Journal of Vegetation Science 10: 895-902, 1999

The bootstrapped ordination re-examined

Pillar, Valério DePatta

Departamento de Ecologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 91540-000, Brazil; Fax +55 51 3191568; E-mail: vpillar@ecologia.ufrgs.br

Abstract. A method is described to determine the number of significant dimensions in metric ordination of a sample. The method is probabilistic, based on bootstrap resampling. An iterative algorithm takes bootstrap samples with replacement from the sample. It finds in each bootstrap sample ordination coordinates and computes, after Procrustean adjustments, the correlation between observed and bootstrap ordination scores. It compares this correlation to the same parameter generated in a parallel bootstrapped ordination of randomly permuted data, which upon many iterations will generate a probability. The method is assessed in principal coordinates analysis of simulated data sets that have varying number of variables and correlation levels, uniform or patterned correlation structure. The results suggest the method is more reliable than other available methods in recovering the true intrinsic dimensionality. Examples with grassland data illustrate utility.

Keywords: Dimensionality; Eigenvalue; Metric ordination; Multidimensional scaling; Multivariate; Principal coordinates analysis; Procrustes; Resampling; Sample size; Stopping rule.

Introduction

Ordination is often used by ecologists as a tool in multivariate data exploration. The objective is to reveal a synthetic, more interpretable view of patterns that otherwise would be too complex to observe. In vegetation ecology, this is particularly important for the detection of community patterns (e.g., see reviews by Orlóci 1978; Pielou 1984; Kenkel & Orlóci 1986; Digby & Kempton 1987; Podani 1994). A problem in ordination is finding significant, nontrivial ordination dimensions to be retained for further interpretation. Although ordination is an exploratory tool, we do need an objective, preferably probability-based approach to test how significant in stability terms is the observed pattern in a given ordination. On the one hand, this is important considering that interpretations may be misleading when based on dimensions depicting unstable trends of variation that would not reappear in the analysis of other samples from the same sampling universe. On the other hand, there may be loss of meaningful information by wrongly discarding significant dimensions.

Jackson (1993) has examined the problem. He evaluated with simulated data the reliability of several heuristical and statistical stopping rules (see also Mehlman et al. 1995; Jackson 1995). He concluded that the broken-stick (Frontier 1976), bootstrapped eigenvalues and bootstrapped eigenvectors were among the most reliable methods. Yet, they were not completely reliable in all the cases that were studied. Among other methods, he evaluated the Kaiser-Guttman criterion of retaining in principal components analysis eigenvalues exceeding the average eigenvalue, and Bartlett's statistical tests based on the χ^2 distribution (see Jackson 1993 and references therein). Bootstrap resampling has earlier been used by others to examine sampling adequacy in ordinations (Stauffer et al. 1985; Knox & Peet 1989). The bootstrap resampling method (Efron 1979; Efron & Tibshirani 1993) relies on the principle that, not having better information, the frequency distribution of variables in a sample is the best indicator of the distribution in the sampling universe. If we accept this as true, the sample may be regarded as a pseudo sampling universe, and resampling the sample with replacement (getting a bootstrap sample) will mimic resampling the sampling universe.

The quantity applied to measure the state of the ordination pattern is critical. Eigenvalues (Jackson 1993) and the ratios between eigenvalues and total inertia, the total sum of the eigenvalues (Stauffer et al. 1985; Efron & Tibshirani 1993: 66) are typical examples. Other instances include the eigenvectors (Jackson 1993; Efron & Tibshirani 1993: 69). It is, however, unclear how eigenvalues as general variances could be taken as direct measures of the pattern in an ordination (Orlóci 1993; Økland 1999). Knox & Peet (1989) evaluated sample stability in detrended correspondence analysis by correlating ranked ordination scores (see also Knox 1989), which seems to be a better measure.

I further examine the problem of ordination stability and offer a new solution that uses bootstrap resampling. The method generates probabilities based on which the significance of ordination dimensions is evaluated. The solution is flexible, applicable to metric ordination techniques. I test the method on principal coordinates analy-

sis of simulated and real data sets. Some of the simulated data sets have similar properties to those evaluated by Jackson (1993).

Methods

Bootstrapped ordination

A sample of n sampling units and p variables is considered. The sample in vegetation ecology is usually a set of relevés, the sampling units, but the method is applicable to other types of samples as well. The variables may be of any kind provided suited for ordination by the chosen method.

The algorithm is illustrated by a numerical example in App. 1. It starts with the application of the chosen ordination method to the sample (pseudo sampling universe). Only metric ordination methods are directly applicable. The ordination scores of the sampling units are saved as reference scores. Many iterations of resampled sample sets then follow. At each iteration, a bootstrap sample with size n is subjected to ordination by the same method. The matrix X^* holds the ordination scores of the bootstrap sample for a given number of dimensions. Note that since sampling is with replacement, sampling units may be repeated in the bootstrap sample. The matrix X holds the scores of the sampling units that are in the bootstrap sample, but extracted from the reference scores.

A Procrustean adjustment is needed to bring the scores of the two ordinations comparable. The adjustment involves a rotation, a translation and a dilation of the ordination subspace in the bootstrap sample so to maximize the fit with the ordination of the pseudo sampling universe (see details in Schönemann & Carroll 1970). Note that this operation is neutral in the sense that it would not change conclusions on trends in the point configuration of each ordination. The need for the Procrustean adjustment arises from the way in which ordination algorithms position axes in different samples. First, mirrored configurations are trivial and arbitrary. Second, variation between bootstrap samples in the order of expression of the same intrinsic ordination dimensions is possible because these may be of very similar importance (see Jackson 1993). The similarity of raw ordination scores under such conditions are likely to be underestimated.

After a Procrustean adjustment involving the first i ordination dimensions, the scores on ordination dimension i in X^* and X are compared by the correlation:

$$\theta_i^* = r(x_i^*, x_i) \quad (1)$$

In this, r is the Pearson product moment correlation, but other measures could be used, such as rank correlation (Kendall & Gibbons 1990). The higher the correlation, the better is the agreement between bootstrap and reference ordination scores and the more stable is the ordination configuration in the sample.

At each iteration the correlation θ_i^* is compared to a correlation θ_i^o generated under the null hypothesis of unstructured data. We could take zero correlation as the expectation θ_i^o, which would allow conventional interpretation of confidence intervals, but it is easily verified that bootstrapping a random data set will not give an expected correlation $r(x_i^*, x_i)$ of zero. Therefore, θ_i^o is computed as in Eq. 1, but with score matrices X^* and X coming from a parallel bootstrap resampling and ordination performed on the observed data set with the observations randomly permuted within variables. If $\theta_i^o \geq \theta_i^*$, the algorithm will add one to the cumulative frequency $F(\theta_i^o \geq \theta_i^*)$. After B bootstrap iterations, the probability $P(\theta_i^o \geq \theta_i^*)$ is the proportion $F(\theta_i^o \geq \theta_i^*)/B$. The number of iterations is only limited by available computation time. The larger is B, the more stable results will be obtained.

Interpretation

The probability $P(\theta_i^o \geq \theta_i^*)$ is an indicator of the strength of the structure in an ordination, as compared to a null data set containing variables with the same observed distribution but zero expected association. We may set an α probability threshold to help the interpretation of $P(\theta_i^o \geq \theta_i^*)$. A small $P(\theta_i^o \geq \theta_i^*)$, that is, smaller than α, will indicate that the ordination dimension in consideration is significantly more stable than that would be expected for the same dimension in the ordination of a random data set. In this case we reject the null hypothesis and conclude, with a probability $P(\theta_i^o \geq \theta_i^*)$ of being wrong, that the given ordination dimension is nontrivial and worthy of interpretation. Otherwise, we accept the null hypothesis and consider the ordination dimension unstable and indistinguishable from a random data ordination. Alternatively, we may adopt a more flexible view on the significance, and consider the resulting $P(\theta_i^o \geq \theta_i^*)$ values as representative of a gradient of reliability.

An insufficient sample size may cause type II error; that is, the test may not detect significance of dimensions that would be found significant if the sample were larger. Sample size sufficiency is indicated by stability of probabilities under increase of sample sizes. A method that uses bootstrapped ordination similarly as in this paper, but that also assesses the significance of ordination axes in bootstrap samples with sizes smaller than n sampling units, is described by Pillar (1998). If sample

size is deemed sufficient and $P(\theta_i^o \geq \theta_i^*)$ is larger than α, the ordination axis is truly irrelevant. Otherwise, we may say that a larger sample size would be needed to reach a more confident conclusion.

The number of dimensions involved in Procrustean adjustments is likely to affect the fitted scores and their correlation with the reference scores. Since the number of dimensions must be set in advance, when we do not yet know which dimensions are statistically significant, an apparent problem of circularity arises. However, in metric ordination the dimensions are uncorrelated and meaningfully ordered by relevance. If the second dimension is significant, it is expected that the first will also be significant. Likewise, if the third dimension is significant, the second and first ones will be as well, and so on. We should therefore start testing the significance from the highest (the least relevant) to the first ordination dimension. This is why the computation of θ_i^* (and θ_i^o) uses Procrustean adjusted scores involving the first i dimensions. Once dimension i is deemed significant, the test may stop and all lower and more relevant $i - 1$ dimensions are also considered significant irrespective of the corresponding probabilities. On the one hand, by following this sequence, we avoid considering probabilities generated after Procrustean adjustments that left out nontrivial dimensions; these probabilities are likely to be overestimated. On the other hand, by sticking to no more than i dimensions the Procrustean fitting is not weakened by least relevant dimensions that may not be significant.

Testing the method

The method was tested using simulated data sets with known intrinsic dimensionality. Similarly to Jackson (1993), dimensionality is given by different levels of correlation between variables, numbers of variables and correlation structures. The procedure applied to generate simulated data sets from correlation matrices is described in Ganeshanandam & Krzanowski (1990), modified by Peres-Neto & Jackson (1999 mscr.). The procedure is based on the Cholesky decomposition of the correlation matrix; the result is then premultiplied by an n by p matrix with PCA scores of a random data set, which yields the $p \times n$ data matrix sought. All simulated data sets contain 40 sampling units. Matrix types named as 4R and 32R have four and 32 variables with uniform correlation varying from 0 to 0.8. Matrices of type 12S1, 12S2 and 12S3 contain 12 variables, with the correlation structured in three groups of variables with different correlation levels within and between groups. In matrices of type 12S1 the groups contain four variables each, the correlation level is 0 in between groups comparisons and varies, among the simulated data sets, from 0.1 to 0.8 in within groups comparisons. Type 12S2 is similar to 12S1 but the correlation level in between groups comparisons is 0.3. In matrices of type 12S3, the three groups contain 5, 4 and 3 variables and the correlation levels are similar to 12S1. RND is a 40 by 12 matrix with random numbers ranging from 0 to 1. To further evaluate the effect of number of variables, the tests used simulated matrices generated with 4 to 32 variables and a uniform correlation level of 0.2.

All simulated matrices defined by Jackson (1993) have a counterpart in the foregoing defined matrices. Yet Jackson (1993) defined the data matrices with 1000 sampling units and in the analysis used three random samples with 40 sampling units each, which due to sampling errors may not have the intended correlation structures. Perhaps, it would have been more appropriate to increase the number of sampling units and define the reference distribution, but I did not use this approach here. Instead, by doing the tests with the original matrices with known correlation structure, I avoid the effect of sample size when evaluating the reliability of the method.

Tests were also made on two field data sets from grassland vegetation in the south of Brazil. The data set VEG contains cover-abundance estimates (1-9 scale) of 60 species in 60 0.5×0.5 m quadrats. The data set ENV describes the same quadrats by 20 environmental variables. These include availability of soil nutrients, soil texture, soil moisture, elevation and grazing intensity. Pillar et al. (1992) describe the survey procedure and the variables in more detail.

I implemented the method in a C++ application program (SAMPLER), available upon request, which was used for all analyses. The ordination method in the tests was principal coordinates analysis. Probabilities $P(\theta_i^o \geq \theta_i^*)$ were generated by 1000 bootstrap iterations.

Results

The tests depicted in Table 1(a-f) used an α threshold of 0.1. Simulated matrices with low correlation levels between variables were expectedly identified as not having any interpretable dimension. The test correctly detected the intended dimensionality if intervariable uniform correlation was at least 0.3 in matrices 4R and 0.1 in matrices 32R (Table 1a-b). There was an effect of number of variables (Fig. 1); for a given correlation level (0.2), the larger the number of variables, the lower was the probability P for axis 1. In simulated matrices with structured correlation, the intrinsic dimensionality (3) was correctly identified if intervariable correlation within groups was at least 0.4 in matrices 12S1 and 12S3 (Table 1c, 1e). In matrices

Table 1. Bootstrap probabilities $P(\theta_i^o \geq \theta_i^*)$ testing the stability of ordination subspaces in simulated and real data sets. Each probability was generated in 1000 resampling iterations. For description of the data sets see text. Probabilities in italics are significant at $\alpha = 0.1$. Evaluation of probabilities is shown from the right (axis 4 or 5) to the left (axis 1). Once an axis is considered significant, all axes to the left and corresponding probabilities are ignored (in parentheses). Probabilities (not shown) for higher axes in data sets with more than five variables were all as large as on axis 5.

(a) Matrices of type 4R with four variables and uniform correlation varying from 0 to 0.8.

Correlation level	Ordination dimensions evaluated				Interpreted dimensionality
Axis:	**1**	**2**	**3**	**4**	
0	0.795	0.672	0.609	0.494	0
0.1	0.510	0.736	0.644	0.478	0
0.2	0.233	0.759	0.655	0.477	0
0.3	*0.088*	0.777	0.660	0.495	1
0.4	*0.043*	0.765	0.665	0.512	1
0.5	*0.019*	0.806	0.680	0.492	1
0.6	*0.003*	0.758	0.643	0.485	1
0.8	*0.001*	0.762	0.651	0.484	1

(b) Matrices of type 32R with 32 variables and uniform correlation varying from 0 to 0.8.

Correlation level	Ordination dimensions evaluated					Interpreted dimensionality
Axis:	**1**	**2**	**3**	**4**	**5**	
0	0.835	0.818	0.804	0.747	0.730	0
0.1	*0.045*	0.880	0.880	0.870	0.858	1
0.2	*0.000*	0.880	0.872	0.859	0.854	1
0.3	*0.000*	0.874	0.821	0.797	0.727	1
0.4	*0.000*	0.879	0.854	0.868	0.872	1
0.5	*0.000*	0.848	0.868	0.859	0.854	1
0.8	*0.000*	0.808	0.815	0.721	0.728	1

(c) Matrices of type 12S1 with three groups of four variables each. The correlation varies from 0.1 to 0.8 between variables in the same group (1st column), and is 0 between variables in different groups.

Correlation level	Ordination dimensions evaluated					Interpreted dimensionality
Axis:	**1**	**2**	**3**	**4**	**5**	
0.1	0.716	0.637	0.528	0.828	0.782	0
0.2	0.683	0.476	0.342	0.851	0.825	0
0.3	0.651	0.406	0.122	0.875	0.836	0
0.4	(0.661)	(0.423)	*0.040*	0.861	0.792	3
0.5	(0.636)	(0.438)	*0.004*	0.905	0.844	3
0.6	(0.655)	(0.385)	*0.000*	0.884	0.818	3
0.8	(0.624)	(0.373)	*0.000*	0.890	0.876	3

(d) Matrices of type 12S2 with three groups of four variables each. The correlation varies from 0.1 to 0.8 between variables in the same group (1st column), and is always 0.3 between variables in different groups.

Correlation level	Ordination dimensions evaluated					Interpreted dimensionality
Axis:	**1**	**2**	**3**	**4**	**5**	
0.1	*0.008*	0.860	0.745	0.755	0.667	1
0.2	*0.009*	0.843	0.785	0.745	0.700	1
0.3	*0.002*	0.849	0.811	0.791	0.748	1
0.4	*0.005*	0.682	0.463	0.862	0.781	1
0.5	*0.005*	0.526	0.146	0.876	0.855	1
0.6	(0.008)	(0.485)	*0.033*	0.890	0.853	3
0.7	(0.008)	(0.532)	*0.002*	0.902	0.824	3
0.8	(0.032)	(0.488)	*0.000*	0.883	0.857	3

(e) Matrices of type 12S3 with 3 groups of five, four and three variables. The correlation varies from 0.1 to 0.8 between variables in the same group (1st column), and is 0 between variables in different groups.

Correlation level	Ordination dimensions evaluated					Interpreted dimensionality
Axis:	**1**	**2**	**3**	**4**	**5**	
0.1	0.697	0.612	0.623	0.802	0.791	0
0.2	0.513	0.456	0.427	0.872	0.786	0
0.3	0.462	0.317	0.250	0.863	0.800	0
0.4	(0.420)	(0.304)	*0.073*	0.892	0.862	3
0.5	(0.397)	(0.301)	*0.016*	0.892	0.826	3
0.6	(0.375)	(0.231)	*0.004*	0.891	0.829	3
0.8	(0.340)	(0.275)	*0.000*	0.858	0.871	3

(f) Random data set RND and grassland data sets VEG and ENV.

Data set	Ordination dimensions evaluated					Interpreted dimensionality
Axis:	**1**	**2**	**3**	**4**	**5**	
RND	0.541	0.538	0.478	0.520	0.453	0
VEG	(0.042)	(0.414)	*0.081*	0.543	0.502	3
ENV	(0.196)	*0.032*	0.648	0.405	0.412	2

12S2 axes 1-3 were deemed significant only if intervariable correlation within groups was at least 0.6; axis 1, however, was significant in all levels that were considered of intervariable correlation within groups (Table 1d). The ordination of the completely random data set RND was correctly identified as not having any interpretable dimension; the probabilities were around 0.5, as expected (Table 1f). For a comparison with the usual method of assessing the relevance of ordination axes by the percentage of each eigenvalue (of the sum of eigenvalues), see Table 2.

The analysis of field data sets revealed three nontrivial ordination dimensions in the vegetation data set VEG

Table 2. Principal coordinates analysis results (percentages of total Sum of eigenvalues) of selected simulated and real data sets. For description of the data sets see text.

Data	Correlation level	Eigenvalues 1	2	3	4
4R	0	25	25	25	25
4R	0.3	47.5	17.5	17.5	17.5
4R	0.8	85	5	5	5
32R	0	3.125	3.125	3.125	3.125
32R	0.1	12.82	2.81	2.81	2.81
32R	0.8	79.83	0.625	0.625	0.625
12S1	0.1	10.83	10.83	10.83	7.50
12S1	0.4	18.33	18.33	18.33	5.00
12S1	0.8	28.33	28.33	28.33	1.67
12S2	0.1	30.83	7.50	7.50	7.50
12S2	0.6	43.33	13.33	13.33	3.33
12S2	0.8	48.3	18.3	18.3	1.67
12S3	0.1	11.67	10.83	10.0	7.50
12S3	0.4	21.67	18.33	15.0	1.67
12S3	0.8	35	28.3	21.67	1.67
RND		15.9	13.8	12.6	10.4
ENV		38.1	20.2	7.77	7.17
VEG		21.6	11.3	9.63	5.35

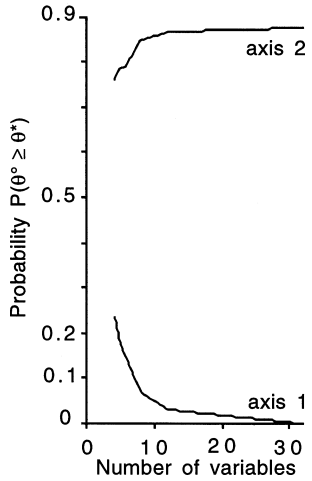

Fig. 1. The effect of number of variables in testing the significance of ordination axes in simulated data sets. The data sets contain 40 sampling units and four to 32 variables (extremes are in matrices 4R and 32R in Table 1a-b). Intervariable correlation is uniform (0.2).

and two in data set ENV (Table 1f). The first two eigenvalues in the ordination of ENV corresponded to 58% of the total inertia, while the first three in the ordination of VEG made up 43% of the total inertia (Table 2).

Discussion

The tests with simulated data sets consistently detected the intrinsic dimensionality if the intervariable correlation was above a threshold level. The lower correlation threshold in matrices with 32 variables compared to matrices with four variables suggests an effect of the number of variables. That is, for the same correlation level and number of sampling units, the larger the number of variables, the more significant was the first ordination dimension (Fig. 1). Interestingly, this observation for the bootstrap test contradicts conclusions by Grossman et al. (1991) that Bartlett's and other eigenvalues tests perform well if the ratio of the number of observations relative to variables is 3:1 or greater.

The method was reliable in disentangling the true dimensionality in data sets for which other methods evaluated by Jackson (1993) failed. This is especially the case in matrices of type 12S2. Matrix 12S2 with intervariable correlation of 0.8 within groups is equivalent to Jackson's matrix S-III. According to his results, the best methods were not completely reliable in detecting the true number of ordination components in this matrix; that is, two to three components were indicated by the broken-stick method, 1 to 3 by the bootstrap eigenvalue method, and 0 to 3 by the bootstrap eigenvector method. The probability $P(\theta_i^o \geq \theta_i^*)$ for matrix 12S2 with this level of correlation, and even for the lower level (0.6), doubtlessly indicated three significant ordination dimensions (Table 1d). With levels of correlation lower than 0.6, however, the first dimension only was deemed significant.

The tests with simulated matrices of type 12S1 detected the intrinsic dimensionality only when intervariable correlation within groups was 0.4 or larger (Table 1c). Matrix 12S1 with a correlation level of 0.3 had no significant dimensions. The uniqueness of the method, however, is in providing a probability, which can be judged as such. When a probability is given, the decision on the significance of a given ordination subspace has an associated error, which renders the level of being right or wrong measurable. That is, at a larger type I error (α) the test could indicate three, albeit weak ($P = 0.122$), nontrivial dimensions in this matrix. The magnitude of the threshold α probability is a matter of choice. Scientists usually by convention choose levels of 0.01 or 0.05, but there is nothing sacrosanct in this respect. We could reverse the problem: At which probability level the test would detect a given number of intrinsic dimensions? In this case, we express the results in probabilistic terms (the values in italics in Table 1). It is obvious that the weaker the correlation levels in the data structure, the greater will the probability be that the patterns revealed in ordination might as well be found in an ordination of random data.

In Jackson (1993), results for a matrix equivalent to matrix 12S1 with a correlation level of 0.3 (Jackson's matrix S-II), indicated one nontrivial component with the broken-stick method, zero components with the bootstrap eigenvalue method and 0 to 2 components with the bootstrap eigenvector method. The inconsistency in Jackson (1993) among results with the same matrix and method is partly due to sampling errors, since only three samples with 40 sampling units were taken from the larger matrix with 1000 sampling units and known correlation structure. Furthermore, the bootstrap methods in Jackson (1993) may have been affected by the occurrence of reversals between very similar (in strength) ordination dimensions across bootstrap samples, since his method did not use Procrustes fitting.

When applied to the grassland data, the method revealed three dimensions worthy of interpretation in the community ordination. Interestingly, this number of ordination dimensions was intuitively considered in the analysis of the same data in Pillar et al. (1992). In general, ecologists interpret no more than two or three community ordination dimensions. However, it remains an open question whether this practice would be sup-

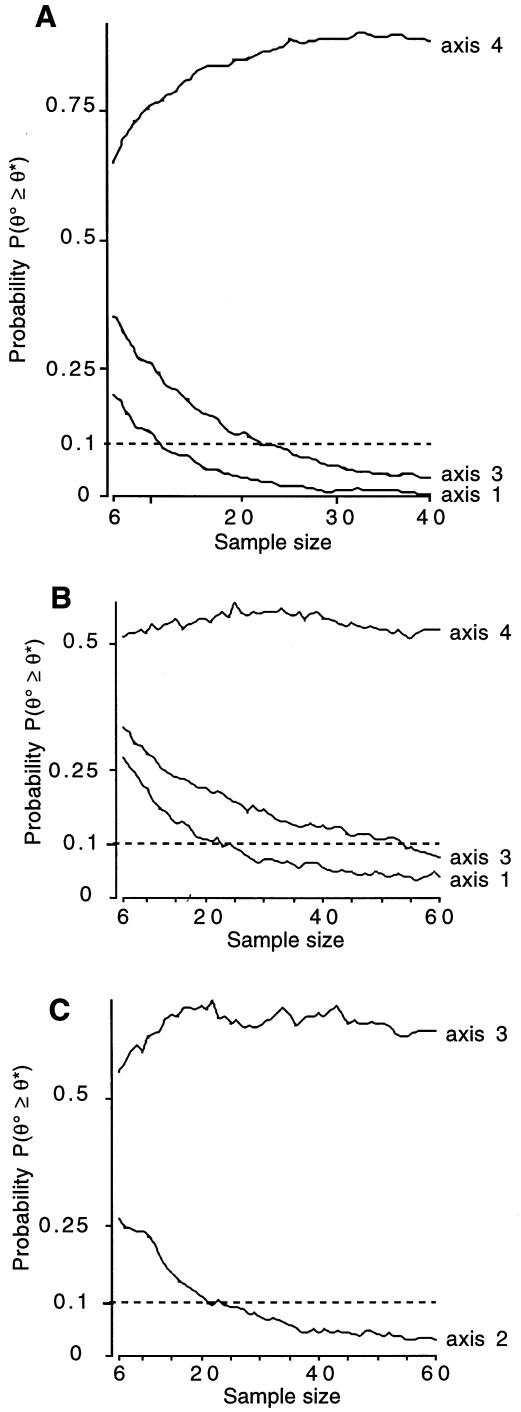

Fig. 2. The effect of sample size on the significance of ordination axes. The method involved bootstrap resampling with increasing sample size (Pillar 1998). In (A) the analysis used matrices of type 12S2 (see Table 1d) with three groups of four variables each and 40 sampling units; intervariable correlation was 0.6 within groups and 0.3 between groups. In (B) the analysis used data set VEG, with 60 species and 60 relevés from grassland in the south of Brazil, and in (C) the data set ENV with the same relevés described by 20 environmental variables (Table 1f).

ported by probability estimates obtained by bootstrapped ordination.

The effect of sample size on test results is illustrated for some typical cases (Fig. 2). Bootstrap sample sizes of 40 sampling units were sufficient to conclude in the test with data set 12S2 that ordination axis 3 is significant (α = 0.1) and that axis 4 is trivial (Fig. 2A). However, bootstrap resampling with 20 sampling units taken from the same pseudo sampling universe is indicating that only axis 1 is significant, while axis 3 is unstable. That is, by taking samples with size 20 from a real sampling universe described with variables of the given correlation structure, we would be likely to make a type II error, concluding that axis 3 is not significant. A similar problem would arise if we were using an α = 0.05 level to interpret the results with the grassland data set VEG (Fig. 2B): only the first dimension would then be interpreted as significant. Since the probabilities for axis 3 are still decreasing up to sample size 60, only by increasing the sample size of the field data set a more definite conclusion could be reached at an α = 0.05 level. In contrast, this is not the case for data set ENV (Fig. 2C), where the probabilities for the non-significant axis 3 are stable and high over a wide range of sample sizes.

The method may have limitations still open to scrutiny. Spatial autocorrelation in the data may hinder the use of bootstrap, but its effect is not clearly understood (see Efron & Tibshirani 1993: 396). Furthermore, I have not investigated how the bootstrap method performs with nonlinear data. The performance of metric and nonmetric ordination with nonlinear data structures is well known (Kenkel & Orlóci 1986). Detrended ordination or nonmetric ordination would be indicated in this case. I do, however, not see any limitations in applying the method to detrended ordination. The application to nonmetric ordination (NMDS) remains to be examined, since NMDS scores in two or more dimensions may not be meaningfully ordered by relevance, which is a requirement in the present method.

Acknowledgements. This work was financially supported by CNPq and CAPES (Brazil), and was in part executed during a leave at the Department of Plant Sciences, University of Western Ontario, Canada. I thank L. Orlóci for comments and hospitality. This paper benefitted from suggestions of Pedro Peres-Neto, Rune Økland and two anonymous reviewers.

References

Digby, P.G.N. & Kempton, R.A. 1987. *Multivariate analysis of ecological communities*. Chapman & Hall, London.

Efron, B. 1979. Bootstrap methods: another look at the jackknife. *Ann. Statist.* 7: 1-25.

Efron, B. & Tibshirani, R. 1993. *An introduction to the bootstrap*. Chapman & Hall, London.

Frontier, S. 1976. Étude de la decroissance des valeurs propres dans une analyze en composantes principales: comparison avec le modèle de baton brisé. *J. Exp. Mar. Biol. Ecol.* 25: 67-75.

Ganeshanandam, S. & Krzanowski, W.J. 1990. Error-rate estimation in two-group discriminant analysis using linear discriminant function. *J. Statist. Comp. Simul.* 36: 157-175.

Grossman, G.D., Nickerson, D.M. & Freeman, M.C. 1991. Principal components analyses of assemblage structure data: utility of tests based on eigenvalues. *Ecology* 72: 341-347.

Jackson, D.A. 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* 74: 2204-2214.

Jackson, D.A. 1995. Bootstrapped principal components analysis – reply to Mehlman et al. *Ecology* 76: 644-645.

Kendall, M. & Gibbons, J.D. 1990. *Rank correlation methods*, 5th ed. Edward Arnold, London.

Kenkel, N.C. & Orlóci, L. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. *Ecology* 67: 919-928.

Knox, R.G. 1989. Effects of detrending and rescaling on correspondence analysis: solutions on stability and accuracy. *Vegetatio* 83: 129-136.

Knox, R.G. & Peet, R.K. 1989. Bootstrapped ordination: a method for estimating sampling effects in indirect gradient analysis. *Vegetatio* 80: 153-165.

Mehlman, D.W., Shepherd, U.L. & Kelt, D.A. 1995. Bootstrapping principal components analysis – a comment. *Ecology* 76: 640-643.

Økland, R.H. 1999. On the variation explained by ordination and constrained ordination axes. *J. Veg. Sci.* 10: 131-136.

Orlóci, L. 1978. *Multivariate analysis in vegetation research*, 2nd ed. Junk, The Hague.

Orlóci, L. 1993. The complexities and scenarios of ecosystem analysis. In: Patil, G.P. & Rao, C.R. (eds.) *Multivariate environmental statistics*, pp. 423-432. Elsevier, Amsterdam.

Pielou, E.C. 1984. *The interpretation of ecological data. a primer on classification and ordination*. Wiley-Interscience, New York, NY.

Pillar, V.D. 1998. Sampling sufficiency in ecological surveys. *Abstr. Bot.* 22: 37-48.

Pillar, V.D., Jacques, A.V.A. & Boldrini, I.I. 1992. Fatores de ambiente relacionados à variação da vegetação de um campo natural. Pesqui. *Agropecu. Bras.* 27: 1089-1101.

Podani, J. 1994. *Multivariate data analysis in ecology and systematics*. SPB, The Hague.

Schönemann, P.H. & Carroll, R.M. 1970. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika* 35: 245-256.

Stauffer, D.F., Garton, E.O. & Steinhorst, R.K. 1985. A comparison of principal components from real and random data. *Ecology* 66: 1693-1698.

Received 6 April 1998;
Revision received 9 February 1999;
Final version received 17 March 1999;
Accepted 17 March 1999.

For App. 1, see next page.

App. 1. The bootstrap resampling algorithm for testing the significance of subspaces in ordination of sampling units, explained with a numerical illustration.

(1) Data set (variables in rows, sampling units in columns):

17	14	27	21	16
5	9	8	5	0
5	8	0	0	10

(2) Compute reference ordination with the data set:
PCOA eigenvalues: 183.159, 43.5497, 11.6911
Scores of sampling units (first 2 axes):

	1	2	3	4	5
Axis 1:	1.80579	4.9865	–9.39389	–4.32964	6.93124
Axis 2:	0.288276	4.93864	–0.0855575	–0.860073	–4.28128

(3) Get a bootstrap sample from (1):
Sampling units: 3 4 2 1 2

(4) Scores of bootstrap sampling units on the first 2 reference ordination axes, taken from (2):

	3	4	2	1	2
Axis 1:	–9.39389	–4.32964	4.9865	1.80579	4.9865
Axis 2:	–0.0855575	–0.860073	4.93864	0.288276	4.93864

(5) Ordination with bootstrap sample:
PCOA eigenvalues: 181.712, 19.1781, 1.91008
Scores of sampling units (first 2 axes):

	3	4	2	1	2
Axis 1:	–9.1088	–4.63347	6.12578	1.49071	6.12578
Axis 2:	–2.40637	2.46444	–1.12336	2.18864	–1.12336

(6) Scores on axis 2 from (5) after Procrustean adjustment involving axes 1 and 2 to maximize fit with similar ordination subspace in (4):

	3	4	2	1	2
Axis 2:	0.830199	–2.00955	5.00678	0.385703	5.00678

(7) Correlation θ_2^* between bootstrap (6) and reference scores (4) on axis 2: 0.971997

(8) Scores on axis 1 from (5) after Procrustean adjustment involving only axis 1 to maximize fit with similar ordination subspace in (4):

	3	4	2	1	2
Axis 1:	–8.88122	–4.7088	5.32221	1.00086	5.32221

(9) Correlation θ_1^* between bootstrap (6) and reference scores (4) on axis 1: 0.995973

(10) Data set from step (1) randomly permuted within variables:

27	14	17	16	21
5	5	0	8	9
5	10	8	0	0

(11) Reference ordination of permuted data set:
PCOA eigenvalues: 139.413, 85.5414, 13.4454
Scores of sampling units (first 2 axes):

	1	2	3	4	5
Axis 1:	4.56838	–6.84347	–5.75821	2.24378	5.78952
Axis 2:	6.56064	–1.2166	2.3469	–5.56826	–2.12269

The following steps (12) to (18) are similar to steps (3) to (9), but performed with permuted data from (10):
(12) Get a bootstrap sample from permuted data (with same number of repeats as in (3)):
Sampling units: 3 4 2 1 2

(13) Scores of bootstrap sampling units on the first 2 reference ordination axes of permuted data, taken from step (11):

	3	4	2	1	2
Axis 1:	–5.75821	2.24378	–6.84347	4.56838	–6.84347
Axis 2:	2.3469	–5.56826	–1.2166	6.56064	–1.2166

(14) Ordination of bootstrap sample taken from permuted data:
PCOA eigenvalues: 132.979, 71.845, 16.776
Scores of sampling units (first 2 axes):

	3	4	2	1	2
Axis 1:	1.69214	–2.07875	4.7068	–9.02699	4.7068
Axis 2:	–3.13995	7.26744	–0.613028	–2.90144	–0.613028

(15) Scores on axis 2 from (14) after Procrustean adjustment involving axes 1 and 2 to maximize fit with similar ordination subspace in (13):

	3	4	2	1	2
Axis 2:	2.3248	–5.53962	–1.20902	6.53895	–1.20902

(16) Correlation θ_2^o between bootstrap (15) and reference scores (13) on axis 2: 0.999999, which is larger than the correlation θ_2^* found in step (7).

(17) Scores on axis 1 from (14) after Procrustean adjustment involving only axis 1 to maximize fit with similar ordination subspace in (13):

	3	4	2	1	2
Axis 1:	–4.05445	–0.649667	–6.77643	5.62398	–6.77643

(18) Correlation θ_1^o between bootstrap (17) and reference scores (13) on axis 1: 0.947299, which is smaller than the correlation θ_1^* found in step (9).

(19) Repeat steps (3) to (18) up to B iterations.

(20) A run with $B = 1000$ iterations gave the following results:

Axis	$P(\theta_i^o \geq \theta_i^*)$	Average θ_i^*
1	0.391	0.982227
2	0.567	0.870788

The probabilities indicate none of the ordination dimensions examined is significant.