

Bootstrapped ordination: a method for estimating sampling effects in indirect gradient analysis

Robert G. Knox¹ & Robert K. Peet

Department of Biology, University of North Carolina, Chapel Hill, NC 27599-3280, USA; ¹ Present address: Department of Ecology and Evolutionary Biology, Rice University, P.O. Box 1892, Houston, TX 77251, USA

Accepted: 30.1.1989

Keywords: Detrended correspondence analysis, Ordination, Procrustes rotation, Resampling, Rotational instability, Scaled rank variance

Abstract

Indirect gradient analysis, or ordination, is primarily a method of exploratory data analysis. However, to support biological interpretations of resulting axes as vegetation gradients, or later confirmatory analyses and statistical tests, these axes need to be stable or at least robust into minor sampling effects. We develop a computer-intensive bootstrap (resampling) approach to estimate sampling effects on solutions from nonlinear ordination.

We apply this approach to simulated data and to three forest data sets from North Carolina, USA and examine the resulting patterns of local and global instability in detrended correspondence analysis (DCA) solutions. We propose a bootstrap coefficient, scaled rank variance (SRV), to estimate remaining instability in species ranks after rotating axes to a common global orientation. In analysis of simulated data, bootstrap SRV was generally consistent with an equivalent estimate from repeated sampling. In an example using field data SRV, bootstrapped DCA showed good recovery of the order of common species along the first two axes, but poor recovery of later axes. We also suggest some criteria to use with the SRV to decide how many axes to retain and attempt to interpret.

Abbreviations: DCA = detrended correspondence analysis, SRV = scaled rank variance

Introduction

Gradient analysis is widely used to explore and describe relationships between vegetation pattern and environment. Two different modes of analysis termed 'indirect' and 'direct' gradient analysis (Whittaker 1978), or 'vegetational ordination' and 'environmental ordination' (Austin 1968), focus primarily on two different features of vegetation data. The patterns of covariation in species dis-

tributions form the basis of indirect gradient analysis, while direct gradient analysis is based on the relationships of species distributions to measured environmental gradients.

Vegetation data suitable for indirect gradient analysis rarely meet the assumptions required to use parametric statistical tests developed for standard applications of multivariate regression, canonical correlation analysis, or principal components analysis (Legendre & Legendre 1983;

Williams 1983). Although features such as non-normality of measurement errors cause some difficulty, a more important difficulty is that species frequently show non-monotonic or peaked distributions along environmental gradients (see Whittaker 1967). Thus, when standard models that assume linear responses are used, the associated confidence limits are of doubtful utility. Frequently-used non-linear ordination methods, such as detrended correspondence analysis (DCA; see Hill 1979; Hill & Gauch 1980), have no established criteria for retaining stable or 'significant' axes. In many indirect gradient analyses the number of species (descriptors or variables) exceeds the number of measured quadrats (objects or observations). In contrast, criteria for insuring that only significant factors are retained in a factor analysis (e.g. Baggaley 1982, 1986), suggest minimum ratios of 'subjects' to variables ranging from 1.37 to 6.0 for various degrees of structure. However, Baggaley (1982) does express the opinion that 'a researcher who wishes to analyze a large number of variables but can obtain only a modest number of subjects should be able to proceed by dealing with only a few retained factors.'

As a prelude to confirmatory analysis or statistical tests, an investigator needs to decide which axes to retain and to establish that the principal gradients under study are stably represented in the data set. In this article we present a method for measuring sampling effects in one of the most commonly used techniques for vegetation ordination, DCA. In principle, this approach could be applied to almost any ordination method. Measurements of sampling effects on the consistency of gradient analyses of field data sets can also be used to compare ordination methods (e.g. Wilson 1981).

Ecological applications of the statistical technique called the 'bootstrap' (e.g. Weiner 1985), suggested to us that computer-intensive resampling methods (Efron 1982) might provide a means of evaluating sampling error in indirect gradient analysis. Indeed, the bootstrap had already been informatively applied to principal components analysis (Diaconis & Efron 1983),

correspondence analysis (Greenacre 1984), and multidimensional scaling (see Weinberg *et al.* 1984). Oksanen (1988) has also applied the bootstrap to examine instability in DCA, implementing resampling in a different manner (see Methods, Bootstrap resampling).

A resampling method does not automatically provide a useful measure of variation. A good measure should respond accurately and primarily to variation in structural features of interest, not other features of the configuration (e.g. orientation) that may be globally unstable because of methodological artifacts. We used the bootstrap to explore the stability (under sampling variation) of multivariate structural features extracted from field data with DCA. We then developed a coefficient suitable for comparing the consistency of species orderings recovered on different axes and among different data sets. The accuracy of this coefficient was assessed with artificial data simulating repeated samples of sets of vegetation gradients.

The bootstrap

In parametric statistics the distribution of values in a sample is used to calculate the expected error in estimates of parameters of an assumed underlying distribution. For example, when a normal distribution is assumed, the observed standard deviation is used to estimate the expected error of the sample mean. Unfortunately, observed data frequently do not come from a known, well-characterized family of distributions. In such cases the bootstrap, developed by Efron (1979), can sometimes be used to estimate errors in sample statistics. With this method the observations themselves provide an estimate of the shape of their parent distribution.

For most of the interesting applications of the bootstrap a Monte Carlo sampling procedure is used to estimate the expected error of calculated statistic. A large number of random samples, called bootstrap samples, are drawn *with replacement* from the original data or empirical sample. Each observation in the original sample is given

an equal probability of selection. A bootstrap sample has as many data points as the original sample. By chance, some observations are left out while others appear multiple times. The statistic of interest is calculated for each bootstrap sample, and the expected error in the value of sample statistic is calculated from the distribution of the bootstrap replicate values.

The bootstrap has been shown to provide good estimates of error with a variety of known distributions (Efron 1982; Diaconis & Efron 1983; Efron & Tibshirani 1986), and to provide useful information in much more complex problems (Diaconis & Efron 1983; Efron & Tibshirani 1986; but see Hartigan 1986). Bootstrap methods can also be used to produce approximate confidence intervals for sample statistics, but with a less secure theoretical foundation than for bootstrap measures of statistical error (Efron 1982) and requiring many more bootstrap samples (Efron 1987).

The sample statistics in most ordinations include a very large number of species and/or stand scores, which precludes separately measuring the error associated with each score. Nonetheless, a bootstrap error envelope could be produced for the entire configuration. The concentration of information in the first few axes (Gauch 1982a) suggests a useful emphasis on the stability of those axes, as consistent species or stand 'orderings'. Bootstrapping provides a way to exploit the redundancy in a large data set to produce an estimate of how the major axes could have differed if a different data set were collected on the same species in the same area.

Methods

Example data

Two data sets were derived from a study of hardwood forest vegetation of the North Carolina Piedmont (Peet & Christensen 1980). Data were included from 0.1 ha sampling units in 106 predominantly hardwood forest stands with minimal recent disturbance. Tree data were summarized

as importance values, calculated as the average of relative density and relative basal area, based on tallies of all stems greater than 1.25 cm dbh. 56 species of trees were included. Herb-layer data were species importance values calculated from the average of relative cover below 1 m in height and relative frequency, in 25 quadrats within the sampling unit (see Peet & Christensen 1980 for details). Herb-layer data included 328 species of vascular plants. As most of the samples were within parts of the Duke University Forest, these data sets are referred to below as 'Duke Forest' data.

A third data set was collected for a study of the flora and vegetation of the Black Mountains and Craggy Mountains in the Blue Ridge Province of North Carolina (McLeod 1988). The Black Mountains include the highest peak in eastern United States, Mt. Mitchell (2037 m). A roughly 35000 ha study area was surveyed with 0.1 ha plots. More intensive altitude and aspect stratified sampling was done in 6 relatively undisturbed reference areas. The 156 plots were distributed from 760 to 1980 m elevation and across slopes of all aspects. Vascular plants occurring in the plots, less a few with single presences, totaled 327 species. Cover under 1 m in height was estimated and recorded using the Braun-Blanquet scale (Mueller-Dombois & Ellenberg 1974). For analysis these values were converted to a scale from 1 to 7.

Simulated data

To evaluate the accuracy of our bootstrap estimates of sampling effects we used a generalized computer program for simulation of multi-dimensional community patterns (COMPAS; see Minchin 1987a). 3 underlying gradients were specified, with beta diversity highest on the first gradient and decreasing to the third. Species modes and sampling locations were placed randomly. Mixtures of approximately Gaussian symmetric and randomly selected asymmetric species response curves were generated.

3 types of simulated vegetation data sets were

produced. 1 had 250 species in the simulated sampling universe, and a second 25 species, with alpha diversity values comparable to our 0.1 ha herb data and tree data respectively. Distributions of these species were sampled at 4 intensities: 10, 25, 50 and 100 sample plots. At each sampling intensity, results from groups of 20 replicate data sets with independent sampling locations were compared with groups of 20 bootstrap replicate analyses. 5 replicate series were generated for the 25 species sets. For information about the stability of axes with no interpretable gradients, a third series of approximately random data sets was generated by specifying nearly flat species response curves and introducing very high levels of random noise.

Bootstrap resampling

Choosing the appropriate probability model 'is the crucial step for the bootstrap' (Efron & Tibshirani 1986). Bootstrap resampling for ordination could conceivably be done at the level of individual measurements (species abundances), sample units (quadrats, plots or stands), or variables (species), or on some derived portion of the ordination solution (residuals, for example). Resampling at different levels makes different assumptions about the structure of the data. Treating measurements or variables as independent implicitly assumes an individualistic model for species distributions, whereas dependences among the abundances of different species (e.g. site carrying capacities or productivity limits) would be preserved by keeping sample units intact. This distinction might be employed to develop tests of fairly subtle effects of community organization (cf. Noy-Meir & van der Maarel 1987; Wilson 1987; Wilson *et al.* 1987). Oksanen (1988) pooled measurements from related plots. The sample locations are generally selected by an ecologist according to some systematic, randomized or subjective scheme, whereas abundances of taxa under study, in those locations, are then recorded as response variables. Thus, we chose to resample complete sample units or

quadrats. Bootstrap samples then simulate data sets collected by a similar plan, but differing in details of plot placement and relative representation of various vegetation types or habitats.

Implementation with DCA

Detrended correspondence analysis (DCA) is a modification of correspondence analysis designed to remove perceived flaws involving polynomial relationships between axes and compression of the ends of axes (Hill 1974, 1979; Hill & Gauch 1980; Gauch 1982b, cf. Peet *et al.* 1988). For analyzing plant communities, DCA has been found superior to several alternative techniques of indirect gradient analysis (Gauch *et al.* 1981; Gauch 1982b, but see Minchin 1987b), and is currently one of the most widely used ecological methods. The FORTRAN source code for the Cornell implementation of DCA, DECORANA (Hill 1979), is readily available for mainframe and micro-computers, and the computational algorithm used is relatively fast. In our implementation, program output for bootstrap replicates was restricted to sets of 'sample' (hereafter 'stand') and species scores. These sets of scores are relative weights or positions along each of the first 4 ordination axes and are computed from the dominant and early sub-dominant eigenvectors of the solution. (Further analyses of the scores were performed using generalized statistical programs.) For analysis of field data, 100 bootstrap replicate solutions were generated. A version of DECORANA modified to perform bootstrap resampling and reanalysis, titled BOOTDCA, is available on request.

Finding a common frame of reference

In initial analysis of stand scores, we encountered two well-known features of axes from eigenanalysis ordinations of vegetation data (Hill & Gauch 1980). First, the direction of an axis is arbitrary and unstable. Axis scores provide a relative, but not an absolute, ranking. Second, when

2 axes account for nearly the same amount of variation, small changes in the data set can change which axis appears first in the analysis (even if the eigenvalues appear distinct in each analysis). A bootstrap assessment of the stability of the useful information in the results of an ordination thus must be insensitive to minor changes in axis order and reversals in direction.

We adopted 2 approaches to these global instabilities. One was to reorder and/or reverse the individual axes to provide a pair-wise 'best fit' to a reference configuration. A second, capable of correcting for rotations other than 90 degrees, was provided by orthogonal Procrustes rotation (see Fasham 1977; Kenkel & Orlóci 1986; Minchin 1987b for ordination applications) of the major axes to a reference 'target configuration'. Sets of axes that are rotated relative to one another (appearing in a rotated coordinate system, e.g. Gauch *et al.* 1977), as well as reversed in direction or in a different order, can be reconciled using Procrustes rotation. There are several versions, employing dilations and translations as well as rotations (Schöneman & Carroll 1970). We employed rotation and a simple translation, produced by centering the scores before rotation and adding the target means to both sets after rotation. In comparisons of ordination methods, these issues have also been addressed with matrix correlations and subjective assessments (Gauch *et al.* 1981).

For the axis-by-axis approach, the first 4 axes produced by applying DCA to the full data set were designated 'reference axes'. Then for each bootstrap replicate solution the axis order and directions that maximized the agreement (rank collection) with the reference axes were chosen. Due to non-linear detrending in DCA, later axes sometimes have spurious linear correlations with earlier, more important, axes. To prevent matching of axes accounting for large amounts of variation with much less important axes, correlations were weighted by the importance of the axes involved. The order and direction for the 4 bootstrap axes giving the highest total weighted correlation with the 4 reference axes was selected. These new scores are referred to as 'remapped' scores below.

Axis reversal, or 'reflection' (Gauch *et al.* 1981), was apparent in many bootstrap replicate analyses of data from the Duke Forest hardwood plots. In Fig. 1A–C it is apparent that in many replicates the stand configuration was reversed, relative to the stand ordering on 'reference axes' derived from the complete data set. Reversals were observed in as many as 51 of 100 replicates in the herb data from the Duke Forest hardwoods, and as few as none of 20 replicates in a subset of the tree data. Later axes were more likely to have nearly 50% reversals than the first axis. Mapping the bootstrap axes back onto a set of reference axes resulted in a marked improvement in the apparent recovery of the original stand configuration in bootstrap replicates of tree data from Duke Forest (Fig. 1D–F).

Stability

To assess the stability of the results of a DCA ordination, we examined the variability of (axis-by-axis) remapped bootstrap stand scores and the level of rank correlations of stand scores along bootstrap axes with scores on reference axes. To compare variabilities in species rankings, we computed coefficients from the average of variances in rank seen in individual species. For a data set, the species occurring in all the bootstrap samples were selected (thereby omitting most species that occur only infrequently). Procrustes rotated scores for these common species were ranked along each axis. This variance in rank, across bootstrap replicates, was computed for each species. The variances were averaged across the n species for each axis. Then to form a coefficient for comparison across axes and data sets, this mean variance was scaled by the expected variance of a set of n random-ranks:

$$\text{Var} = (n^2 - 1)/12 \quad (1)$$

'Scaled rank variance' (SRV) values near zero indicate very consistent species rankings, whereas values near 1.0 indicate species ranks as variable as random ranks.

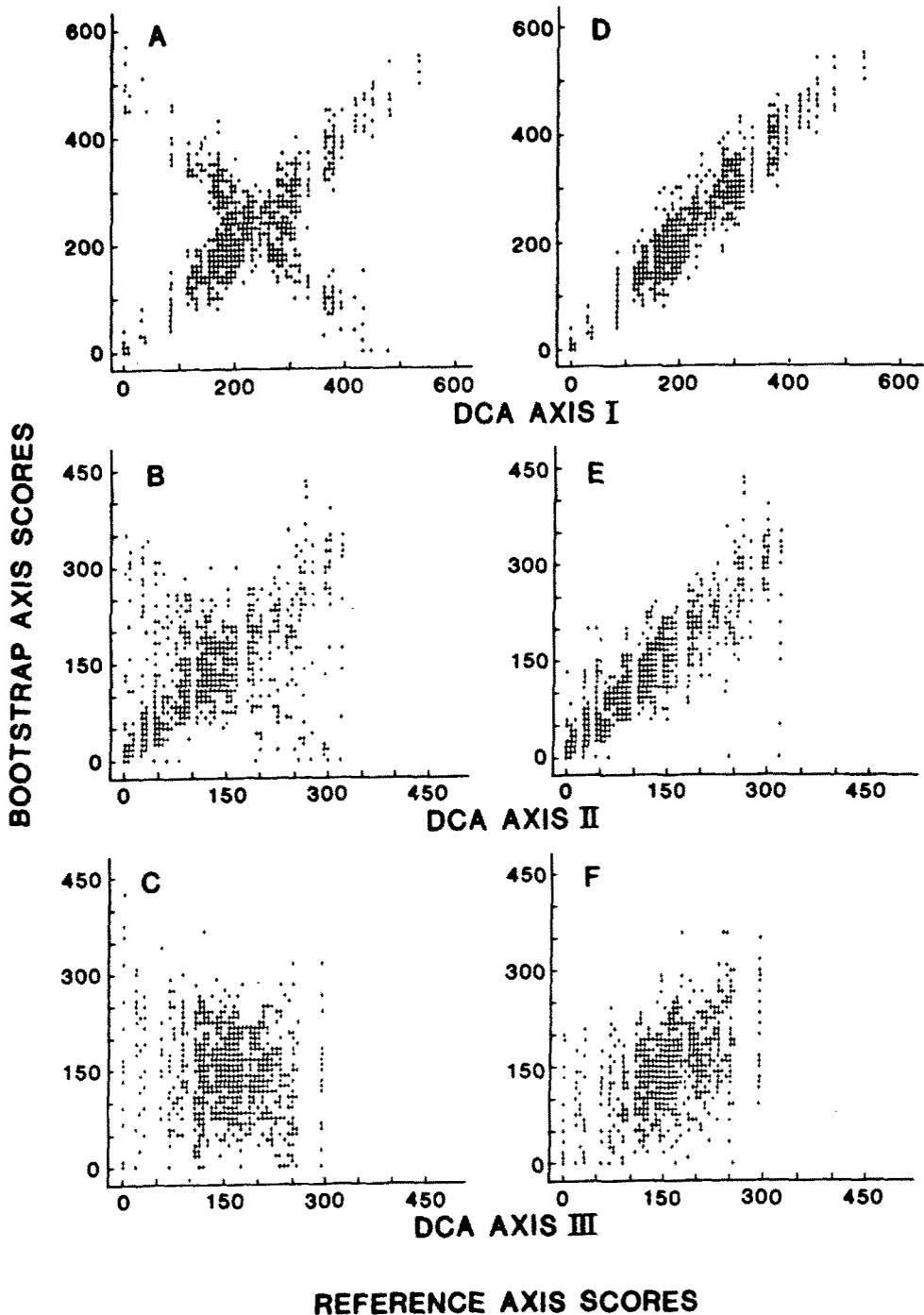


Fig. 1A-F. Stand scores from bootstrapped ordination plotted against conventional DCA scores, before and after remapping them to a common frame or reference. Data are for tree abundances in the Duke Forest hardwoods stands. For visual clarity, only scores from a 10% random sample of the 10 600 scores, for 100 bootstrap replicates each with 106 stands, are shown: (A-C) Axes I, II and III, before remapping; (D-F) Axes I,II, and III, after remapping.

Results and discussion

Variation in stand scores

To illustrate the variation in stand scores under bootstrapping, we chose from the Duke Forest hardwoods herb data set representative pairs of stands bracketing the central 50% interval in each of the first 3 axes of the DCA reference analysis. For these stands, distributions of remapped bootstrap stand scores overlap on each of the first 3 axes (Fig. 2). As expected, the greatest overlap is on the third axis, and third axis scores are more variable. Comparable results were obtained for other similar pairs of stands (results not shown).

Because the zero point of a DCA axis is set by the stand with the lowest score (Hill 1979), DCA axes change when different stands define the axis origin in different bootstrap replicates. The scores on an axis shift up or down when an extreme stand is included or excluded. Thus overlaps in the distribution of scores across bootstrap replicates do not necessarily indicate changes in relative order within replicates. Relative ordering of scores is less subject to the influence of extreme stands. Reanalysis of relative scores for stands at the 25th and 75th percentiles on the first reference axis, showed that, despite the overlap in scores seen in Fig. 2a, when both stands were present in the same bootstrap sample, inter-changes in their order on the first axes were quite rare. Inter-changes on the first axis occurred in none of the 38 bootstrap replicate analyses that included both stands. No inter-changes in the order of the corresponding stands for the second axis were found either. On the third axis, the 25th and 75th percentile stands exchanged order in 11 of the 41 replicates that included both stands. We focused further analysis of the overall stability of ordination axes on the consistency of relative ordering along the principal ordination axes, rather than the scores themselves.

For the 3 data sets described above, we used Spearman rank correlations of scores on remapped bootstrap axes with scores on DCA reference axes to measure the stability of the original DCA axes. Table 1 reports medians and

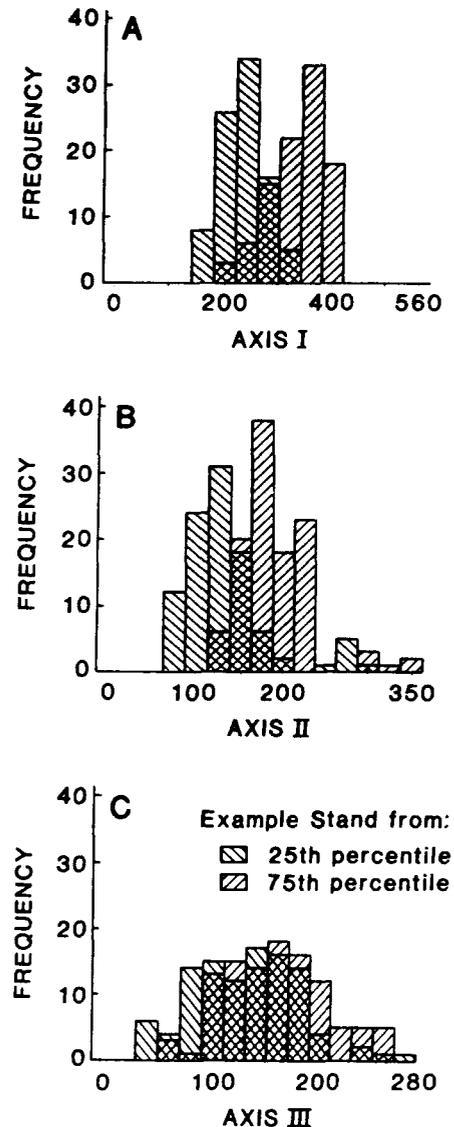


Fig. 2A–C. Frequency distributions of stand scores under bootstrapping, for 6 example stands from the Duke Forest hardwoods herb data set. The stands shown fall at the 25th or 75th percentiles in the distribution of values on one of the first 3 axes from conventional DCA of this data set. Scores are from 100 bootstrap replicates.

50% central intervals for rank correlations between reference axes and corresponding bootstrap axes, based on 100 bootstrap replicates for each of 3 data sets. Results differed by axis number and between the data sets. The first DCA axis appeared quite consistent in replicates from the 2 Duke Forest hardwood data sets. In more

Table 1. Stability of DCA axes with axis-by-axis remapping, as assessed by rank correlations between ordering of stands on remapped bootstrap axes and reference axes, from 100 bootstrap replicates on each of 3 data sets.

Data set	Rank correlation with DCA reference axes		
	Axis I Median* (75%, 25%)+	Axis II Median (75%, 25%)	Axis III Median (75%, 25%)
Duke Forest			
Hardwood	0.976	0.862	0.336
Herb-layer data 106 stands, 328 spp.	(0.99, 0.95)	(0.93, 0.76)	(0.50, 0.22)
Duke Forest			
Hardwood	0.982	0.923	0.569
Tree data 106 stands, 56 spp.	(0.99, 0.97)	(0.95, 0.85)	(0.67, 0.46)
Black Mountains			
Herb-layer data 156 stands, 327 spp.	0.959 (0.99, 0.84)	0.937 (0.98, 0.79)	0.628 (0.72, 0.45)

* Median rank correlation out of 100 bootstrap replicates.
+ 75th and 25th percentile points from the distribution of rank correlations in 100 bootstrap replicate analyses.

than 75% of the replicates, the first axis ordering was recovered well-enough to produce rank correlations of 0.95 or above. The second axes were not recovered as consistently, but still generally produced high correlations. The third axes were recovered poorly, especially considering that a remapping procedure generates small, consistently positive correlations from sets of random scores. The length of the inter-quartile interval for these correlations also increased as typical correlations decreased, from axis I to axis III. Recovery of the third axes was both lower on average and more erratic. The Black Mountains data showed little difference between the stability of the first and second axes (note the overlap of intervals in Table 1). This was consistent with evidence from correlations of environmental factors with the reference DCA axes (McLeod 1988) which suggested that at least 2 strong environmental gradients structure this data set.

Rank correlations of stand scores on Procrustes rotated axes, with reference axes used

as target axes, were higher than corresponding rank correlations for remapped axes (see Table 2). Recovery of major axes was better than with axis-by-axis remapping, suggesting that rotations other than 180 or 90 degrees were sometimes important.

In bootstrapping to assess sampling effects on the ordering of stands, we assumed that important uncertainties in estimated gradient position for a stand can be attributed to the context of other stands in the sample. Because the composition of a stand is measured only once and as a sampling unit, independent samples could share species but not stands. The results presented thus far deal with a data-analytic stability rather than some statistical stability. (See de Leeuw & Meulman 1986, for an application of the jackknife to multi-dimensional scaling carried out in this spirit.) In contrast, species are descriptors or variables instead of data objects and a properly collected set of vegetation quadrats could be seen as a statistical sample of the joint distributions of the component species. Leaving aside spatial autocorrelation issues (see Palmer 1988) and the question of whether vegetation data sets are typi-

Table 2. Stability of DCA axes with orthogonal Procrustes rotation, as assessed by rank correlations between ordering of stands on rotated bootstrap axes and reference axes, from 100 bootstrap replicates on each of 3 data sets.

Data set	Rank correlation with DCA reference axes		
	Axis I Median* (75%, 25%)+	Axis II Median (75%, 25%)	Axis III Median (75%, 25%)
Duke Forest			
Hardwood	0.984	0.918	0.466
Herb-layer data	(0.989, 0.976)	(0.95, 0.88)	(0.62, 0.35)
Duke Forest			
Hardwood	0.987	0.956	0.762
Tree data	(0.991, 0.981)	(0.97, 0.93)	(0.84, 0.65)
Black Mountains			
	0.995 (0.998, 0.987)	0.981 (0.99, 0.96)	0.718 (0.86, 0.58)

* Median rank correlation out of 100 bootstrap replicates.
+ 75th and 25th percentile points from the distribution of rank correlations in 100 bootstrap replicate analyses.

cally good samples of an identifiable sampling universe, bootstrapped species scores can provide distributions for estimates of species gradient optima. In direct gradient analysis these distributions are like bootstrapped regression coefficients. In indirect gradient analysis, with an abundance of estimated parameters, we are again reduced to relative scores and their ranks.

Simulated gradients

Without appropriate significance levels, internal correlations are awkward for comparing data sets and provide little guidance as to what is or is not a stably estimated axis. To better fit those requirements we derived the scaled rank variance from the bootstrap variance in species ranks. Before presenting results from SRV calculations for field data, we needed to calculate SRV from unstructured data as well as test the accuracy of the bootstrap estimate in a situation where an estimate from true replicate samples could be produced for comparison with it.

Before reporting results for SRV, we note that some users of DCA might be tempted to rely on the reported eigenvalues in deciding which axes to retain. Work with random data sets (cf. Gauch 1982a) showed that this could be inappropriate. Even when there was no interpretable simulated structure very early eigenvalues from DCA increased with decreasing sample size (Fig. 3),

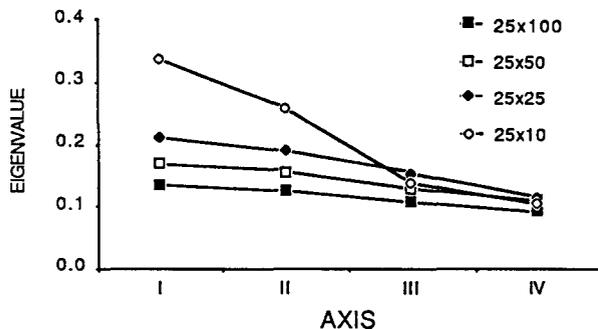


Fig. 3. Eigenvalues from DCA of 25 species simulated data sets with minimal simulated structure, sampled at 4 intensities (10, 25, 50 and 100 stands).

yielding seemingly impressive eigenvalues in the smallest samples. This sample-size effect also biases eigenvalues and axis length estimates from bootstrap replicates. Values for first and second axes were nearly always higher on average in a set of bootstrap replicates than in analysis of a complete data set.

In contrast, SRV among bootstrap axes was consistent across a 10-fold range in the size of samples from unstructured data (Fig. 4). These results also support application of a 50% (0.5) criterion for retention of third and fourth axes, i.e. interpretable rankings ought to have less than half the mean variability of random ranks. (This was suggested to us by the 'rule of 2' for ratios of mean squares, advocated by Tukey (1988 mscr.) for 'modern ANOVA'). It appears that more strict criteria should be applied to first and second axes, perhaps 0.3 and 0.4 respectively.

We found good agreement between bootstrap and repeated sampling SRV estimates of the variability of species orderings on the first 3 axes of a 3 gradient simulation with 250 species (Fig. 5A–C). Both estimates suggested recovery of 3 fairly stable axes in samples of 25 stands or more, and 2 axes with only 10 samples. The bootstrap estimate was low for the fourth axes (Fig. 5D), but a user following a 0.5 rule would have made the same decisions employing either estimate.

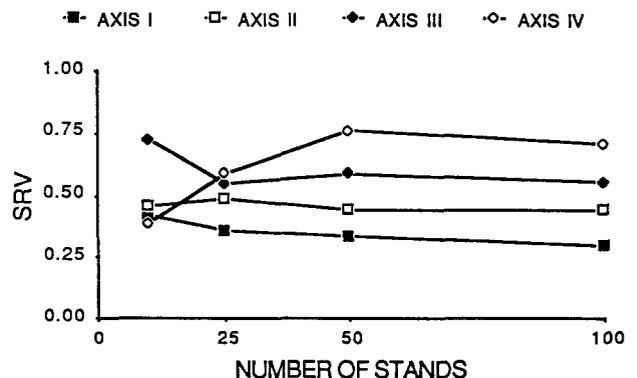


Fig. 4. Scaled rank variances (SRV) from bootstrapped DCA of 25 species simulated data sets with minimal simulated structure, sampled at 4 intensities (10, 25, 50 and 100 stands). Each estimate is calculated from 20 bootstrap replicates rotated to best match a single reference analysis.

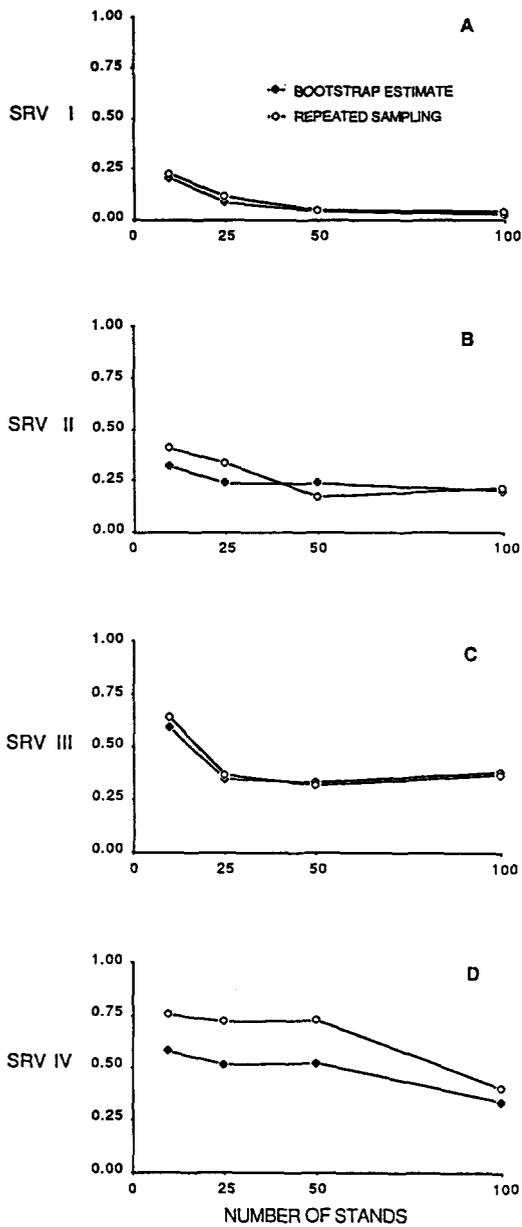


Fig. 5A-D. Bootstrap and repeated sampling estimates of SRV for 250 species simulated vegetation data sets with 3 simulated gradients. Estimates associated with reference analysis: (A) Axis I; (B) Axis II; (C) Axis III; (D) Axis IV. The species configuration was sampled 21 times at each of 4 intensities and 20 bootstrap replicates were generated from the first sample in each series. Each estimate is calculated from 20 replicates rotated to best match the same reference analysis.

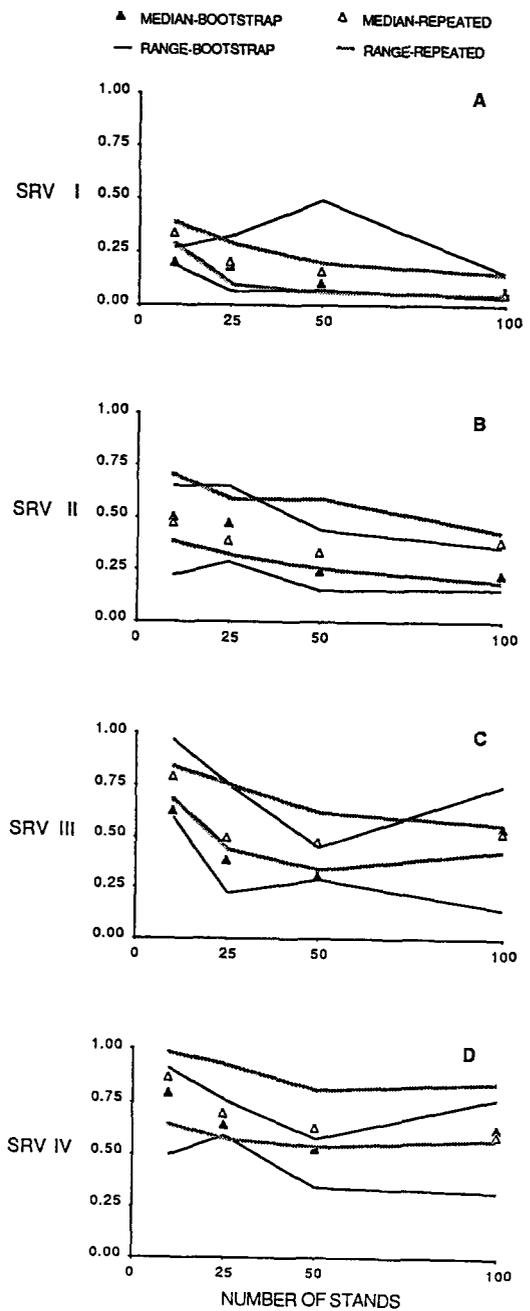


Fig. 6A-D. Median and range for bootstrap and repeated sampling estimates of SRV from 5 configurations of 25 species, repeatedly sampled at 4 intensities (as above). Estimates from reference axes associated with: (A) Gradient 1; (B) Gradient 2; (C) Gradient 3; (D) No gradient.

Producing multiple (5) groups of stimulated data sets was feasible with 25-species gradients, providing replicated assessments of the bootstrap SRV. Agreement with SRV from repeated sampling was good for axes associated with the first gradient (Fig. 6A) and moderate for the other axes (Fig. 6B–D). Generally at most one axis would be retained from analyses of only 10 stands. With only 25 species, recovery of a stable third axis was erratic. Spurious fourth axes would be discarded under a 0.5 rule in all but 2 cases. Again, there was a hint of downward bias in the bootstrap estimate. It was lower for corresponding axes in 56 of 80 comparisons, but without many more independent simulations the difference could not be distinguished from chance variation. With additional simulation work a bias correction might be developed, but the bias from bootstrapping appeared smaller than the bias from needed corrections for rotational instabilities. In any case, our suggested criteria for axis retention refer to bootstrap SRV estimates.

SRV for field data

These criteria were then applied to bootstrap SRV estimates for DCA axes from our example field

Table 3. Variability in species ordering from DCA axes, after orthogonal Procrustes rotation, on each of 3 field data sets. Coefficients (SRV) are variances in ranks, averaged across species frequent enough to be present in all 100 bootstrap samples, and divided by the expected variance of a set of n random ranks.

Data set	Axis I	Axis II	Axis III	Axis IV
Duke Forest Hardwood Herb-layer data 192 spp.	0.234	0.341	0.862	0.776
Duke Forest Hardwood Tree data 46 spp.	0.123	0.302	0.610	0.737
Black Mountains Herb-layer data 237 spp.	0.090	0.153	0.627	0.551

data sets. Despite fairly high correlations with reference axes, the third and fourth axes in these data were not particularly robust to sampling variation (see Table 3). DCA of each field data set produced 2 axes with low SRV. The Black Mountains data produced more stable estimates of species order on the first two gradients than the Duke Forest data. The occasional instability in DCA identified by Oksanen (1988) apparently did not occur frequently enough to markedly increase variation among recovered solutions in the Black Mountains data, despite 2 primary gradients of nearly equal length and frequent reversals in their order of appearance.

Ordination evaluation

An important potential application for bootstrapped ordination is in evaluation and comparison of ordination methods (Knox 1987). Ordination methods should reliably and conservatively reveal vegetation structure without major distortion (Gauch *et al.* 1981). The ability to use a bootstrap to estimate sampling effects for stand and species scores and estimate recovery of axis ordering from either real or simulated data, would complement the usual evaluations based on interpretations of field data and recovery of simulated structure in simulated data. Unlike random partitions of the data set (e.g., Wilson 1981), bootstrapping measures performance in samples of a size comparable to the original data set and a large number of replicate analyses contribute to the performance estimate. Unlike simulated data, no assumptions about the shape of plant distributions on gradients (see Gauch & Whittaker 1972, 1976; Austin 1985, 1987; Minchin 1987a, b) are required.

The results presented above suggest that in samples of 100 to 150 standard 0.1 ha forest plots the third and fourth DCA axes may be unstable and include spurious information. Other ordination methods might stably recover more dimensions from this type of data set.

Conclusions

The bootstrap provides a means of characterizing the major kinds of sampling variability affecting indirect gradient analysis. In DCA ordination, overall directional instabilities in the axes are prominent. The usual DCA procedure for axis scaling causes another level of spurious instability because axis origins are set by stands with extreme values. These effects have to be removed before variation in axis scores can be measured in a meaningful way.

Measures of ranks agreement and scaled rank variance (SRV) based on species rank show substantial differences in stability among the first 4 DCA axes. Third and fourth axes from field data were highly unstable. Even if the gradients represented on these axes are in some senses 'real', these data sets do not provide robust estimates of their effects. In contrast, relative orderings of stands and species on first axes were quite reliably estimated in all 3 field data sets.

Bootstrap SRV coefficients could be used to develop more precise guidelines for axis interpretation, so as to help avoid fruitless attempts to explain noise variation. Until additional results are available and we have more experience with SRV in ordination, we suggest that interpreted axes have less than half (0.5) the variance of random ranks and, more stringently, first axes have less than a third the variance of random ranks. Researchers drawing contrasts between gradients in different environments or at different scales could also use bootstrapped ordination to focus attention on robust properties of the sample data and away from potential sampling artifacts.

Acknowledgements

We thank R.J. Carroll, J.S. Case, R.B. Allen, I.C. Prentice, N.L. Christensen, F. Young, and J. White for helpful discussions, and D. McLeod for permission to use unpublished data. M.B. Dale, I.C. Prentice and C.J.F. ter Braak provided valuable comments on an earlier draft. This work was supported by grants from the U.S. National

Science Foundation to R.K.P. and a National Science Foundation Predoctoral Fellowship to R.G.K.

References

- Austin, M.P. 1968. An ordination study of a chalk grassland community. *J. Ecol.* 56: 739–757.
- Austin, M.P. 1985. Continuum concept, ordination methods, and niche theory. *Ann. Rev. Ecol. Syst.* 16: 39–61.
- Austin, M.P. 1987. Models for the analysis of species' response to environmental gradients. *Vegetatio* 69: 35–45.
- Baggaley, A.R. 1982. Deciding on the ratio of number of subjects to number of variables in factor analysis. *Multivar. Exp. Clin. Res.* 6: 81–85.
- Baggaley, A.R. 1986. Heterogeneous correlations and estimates of required sample size in factor analysis. *Multivar. Exp. Clin. Res.* 8: 161–163.
- de Leeuw, J. & Meulman, J. 1986. A special jackknife for multidimensional scaling. *J. Classification* 3: 97–112.
- Diaconis, P. & Efron, B. 1983. Computer-intensive methods in statistics. *Sci. Am.* 248(5): 116–130.
- Efron, B. 1979. Bootstrap methods: another look at the jackknife. *Ann. Statist.* 7: 1–26.
- Efron, B. 1982. The jackknife, the bootstrap, and other resampling plans. *SIAM Monograph* 38.
- Efron, B. 1987. Better bootstrap confidence intervals. *J. Amer. Statist. Assoc.* 82: 171–185.
- Efron, B. & Tibshirani, R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.* 1: 54–77.
- Fasham, M.R.J. 1977. A comparison of nonmetric multidimensional scaling, principal components and reciprocal averaging for the ordination of simulated coenoclines, and coenoplanes. *Ecology* 58: 551–561.
- Gauch Jr, H.G., 1982a. Noise reduction by eigenvector ordinations. *Ecology* 63: 1643–1649.
- Gauch Jr, H.G., 1982b. *Multivariate analysis in community ecology*. Cambridge University Press, Cambridge.
- Gauch Jr, H.G. & Whittaker, R.H. 1972. Coenocline simulation. *Ecology* 53: 446–451.
- Gauch Jr, H.G. & Whittaker, R.H. 1976. Simulation of community patterns. *Vegetatio* 33: 13–16.
- Gauch Jr, H.G. & Whittaker, R.H. & Singer, S.B. 1981. A comparative study of nonmetric ordinations. *J. Ecol.* 69: 135–152.
- Gauch Jr, H.G. & Whittaker, R.H. & Wentworth, T.R. 1977. A comparative study of reciprocal averaging and other ordination techniques. *J. Ecol.* 65: 157–174.
- Greenacre, M.J. 1984. *Theory and applications of correspondence analysis*. Academic Press, London.
- Hartigan, J.A. 1986. Comment (on: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy.) *Statist. Sci.* 1: 75–77.

- Hill, M.O. 1974. Correspondence analysis: a neglected multivariate method. *Applied Statistics* 23: 340–354.
- Hill, M.O. 1979. DECORANA—a FORTRAN program for detrended correspondence analysis and reciprocal averaging. Cornell University, Ithaca, New York.
- Hill, M.O. & Gauch Jr, H.G., 1980. Detrended correspondence analysis: an improved ordination technique. *Vegetatio* 42: 47–58.
- Kenkel, N.C. & Orlóci, L. 1986. Applying metric and non-metric multidimensional scaling to ecological studies: some new results. *Ecology* 67: 919–928.
- Knox, R.G. 1987. Hypothesis testing in plant community ecology: analysis of long-term experiments and regional vegetation data. Ph. D. Diss. University of North Carolina, Chapel Hill, North Carolina.
- Legendre, L. & Legendre, P. 1983. *Numerical ecology*. Elsevier, Amsterdam.
- McLeod, D.E. 1988. Vegetation patterns, floristics, and environmental relationships in the Black and Craggy Mountains of North Carolina. Ph. D. Diss., University of North Carolina, Chapel Hill, North Carolina.
- Minchin, P. 1987a. Simulation of multidimensional community patterns: towards a comprehensive model. *Vegetatio* 71: 145–156.
- Minchin, P. 1987b. An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* 69: 89–107.
- Mueller-Dombois, D. & Ellenberg, H. 1974. *Aims and methods of vegetation ecology*. John Wiley and Sons, New York.
- Noy-Meir, I. & van der Maarel, E. 1987. Relations between community theory and community analysis in vegetation science: some historical perspectives. *Vegetatio* 69: 5–15.
- Oksanen, J. 1988. A note on the occasional instability of detrending in correspondence analysis. *Vegetatio* 74: 29–32.
- Palmer, M.W. 1988. Fractal geometry: a tool for describing spatial patterns of plant communities. *Vegetatio* 75: 91–102.
- Peet, R.K. & Christensen, N.L. 1980. Hardwood forests of the North Carolina piedmont. *Veröff. Geobot. Inst. Rübel* 69: 14–39.
- Peet, R.K., Knox, R.G., Case, J.S. & Allen, R.B. 1988. Putting things in order: the advantages of detrended correspondence analysis. *Amer. Nat.* 131: 924–934.
- Schöneman, P.H. & Carroll, R.M. 1970. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika* 35: 245–255.
- Weinberg, S.L., Carroll, J.D. & Cohen, H.S. 1984. Confidence regions for INDSCAL using the jackknife and bootstrap techniques. *Psychometrika* 49: 475–491.
- Weiner, J. 1985. Size hierarchies in experimental populations of annual plants. *Ecology* 66: 743–752.
- Whittaker, R.H. 1967. Gradient analysis of vegetation. *Biol. Rev.* 42: 207–264.
- Whittaker, R.H. (ed.) 1978. *Ordination of plant communities*. *Handbook of Vegetation Science* 5. 2nd ed. Junk, The Hague.
- Williams, B.K. 1983. Some observations on the use of discriminant analysis in ecology. *Ecology* 64: 1283–1291.
- Wilson, J.B. 1987. Methods for detecting non-randomness in species co-occurrences: a contribution. *Oecologia* 73: 579–582.
- Wilson, J.B., Gitay, H., Agnew, A.D.Q. 1987. Does niche limitation exist? *Funct. Ecology* 1: 391–397.
- Wilson, M.V. 1981. A statistical test of the accuracy and consistency of ordinations. *Ecology* 62: 8–12.