2-1995

Parallel Analysis: a Method for Determining Significant Principal Components

Scott B. Franklin
Southern Illinois University Carbondale

David J. Gibson
Southern Illinois University Carbondale, dgibson@plant.siu.edu

Philip A. Robertson
Southern Illinois University Carbondale

John T. Pohlmann
Southern Illinois University Carbondale

James S. Fralish
Southern Illinois University Carbondale

Published in *Journal of Vegetation Science* Vol. 6, No. 1 (1995): 99-106.
See also the SAS code.

Journal of Vegetation Science 6: 99-106, 1995
© IAVS; Opulus Press Uppsala. Printed in Sweden

Parallel Analysis: a method for determining significant principal components

Franklin, Scott B.[1*], Gibson, David J.[1], Robertson, Philip A.[1], Pohlmann, John T.[2] & Fralish, James S.[3]

[1]*Department of Plant Biology, Southern Illinois University, Carbondale, IL 62901, USA;* [2]*Education Psychology and Special Education Department, Southern Illinois University, Carbondale, IL 62901, USA;*
[3]*Department of Forestry, Southern Illinois University, Carbondale, IL 62901, USA;*
[*]*Author for correspondence; Tel. +1 618 4533236; Fax +1 618 4533441; E-mail GR3930@SIUCVMB.SIU.EDU*

Abstract. Numerous ecological studies use Principal Components Analysis (PCA) for exploratory analysis and data reduction. Determination of the number of components to retain is the most crucial problem confronting the researcher when using PCA. An incorrect choice may lead to the underextraction of components, but commonly results in overextraction. Of several methods proposed to determine the significance of principal components, Parallel Analysis (PA) has proven consistently accurate in determining the threshold for significant components, variable loadings, and analytical statistics when decomposing a correlation matrix. In this procedure, eigenvalues from a data set prior to rotation are compared with those from a matrix of random values of the same dimensionality (*p* variables and *n* samples). PCA eigenvalues from the data greater than PA eigenvalues from the corresponding random data can be retained. All components with eigenvalues below this threshold value should be considered spurious. We illustrate Parallel Analysis on an environmental data set.

We reviewed all articles utilizing PCA or Factor Analysis (FA) from 1987 to 1993 from *Ecology, Ecological Monographs, Journal of Vegetation Science* and *Journal of Ecology*. Analyses were first separated into those PCA which decomposed a correlation matrix and those PCA which decomposed a covariance matrix. Parallel Analysis (PA) was applied for each PCA/FA found in the literature. Of 39 analyses (in 22 articles), 29 (74.4 %) considered no threshold rule, presumably retaining interpretable components. According to the PA results, 26 (66.7 %) overextracted components. This overextraction may have resulted in potentially misleading interpretation of spurious components. It is suggested that the routine use of PA in multivariate ordination will increase confidence in the results and reduce the subjective interpretation of supposedly objective methods.

Keywords: Literature research; Overextraction; Principal Components Analysis; Spurious component.

Introduction

Numerous ecological studies have used some variant of Principal Components Analysis (PCA) since its introduction as an analytical method of classification and ordination by Goodall in 1954. Factor Analysis (FA) is a variant of PCA when communality estimates are incorporated into the matrix. These techniques are used mainly for reduction of data in the exploratory analysis of data sets. PCA may be used to decompose a correlation matrix or a covariance matrix (Noy-Meir et al. 1975; Ludwig & Reynolds 1988). While transformations may have little effect on correlation matrices, they will strongly affect covariance matrices (Noy-Meir et al. 1975).

Principal Components Analysis, which includes FA for the following discussion, assumes the data to be linear and normally distributed. Ecological data are notoriously non-normal, most notably species data (Pielou 1984; Austin 1987; Palmer 1993). However, environmental variables often have linear relationships and are normally distributed, especially when small ranges are sampled. These data must be tested for linearity and normality before continuing with PCA. If the data do not substantially violate linear relationships and normal distributions, use of PCA on a correlation matrix is an appropriate and valuable data reduction or exploratory technique. When analyzing species abundance data, another method (e.g. Canonical Correspondence Analysis; Palmer 1993), PCA of a covariance matrix, or properly transformed data (Karadžič & Popovič 1994) is recommended over PCA of a correlation matrix.

The problem of the 'number of components'

The problem of the 'number of components (factors)' (Howard & Gordon 1963) is the most critical one the researcher faces when using PCA (Frane & Hill

1976; Zwick & Velicer 1986; Fava & Velicer 1992; Greig-Smith 1980). Although methods are available for testing component significance, the general practice has been to rely on intuition (Frane & Hill 1976) or 'rules of thumb' (Singh & West 1971). Using the number of components the researcher wants to display or interpret has often been the 'rule' used (Gauch 1982; Pielou 1984; Kershaw & Looney 1985). An incorrect choice may lead to the underextraction of components (i.e. loss of information), but usually results in overextraction (i.e. inclusion of spurious components). Overextraction of components attaches meaning to noise and results in the interpretation of random variation in the data, thus affecting subsequent analyses or component rotations (Zwick & Velicer 1986). Several methods for determining the number of retained components are used in other disciplines, however there is little consensus in ecology (Legendre & Legendre 1983; Jackson 1993).

Our objective is to present Parallel Analysis (PA) as a technique for determining the number of retained components when using PCA on a correlation matrix. We summarize current methods for determining component significance, describe PA, apply PA to a research data set as an example, and apply PA to published analyses to illustrate the overinterpretation of PCA components in ecological literature.

Parallel Analysis, a Monte-Carlo test for determining significant Eigenvalues

Horn (1965) developed PA as a modification of Cattell's scree diagram to alleviate the component indeterminacy problem. Parallel Analysis is a "sample-based adaptation of the population-based [Kaiser's] rule" (Zwick & Velicer 1986), and allows the researcher to determine the significance of components, variable loadings, and analytical statistics. The rationale is that sampling variability will produce eigenvalues > 1 even if all eigenvalues of a correlation matrix are exactly one and no large components exist (as with independent variates) (Zwick & Velicer 1986; Buja & Eyuboglu 1992). The eigenvalues (EV) from research data prior to rotation are compared with those from a random matrix (actually normal pseudorandom deviates) of identical dimensionality to the research data set (i.e. same number of p variables and n samples). Component PCA eigenvalues which are greater than their respective component PA eigenvalues from the random data would be retained. All components with eigenvalues below their respective PA eigenvalue threshold probably are spurious. Frane & Hill (1976) suggested that research data be subsequently reanalyzed (run through PCA/FA again) using only the 'correct' number of components.

Parallel Analysis can be performed by running simulations (App. 1), referencing published work which presents regression models or tables of threshold values to test the significance of components, or readily available programs (Allen & Hubbard 1986; Lautenschlager 1989; Buja & Eyuboglu 1992; Pohlmann unpubl. - available from the author upon request). Longman et al. (1989) provided models that generate mean and 95th percentile eigenvalues. With these models, p and n sizes of the research data can be incorporated to calculate PA threshold eigenvalues. To date, the published works are entirely for PCA decomposing a *correlation matrix*. When decomposing a *covariance matrix* with PCA, the PA must restrict random matrices to have variable means and standard deviations identical to collected data, and include transformations performed on the variables.

Determining significant loadings

Parallel Analysis determines which variable loadings are significant for each component (Buja & Eyuboglu 1992; Pohlmann unpubl.), thus parsimoniously simplifying structure and reducing the analysis of noise. The PA procedure would replace subjectively determined thresholds (e.g. common thresholds are 0.5 and 0.8), and the inappropriate interpretation of correlation significance between variables and components. PCA extracts as much variance as possible out of the data. Even when the variables are uncorrelated, PCA will produce non-zero component correlations. If a matrix of zero correlations, with values of one along the diagonal, is subjected to PCA, all eigenvalues (sum of the squared variable-component correlations) will equal one. Hence, the average squared variable-component correlation is the reciprocal of the number of variables. Any inferential analysis of variable-component correlations must consider this bias. Correlation tables fail to provide guidance in the distribution of variable loadings.

A PA procedure applying the same methodology (e.g. rotations) as PCA can be used to derive random variable loadings. Multiplying the total number of variable loadings (number of variables × number of extracted components) by the significance level (i.e. 0.05 = 95th percentile) results in an empirical estimate of the 95th percentile. This empirical estimate is an objectively determined threshold for significant loadings and is appropriate for either correlation or covariance matrix PCA loadings. Buja & Eyuboglu (1992) also report a series of loadings tables (median, 90th, 95th, and 99th quantiles) for determining the significant variable loadings prior to rotation for a correlation matrix. The determination of significant loadings may seem cumbersome, but it is necessary when using a technique without objective stopping rules.

Material and Methods

Example use of Parallel Analysis with ecological data

Environmental data were collected from Land Between The Lakes, a National Recreation Area in western Kentucky and Tennessee, USA. (Franklin et al. 1993). Data were visually tested for linearity with scattergrams. Factor Analysis was performed on 15 environmental variables (p) in 133 stands (n) (Anon. 1990). Parallel Analysis was employed using the models derived by Longman et al. (1989) (App. 1). Factor Analysis was executed again using the correct number of components. Loadings were tested for significance using the Parallel Analysis program (App. 2).

Application of Parallel Analysis to published analyses

From 1987 to 1993, 61 articles utilizing PCA or FA were published in the *Journal of Vegetation Science*, *Journal of Ecology*, *Ecology* and *Ecological Monographs*. However, only 50 of the articles contained the necessary information (i.e. sample size, number of variables used in the analysis, and either the percent variance accounted for or eigenvalues for each factor) to run Parallel Analysis (PA). Of these, only 30 articles documented the use of a correlation matrix (22 articles, 73.3 %) or covariance matrix (8 articles, 26.7 %). Parallel Analysis (equations given by Longman et al. 1989; App. 1) was applied for each PCA/FA found in the literature that used a correlation matrix. The PA results were then compared with the published eigenvalues to determine the number of significant components (i.e. those components that should have been retained for subsequent analysis and interpretation).

Results

Example use of Parallel Analysis with ecological data

The EVs for the 4th and subsequent components [factors] were greater in the PA than in the FA analysis, indicating EVs of this magnitude could have been derived from sample noise (Table 1). The same conclusion resulted from using the tables of Buja & Eyuboglu (1992). Therefore, only three components were retained for further analysis. Cattell's scree test (Fig. 1) matched the above results while Kaiser's rule would have retained five components.

Parallel Analysis (App. 2) was performed using three components and the same rotation methods as FA to generate a random set of variable loadings (Table 2). As described above, the total number of loadings (3 factors × 15 variables = 45) was multiplied by the selected significance level (0.05 × 45 = 2.25, or 2) providing an empirical estimate of the 95th percentile loading, because two of the total number of loadings (i.e. 5 %) are expected to fall outside two standard deviations of the normal distribution. Thus, the second highest random structure loading is an estimate of the 95th percentile value. The absolute value of the second highest loading was |0.545| and thus all loadings below this value were considered insignificant in our analysis (Table 2).

Table 1. Comparison of Factor Analysis and Parallel Analysis eigenvalues for data from Land Between The Lakes. The eigenvalues for the first three FA components are larger than the corresponding PA eigenvalues and are thus significant at ρ = 0.05. Retaining these components for interpretation and subsequent analysis is appropriate.

Component	FA eigenvalue	PA eigenvalue
1	4.360	1.832
2	2.630	1.627
3	2.208	1.488
4	1.300	1.357
5	1.131	1.260
6	0.884	1.175

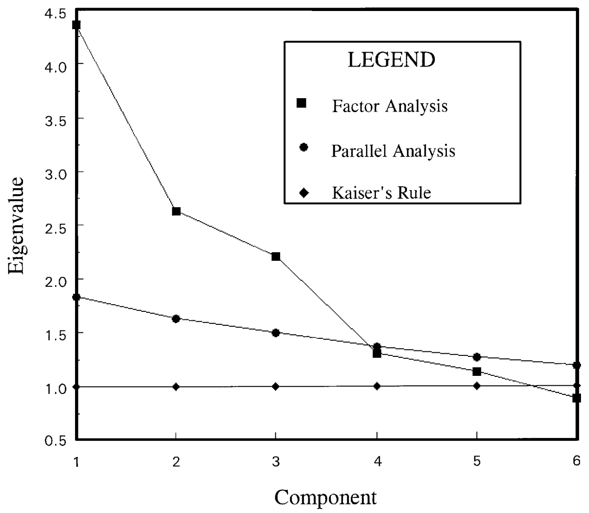

Fig. 1. Scree diagram comparing methods for determining the number of components to retain. The scree test uses the eigenvalues from Principal Components Analysis or Factor Analysis, drawing a straight line through the lowest eigenvalues. The threshold is where this line separates from the eigenvalue line, which can be a subjective decision (decision = retain three components). The Parallel Analysis threshold is when the eigenvalues from PA are greater than those from PCA/FA (decision: retain three components). Kaiser's rule retains all components with eigenvalues > 1, and would retain five components in this example. Analyses are of environmental data taken at Land Between The Lakes (Franklin et al. 1993).

Table 2. Random structure loadings (* = top two variable loadings) and variable structure correlations on three components from Parallel Analysis of 15 variables and 133 stands from Land Between The Lakes. CEC = cation exchange capacity; AWC = available water capacity; ESD = effective soil depth (Franklin et al. 1993). Variable structure loadings > |0.545| (in bold) are considered significant at $\rho = 0.05$.

Variable	Comp. 1	Comp. 2	Comp. 3
Random structure loadings			
1	– 0.089	– 0.482	– 0.254
2	0.525	0.028	0.147
3	0.115	– 0.146	– 0.189
4	0.032	– 0.178	0.637 *
5	0.313	– 0.039	– 0.260
6	– 0.023	0.228	0.348
7	– 0.005	– 0.177	0.255
8	– 0.044	0.008	0.229
9	0.190	– 0.043	– 0.402
10	0.185	– 0.451	0.105
11	– 0.424	0.055	0.213
12	0.518	– 0.093	– 0.039
13	0.329	0.167	– 0.226
14	– 0.180	0.327	– 0.070
15	0.158	0.545 *	0.037
Environmental variable structure loadings			
Calcium	**0.851**	0.301	– 0.208
Magnesium	**0.816**	0.380	– 0.074
pH	**0.738**	**0.554**	– 0.274
Potassium	**0.786**	0.182	– 0.058
CEC	**0.720**	0.016	– 0.038
Phosphorus	0.470	– 0.004	0.347
% Organic matter	**0.546**	– 0.401	0.386
% sand A-horizon	– 0.086	– 0.013	**0.804**
% clay A-horizon	– 0.085	– 0.188	**– 0.709**
% rock	– 0.018	– 0.126	**0.759**
AWC (cm)	0.307	**– 0.548**	**– 0.577**
Slope position	0.094	**0.846**	0.005
Distance to opposing slope	– 0.127	**– 0.740**	– 0.059
Elevation	– 0.142	**– 0.654**	0.122
ESD	0.373	0.490	0.170

Application of PA to published analyses

Of the analyses reviewed (39 analyses in 22 articles), 8 (20.5 %) used Kaiser's Rule, 2 (5.1 %) used a percent variance explained threshold, 1 (2.6 %) used Parallel Analysis, and 29 (74.4 %) retained components based on interpretability (Table 3). Parallel Analysis of the 39 PCAs that decomposed a correlation matrix indicated that 26 (66.7 %) overextracted components (Table 3). We could not determine if components had been underextracted. It appears that better criteria are needed for determining the number of retained components when applying PCA to ecological data.

Table 3. A comparison of retained Principal Components in the literature with Parallel Analysis. Ret. Comp. are the number of components retained by the author(s). PA Ret. is the number of components that should be retained based on spurious eigenvalues.

Ret. Comp.	PA Ret.	Rule	Data	Method	Reference
3	1	I	E	PCA	Abdel-Razik & Ismail 1990
3	2	I	E	PCA	Blinn 1993
4	3	I	E	PCA	Bornette & Amoros 1991
5	4	K	E	PCA	Finch 1989
3	3	PA	E	FA	Franklin et al. 1993
4	4	K/ %	S	PCA	Ganter & Starmer 1992
9	1	I	M	PCA	Gascon 1991
5	3	I	E	PCA	Gascon 1991
5	4	I	E	FA	Hayati & Proctor 1990
3	3	K	M	FA	Herrera 1987
4	2	I	M	FA	Herrera 1987
8	2	I	E	FA	Herrera 1987
2	1	I	M	PCA	Losos 1990
2	1	I	M	PCA	Losos 1990
3	1	I	M	PCA	Losos 1990
3	3	I	M	PCA	Losos 1990
2	1	I	M	PCA	Losos 1990
4	4	I	M	FA	McPeek 1990
4	2	I	E	PCA	Meffe & Sheldon 1990
3	2	I	E	PCA	Nakashizuka 1989
3	2	I	M	PCA	Sallabanks 1993
2	2	I	M	PCA	Sallabanks 1993
2	2	I	M	PCA	Sallabanks 1993
7	2	I	E	FA	Scheibe 1987
3	2	I	M	FA	Scheibe 1987
2	2	I	E	PCA	Schieck & Hannon 1993
3	1	I	M	PCA	Schwaegerle & Bazzaz 1987
3	1	I	E	PCA	Schwaegerle & Bazzaz 1987
4	1	I	E	PCA	Schwaegerle & Bazzaz 1987
3	3	I	M	PCA	Smith 1987
3	2	%	E	PCA	Sun & Feoli 1992
3	2	K	E	PCA	Wiens 1991
2	2	K	E	FA	Wikramanayake 1990
2	2	K	M	FA	Wikramanayake 1990
2	1	K/I	M	FA	Wikramanayake 1990
2	2	K	M	FA	Wikramanayake 1990
3	2	I	E	PCA	Wilson & Hebert 1992
3	2	I	E	PCA	Wilson & Hebert 1992
3	3	I	E	PCA	Wilson & Hebert 1992

* Rule: K = Kaiser's, PA = Parallel Analysis, % = percent variance accounted for, I = interpretability; Data: E = environmental, S = species, M = measurements on object of study; Method: PCA = Principal Components Analysis, FA = Factor Analysis (using communality estimates).

A few additional observations warrant discussion. Several authors were vague concerning one or more of the necessary criteria for applying PA. Only 60 % of the articles distinguished between the use of a correlation matrix or a covariance matrix, an important difference when using PCA. The size of the matrix was often difficult to discern. More importantly, in some cases (~20 %) neither eigenvalues nor percent variance values were given for each extracted component. This information is necessary to determine the robustness of results. In several articles, correlations of variables with the extracted components were inappropriately used to determine the significance of each variable loading.

As noted earlier, known non-normal distributions of abundance data exclude the application of PCA decomposing a correlation matrix. Almost all articles using PCA of a correlation matrix analyzed environmental data or measurement data which likely conform more closely to PCA assumptions than species abundance data. Nevertheless, few authors mention whether they tested the linearity of their data. Except for a few rare cases, PCA was only applied to species data when decomposing a covariance matrix.

Discussion

Parallel Analysis is an efficient and robust means for determining the number of principal components to retain for further analysis and interpretation when decomposing a correlation matrix. The example analysis of environmental data from Land Between The Lakes shows the capability of PCA to extract meaningful information from a data matrix when the data have a linear relationship and are normally distributed. The example PA demonstrates that significant eigenvalues and variable loadings may be objectively determined. This simple technique leads to parsimonious results, the purpose for analyzing data with PCA.

Our review of the ecological literature indicated that objective criteria to determine retained components often are not used with PCA or FA. This has resulted in potentially misleading interpretation of spurious components. We strongly recommend PA for determining component significance when using PCA to decompose a correlation matrix.

Reliability of PA and other stopping rules

Several methods available for determining the number of components to extract from PCA were tested by Zwick & Velicer (1986) and Jackson (1993). One common rule is Kaiser's 'eigenvalue greater than 1' method (Kaiser 1960). A component eigenvalue of one accounts for as much significance as a single variable. If data reduction is one objective of the analysis, retaining components with eigenvalues less than one is inappropriate and not parsimonious (i.e. retained components have less summarizing power than the original variable alone). This popular rule often overextracts components (Zwick & Velicer 1986). The Maximum Likelihood test (Lawley 1940, 1941), Bartlett's chi-square test (Bartlett 1950, 1951), and the Asymptotic Theory (Anderson 1963) are similar in that they test the equality of eigenvalues. Zwick & Velicer (1986) found Bartlett's test to be highly variable because of its sensitivity to a number of influences (e.g. sample size), and proposed the same limitation for the Maximum Likelihood test. Cattell's scree diagram (Cattell 1966) also may be used as a stopping rule, but is known to overestimate the number of components and is prone to subjective bias (Zwick & Velicer 1986; Jackson 1993). However, the scree test was found to be more accurate than Kaiser's rule or Bartlett's test. Zwick & Velicer (1986) argued against the use of Kaiser's rule, Bartlett's test, or the scree test as methods of choice for determining the number of components. Jackson (1993) similarly found Kaiser's rule, the scree test, Bartlett's test and the Maximum Likelihood test to be inaccurate measures for determining the number of retained components. However, these are the most commonly available threshold techniques in popular statistical packages - scree test and Maximum Likelihood in SAS/STAT (Anon. 1990); scree, Bartlett's, Anderson-Rubin (asymptotic), and Maximum Likelihood in SPSS-X (Anon. 1988).

The Minimum Average Partial, MAP (Velicer 1976; Reddon 1985 provides FORTRAN program subroutines) based on partial correlations was more accurate than the above methods but it tended to underextract components (Zwick & Velicer 1986). The final method tested by Zwick & Velicer (1986), Parallel Analysis (Horn 1965), proved consistently accurate with only a slight tendency to overextract components. The MAP and PA techniques were found to be the most accurate methods for determining the number of components (Zwick & Velicer 1986). Another form of PA involves adding one random variable into a research data set (Ibanez 1973 in Legendre & Legendre 1983). Interpretation ends when the random variable has the most important loading on a component. This PA technique also was considered more appropriate than Bartlett's test for ecological data (Legendre & Legendre 1983).

Jackson (1993) concluded that the broken stick method (Frontier 1976) and the bootstrapped eigenvector-eigenvalue method (Lambert et al. 1990) appear more promising than the above methods, excluding MAP which was not tested. The broken stick method is similar to the PA described and tested in this article, but does not consider sample size and, thus, cannot really model sampling distributions of eigenvalues (Horn 1965). Fisher's proportion test (Fisher 1958) has the same limitation; it does not consider sample size.

Jackson (1993) found the bootstrap eigenvalue-eigenvector method to be a reliable assessment of 'meaningful' components. Monte Carlo permutation tests are available in the commonly used CANOCO (ter Braak 1988) and MRPP programs (Biondini et al. 1988). Nevertheless, permutation methods are based on repeated samplings of randomizations of collected data and thus are restricted to the range of values in the data set. There are two arguments here. First, it can be argued that because ecological data are notoriously skewed, use of anything besides the collected data would render a useless comparison. The alternative view rests on known properties of sample distributions. An entire population is rarely sampled. For this reason, sample data are biased to what was collected. Bootstrapping is a good method for estimating expected values, but often the parameter of interest is not an expected value (i.e. biased estimators will yield biased bootstrapped inferences).

There may be concern when using normally distributed data (i.e. normal pseudorandom deviates) in comparison to ecological data, which may be nonlinear and skewed (Lambert et al. 1990; Jackson 1993). However, Buja & Eyuboglu (1992) state that permutations offer little advantage over normal assumption techniques except in more complex situations where tabulations are impossible. Skinner (1979) found little difference between his parallel analysis and permutation results. In addition, PA is a much simpler approach to the ‘number of components’ problem than permutation calculations.

Some authors interpret only components with at least 2 or 3 significant loadings (Zwick & Velicer 1986; Jackson 1993). In the final analysis, the retained components must make good scientific sense (Frane & Hill 1976; Legendre & Legendre 1983; Pielou 1984; Zwick & Velicer 1986; Ludwig & Reynolds 1988; Palmer 1993).

Use of PA for other multivariate procedures

Parallel Analysis also may be used for PCA decomposing a covariance matrix by restricting the random matrix to variable means and standard deviations identical to collected data, as well as any transformations. Means and standard deviations were not given in the articles where a covariance matrix was decomposed, thus we could not test their results. However, it is likely that overextraction of components and hence overinterpretation exists in these studies as in the majority of studies that used PCA. Indeed, 64 % of the analyses which used PCA to decompose a covariance matrix did not use an objective method for determining the number of retained components. Although not explored in this report, we suggest that PA could be adapted for many other eigenanalysis techniques used in ecological ordination (i.e. DCA, RA, CCA). All are essentially similar, matrix based procedures.

Generally, we recommend using more than one rule, e.g. Parallel Analysis (randomization) and a permutation test) for determining the number of components to retain for use in any PCA analysis. Use of two rules would add robustness to the ‘number of components’ decision and subsequent interpretation (Frane & Hill 1976; Zwick & Velicer 1986). Ultimately, the routine use of PA and other stopping rules for users of multivariate techniques will allow greater confidence in the results and lessen the subjective interpretation of supposedly objective methods.

Acknowledgements. We thank Fred B. Crooks for help with field collection and summarization of stand data. A special thanks to Jari Oksanen and anonymous reviewers for their constructive evaluation and comments. Support for SBF and JSF was provided in part by the Center for Field Biology, Austin Peay State University, Clarksville, TN. DJG was supported in part by NSF grant DEB-9317976.

References

Anon. 1988. *SPSS-X user’s guide, 3rd ed.* SPSS Inc., Chicago, IL.

Anon. 1990. *SAS/STAT user’s guide, version 6, vol. 2*, GLM-VARCOMP. SAS Institute Inc., Cary, NC.

Abdel-Razik, M.S. & Ismail, A.M.A. 1990. Vegetation composition of a maritime salt marsh in Qatar in relation to edaphic features. *J. Veg. Sci.* 1: 85-88.

Allen, S.J. & Hubbard, R. 1986. Regression equations for the latent roots of random data correlation matrices with unities on the diagonal. *Multi. Behav. Res.* 21: 393-398.

Anderson, T.W. 1963. Asymptotic theory for principal component analysis. *Ann. Math. Stat.* 34: 122-148.

Austin, M.P. 1987. Models for the analysis of species’ response to environmental gradients. *Vegetatio* 69: 35-45.

Bartlett, M.S. 1950. Tests of significance in factor analysis. *Br. J. Psychol.* 3: 77-85.

Bartlett, M.S. 1951. A further note on tests of significance in factor analysis. *Br. J. Psychol*. 4: 1-2.

Biondini, M.E., Mielke, P.W. & Redente, E. F. 1988. Permutation techniques based on euclidean analysis spaces: a new and powerful statistical method for ecological research. *Coenoses* 3: 155-174.

Blinn, D.W. 1993. Diatom community structure along physicochemical gradients in saline lakes. *Ecology* 74: 1246-1263.

Bornette, G. & Amoros, C. 1991. Aquatic vegetation and hydrology of a braided river floodplain. *J. Veg. Sci.* 2: 497-512.

Buja, A. & Eyuboglu, N. 1992. Remarks on parallel analysis. *Multi. Behav. Res.* 27: 509-540.

Cattell, R.B. 1966. The scree test for the number of factors. *Multi. Behav. Res.* 1: 245-276.

Fava, J.L. & Velicer, W.F. 1992. An empirical comparison of factor, image, component, and scale scores. *Multi. Behav. Res.* 27: 301-322.

Finch, D.M. 1989. Habitat use and overlap of riparian birds in three elevational zones. *Ecology* 70: 866-880.

Fisher, R.A. 1958. *Statistical methods for research workers, 13th ed*. Hafner, New York, NY.

Frane, J.W. & Hill, M. 1976. Factor analysis as a tool for data analysis. *Commun. Stat. Theor. Meth.* A5: 507-527.

Franklin, S.B., Robertson, P.A., Fralish, J.S. & Kettler, S.M. 1993. Overstory vegetation and successional trends of Land Between The Lakes, USA. *J. Veg. Sci.* 4: 509-520.

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

Ganter, P.F. & Starmer, W.T. 1992. Killer factor as a mechanism of interference competition in yeasts associated with cacti. *Ecology* 73: 54-67.

Gascon, C. 1991. Population- and community-level analyses of species occurrences of central Amazonian rainforest tadpoles. *Ecology* 72: 1731-1746.

Gauch, H.G. Jr. 1982. Noise reduction by eigenvector ordinations. *Ecology* 63: 1643-1649.

Goodall, D.W. 1954. Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. *Aust. J. Bot.* 2: 304-324.

Greig-Smith, P. 1980. The development of numerical classification and ordination. *Vegetatio* 42: 1-9.

Hayati, A.A. & Proctor, M.C.F. 1990. Plant distribution in relation to mineral nutrient availability and uptake on a wet-heath site in south-west England. *J. Ecol.* 78: 134-151.

Herrera, C.M. 1987. Vertebrate-dispersed plants of the Iberian Peninsula: a study of fruit characteristics. *Ecol. Monogr.* 57: 305-331.

Horn, J.L. 1965. A rationale and test for the number of factors in factor analysis. *Psychometrica* 30: 179-185.

Howard, K.I. & Gordon, R.A. 1963. Empirical note on the 'number of factors' problem in factor analysis. *Psychol. Rep.* 12: 247-250.

Ibanez, F. 1973. Méthode d'analyse spatio-temporelle du processus d'échantillonnage en planctologie, son influence dans l'interprétation des données par l'analyse en composantes principales. *Ann. Inst. Océanogr. Paris* 49: 83-111.

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

Kaiser, H.F. 1960. The application of electronic computers to factor analysis. *Ed. Psychol. Meas.* 20: 141-151.

Karadžič, B. & Popovič, R. 1994. A generalized standardization procedure in ecological ordination: tests with principal components analysis. *J. Veg. Sci.* 5: 259-262.

Kershaw, K.A. & Looney, J.H.H. 1985. *Quantitative and dynamic plant ecology, 3rd edition*. Edward Arnold, London, U.K.

Lambert, Z.V., Wildt, A.R. & Durand, R.M. 1990. Assessing sampling variation relative to number-of-factors criteria. *Educ. and Psychol. Meas.* 50: 33-49.

Lautenschlager, G.J. 1989. A comparison of alternatives to conducting Monte Carlo analyses for determining Parallel Analysis criteria. *Multi. Behav. Res.* 24: 365-395.

Lawley, D.M. 1940. The estimation of factor loadings by the methods of maximum likelihood. *Proc. R. Soc. Edinb.* 60: 64-82.

Lawley, D.M. 1941. Further investigation in factor estimation. *Proc. R. Soc. Edinb. Sect. A* 61: 176-185.

Legendre, L. & Legendre, P. 1983. *Numerical ecology*. Elsevier Scientific Publishing Co., New York, NY.

Longman, R.S., Cota, A.A., Holden, R.R. & Fekken, G.C. 1989. A regression equation for the parallel analysis criterion in principal components analysis: mean and 95th percentile eigenvalues. *Multi. Behav. Res.* 24: 59-69.

Losos, J.B. 1990. Ecomorphology, performance capability, and scaling of West Indian *Anolis* lizards: an evolutionary analysis. *Ecol. Monogr.* 60: 369-388.

Ludwig, J.A. & Reynolds, J.F. 1988. *Statistical ecology: a primer on methods and computing*. John Wiley and Sons, New York, NY.

McPeek, M.A. 1990. Behavioral differences between *Enallagma* species (Odonata) influencing differential vulnerability to predators. *Ecology* 71: 1714-1726.

Meffe, G.K. & Sheldon, A.L. 1990. Post-defaunation recovery of fish assemblages in southeastern blackwater streams. *Ecology* 71: 657-667.

Nakashizuka, T. 1989. Role of uprooting in composition and dynamics of an old-growth forest in Japan. *Ecology* 70: 1273-1278.

Noy-Meir, I., Walker, D. & Williams, W.T. 1975. Data transformations in ecological ordination. *J. Ecol.* 63: 779-800.

Palmer, M.W. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* 74: 2215-2230.

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

Reddon, J.R. 1985. MAPF and MAPS: subroutines for the number of principal components. *Appl. Psychol. Meas.* 9: 97.

Sallabanks, R. 1993. Hierarchical mechanisms of fruit selection by an avian frugivore. *Ecology* 74: 1326-1336.

Scheibe, J.S. 1987. Climate, competition, and the structure of temperate lizard communities. *Ecology* 68: 1424-1436.

Schieck, J.O. & Hannon, S.J. 1993. Clutch predation, cover, and the overdispersion of nests of the willow ptarmigan. *Ecology* 74: 743-750.

Schwaegerle, K.E. & Bazzaz, F.A. 1987. Differentiation among nine populations of *Phlox*: response to environmental gradients. *Ecology* 68: 54-64.

Singh, T. & West, N.E. 1971. Comparison of some multivariate analyses of perennial *Atriplex* vegetation in southeastern Utah. *Vegetatio* 23: 289-313.

Skinner, H.A. 1979. Dimensions and clusters: a hybrid approach to classification. *Appl. Psychol. Meas.* 3: 327-341.

Smith, T.J. III. 1987. Seed predation in relation to tree dominance and distribution in mangrove forests. *Ecology* 68: 266-273.

Sun, C.Y. & Feoli, E. 1992. Trajectory analysis of Chinese vegetation types in a multidimensional climatic space. *J. Veg. Sci.* 3: 587-594.

ter Braak, C.J.F. 1988. *CANOCO - a FORTRAN program for canonical community ordination by [partial] [detrended] [canonical] correlation analysis, principal components analysis and redundancy analysis (version 2.1).* Technical Report LWA-88-02, GLW, Wageningen.

Velicer, W.F. 1976. Determining the number of components from the matrix of partial correlations. *Psychometrica* 41: 321-327.

Wiens, J.A. 1991. Ecological similarity of shrub-desert avifaunas of Australia and North America. *Ecology* 72: 479-495.

Wikramanayake, E.D. 1990. Ecomorphology and biogeography of a tropical stream fish assemblage: evolution of assemblage structure. *Ecology* 71: 1756-1764.

Wilson, C.C. & Hebert, P.D.N. 1992. The maintenance of taxon diversity in an asexual assemblage: an experimental analysis. *Ecology* 73: 1462-1472.

Zwick, W.R. & Velicer, W.F. 1986. Comparison of five rules for determining the number of components to retain. *Psychol. Bull.* 99: 432-442.

Received 12 October 1993;
Revision received 11 October 1994;
Accepted 15 October 1994.

App. 1. Parallel Analysis, PA. SAS program giving the 95th percentile eigenvalue loading utilizing equations derived by Longman et al. (1989).

```
Data Longman;  options LS=73;
******************************
This program produces estimates of the 95th percentile eigenvalues

From a parallel analysis, using the work of Longman et al. (1989).
Change the values of n and p to those of your data matrix.
******************************;
N=Xx;  P=Xx;  * N = Sample Size, P = No. Of Variables;
Ln = Log (N);  Lp = Log(P);
Leig1 = 0.0316*Ln +0.7611*Lp -0.0979 *(Ln*Lp) -0.3138; Lam1 =Exp(Leig1);
Leig2 = 0.1162*Ln +0.8613*Lp -0.1122 *(Ln*Lp) -0.9281; Lam2 =Exp(Leig2);
Leig3 = 0.1835*Ln +0.9436*Lp -0.1237 *(Ln*Lp) -1.4173; Lam3 =Exp(Leig3);
Leig4 = 0.2578*Ln +1.0636*Lp -0.1388 *(Ln*Lp) -1.9976; Lam4 =Exp(Leig4);
Leig5 = 0.3171*Ln +1.1370*Lp -0.1494 *(Ln*Lp) -2.4200; Lam5 =Exp(Leig5);
Leig6 = 0.3809*Ln +1.2213*Lp -0.1619 *(Ln*Lp) -2.8644; Lam6 =Exp(Leig6);
Leig7 = 0.4492*Ln +1.3111*Lp -0.1751 *(Ln*Lp) -3.3392; Lam7 =Exp(Leig7);
Leig8 = 0.5309*Ln +1.4265*Lp -0.1925 *(Ln*Lp) -3.8950; Lam8 =Exp(Leig8);
Leig9 = 0.5734*Ln +1.4818*Lp -0.1986 *(Ln*Lp) -4.2420; Lam9 =Exp(Leig9);
Leig10= 0.6460*Ln +1.5802*Lp -0.2134 *(Ln*Lp) -4.7384; Lam10=Exp(Leig10);
Proc Print; Var N P Lam1-Lam10;
Run;
Endsas;
```

*Note: Exp raises *e* (2.71828) to a specified power.

App. 2. Parallel Analysis, PA. SAS program for determining variable loading significance.

```
 1   Data loadings; options LS = 73;
 2   ** this program will generate parallel analysis significant loadings**
 3   ** generalized parallel analysis procedure:
 4           1. Number of variables is set with the www index value,
 5           2. Number of observations is set with the yyy index for j in the
 6                   First do statement and in the var statement,
 7           3. Number of analyses is set with the zzz index,
 8           4. Lastly, perform the same factor analysis on the simulated
 9                   Data matrix that you performed on the actual data matrix.
10   ****** Warning - this program generates a big listing ****** ;
11   Array x (i) x1 – xwww; * set the number of variables (www);
12   Do k = 1 to zzz; * set the number of analyses (zzz);
13           Do j = 1 to yyy; * set the sample size (yyy);
14                   Do over x;          ** x1 = normal(0) * std + mean;
15                   X = normal (0);     ** x2 = normal(0) * std + mean;
16                   End;                ** .    .        .    .  ;
17           Output;                     ** .    .        .    .  ;
18           End;                        ** xwww = normal(0) * std + mean;
19   End;                                ** only for use with covariance matrices;
20   Data two; set one;
21   ** set the number of factors with the n = parameter;
22   ** use the same methods (i.e. PCA or FA, rotations, etc.) As with real
23           data analysis;
24   Proc factor method = p rotate = promax prerotate = varimax nfact = 3
25   score outstat = new plot nplot = 2; var x1-xwww;
26   Proc print;
27   Run;
28   Endsas;
```

Note: For analyses decomposing a covariance matrix, lines 14-16 must be replaced with the given equations for X1-XWWW (as shown). This will perform a PA (randomization technique) using known parameters of each variable. If the variables are transformed when analyzing the actual data, they must also be transformed for PA.