with species composition into three temporal categories. Composition may be compared among groups of sites at a single time, the amounts and directions of change in composition between pairs of samples may be compared among groups of sites, and trends or progressive change across a series of dates may be compared among sites. Each class of change corresponds to a comparison of dissimilarity between types of pairs (e.g., temporal pairs to replicate pairs).

**Comparison Among Sites at a Single Time**

If species composition changes between impacted and unimpacted sites, then the pairwise dissimilarities among impacted sites and among unimpacted sites will be smaller than the pairwise dissimilarities between impacted and unimpacted sites. These dissimilarities can be summarized either by the mean or median pairwise intergroup dissimilarity, or by a test statistic such as a Mann-Whitney U or Clarke’s (1993) R that compares between-group to within-group dissimilarities. Clarke’s R is computed by ranking all n(n−1)/2 pairwise dissimilarities among the n sites, and then computing

\[
R = \frac{4(r_b - r_w)}{n(n-1)}
\]

where \(r_b\) is the average rank of impacted–unimpacted (between-group) pairs, and \(r_w\) is the average rank of within-group pairs. If the smallest difference between groups is larger than every difference within groups, then \(R = 1\), but if differences between groups are similar to those among groups, \(R\) will be close to 0.

Randomization tests are required to evaluate whether dissimilarities between groups are larger than expected by random chance (Field et al. 1982, Smith et al. 1990, Clarke and Warwick 1994). Traditional significance tests are invalid because dissimilarities are computed from all pairs of samples. Even if each of the n sites is considered to be independent, the n(n−1)/2 possible pairwise dissimilarities are not. In each randomization, labels (e.g., impacted or unimpacted) are randomly reassigned (without replacement) to sites, and the measure is recomputed. The calculation is repeated for all permutations of labels, or for a sample of permutations if there are many. Note that the dissimilarity matrix need be computed only once. The P value for the hypothesis test is \((x + 1)/(n + 1)\), where \(x\) is the number of more extreme values found in \(n\) randomizations. This randomization test maintains the species composition in each sample, but it randomizes labels (e.g., impacted/unimpacted) among samples. For most questions, this is the most appropriate randomization scheme, but it is not the only possible scheme. It is also possible to randomly shuffle species among samples (Dixon...
1994) or to randomly shuffle individuals among species and samples (Solow 1993).

**Evaluation of Temporal Change**

Three major classes of questions about temporal change are as follows. Is temporal dissimilarity greater than sampling variation between replicates? Does one group of sites change more than another group, and does between-group dissimilarity increase or decrease over time? Is there a temporal trend in dissimilarity at a given site? Answering these questions is more difficult than answering similar questions about scalar quantities. If there is no change in a scalar environmental variable, the expected difference between two samples is zero, because sampling errors are both positive and negative. With a sample of compositional data, all dissimilarities are positive, so the expected dissimilarity is nonzero, even if there is no true difference between two samples. The expected dissimilarity can be estimated with monitoring designs that include replicate samples at each site and time.

For the first class of questions, if temporal change is larger than sampling variation, then the temporal dissimilarities within replicates are larger than the replicate dissimilarities computed among replicates within each time. This is a comparison between temporal similarity and replicate similarity, and it can be tested using Clarke’s R and randomization tests (Clarke 1993). If multiple sites are sampled, the test can be performed separately for each site, or results from all sites can be pooled (Clarke and Warwick 1994).

Questions about differences in the magnitude of temporal dissimilarity between groups of sites are appropriate when the monitoring follows a Before–After, Control–Impact (BACI) design with replicate sites (Underwood 1992), where impacted sites may be expected to change more than control sites. The appropriate comparisons and tests depend on the details of the sampling design. One common design is to sample multiple, permanently marked sites once before and once after the impact. A single dissimilarity measure, comparing before and after samples, can be computed from each site. These dissimilarity measures are independent, if sites are independent; thus, the difference between control and impacted sites can be tested by standard statistical tests, e.g., the Mann-Whitney U test.

Questions about divergence or convergence among groups of sites over time arise if the monitoring design has repeated sampling during the potential response to impact (divergence) or recovery (convergence), or if restoration and remediation efforts are monitored. The form of possible tests depends on the number of sampling times and pattern of within-group dissimilarities over time. With many temporal samples, a simple rank correlation between date and between-group dissimilarity may be used (possibly corrected for within-group dissimilarity via Clarke’s R). With fewer temporal samples, there is no generally applicable test, although situation-specific randomization tests may be possible.

Temporal trend in a scalar quantity is relatively easy to define: a tendency to increase or decrease over a specific time period. Although it is easy to test for change in species composition, it is more difficult to define trend in species composition. If species composition can be used to estimate some environmental scalar (e.g., by calibration), then the problem is reduced to evaluating a trend in the calibrated scalar. However, a more general definition of trend in species composition is not clear. Requiring a trend in the abundance of every species is overly restrictive. We suggest a less restrictive definition, which we call progressive change, that the dissimilarity in species composition between two samples tends to increase with their temporal separation.

Progressive change can be tested using at least three different test statistics. Each is sensitive to subtly different types of change. In all cases, the concern is with trend over time, not simply variation among times, so the hypothesis of no trend should be tested by randomizing each set of replicates together. Alternatively, one can randomize labels (times) to the average dissimilarities. The simplest test for progressive change focuses on change from a specified baseline condition, and considers the correlation or rank correlation between time and dissimilarity with the baseline. Suitable nonparametric test statistics include Spearman’s rank correlation and Kendall’s tau correlation (Kendall and Gibbons 1990). These statistics test the hypothesis that $D_{i1} = D_{i2} = D_{i3} = D_{i4} = D_{i5}$ against the alternative that $D_{i1} < D_{i2} < D_{i3} < D_{i4} < D_{i5}$, where $D_i$ is the dissimilarity between dates $i$ and $j$, and date 1 is the chosen baseline condition. The second test generalizes the possibly arbitrary choice of a single baseline by computing the correlation between date and dissimilarity using each date as the baseline to which subsequent or previous dates are compared. This test is sensitive to alternatives of the form: $D_{i1} < D_{i2} < D_{i3} < D_{i4} < D_{i5}$, $D_{i2} < D_{i3} < D_{i4} < D_{i5}$, $D_{i3} < D_{i4} < D_{i5}$, $D_{i4} < D_{i5}$, $D_{i5} < D_{i5}$, $D_{i3} < D_{i4}$, $D_{i4} < D_{i5}$, $D_{i5} < D_{i5}$, $D_{i2} < D_{i5}$, $D_{i3} < D_{i4}$, $D_{i4} < D_{i5}$, $D_{i5} < D_{i5}$, $D_{i1} < D_{i2}$. This method uses all dates as baselines, but only compares species composition in overlapping time periods (e.g., between dates 1 and 2 and between dates 1 and 3). The third test calculates the correlation between dissimilarity in species composition and difference in times, using all pairs of samples. This is a Mantel test of the association between two distance matrices (Manly 1991). This tests alternatives of the form $(D_{i1}, D_{i2}, D_{i3}, D_{i4}, D_{i5}) < (D_{i1}, D_{i2}, D_{i3}, D_{i4}, D_{i5}) < (D_{i3}, D_{i4}, D_{i5}) < (D_{i4}, D_{i5}) < (D_{i5})$. Unlike the second test, the Mantel approach presumes that change in species composition can be compared between non-overlapping time periods.
Dissimilarity Metrics

The preceding tests were presented in terms of unspecified dissimilarities. Although the general principles of similarity and dissimilarity in species composition are simple and clear, the definition of dissimilarity can be troublesome, because of the infinite number of ways to collapse the difference between two vectors into a single number and the great variety of named similarity and dissimilarity metrics (Goodall 1982, Boyle et al. 1990). There is no universally best dissimilarity metric. Choice of a metric should be based on biological knowledge: what forms of compositional differences are expected or considered to be important. In some cases, simulation studies suggest metrics with desirable properties for specific questions (Faith et al. 1987). Metrics vary in the relative weightings they give to many small differences in abundances vs a few large differences in abundances of common species. Asymmetric metrics only treat shared presences as informative of similarity, which often is appropriate because there are many different reasons why a species might be absent from a sample. Finally, dissimilarity measures are usually computed from a sample of individuals, not from the entire community, and are sensitive to sample sizes; those that are independent of sample size are so only for a specific species abundance distribution (e.g., Morisita for log-series abundances, Wolda 1981).

Calibration

If the goal of a monitoring program is to detect response to environmental change, then detection of change or even trend in species composition is not enough. Temporal change in species composition occurs for many reasons, only one of which is a response to a change in the underlying environment. An obvious example is secondary succession following an acute disturbance such as clearing or a hurricane. However, plant species composition naturally changes at temporal scales from seasonal phenology and year-to-year fluctuations (Philippi et al. 1990) through several millennia, e.g., the 6000-yr chronosequence in McAuliffe (1991) and migration since the last ice age in Davis (1986). Therefore, an additional step is required to allow interpretation of the change in species composition.

If control and impacted plots are monitored, then the divergence between the two sets of sample plots reflects different temporal trajectories, and is thus directly interpretable as a difference caused by the impact. Otherwise, a transfer function may be generated via a calibration data set and may be used to translate the observed changes in composition into differences in environmental factors. The calibration data set is usually a set of species compositions sampled at sites differing in environmental factors. Therefore, one caveat is that the space-for-time substitution may not be valid. Until enough temporal data are collected, the space-for-time substitution must remain an untested assumption of all such analyses. A second caveat is that correlation is not causation, and the calibration data will rarely be from matched, experimentally manipulated sites, so the assumption of all else being equal may be crucial. The calibration approach may be used both for inferring the cause of the compositional change and for inferring the effects of the compositional change on subsequent ecosystem properties.

Example: Trends in Zooplankton Species Composition after Disturbance

We will illustrate the dissimilarity-based analysis of trends and differences in species composition with data on recovery from episodic disturbance of zooplankton in a cooling water reservoir on the Savannah River Site, South Carolina, USA (Leeper and Taylor 1995). Pairs of replicate estimates of zooplankton density were made at six times and at three sites in a nuclear reactor cooling-water reservoir. One site (MC-10) was in the middle of the main channel of the reservoir. The other two sites (SC-30 and BD-30) were in protected coves. The first sample (23 May 1986) was taken when the reactor was operating and the reservoir was receiving heated water. Water temperatures in the main channel were hotter (up to 58°C) than those in the coves (38°C–49°C). No zooplankton were present in the main channel on 23 May 1986. The next four samples (June–late August 1986) were taken during an extended reactor outage that began on 1 June 1986. Water temperatures at the three sites were equivalent (25°C–35°C) when the reactor was not operating. The data used here are a subset from a much larger data set; details of sample collection, processing, and zooplankton counting and analysis of trends in individual species are given in Leeper and Taylor (1995).

We ask four questions. What are the relative magnitudes of dissimilarity between replicate, temporal, spatial, and unrelated pairs? Does the species composition differ between the three sites? Does the species composition change over time? Is there a trend in species composition over time? All four questions can be addressed by comparing dissimilarity in species composition among sets of samples. Because the sampling program included multiple sites and multiple dates, there are four different types of pairs of samples. Replicate pairs are those collected at the same site on the same date. Temporal pairs are collected at the same site on different dates. Spatial pairs are collected from different sites on the same date, and unrelated pairs are collected from different sites on different dates. Pairwise dissimilarities were calculated using the Bray-Curtis metric on fourth-root transformed data, using the DISTANCE macro in SAS version 6.12 (METH- OD=NONMETRI; Kuo 1995).

A visualization of the matrix of pairwise distances
was obtained using nonmetric multidimensional scaling (SAS PROC MDS LEVEL=ORDINAL, SAS version 6.12). NMMDS arranges points in 1, 2, 3, or perhaps more dimensions, so that the ranking of Euclidean distance between points is maximally correlated with the ranking of pairwise dissimilarity in species composition (Kenkel and Orloci 1986, Clarke 1993). A two-dimensional plot provides a reasonable summary of the pairwise distances (Fig. 1). Each pair of replicates lies close together, and points for samples taken at the same time tend to form clusters. Samples taken at different times (temporal clusters) tend to be well separated from each other. Although such plots are very useful for portraying relationships among samples, distances between points on the plots should not be used as a measure of the dissimilarity in species composition (Clarke 1993). MDS, like all other ordination techniques, distorts the actual pairwise distance between samples into an artificial distance in a small number of dimensions. Hence, comparisons and tests should use the observed dissimilarities.

We found that species compositions in replicate pairs collected at the same time and place were quite similar. Between-replicate dissimilarities were small (median = 0.14). The spatial dissimilarities, among sites at the same time, were larger (median = 0.30), but temporal dissimilarity, among sites at the same site, was larger still (median = 0.60). Dissimilarity between samples collected in different times and different sites (median = 0.62) was similar to temporal dissimilarity. The same pattern of replicate dissimilarity (median = 0.12) < spatial dissimilarity (median = 0.33) < temporal dissimilarity (median = 0.65) was found among the subset of samples taken when the reactor was not operating.

The MDS plot suggests that spatial variability among sites is highest on 23 May 1986 and smaller on subsequent dates, and that dissimilarities among sites are larger than those between replicates. Both of these patterns can be confirmed using the observed dissimilarities for spatial and replicate pairs of samples (Table 1). The mean dissimilarities among the spatial pairs are considerably larger on 23 May 1986 than on any subsequent date. On all dates, the mean spatial dissimilarity is larger than the mean replicate dissimilarity. On most dates, the smallest pairwise dissimilarities are associated with replicate pairs of samples, not spatial pairs of samples; thus, Clarke’s R is 1. Because of the small number of samples on any single date, the number of possible permutations in the randomization test is small. When two replicates are taken at three sites, there are 6!/3! x 2! x 2! = 15 possible permutations of labels. None of the tests has P values <0.05, even though the observed results for these data are generally the most extreme possible. If results from all dates are combined, either by Fisher’s method of combining P values (Sokal and Rohlf 1981), or by stratifying by date, the spatial dissimilarity is significantly larger than the replicate dissimilarity (x² = 21.664, df = 10, P = 0.016 for combining approach). Although there are differences in species composition among the sites, these spatial differences are smaller when the reactor is not operating.

Temporal differences in composition can be assessed by comparing the temporal dissimilarities to the replicate dissimilarities (Table 2). Because there are data from five dates at two sites (BD and SC) and from four dates at the other (MC), there are many possible permutations of temporal and replicate dissimilarities. For example, at the BD10 site, there are 10!/5! x 2! x 2! x 2! x 2! = 945 arrangements of the pairwise dissimilarities into temporal and replicate groups. At each of the three stations, the observed between-rep-

---

TABLE 1. Dissimilarity in zooplankton species composition among samples on the Savannah River site, South Carolina, USA. Mean spatial, mean replicate dissimilarity, Clarke’s R, and the probability of observing an equal or larger Clarke’s R under randomization are given for each date and for all dates, stratifying by date.

<table>
<thead>
<tr>
<th>Date</th>
<th>Mean spatial</th>
<th>Mean replicate</th>
<th>Clarke’s R</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>23 May 1986</td>
<td>0.732</td>
<td>0.091</td>
<td>1.0</td>
<td>0.33</td>
</tr>
<tr>
<td>6 June 1986</td>
<td>0.311</td>
<td>0.105</td>
<td>1.0</td>
<td>0.067</td>
</tr>
<tr>
<td>20 June 1986</td>
<td>0.368</td>
<td>0.218</td>
<td>0.89</td>
<td>0.20</td>
</tr>
<tr>
<td>8 July 1986</td>
<td>0.286</td>
<td>0.152</td>
<td>1.00</td>
<td>0.067</td>
</tr>
<tr>
<td>25 July 1986</td>
<td>0.269</td>
<td>0.157</td>
<td>1.00</td>
<td>0.067</td>
</tr>
<tr>
<td>All dates</td>
<td>0.344</td>
<td>0.122</td>
<td>0.973</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Configuration of zooplankton samples in a two-dimensional nonmetric multidimensional scaling (NMMDS) representation of the Bray-Curtis distances. Symbols identify the site: □, BD30; △, SC30; ×, MC10. Solid symbols indicate when the reactor was operating (23 May 1986). Convex hulls enclose all samples taken on the same date. Dates are coded as: 1, 23 May 1986; 2, 6 Jun 1986; 3, 20 June 1986; 4, 8 July 1986; 5, 25 July 1986.
TABLE 2. Replicate and temporal dissimilarity at SC30, BD30, and MC10 stations in 1986, from the dates in column 1 to the dates in the column heads. Values on the diagonal are between-replicate dissimilarities. Values above the diagonal are average pairwise dissimilarities between two dates. Only one sample was counted at SC30 on 6 June 1986, so there is no between-replicate dissimilarity for that date. No species were present at MC10 on 23 May 1986.

<table>
<thead>
<tr>
<th>From:</th>
<th>23 May</th>
<th>6 June</th>
<th>20 June</th>
<th>8 July</th>
<th>25 July</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 May</td>
<td>0.114</td>
<td>0.727</td>
<td>0.896</td>
<td>0.952</td>
<td>0.874</td>
</tr>
<tr>
<td>6 June</td>
<td>...</td>
<td>0.588</td>
<td>0.701</td>
<td>0.724</td>
<td></td>
</tr>
<tr>
<td>20 June</td>
<td>0.278</td>
<td>0.424</td>
<td>0.525</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 July</td>
<td>0.174</td>
<td>0.387</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 July</td>
<td>0.206</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BD30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 May</td>
<td>0.067</td>
<td>0.672</td>
<td>0.746</td>
<td>0.779</td>
<td>0.750</td>
</tr>
<tr>
<td>6 June</td>
<td>0.141</td>
<td>0.699</td>
<td>0.805</td>
<td>0.616</td>
<td></td>
</tr>
<tr>
<td>20 June</td>
<td>0.144</td>
<td>0.285</td>
<td>0.527</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 July</td>
<td>0.125</td>
<td>0.504</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 July</td>
<td>0.101</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 May</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>6 June</td>
<td>0.069</td>
<td>0.587</td>
<td>0.668</td>
<td>0.559</td>
<td></td>
</tr>
<tr>
<td>20 June</td>
<td>0.232</td>
<td>0.391</td>
<td>0.598</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 July</td>
<td>0.158</td>
<td>0.448</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 July</td>
<td>0.165</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Replicate dissimilarities are smaller than any of the temporal dissimilarities; hence Clarke’s R is 1.0 for each site. The observed pattern of rank dissimilarity represents the most extreme temporal difference that can be found by randomizing samples, so P values for tests of no temporal differences are small (P = 0.0011 for BD and SC sites; P = 0.0095 for MC site). However, temporal differences between samples do not imply a trend in species composition.

With these data, there is weak evidence of progressive change. Between 6 June 1986 and 8 July 1986, the species composition is increasingly dissimilar to the species composition on 23 May 1986 (Fig. 1, Table 2). However, this trend does not continue on 25 July 1986. The tests of progressive trend provide equivocal answers (Table 3). The first and second tests suggest that the correlations of dissimilarity and time are moderately large, but not statistically significant. If only data from the first four time periods are used, the sample size is too small to get P values < 0.05, even though there is a perfect correlation between dissimilarity and time (Table 3). If it is reasonable to compare dissimilarities among non-overlapping time periods, the Mantel test is appropriate. It indicates a large and statistically significant correlation of dissimilarity and time (Table 3). What is probably happening to the zooplankton species composition is simultaneous repopulation of the entire reservoir, followed by seasonal succession during July.

EXAMPLE: PROGRESSIVE CHANGE IN THE LAND-BIRD COMMUNITY OF SKOKHOLM ISLAND

As a second example of testing progressive change, we examine changes in the species composition of the land-bird community of Skokholm Island (off the southwest coast of Wales) from 1928 until 1979, using the data given by Williamson (1983). There are no replicate samples, because the data are a complete census of all birds on the entire island, with the exception of some interpolated missing values. NMDS of the Bray-Curtis distance in species composition shows three periods of relative small and erratic changes in species composition from 1928 until 1944, from 1951 until 1960, and from 1975 until 1979 (Fig. 2). In between, there are repeated and consistent trends in species composition, especially from 1963 until 1975. These patterns are similar, but not identical, to the patterns seen by Williamson using a different dissimilarity measure and ordination technique. The progressive trend statistics (Table 4) indicate that there is a highly significant trend in species composition across the entire period. If individual decades are considered, there is strong evidence of progressive change in all decades except 1950–1959, where the evidence is slightly weaker. Although all of these tests indicate that there is progressive change, the test statistics and P values do not indicate the amount or importance of that change. One measure of the magnitude of change is the Bray-Curtis distance between the beginning and end of the period (Table 4). By that measure, the change during the 1930s and 1950s is smaller than that during the 1960s and 1970s, matching Williamson’s (1983) interpretation.

TABLE 3. Spearman rank correlations and randomization P values (in parentheses) for tests of progressive change. Details of each test are given in the text. P values are computed by permuting date labels to the matrices of average dissimilarity in Table 2.

<table>
<thead>
<tr>
<th>Test</th>
<th>SC site</th>
<th>BD site</th>
<th>MC site</th>
</tr>
</thead>
<tbody>
<tr>
<td>A) From 23 May 1986 through 25 July 1986</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change from baseline</td>
<td>0.60 (0.33)</td>
<td>0.80 (0.17)</td>
<td>0.33 (0.83)</td>
</tr>
<tr>
<td>All changes</td>
<td>0.88 (0.017)</td>
<td>0.79 (0.067)</td>
<td>0.78 (0.25)</td>
</tr>
<tr>
<td>Matrix correlation</td>
<td>0.86 (0.017)</td>
<td>0.86 (0.033)</td>
<td>0.85 (0.25)</td>
</tr>
<tr>
<td>B) From 23 May 1986 through 8 July 1986</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change from baseline</td>
<td>1.00 (0.17)</td>
<td>1.00 (0.17)</td>
<td>...</td>
</tr>
<tr>
<td>All changes</td>
<td>1.00 (0.083)</td>
<td>1.00 (0.083)</td>
<td>...</td>
</tr>
<tr>
<td>Matrix correlation</td>
<td>0.92 (0.083)</td>
<td>0.92 (0.083)</td>
<td>...</td>
</tr>
</tbody>
</table>
FIG. 2. Configuration of Skokholm land-bird communities in a two-dimensional nonmetric multidimensional scaling representation of the Bray-Curtis distances. Selected dates are marked at the side of the points.

OTHER APPROACHES

Austin (1977) proposed that indirect ordination could be performed on the entire set of spatial and temporal samples to reduce variation to two dimensions, and then trajectories of individual plots over time in this ordination space could be interpreted to investigate succession. The trajectories over time are presented as arrows tracking the change in a plot’s ordination scores over consecutive time periods. The magnitudes and directions of such arrows are interpreted and perhaps tested, in effect treating Euclidean distances in the two-dimensional ordination space as dissimilarities. This approach has also been widely recommended for other groups of species (e.g., Environment Canada 1995), but has several drawbacks if the ordination is done by a projection method (e.g., PCA, RDA, CA, DCA, CCA). Samples that are very different may be adjacent in a two-dimensional ordination plot if their differences were loaded onto the third and higher ordination axes (Digby and Kempton 1987). This problem can be quite serious, because the first two axes might be determined by random changes in a few rare species (Gauch 1982), in which case the biologically important variation is completely missed, or by two strong spatial environmental gradients, in which case most temporal change might fall along higher axes. The detrending and corresponding rescaling of the first axis in detrended correspondence analysis, the most commonly used variant, exacerbate the problem of treating arrow lengths as dissimilarities.

Aitchison (1986) popularized an approach to composition data analysis that may be appropriate when all or most species are present in every sample. In this approach, the observed species compositions are log-ratio transformed: $Y_i = \log(X_i/X_1)$, where $X_i$ is the proportion of species $i$ in the sample. The proportion of the first species is arbitrarily chosen as a reference proportion. Principal components analysis or multivariate regression modeling of the transformed proportions is used to summarize the multidimensional structure or to model temporal trends (Grunwald et al. 1993). This approach is consistent with the general dissimilarity approach suggested here when Euclidean distance between transformed proportions is used as the dissimilarity measure (Aitchison 1986:193).

The dissimilarity approach described here evaluates change in composition. If the species–environment relationship were known, then calibration could be used to infer environmental change from the species composition change. Another family of approaches reverses the order of those steps, first inferring environmental measures from the species compositions, and then estimating change or trend in those measures. Such procedures go by many names: direct ordination (Gauch 1982), weighted-average calibration (Oksanen et al. 1988, Birks et al. 1990), or an analysis of indices of biotic integrity (Karr 1981, 1991). Although the details of each procedure differ, all require information about the relationship between each species and an environ-

---

**TABLE 4.** Tests and measures of change in the Skokholm land-bird community. Values are given for the entire data set and then for each decade. The 1940s decade was not analyzed because no data were collected during 1941–1945.

<table>
<thead>
<tr>
<th>Period</th>
<th>Dissimilarity to initial sample correl. ($P$)</th>
<th>Dissimilarity in nested sets of years correl. ($P$)</th>
<th>Mantel test correl. ($P$)</th>
<th>Dissimilarity (first to last)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire sequence</td>
<td>0.81 (&lt;0.0001)</td>
<td>0.75 (&lt;0.0001)</td>
<td>0.82 (&lt;0.0001)</td>
<td>0.60</td>
</tr>
<tr>
<td>Individual decades</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1930–1939</td>
<td>0.56 (0.027)</td>
<td>0.70 (&lt;0.0001)</td>
<td>0.80 (0.0015)</td>
<td>0.13</td>
</tr>
<tr>
<td>1950–1959</td>
<td>0.44 (0.067)</td>
<td>0.68 (&lt;0.0001)</td>
<td>0.66 (&lt;0.001)</td>
<td>0.25</td>
</tr>
<tr>
<td>1960–1969</td>
<td>0.83 (&lt;0.0001)</td>
<td>0.84 (&lt;0.0001)</td>
<td>0.89 (&lt;0.0001)</td>
<td>0.33</td>
</tr>
<tr>
<td>1970–1979</td>
<td>0.89 (&lt;0.0001)</td>
<td>0.86 (&lt;0.0001)</td>
<td>0.81 (&lt;0.0001)</td>
<td>0.31</td>
</tr>
</tbody>
</table>
mental variable. For example, if the optimum and tolerance for an important environmental factor are available for each taxa, weighted-averaging calibration will estimate environmental scores for each site. The analysis of trends is then an analysis of the estimated environmental scores.

**Implications for the Design of Monitoring Programs**

The design of a monitoring program for detecting change in species composition suffers from the same conflicting goals as any other ecological monitoring program. We simultaneously want to have widespread, fine-grained spatial coverage and long-term, detailed, temporal sampling at individual sites. Unfortunately, such detailed sampling of so many sites will be too costly to implement for most ecological questions.

One component of a monitoring design for species composition that should not be sacrificed is the estimation of among-replicate similarity. Ideally, this comes from replicate samples. Two replicates are the minimum, but power of tests that compare among-replicate similarity to among-time or among-site similarity is greatly enhanced if more than two replicates are available. With \( R \) replicates, the number of pairs of among-replicate similarities is \( \binom{R^2}{2} = R(R-1)/2 \), i.e., \( R \) for three replicates and 6 for four replicates. If true replicate plots are not possible, spatially clustered or neighboring plots might be usable to estimate within-site dissimilarity; the tighter the clustering, the smaller the additional component of spatial variation inflating the estimated within-site dissimilarity.

Using permanent plots, where possible, gives better measures of change because they reduce the confounding of temporal and small-scale spatial variation (Stewart-Oaten et al. 1995, Bakker et al. 1996), although the subsequent analysis may be more complicated. Permanent plots are especially useful in systems with considerable spatial and random variation in species composition (Thrush et al. 1994). The appropriate size, number, and clustering of such plots are not yet determined, but they will almost certainly differ in different habitats. If sampling of sentinel sites is contemplated, then non-impacted or control sites also should be sampled to allow tests of divergence. If a network of stratified random sites is established, these sites may be able to form a calibration set for interpreting changes in species composition. At least for terrestrial vegetation, there is no natural spatial scale for sampling, so some form of nested sampling, even with cruder estimates of abundances, may be more informative than detailed information from a single size of plots. Finally, plots for compositional analysis should not be so large as to include heterogeneous vegetation within individual plots, because such variation would greatly reduce the power to detect change, especially if the response to change is a shift in a gradient within a plot.

Finally, very little-to-nothing is known about appropriate sample sizes and the trade-off among number of replicates, number of sites, and number of times to sample. The power to detect a trend in species composition depends on the sample size, the magnitude of the trend, and the magnitude of random variation, just as it does when sampling scalar quantities (Urquhart 1998). A detailed study of sample size and the power of tests of change in species composition has not yet been done, but it is possible to suggest some minimum sample sizes. These are determined by the number of possible permutations in the randomization tests, as illustrated by the zooplankton data analysis. Two replicates at three sites were not enough to identify significant differences among sites, but two replicates at four times were sufficient to identify significant differences between times. Five sampling dates were barely sufficient to test for progressive change, but 10 sampling dates were sufficient to detect relatively small changes in the Skokholm data.

**Acknowledgments**

This research was supported by Financial Assistance Award DE-FC09-96SR18546 from the U.S. Department of Energy to the University of Georgia Research Foundation. The very helpful comments of Tony Olsen and two anonymous reviewers were much appreciated.

**Literature Cited**


Dixon, P. M. 1994. Resampling techniques: size hierarchies and similarity indices. Pages 290–318 in S. Scheiner and

Environment Canada and Department of Fisheries and Oceans. 1995. Further guidance for the invertebrate community survey for aquatic environmental effects monitoring related to federal Fisheries Act requirements. Environment Canada, Environmental Effects Monitoring Program, Ottawa, Canada.


