Notes

INDICES FOR DETECTING DIFFERENCES IN SPECIES COMPOSITION: SOME SIMPLIFICATIONS OF RDA AND CCA

Michael W. Palmer,[1,3] Daniel J. McGlinn,[1] Lars Westerberg,[2] and Per Milberg[2]

[1]*Department of Botany, Oklahoma State University, Stillwater, Oklahoma 74078 USA*
[2]*IFM Biology, Division of Ecology, Linköping University, SE-581 83 Linköping, Sweden*

Abstract. We provide algebraic simplifications for the redundancy analysis (RDA) eigenvalue and the canonical correspondence analysis (CCA) eigenvalue in the special case of permanent plots sampled twice. The indices for RDA and CCA are interrelated and are intuitively interpretable. These simplifications also apply to simple split-plot designs and to a balanced design with two independent samples.

Key words: canonical correspondence analysis; monitoring; ordination eigenvalues; permanent plots; redundancy analysis; split-plot design.

INTRODUCTION

While much ecological monitoring focuses on univariate analyses of single species or variables (Niemi and McDonald 2004, Plattner et al. 2004, Weber et al. 2004, Gray and Azuma 2005), there is widespread interest in using species composition data. Ordination methods are useful for visualizing and evaluating compositional patterns. However, permutation tests of ordination eigenvalues may lack intuitive appeal, hindering their widespread use. Here, we provide algebraic simplifications of the eigenvalues for redundancy analysis (RDA) and canonical correspondence analysis (CCA), in the special case where permanent plots are examined at two points in time (or equivalently, a split-plot design with two treatments; see Plate 1).

RDA eigenvalue

RDA is a direct gradient analysis extension of principal components analysis (PCA; Lepš and Šmilauer 2003). Like PCA, RDA identifies orthogonal axes that maximally "explain" variation in species composition (Legendre and Legendre 1998). Unlike PCA, the axes are constrained to be linear combinations of explanatory variables, and the eigenvalue associated with each axis is a measure of this variation. Thus, RDA is a form of multivariate regression (Jongman et al. 1995).

For simplicity, we will assume we have fixed plots sampled twice, although we realize that our derivations apply to any paired-sample test. Therefore, we refer to

pairs of samples as "plots" and the variable as "time." We have one explanatory variable (time) that can be dummy coded as 1s and 0s. A direct gradient analysis with one explanatory variable has only one constrained eigenvalue (ter Braak 1986). The RDA eigenvalue simplifies (Appendix A) to

$$\lambda_{\mathrm{RDA}} = \frac{\sum_{j}^{r} \left(\sum_{k}^{n/2} d_{kj} \right)^2}{n(n-1)} \tag{1}$$

where d_{kj} is the difference in abundance of species j in plot k between the two times. Note that although we are summing the differences between pairs, the associative property of addition makes it equivalent to the differences of the sums. Thus, the RDA eigenvalue is also equivalent to the test for two balanced, independent samples.

This index has intuitive appeal in the case of paired samples or plots sampled twice: the quantity in the parentheses is a direct measure of the degree to which each species has changed. In such cases, significance (with the H_0 being no change through time) is assessed using a permutation test (ter Braak and Šmilauer 2002, Lepš and Šmilauer 2003) conditioned on the plots. That is, the two times are randomly swapped for each plot. The lowest significance level attainable by this permutation scheme, not accounting for identical but inverse combinations, is $2/2^{n/2}$, where $n/2$ is the number of plots. McArdle and Anderson (2001) appear to prove that such an analysis is equivalent to an MRPP (multiresponse permutation procedure; McCune and Grace 2002) based on Euclidean distances. As part of a

PLATE 1. Witness photos for a 0.01-ha permanent plot at the Tallgrass Prairie Preserve, Osage County, Oklahoma (USA). The photo on the left was taken in 2005, the photo on the right in 2006. There are a total of 20 such plots at the preserve with annual records spanning more than a decade. Photo credits: M. W. Palmer.

simulation study on power analysis of multivariate techniques, we have independently proven this (results not shown).

For independent samples, there is no condition to the permutation; samples are randomly reassigned to the two groups.

In our analysis, the $n/2$ plot identities can be considered confounding variables. Partial RDA (pRDA) allows the factoring out of such covariables. However, in our case the covariables are completely orthogonal to the sole explanatory variable (since all plots are sampled once at each time). Thus, the pRDA eigenvalue will be identical to the RDA eigenvalue, and Eq. 1 holds.

CCA eigenvalue

Canonical correspondence analysis (CCA) is the direct gradient analysis extension of correspondence analysis (ter Braak 1986, Palmer 1993, 1998). It is numerically allied to RDA except that species abundances are rescaled to a $\overline{Q}$ matrix, and weights derived from the species abundance matrix are applied to the explanatory variables (ter Braak and Verdonschot 1995, Legendre and Legendre 1998). In our case, the CCA eigenvalue simplifies (Appendix B) to

$$\lambda_{CCA} = \frac{1}{y_{++}p(1-p)} \sum_{j}^{r} \frac{(h_{+j} - py_{+j})^2}{y_{+j}} \qquad (2)$$

where h_{+j} is the abundance of species j summed for all plots at the second sampling (hereafter, "time 2"), y_{+j} is the total abundance of species j for both times combined, y_{++} is the grand total for all species at both times combined, and p is the proportion of the total abundance encountered in time 2 (that is, $p = h_{++}/y_{++}$).

In essence, we still have a sum of differences as in Eq. 1, but now it is expressed in proportions. Since CCA eigenvalues are also correlation coefficients (ter Braak

1986), it is not surprising we can recognize variance due to time in the numerator of the right half of Eq. 2 and total variance in the denominator of the left half: $p(1 - p)$ is the binomial variance. Thus, the eigenvalue is a ratio of explained variance to total variance and can therefore be interpretable as "proportion of variation explained" in a binary setting. Also note that we can rewrite Eq. 2 as

$$\lambda_{CCA} = \frac{1}{y_{++}(1-p)} \sum_{j}^{r} \frac{(h_{+j} - py_{+j})^2}{py_{+j}}. \qquad (3)$$

The summation part of the equation is, consistent with the chi-square basis of correspondence analysis (Legendre and Legendre 1998), a measure of chi-square: $(O - E)^2/E$, where O is the observed abundance at time 2, and E is the expected abundance at time 2 if all species change identically through time. We can test for a time effect using the same permutation scheme previously described for RDA.

We were unable to find a solution for partial CCA because unlike with RDA, time and plots are not necessarily orthogonal. This is because of the weightings of the explanatory variables by the species data. However, if we relativize the abundance data (e.g., to unity or to 100%), then the eigenvalue for partial CCA (pCCA) will equal that of CCA. Furthermore, since in this case $p = 0.5$, it allows further simplification (Appendix C) to

$$\lambda_{CCA-REL} = \frac{1}{y_{++}} \sum_{j}^{r} \frac{\left(\sum_{k}^{n/2} d_{kj}\right)^2}{y_{+j}}. \qquad (4)$$

If the samples are relativized to unity, then y_{++} simplifies to n and the CCA eigenvalue becomes similar to the RDA eigenvalue.

CONCLUSIONS

In conclusion, eigenvalues for commonly used direct gradient analysis techniques can be simplified into intuitively satisfying equations that do not require complex eigenanalyses in the simple cases of plots monitored twice, split-plot designs, and balanced independent samples.

LITERATURE CITED

Gray, A. N., and D. L. Azuma. 2005. Repeatability and implementation of a forest vegetation indicator. Ecological Indicators 5:57–71.

Jongman, R. H. G., C. J. F. ter Braak, and O. F. R. van Tongeren. 1995. Data analysis in community and landscape ecology. Cambridge University Press, Cambridge, UK.

Legendre, P., and L. Legendre. 1998. Numerical ecology. Second English edition. Elsevier, Amsterdam, The Netherlands.

Lepš, J., and P. Šmilauer. 2003. Multivariate analysis of ecological data using Canoco. Cambridge University Press, Cambridge, UK.

McArdle, B. H., and M. J. Anderson. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology 82:290–297.

McCune, B., and J. B. Grace. 2002. Analysis of ecological communities. MjM Software Design, Gleneden Beach, Oregon, USA.

Niemi, G. J., and M. E. McDonald. 2004. Application of ecological indicators. Annual Review of Ecology, Evolution, and Systematics 35:89–111.

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

Palmer, M. W. 1998. Canonical correspondence analysis: some advice and guidelines. Pages 171–187 in E. Kazmierczak, A. Nienartowicz, A. Piernik, and J. Wilkon-Michalska, editors. Computer methods in investigation of the structure and functioning of the vegetation cover. Nicholas Copernicus University, Torun, Poland.

ter Braak, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67:1167–1179.

ter Braak, C. J. F., and P. Šmilauer. 2002. Canoco reference manual and CanoDraw for Windows user's guide: software for canonical community ordination (version 4.5). Microcomputer Power, Ithaca, New York, USA.

ter Braak, C. J. F., and P. F. M. Verdonschot. 1995. Canonical correspondence analysis and related multivariate methods in aquatic ecology. Aquatic Sciences 57:255–289.

Weber, D., U. Hintermann, and A. Zangger. 2004. Scale and trends in species richness: considerations for monitoring biological diversity for political purposes. Global Ecology and Biogeography 13:97–104.

APPENDIX A

Simplification of RDA eigenvalue (*Ecological Archives* E089-105-A1).

APPENDIX B

Simplification of CCA eigenvalue (*Ecological Archives* E089-105-A2).

APPENDIX C

Simplification of CCA eigenvalue for relativized species data (*Ecological Archives* E089-105-A3).