Principal Components Analysis[☆]

Craig Syms, James Cook University, Townsville, QLD, Australia

Glossary

Biplot An ordination diagram that simultaneously presents dependent variables.

Centroid The (weighted) mean of a multivariate data set. Can be represented by a vector.

Correlation coefficient A measure of strength of the relationship between two variables.

Eigenanalysis The process of finding eigenvectors and eigenvalues of a matrix. Sample scores are often eigenvectors while eigenvalues are usually ranked from highest to lowest, and termed the first, second, third, etc. eigenvalues.

Matrix A set of numbers arranged in rows and columns. A correlation matrix consists of correlation coefficients and is an example of a square symmetric mtrix. The rows and the columns represent the variables. A Covariance matrix consists of covariant entries, where the diagonal elements will equal the variances. A similarity matrix reflects the degree of similarity or dissimilarity between groups and are easily produced from, or converted into, a distance matrix.

Ordination The ordering of a set of data points with respect to one or more axes so that relationships among the points in any number of dimensions can be visible on inspection.

Principal Components The axes of a principal components analysis. The first principal component will, ideally, represent the dominant gradient. The second component will be orthogonal to the first, and will explain some of the residual variation. The third will be orthogonal to the first and second components, and so on.

Introduction

Principal components analysis (PCA) is a distance-based ordination technique used primarily to display patterns in multivariate data. It aims to display the relative positions of data points in fewer dimensions while retaining as much information as possible, and explore relationships between dependent variables. In general, it is a hypothesis-generating technique that is intended to describe patterns, rather than test formal statistical hypotheses. Although PCA was originally developed to analyze continuous variables, it can also be used on ordinal and presence–absence data.

PCA is carried out on the response of dependent variables in a multivariate data set. Consequently it is an unconstrained ordination, in which hypothetical causal independent variables are not explicitly included in the analysis. For example, if the abundance of several species of fish (the response or dependent variables) were measured at a range of different sites with different characteristics such as wave exposure (causal or independent variables), the exposure information would not be explicitly included in the analysis. Patterns recovered in PCA are solely a function of relationships between the dependent variables. For this reason, PCA can also be classified as an indirect gradient analysis, in which hypothetical causal processes such as exposure, moisture, etc., are inferred from patterns in the dependent variables. PCA assumes that the relationships between dependent variables are linear. This implies that PCA should be applied when most dependent variables have nonzero values across most of the samples, and that bivariate scatterplots of each variable against each other variable should be linear or at least monotonically increasing or decreasing. PCA is a very useful analytical tool, and is one of the most widely used ordination methods in ecology.

An Informal Explanation

Given a multivariate data set consisting of a number of samples in which many variables have been recorded intuitively, PCA is a process in which the original variable axes are aligned along lines of variation in the data and the values for each sample on those new axes are calculated. For example, consider a data set containing 10 samples of abundances of two species. The relative position of each sample in two dimensions can be displayed with a scatterplot of species A versus species B (Fig. 1A). A new set of ordination axes can be generated by moving the old axes to the center of the data set by subtracting the mean of each variable from the sample values—a process known as centering—then rotating these axes so that they lie along the major lines of variation (Fig. 1B). In PCA, the first axis (principal component 1) lies along the greatest line of variation, the second axis lies along the next greatest line of variation on the condition that it lies at right angles to the first, and so on for subsequent axes. This guarantees a property known as orthogonality; which means that each axis is independent of each other. The next step is to project the sample positions onto these new axes—these axes are called principal components (PCs) (Fig. 1C).

[☆] *Change History*: March 2018. H R Pethybridge included the addition of a glossary and keywords, updates to the abstract, and a revised references list.

This is an update of C. Syms, Principal Components Analysis, In Encyclopedia of Ecology, edited by Sven Erik Jørgensen and Brian D. Fath, Academic Press, Oxford, 2008, pp. 2940–2949.

doi:10.1016/B978-0-12-409548-9.11152-2

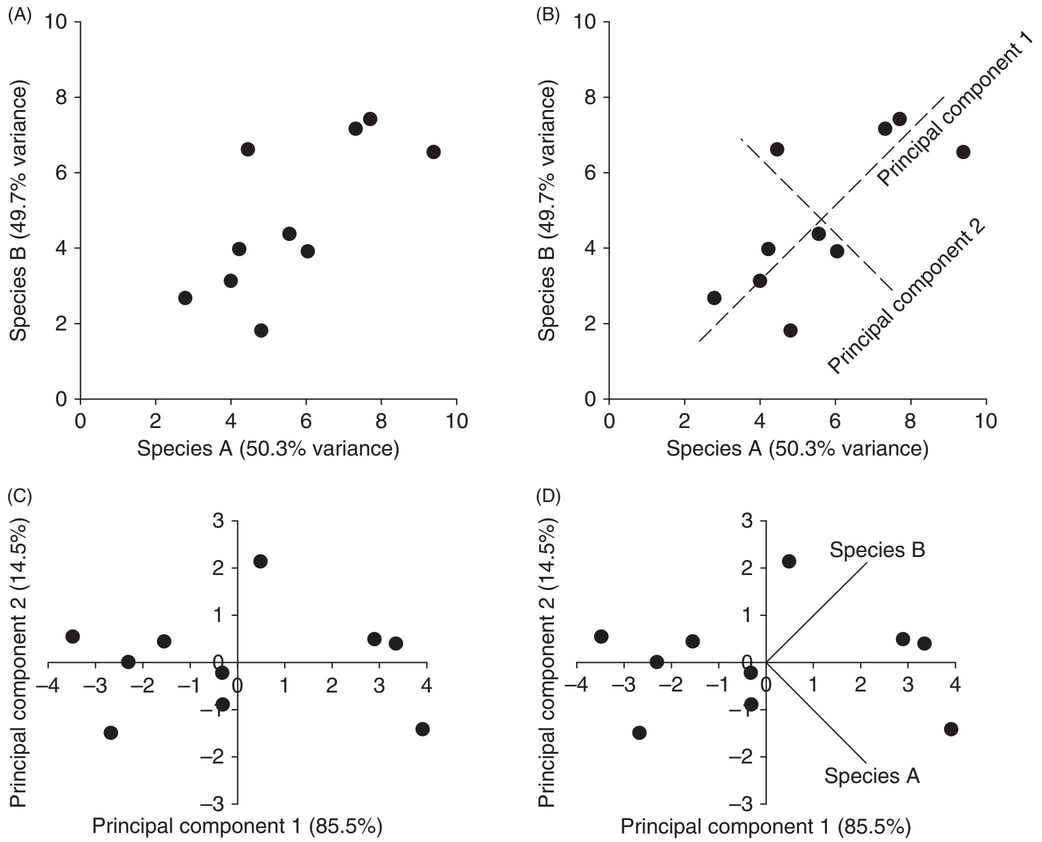

Fig. 1 Deriving principal component (PC) axes of a two-species data set. (A) The original data points can be displayed as a scatterplot of the two species. (B) A new set of axes (PCs) can be derived by placing axes at the center of the data mass, rotating the first axis along the main line of variation in the data set, and rotating the second axis along the next line of variation, conditional on independence with the first axis. (C) The position of the data points on the PCs can be plotted as a reduced space plot. (D) The direction of the original centered species axes can also be projected into the space to generate a biplot.

In this two-species example, it is possible to display all the variation in a two-dimensional (2D) scatterplot of the original variables. However, if the aim was to explain as much variation as possible in only one dimension (i.e., a line) then the PCs have an important advantage over the variable values. In this example the species explain similar amounts of variation because their variance is approximately the same. At best, about 50% of the entire variation in the data could be displayed in one dimension by plotting values for a single species. In contrast, a plot of the first PCs in this example would explain 85.5% of the total variation in the data set. PCA partitions the variation so that the first PCs will explain more variation than any single variable, assuming there is some correlation between variables. The importance of this is apparent when there are more than two species in a sample. More variation can be presented in a plot of two PCs than can be presented by plotting any pair of species. The PCs can also be interpreted in terms of the original species abundances. A projection of the original centered species axes into the reduced space plot can be used to derive a measure of association of that species with the PC axis (Fig. 1D). In this example, samples that lie on the left of the axis had low abundances of both species A and species B, whereas samples that lie to the right of the axis had high numbers of both species.

Calculation of PCA

Mathematically, PCA can be calculated from a mean-centered (i.e., the mean of each variable is subtracted from all values of that variable) data matrix, **Y**. From **Y**, the covariance matrix is calculated using the formula: $1/(n - 1)\mathbf{Y}'\mathbf{Y}$ (i.e., the sums of squares and cross-products matrix, rescaled by the $n - 1$ degrees of freedom). This square, symmetric matrix can be decomposed by an eigenanalysis or "singular value decomposition" into eigenvalues (**L**) and eigenvectors (**U**), which are normalized or scaled to a length of 1. The eigenvalues represent the amount of variation explained by each axis and are usually expressed as a proportion or percentage of the sum of all the eigenvalues. The PCs (**F**) are calculated by projecting the mean-centered data into the ordination space by postmultiplying the centered data by the eigenvectors: $\mathbf{F} = \mathbf{YU}$. An important point to note is that the value of a sample on the PC is a linear combination of the values of the variables in the sample, multiplied by their corresponding eigenvectors. The eigenvectors represent the projection of the old species axes into the new ordination space.

An alternative method of calculating PCA is to use an iterative method such as the two-way weighted summation (TWWS) algorithm. This method starts with a mean-centered data matrix, and arbitrary initial scores on the first PC axis are assigned. The eigenvectors on the initial PC scores are calculated, and then the sample PC scores on these eigenvectors are calculated and rescaled to a length of 1. An estimate of the eigenvalue is obtained from the standard deviation divided by the number of samples, and the procedure is rerun until the eigenvalue does not change with further iterations. Upon convergence, the eigenvectors are scaled to a length of 1, and the PCs are scaled to the eigenvalue. Subsequent axes are calculated in a similar way, except that the PC score estimates at each iteration stage are made uncorrelated with previous ones using the Gram–Schmidt orthogonalization procedure. Both methods yield the same result (within iterative tolerance limits). The eigenanalysis method is easier to program in languages that support matrix operations, whereas the TWWS algorithm can be more efficient for very large data sets because each PC axis is calculated sequentially.

Presentation and Interpretation of PCA Results

A plot of the samples on the PC axes is the primary output of PCA. This reduced space plot displays the relative positions of samples in multivariate space in fewer dimensions. Although a simple scatterplot of samples on the PC axes might provide some useful information on data structure—for example, whether samples are clustered together or occupy a gradient—additional information is usually included on the plot to assist interpretation. This can be illustrated by a PCA of triplefin (Pisces: Tripterygiidae) fish abundance at a range of sites in northeastern New Zealand. The data were collected from sites with different exposure and location characteristics and so graph symbols could be used to reflect these characteristics of the samples (**Fig. 2**A). This contributes to the interpretation of patterns in the samples based on additional information and is an informal exploration that can identify hypotheses about causal processes. There appears to be a gradient in triplefin assemblages across exposure gradients from sheltered to exposed sites, but assemblages on offshore exposed and sheltered mainland sites are distinct from the semiexposed and exposed mainland sites (**Fig. 2**A). Information about the value of individual variables can also be included in the reduced space plot to identify which variables are responsible for the observed patterns. If plot symbols are scaled to reflect the value of a single variable in the analysis we see that the triplefin *Forsterygion varium* was more abundant in semiexposed and exposed mainland sites, but relatively uncommon on sheltered mainland sites and practically absent from exposed offshore sites (**Fig. 2**B). Presenting multiple bubble plots of species abundances is often not a suitable option due to the large number of graphs required; so a more compact and formal presentation of the dependent variables can be generated by plotting the eigenvectors into the reduced space plot (**Fig. 2**C). This presentation is known as a biplot, and follows from the mathematics of PCA in which the samples are projected into the space by premultiplication of the eigenvectors. In this example the importance of *F. varium* in characterizing mainland exposed/semiexposed sites is clear from the length and direction of its eigenvector. *Notoclinops segmentatus* is characteristic of exposed sites, regardless of mainland or offshore status, and *F. lapillum* is characteristic of sheltered sites. Examination of bubble plots and species–frequency histograms at each site support this interpretation.

Numerical Scale: Transformation and Standardization

Like most (if not all) multivariate methods PCA can be sensitive to data transformation and standardization. Because PCA is an eigenanalysis of a variance–covariance matrix, which is dependent on the numerical scale of the data, variables with large absolute values will dominate the data structure. If the data table consisted of variables measured on different scales (e.g., abundance, kilograms, milliliters, pH) then this scale dependency could exert unwanted effects on the analysis. In addition, a quantity such as a volume if measured in milliliters in one sample, for example, would exert more effect than a volume measured in liters in another sample, even if both samples contained the same volume. In the triplefin example, the PCA of the covariance of untransformed triplefin abundance data was dominated by the numerically dominant species, *N. segmentatus* and *F. varium* (**Fig. 3**A) and largely insensitive to less-common species. This may be a problem if the intent of the analysis is to retain information on less-abundant species or, more generally, variables with small but biologically important values.

There are two ways to reduce the effect of variables with large absolute values, and increase the effect or weight of variables with small values. First, the data can be centered and transformed to standard deviation units. This process is called standardization, and is implicit in many software implementations of PCA. If data are standardized to unit variance prior to the analysis, then PCA becomes an eigenanalysis of a correlation rather than a covariance matrix. The effect of this standardization is to give all variables equal weight in the analysis and is commonly used and recommended in ecological applications. In contrast with the covariance matrix PCA, an analysis of the correlation matrix of untransformed triplefin abundance data yields an ordination in which both common and uncommon species are important in defining the ordination space (**Fig. 3**B). Second, data can be numerically transformed using functions such as a square-root, fourth-root, and log transform. Transformations are often used to improve linearity between variables or to reduce the effect of variables with large values in the analysis. However different transformations will alter the importance of different variables in defining the ordination space, and hence may alter the ecological interpretation. In general, increasing levels of transformation (e.g., $x^{0.5}$, $x^{0.25}$, $\log(x)$) will progressively shift analytical emphasis from abundance to compositional aspects of the data. For example, less-abundant triplefin species assume more importance in a covariance matrix PCA with a fourth-root transform ($x^{0.25}$), although not to the same extent as a PCA on the correlation matrix (**Fig. 3**C).

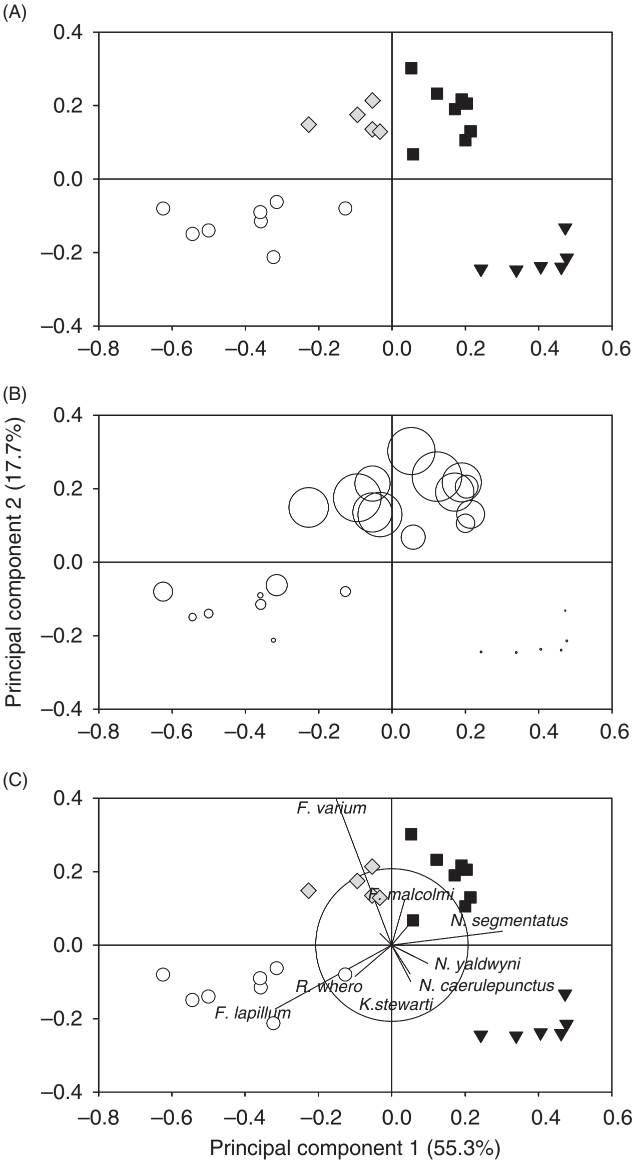

Fig. 2 Reduced space plots of a principal components analysis (PCA) of triplefin fish (Family: Tripterygiidae) abundances. The PCA was calculated using the covariance matrix of the square root of the proportional species abundance in a sample. Percent variance explained is derived from the eigenvalues. Note the equal scaling of the *x* and *y* axes—this ensures the ordination space is not distorted in the plot. (A) Information about wave exposure and location of sites is added by changing *plot* symbols: *circle* indicates sheltered mainland; diamond indicates semiexposed mainland; square indicates exposed mainland; downward triangle indicates exposed offshore. (B) Symbol size can be scaled in proportion to the value of the variables, in this case triplefin species abundance. *Forsterygion varium* is characteristic of exposed and semiexposed mainland sites, uncommon in sheltered sites, and practically absent from offshore exposed sites. (C) A joint presentation of the reduced space and eigenvectors forms a biplot. Distances between sites approximate the Euclidean distance of the transformed data and the eigenvectors are the projection of the original species axes into the space. Eigenvectors have been rescaled to half of their original value for clarity in the plot. The circle is the equilibrium contribution of the eigenvectors. Species outside this circle are influential in defining the ordination space. In this example, *Notoclinops segmentatus* is characteristic of exposed offshore and exposed mainland sites, *F. varium* is characteristic of exposed and semiexposed mainland sites, and *F. lapillum* is characteristic of sheltered sites.

Biplot Scaling

In addition to data-scaling considerations, biplots can also be scaled differently, which may in turn alter their interpretation. Two common biplot scalings are used to display PCA. A superimposed plot of the PCs (**F**) and the normalized eigenvectors (**U**) is known as a distance or Euclidean biplot. In this biplot, the PC scores are scaled so that their sums of squares equals the eigenvalue (λ) for a given axis, the positions of samples in ordination space approximate their distance in Euclidean space, and the eigenvectors represent the projection of the dependent variable axes into the ordination space (Fig. 4A). The length of the

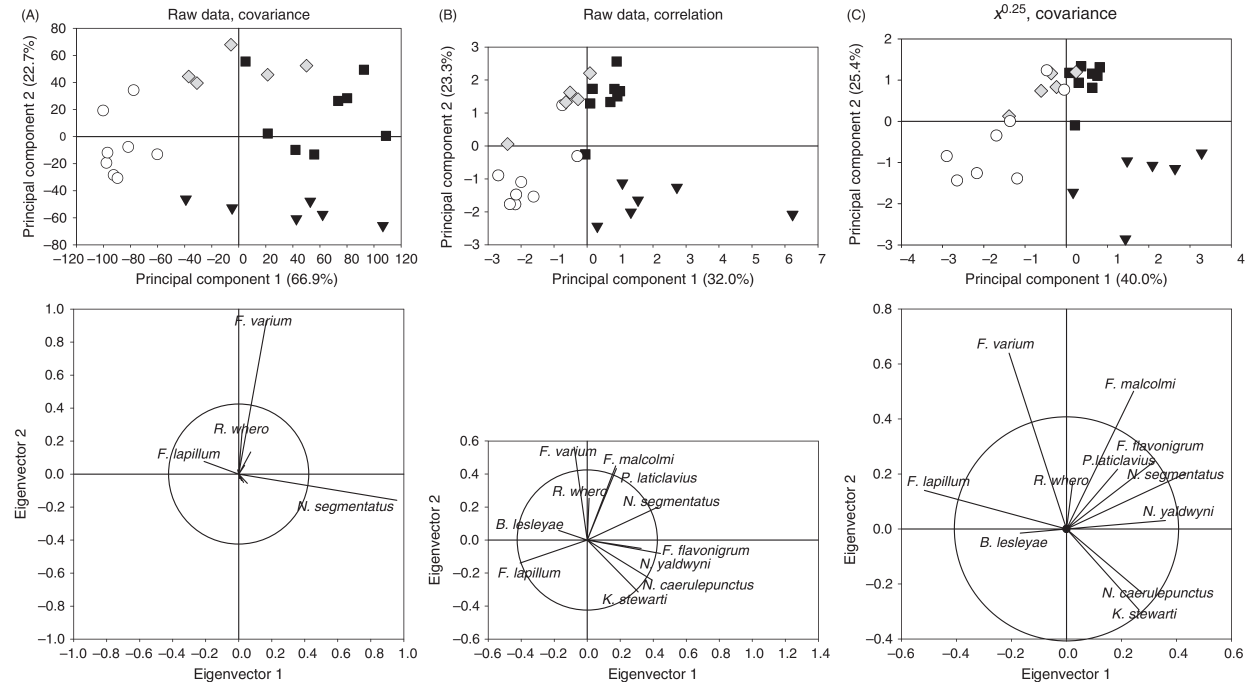

Fig. 3 Effects of data transformation and standardization on PCA of triplefin assemblages. Top row of graphs are the reduced space p corresponding eigenvector plots with equilibrium contribution circles. (A) Untransformed covariance matrix analysis is strongly influenced species, *Notoclinops segmentatus* and *Forsterygion varium*. (B) Untransformed correlation matrix analysis reduces the influence of the nu increases the weight of the rarer species. (C) Covariance analysis of fourth-root transformed data also reduces the influence of the numerically dominant species, and increases the weight of the rarer species.

eigenvector indicates the contribution of the variable to the space—an eigenvector approaching a length of 1 indicates that the variable contributes strongly to defining the ordination space. In addition, the approximate values of the dependent variables in each sample can be reconstructed by projection at right angles of the sample values onto the eigenvector axes. Another common scaling is to scale the eigenvectors to equal their standard deviation ($\mathbf{UL}^{0.5}$) and standardize the PC scores to unit sum of squares ($\mathbf{G} = \mathbf{FL}^{-0.5}$). This is the covariance (or correlation if the data have been standardized) biplot. Unlike the Euclidean biplot, the distances between samples in the reduced space do not approximate their Euclidean distances—they have been standardized by a variance measure. In the covariance biplot the eigenvectors are rescaled to equal the square root of the eigenvalue (cf. normalized in the Euclidean biplot). This projection effectively rescales the eigenvectors to standard deviation units, and has some interesting properties. The length of the vector approximates the standard deviation of the variable, not its contribution to the ordination space. The angle between dependent variable vectors provides a measure of their covariance: covariance $\approx \cos\theta$, where θ is the angle between dependent variable vectors (**Fig. 4**B). If the PCA has been carried out on standardized data, then this angle will represent the correlation. These angles will only provide a good covariance or correlation estimate if the number of samples is large, the vectors are well represented in the analysis, and the variation explained by the axes is large. Both biplots have the property that the centered data can be reconstructed from the sample scores and the variable vectors: $\mathbf{FU}' = \mathbf{G}(\mathrm{UL}^{0.5})' = \mathbf{Y}$.

The correlations between the original variables and the values of the samples on the Euclidean PC axes may also be used to project dependent variables into a reduced space plot. These values are often termed factor loadings or factor patterns, but their usage should be treated with caution. If the PCA has been carried out on standardized data (i.e., the correlation matrix) then the covariance biplot eigenvector scaling ($\mathbf{UL}^{0.5}$) is equal to the factor unimportant. Conversely variables with large variance might appear to be strongly associated with an axis, when in fact they contribute nothing to its construction. Factor loadings describe how important an axis is to a variable, not how important a variable is to an axis (**Fig. 4**C). The rationale for this approach comes from a related method—"factor analysis"—which considers measured variables as a function of a hypothesized causal process represented by the PCs, rather than the variables defining a reduced ordination space.

Adequacy of the PCA Solution

PCA generates as many PC axes as there are variables. The axes with the larger eigenvalues hopefully describe trends in the data, whereas the axes with smaller eigenvalues simply represent random variation. There are no authoritative rules for deciding how many PCs are interpretable. Initial recommendations were based on the cumulative percentage of variation explained by the eigenvalues. However ecological data sets differ in their correlation structure, so defining an arbitrary level of variation (e.g., 75%–95%) is not a biologically

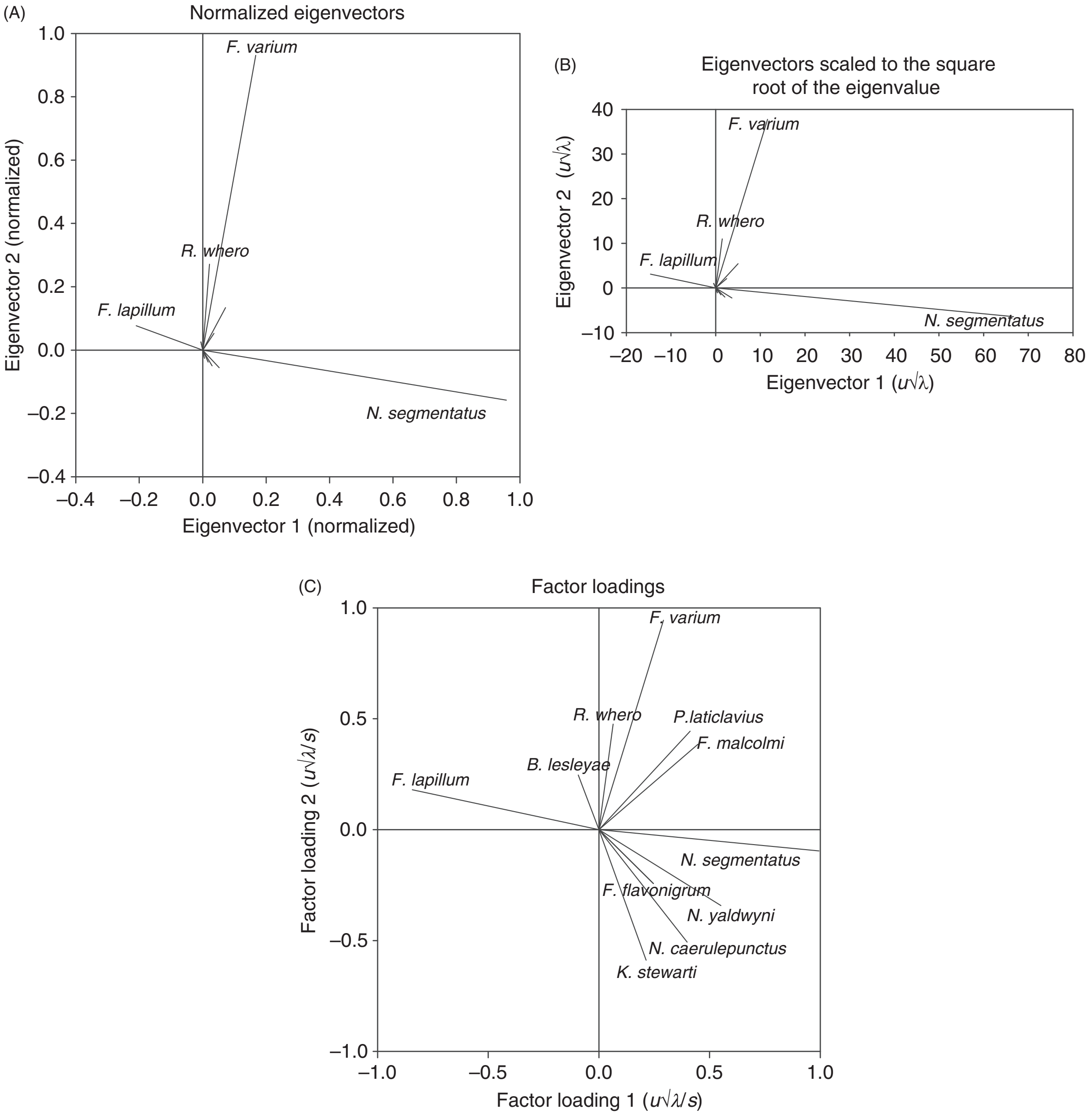

Fig. 4 Effects of eigenvector scaling on a PCA of the raw covariance matrix. (A) Normalized eigenvectors associated with distance biplots are scaled to a length of 1. (B) Covariance biplots scale the eigenvectors so that the length of the vector approximates their standard deviation, and the cosine of the angle between variables approximates their covariance. (C) Factor loadings rescale the covariance biplot scaling by the standard deviation of the variable. This gives an estimate of the importance of the axis in explaining the variance of the variable—it does not represent the importance of the variable in explaining the axis. Note that if the PCA was carried on the correlation matrix the factor loadings would be equal to the covariance biplot because each variable's standard deviation is made equal to 1. In this example the ordination space is defined primarily by two species—*Notoclinops segmentatus* and *Forsterygion varium*.

relevant criterion and its use has been widely disregarded. The plot of the eigenvalues against the axis order (a scree plot) can guide the identification of "important" PCs (**Fig. 5**). Scree plots can be used to visually identify breaks between PCs that potentially explain trends, and those that represent statistical noise. Typically the trivial eigenvectors on the right of the scree plot will form a linear series, and major magnitude changes on the left may represent trends. The efficacy of this visual determination of breaks is dependent on the underlying data structure. The Kaiser–Guttman criterion requires that eigenvalues that exceed the average expectation should be retained. In a PCA of the correlation matrix, all variables have equal variances and hence the sum of eigenvalues is equal to the number of variables. Consequently the Kaiser–Guttman criterion on a PCA on the correlation matrix is that eigenvalues greater than 1 should be interpreted. While intuitively the Kaiser–Guttman criterion seems reasonable, there is sampling variability in ecological data sets and so the average expectation may not be a suitable null model. An alternative approach is to use the "broken stick model" to identify the null distribution of eigenvalues, if there was no structure in the data (**Fig. 5**). Expected eigenvalues for a given axis under the broken stick

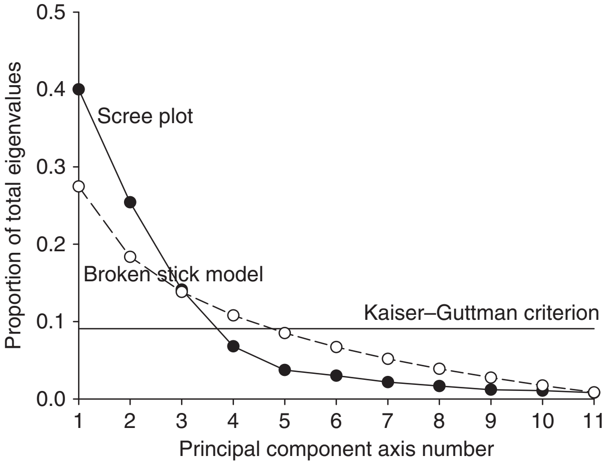

Fig. 5 Scree plot of the proportion of variation explained by each successive principal component (PC) of an analysis on the covariance matrix of $x^{0.25}$-transformed triplefin data (*filled circle*). The Kaiser–Guttman criterion (average proportion explained by eigenvalues) suggests that three axes should be retained for analysis. The intersection of the broken stick model (*open circle*) with the scree plot suggests that two axes should be retained for analysis. The inflection point of the scree plot also suggests that three axes should be retained. While it is clear that at least two axes should be retained, most users of PCA would also examine the third axis to ascertain if it described some ecologically interpretable pattern.

model can be calculated as $b_k = 1/p\sum_{i=k}^{p} 1/i$, where p is the number of variables, and b_k is the expected proportional eigenvalue for the kth component. Computationally intensive randomization tests such as bootstrap confidence intervals can also be used to identify which eigenvalues are nontrivial. Formal statistical tests such as Bartlett's test of sphericity, and both Bartlett's and Lawley's test of homogeneity of the correlation have generally fared poorly in simulations. A general recommendation would be to use the broken stick model to identify nontrivial PC axes if bootstrapping was not available.

Similar issues exist for interpreting which eigenvectors are important in a PCA. When eigenvectors are normalized, their total length is 1. Consequently if an eigenvector on a particular PC axis has a value close to 1 then that variable is well represented on that axis and less well represented on other axes. However if a variable is not strongly associated with any PC, the eigenvectors for that variable should be equal across axes. The expected eigenvector for a variable that is not associated with any PCs is known as the equilibrium contribution, p and is given by $\sqrt{d/p}$, where d is the number of dimensions of interest, and p is the number of variables. Eigenvectors with values larger than the equilibrium contribution for a single axis can be considered to be associated with that axis. Similarly eigenvectors with values larger than the equilibrium contribution for two axes could be considered to be associated with forming a 2D space. If the eigenvectors are not normalized then the equilibrium contribution must be calculated separately p for each variable and is given by $s_j \sqrt{d/p}$, where s_j is the standard deviation of the jth variable. If the eigenvectors are normalized, the equilibrium contribution can be presented on a graph as a circle (e.g., **Fig. 2**C).

Assumptions, Limitations, and Other Considerations

PCA was originally developed to describe patterns in multivariate normal (MVN) data. However, deviations from MVN are generally not as critical to the success of unless there are fewer samples in the data set than there are variables. Most software packages will still calculate the analysis under this condition, with the restriction that the number of PCs will equal the number of samples rather than the number of variables. In general, the first PC axes will still be interpretable but minor axes may not be because of overfitting of the model. This is analogous to multiple regression analysis in which fitting too many variables with too few data points will yield a degenerate solution. Several guidelines have been suggested for establishing appropriate sample sizes to generate robust PCA solutions. Some researchers have suggested that studies should aim to achieve absolute sample sizes ranging from 50 (very poor) through 200 (fair), 300 (good) up to 1000 or (excellent). Others have suggested that the ratio of samples to variables is of more importance, with minimum values of 5:1–10:1. It is important to note that these recommendations have generally been recommended by users of "factor analysis," in which robust covariance estimates are key to identifying stable analytical solutions. In addition, many of these suggestions stem from the social sciences in which raw data, such as questionnaire responses, are often "indicators" of variables, rather than direct measures of the variable itself. In most ecological applications these sample sizes are unrealistic and the focus of the analysis is on description of an assemblage, rather than the "factor analysis" objective of recovering underlying causal factors. For most purposes, a rule of thumb for PCA would be to ensure that there are more replicates than variables. In general, the greater the ratio of replicates to variables, the better. PCA is a data exploration and display tool—not a hypothesis-testing method subject to strict distributional assumptions—so it should yield useful insight into data set structure even if the replication is not as great as desired. Robustness of the PCA solution can always be evaluated by bootstrapping, as described above for eigenvalues and eigenvectors.

It is important to be aware of software idiosyncrasies when calculating PCA. Many software implementations calculate PCA on the correlation matrix by default. In addition, PCA is mathematically related to "factor analysis" (FA), a method used widely in the social sciences. The main conceptual difference between PCA and FA is that PCA considers the ordination axes as a product of the variables—they are a linear combination of the eigenvectors. FA considers the variables as a product of the axes themselves. In this interpretation, the variable values are "caused" by the hypothetical ordination axes rather than the axes simply reflecting patterns in the data. The mathematical similarity of the two approaches has led to many software packages combining PCA routines into FA routines. Implementations that incorporate FA and PCA may, by default, yield a covariance biplot of the correlation matrix, with the sample scores scaled to 1 and the eigenvectors scaled to their standard deviation (the factor loading or pattern). FA uses the correlation matrix by default, so as outlined above an analysis of the covariance matrix may substantially change the interpretation of the resulting biplot. FA software also offers the option of axis rotations. Rotations are intended to align axes so that factor loadings are maximized, that is, to make variables associated with single axes. These procedures should not be used for PCA, unless the intent is to use PCA in an exploratory FA and not as a descriptor of ecological data. As with all ecological data analysis, it is important to ensure the correct technique is being employed.

PCA is a very flexible procedure. In its basic form, it is an eigenanalysis of a covariance or correlation matrix. Consequently, it is possible to calculate PCA on a nonparametric correlation matrix such as Spearman's rank correlation. This approach can be useful to deal with nonlinearity of variables. It also follows that PCA can also be calculated on a partial correlation or covariance matrix. A partial correlation coefficient is one that has been statistically adjusted for another variable, essentially correcting for a covariate. Ecological applications of partial PCA are rare, but morphometric studies frequently use partial PCA to assess relationships between morphometric variables after correcting for size of the organism. The ecological application is clear. Dominant variables such as wave exposure, moisture gradients, etc., could be effectively removed from the analysis prior to PCA to yield an ordination that statistically 'corrects' for dominant variables.

Although PCA was developed as an ordination method to summarize patterns in multivariate data, it also has a range of other uses. PCA can be used in multiple regression to detect collinearity of predictor variables, and as a variable-reduction tool. For example, collinear variables could be replaced in a multiple regression with their first PC. This reduction in number of regression coefficients will increase the power and stability of a multiple regression by reducing the number of variables and improving independence of the coefficients. PCs can themselves be used in data presentations and analyses. For example, a contour plot of PCs of spatially structured data could provide information on a range of variables in a single graph.

See also: Ecological Data Analysis and Modelling: Visualization as a Tool for Ecological Analysis; Statistical Inference; Model Types: Overview. Evolutionary Ecology: Metacommunities. General Ecology: Numerical Ecology

Further Reading

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics 2 (4), 433–459.

Abdi, H., Williams, L.J., Valentin, D., 2013. Multiple factor analysis: Principal component analysis for multitable and multiblock data sets. Wiley Interdisciplinary Reviews: Computational Statistics 5 (2), 149–179.

Budaev, S.V., 2010. Using principal components and factor analysis in animal behaviour research: Caveats and guidelines. Ethology 116 (5), 472–480.

Candès, E.J., Li, X., Ma, Y., Wright, J., 2011. Robust principal component analysis? Journal of the ACM (JACM) 58 (3), 11.

Demšar, U., Harris, P., Brunsdon, C., Fotheringham, A.S., McLoone, S., 2013. Principal component analysis on spatial data: An overview. Annals of the Association of American Geographers 103 (1), 106–128.

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

Jolliffe, I.T., 1986. Principal components analysis. New York: Springer.

Legendre, P., Gallagher, E.D., Ecologically meaningful transformations for ordination of species data. Oecologia 129, 271–280.

Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A 374 (2065), p. 20150202.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38 (8), 904.

Shlens, J., 2014. A tutorial on principal component analysis *arXiv preprint arXiv:1404.1100*.

ter Braak, C.J.F., ter Braak, C.J.F., van Tongeren, O.F.R., 1995. Ordination. In: Jongman, R.H.G. (Ed.), Data analysis in community and landscape ecology. Cambridge: Cambridge University Press, pp. 91–173. ISBN: 0-521-47574-0.

Witten, D.M., Tibshirani, R., Hastie, T., 2009. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. Biostatistics 10 (3), 515–534.

Relevant Websites

http://ordination.okstate.edu/—Ordination methods for ecologist.

https://commons.wikimedia.org/wiki/Category:Principal_component_analysis—Wikimedia commons.