Progress in Physical Geography 30, 3 (2006) pp. 399–408

Progress reports

Numerical classification and ordination methods in biogeography

Martin Kent

School of Geography, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA, UK

I Introduction

Methods of multivariate analysis have always played an important role in detecting and summarizing patterns in biogeographic data. Originating within plant ecology in the 1950s and 60s, methods have developed steadily over the past 40–50 years and are now widely applied across the full spectrum of organism groups and in many different research contexts. A succession of texts has summarized the evolution and progressive refinement of methods (Kershaw, 1973; Greig-Smith, 1983; Kershaw and Looney, 1985; Gauch, 1982; Kent and Coker, 1992; Jongman *et al.*, 1995; Waite, 2000; Quinn and Keough, 2002). Such methods have been used primarily in an inductive manner, to synthesize large quantities of species and environmental data and to look for patterns and order within them.

Although a wide range of techniques have been applied, at the core have been the groups of methods that come under the headings of, first, numerical classification, also widely known as cluster analysis, and, second, ordination, also known under the title of methods of gradient analysis. The trends in the early application of these techniques were summarized by Kent and Ballard (1988), who concluded that, in both areas, a succession of changes of preferred and 'best' techniques could be identified, linked closely to the evolution of new or more refined methods. However, often one (usually the 'latest' technique) would be recommended for a number of years, only to be succeeded by some new and more recent development, which claimed to be an improvement. Subsequent application and testing of methods on a wide range of data sets and in differing contexts, nevertheless, often demonstrated sometimes quite serious problems and limitations.

The aim of this review is, first, to summarize the more recent changing patterns in the application of numerical classification (cluster analysis) and ordination techniques and related techniques in biogeography. Particular branches of ecology and biogeography appear to prefer certain methods. However, this preference is often based on a 'tradition' of use within that branch of the subject, rather than being informed by wider review, critical comment and new developments. Second, this review seeks to synthesize and comment on the various criticisms that have been made regarding almost all the methods in use at the present time. Finally, some pointers to the future are highlighted in terms of very recent developments and new advances in techniques.

 10.1191/0309133306pp489pr

II Changing patterns and problems in the application of numerical classification (cluster analysis)

Numerical classification (cluster analysis) methods are used to define species assemblages and plant community types, as well as being more widely used in taxonomic and phylogenetic research. By far the most widely used methods are those that come under the general heading of 'similarity analysis' – which is really a suite of methods with a range of choices of (dis)similarity coefficient and sorting strategy. Within this range of choices, two methods – the minimum variance (Ward's) method (Ward, 1963) and unweighted pairs group average (UWPGA) technique (Williams *et al*., 1966) – have been applied the most. However, Ward's method has limitations when applied to species data because of the implicit use of the squared Euclidean distance coefficient, which is unsuitable for the often sparse data of most species data matrices (Faith *et al*., 1987; Clarke, 1993; Legendre and Legendre, 1998; Everitt *et al*., 2001; McArdle and Anderson, 2001); although, when relativized, better results may be obtained (McCune and Grace, 2002). Ward's method nevertheless has enjoyed widespread use in the classification of environmental, rather than species, data sets.

A major problem of both of the above methods of cluster analysis is that results are influenced by quite arbitrary choices that are required at various stages of the analysis ([dis]similarity coefficient; sorting strategy, final number of groups) (van Groenewoud, 1992; ter Braak *et al*., 2003).

Given the problems of Ward's method in relation to species data, the main alternative apart from UWPGA has been two-way indicator species analysis (TWINSPAN) (Hill, 1979a). Despite its extremely widespread use, especially in vegetation science, TWINSPAN has been particularly severely criticized by some authors. Foremost among these are McCune and Grace (2002: 97), who state: 'Ecologists should not use TWINSPAN, except in the very special case where a two-way ordered table ... is needed for a data set with a simple, one-dimensional underlying structure. Presence of a single strong underlying gradient can be manifest from prior knowledge or it can be detected with non-metric multi-dimensional scaling ... TWINSPAN cannot represent complex datasets in its one-dimensional framework (van Groenewoud, 1992; Belbin and McDonald, 1993)'. McCune and Grace (2002) also criticize the complexity of the method, the artificial nature of the 'pseudospecies' employed and the fact that, apart from an attempt in Kent and Coker (1992), there is no detailed published description of the method.

Despite these serious criticisms, TWINSPAN is nevertheless still very widely used by biogeographers, plant ecologists, limnologists, palynologists and animal ecologists partly because of the widespread availability of the program (it has just been re-released in a Windows version – Hill and Šmilauer, 2005) and also the attractiveness of its final two-way sorted table. A search on the Web of Science database for the period January 2003 to September 2005 reveals 101 published articles in ecology and biogeography that have recently made use of it. Part of the reason for this is that no clearly improved and widely available methods of numerical classification have emerged in the past two decades, despite various interesting developments (see below).

III Changing patterns and problems in the application of ordination and gradient analysis methods

A wider range of methods are in general use under the heading of ordination and gradient analysis than is the case with numerical classification. Nevertheless, for indirect ordination, most researchers now use either detrended correspondence analysis (DCA) (Hill, 1979b) or non-metric multidimensional scaling (NMS) (Shepard, 1962; Kruskal, 1964a, 1964b; Cox and Cox, 1996). These are recommended for use where environmental gradients are long with high species turnover along the gradient and species have unimodal response curves (Lepš and Šmilauer, 2003; ter Braak and Šmilauer, 2003). Where gradients are short and linearly related to abundances, then principal components analysis (PCA) (Gittins, 1969) is still recommended (Lepš and Šmilauer, 2003; Wagner, 2004).

Analysis of papers using indirect ordination from the Web of Science database for the period 2003–2005 shows a fairly equal split in the application of DCA and NMS, with DCA being used more widely in terrestrial plant ecology (the situation for which it was originally formulated) and NMS being more popular in the analysis of non-vegetation data, bird and particularly marine and freshwater data, in the case of marine data because of the very strong endorsement by Clarke (1993) and its incorporation into the *Primer*™ package (Clarke and Gorley, 2001). Despite its widespread use on ecological and biogeographical data, DCA has been severely criticized by some authors, notably Minchin (1987), who compared DCA and NMS using simulated data in relation to a variety of assumptions concerning the underlying distribution of species along environmental gradients and clearly indicated his preference for NMS. The whole basis of detrending in DCA to reduce the 'arch' and 'axis compression' effects of basic correspondence analysis has been criticized by various authors, including Beals (1984), Jackson and Somers (1991), Legendre and Legendre (1998) and Minchin (1987), all of whose criticisms are very effectively summarized in McCune and Grace (2002), who conclude by stating: 'DCA unnecessarily imposes assumptions about the distributions of sample units and species in environmental space. Other methods, especially non-metric multidimensional scaling, perform as well or better without making those assumptions. There is no need to use DCA' (McCune and Grace, 2002: 160).

Also research in the mid-1990s showed that quite minor changes in the input data in DCA could result in major changes in the resulting ordination (Tausch *et al.*, 1995), which Oksanen and Minchin (1997) demonstrated was due to the lax criterion for the stability of the correspondence analysis solutions. This fault affected both earlier versions of the DCA computer program and also the TWINSPAN program discussed above, since the latter uses a single axis correspondence analysis as part of the method. However, this problem has been corrected in all recent versions of the DCA and TWINSPAN programs.

In addition, two of the most extensively used ordination methods at the present time are those contained within the CANOCO™ software package (Lepš and Šmilauer, 2003; ter Braak and Šmilauer, 2003) that are described as techniques of direct ordination – these are canonical correspondence analysis (CCA) (ter Braak, 1986; 1987; 1988a, 1988b), used on long environmental gradients with high species turnover and hence unimodal species response curves, and redundancy analysis (RDA) applied on shorter linear gradients (Wagner, 2004).

CCA is based on a standard correspondence analysis (CA) of species data with sample ordination axis scores (dependent variable) constrained by a multiple linear regression with a second matrix of environmental or other related (independent) variables. Once again, McCune and Grace (2002) express serious reservations about its use:

> Multivariate analysis is a way of getting messages from a high-dimensional world that we cannot see directly. Using CCA is like getting those messages through a narrow mail slot in a door. The edges of the slot are defined by the measured environmental variables. Messages that do not fit the slot are either deformed to push them through or just left outside . . . CCA is currently one of the most popular ordination techniques in community ecology. It is, however, one of the most dangerous in the hands of people who do not take the time to understand this relatively complex method. (McCune and Grace, 2002: 164)

Their criticisms centre on three features of the method. First, the need to understand all the limitations and assumptions of standard multiple regression, especially those relating to linearity of the relationships explored and colinearity between independent variables (Graham, 2003). Second, the fact that, as the number of independent environmental variables increases, the results become increasingly questionable. Third, data presented on the variance in the species data as summarized by the ordination axes explained by the independent environmental variables are expressed in different ways and are often misinterpreted.

RDA is very similar to CCA, except that the ordination axes are derived from PCA and the method thus assumes a species response model of short linear gradients, rather than the unimodal response model typical of longer gradients underlying correspondence analysis. However, most of the above criticisms apply equally to RDA as to CCA.

Despite these comments, CCA, in particular, is now very widely used, partly because of the limited range of alternative choices and because it has been claimed that it and the underlying multiple regression model are quite 'robust' (Lepš and Šmilauer, 2003). However, many researchers (Clarke, 1993; Clarke and Ainsworth, 1993; McCune and Grace, 2002) believe that, in most situations where CCA ordination is applied inductively, it would almost certainly have been equally as sound or even more appropriate to use the indirect method of NMS with post-analysis use of correlation methods to examine environmental and biotic gradients (Clarke, 1993; Økland, 1996; McCune and Grace, 2002).

An important claim for CCA is that it can be used deductively as well as inductively. Monte Carlo permutation tests for assessing the significance of individual axis eigenvalues and species-environment correlations are available and this, together with the possibilities for using CCA in combination with analysis of variance (ANOVA) in the analysis of manipulative experiments, means that it offers a potentially powerful method for hypothesis testing. In summary, the effectiveness of CCA as an ordination method is still open to question. Its use requires careful planning and a full understanding of the complexity of the method and its assumptions, particularly those of the underlying linear regression model. It should always be seen as a modelling technique where the selection of independent environmental/biotic variables is very carefully made.

The non-linear nature of many species response curves and ordination axes in relation to environmental variables is still a major issue. Kent *et al*. (2006) highlight the potentially exciting new development of the multiplicative habitat modelling package for non-parametric regression by McCune (2004) and programmed in the *Hyperniche* computer package of McCune and Mefford (2004). In principle, this should allow better exploration of the non-linear and often Gaussian response models that typify most species-environment data and could be employed to examine the relationships between axes derived from indirect ordinations (NMS and DCA) and environmental and biotic variables. As yet, however, little use has been made of this method and the computer program in published research. Other developments in fitting non-linear environmental gradients to multivariate community data include the proposal by Makarenkov and Legendre (2002) for non-linear RDA or CCA by applying higher-order terms to the matrix of environmental variables (x^2, x^3, etc) and the development of non-linear canonical correlation analysis (NCCorA) and nonlinear canonical analysis of principal coordinates (NCAP), which can be based on any distance measure (Millar *et al*., 2005). Almost any non-linear form of gradient can be examined using this approach and with virtually any distance measure, offering an extremely flexible means of modelling species response to environmental, spatial and temporal gradients.

IV Spatial autocorrelation, spatial dependence and the application of ordination methods in biogeography and ecology

Legendre (1993) showed that there are two components behind spatial structure in plant and animal communities. First, population dynamics, dispersal processes and interactions within and between species lead to spatial autocorrelation in the communities themselves. Second, spatial dependence occurs between species and controlling environmental/biotic factors, which also invariably show spatial trends. Wagner and Fortin (2005) and Kent *et al*. (2006) have recently summarized the issues and solutions relating to spatial structure and the application of ordination methods in biogeography and ecology. Until recently, most users of ordination methods conveniently ignored the potential problems of spatial autocorrelation and dependence when using either indirect ordination methods (DCA; NMS), followed by correlation or regression to define axis relationships with environmental or biotic data and thus environmental gradients, or the direct methods of CCA or RDA. Where ordination axis scores and any environmental/ biotic variable exhibit spatial autocorrelation or dependency, then Type I errors occur, where the null hypothesis is rejected when it should be accepted (Legendre, 1993; Lennon, 2000; Fortin and Gurevitch, 2001; Dale and Fortin, 2002; Legendre *et al*., 2002; Diniz-Filho *et al*., 2003; Jetz *et al*., 2005). Interestingly, similar problems have also recently been demonstrated for related methods for the derivation of transfer functions in Quaternary ecology (Telford and Birks, 2005). Lennon (2000) emphasizes that, when using Pearson product moment correlation and (multiple) regression, the greater the degree of spatial autocorrelation, within both the dependent and independent variables, the stronger this effect is likely to be. Thus many attempts to show the relative importance of environmental/biotic factors as independent variables in explaining CCA ordination axes and defining environmental gradients may only result in ordering the independent variables in terms of their intensity of positive spatial autocorrelation. Where this occurs, corrective measures should be applied (Kent *et al*., 2006).

The use of multiple regression at the core of CCA and RDA almost certainly accentuates possible problems of spatial autocorrelation and dependency in applying ordination methods and this was realized some years ago by Borcard *et al*. (1992). They and other authors (Legendre, 1993; Palmer, 1993; Borcard and Legendre, 1994; Økland and Eilertsen, 1994; Anderson and Gribble, 1998; Méot *et al*., 1998; Cushman and Wallin, 2002) showed how, through the use of partial canonical correspondence analysis, the explained variance in the regression could be partitioned, to enable the relative contributions attributable to spatial autocorrelation and dependency, as opposed to the environmental/biotic variables, to be determined. However, care is necessary, since in any analysis part of the explained variance may be attributed directly to spatial structure in the data, while another part may represent overlap between spatial and environmental/biotic effects (Wagner and Fortin, 2005). A further aspect is the idea that the unexplained variance may also be partially spatially structured, but the extent of this in any given analysis is obviously unknown.

Application of variance partitioning to assess the spatial component of variation among samples in CCA analyses is increasing. Recent examples are provided by Corney *et al*. (2004) examining landscape-scale controls of vegetation composition in British woodlands, Titeaux *et al*. (2004) working with breeding birds in Belgium, Svenning *et al*. (2004) analysing plant community structure of tropical forests in Panama, and Lee and Rotenberry (2005) studying bird and tree assemblage relationships in eastern North America.

V New developments in numerical classification and ordination methods

1 Numerical classification

Developments in numerical classification are focused in five main areas. First, the potential of Bayesian model-based analysis as an alternative to the classical cluster analysis-discriminant analysis approach has been demonstrated by ter Braak *et al*. (2003), although a number of problems still exist with the method.

Second, techniques of fuzzy classification, although originally devised some years ago (Bezdek, 1974; 1981; Roberts, 1989; Equihua, 1991), are becoming more widely used, partly due to their availability with the computer programs available from Pisces Conservation Ltd (2004). Similarly, methods of probabilistic classification (Goodall, 2002) deserve greater attention, although they have rarely been used. The programs for probabilistic classification are now freely available from the author.

Third, there have been important advances in techniques for determining whether groups of samples derived from cluster/similarity analysis or TWINSPAN differ significantly from each other and also for determining which species or variables are most important in accounting for these differences. For various reasons, particularly the nature of species data, the classical multivariate analysis of variance (MANOVA) methods are inappropriate. The 'analysis of similarities' (ANOSIM) and 'species contributions to similarity' (SIMPER) programs of Clarke and Green (1988) and Clarke and Warwick (1994) are becoming more widely used in this context, along with the multi-response permutation procedures (MRPP) method of Mielke *et al*. (1976; 1981), available in the PC-Ord package (McCune and Mefford, 1999), which performs a similar function.

The fourth important development in this area is that of classification and regression trees (CART) (Breiman *et al*., 1984; Vayssieres *et al*., 2000). CART divides a set of samples recursively into subsets that become increasingly homogeneous in relation to the predefined groups. The result is a tree-like classification based on a dichotomous key. This can then be used to assign unknown samples into the groups. CART is now often used in conjunction with geographic information systems to produce maps of plant communities, habitats or land cover based on spatially referenced environmental or impact data. An excellent introduction is presented in Urban (2002) and a recent biogeographical example of CART combined with GIS in the area of soil classification is given in Scull *et al*. (2004).

Finally, the problems of spatial autocorrelation and dependence in relation to numerical classification have received attention. Hervada-Sala and Jarauta-Bragulat (2004) presented an extension to Ward's classification method in the context of sediment analysis, where multivariate data have been collected as regionalized variables, which are not spatial independent. Previous methods based on correction of Ward's method (Pawlowsky-Glahn *et al*., 1997; Jiménez-Gonzáles, 1999; Jiménez-Gonzáles *et al*., 1998) involved use of Mahalanobis distance but were limited to two variables and placed various restrictions on the variogram model. Hervada-Sala and Jarauta-Bragulat (2004) show that further refinement is possible, which makes the method applicable to a true multivariate situation and which has considerable potential for various types of biogeographic data sets.

2 Ordination

Despite the reservations expressed above about CCA as an ordination technique, most recent developments in ordination are linked to that method and also to the problems of spatial structure in both species and environmental data and linked scale effects discussed above. Wagner (2003; 2004) introduced multiscale ordination (MSO) to allow assessment of how results of multivariate ordination may differ at varying spatial scales. She showed how correspondence analysis (CA) and CCA can be partitioned by distance and integrated with geostatistics using various forms of empirical chi-square variograms. This research has recently been placed in a broader framework for variogram-based multiscale ordinations by Couteron and Ollier (2005).

Cushman and McGarigal (2003) extended ideas of variance partitioning in CCA to examine hierarchically structured landscape data, looking at differing sources of variation at both patch and landscape scales. Borcard and Legendre (2002) and Borcard *et al.* (2004) introduced principal coordinates of neighbour matrices (PCNM), which performs a spectral decomposition of the spatial relationships among a set of sampling sites, creating variables that correspond to all detectable spatial scales in a data set. Important PCNM variables can be directly interpreted in terms of spatial scales or included in a variance partitioning of both spatial and environmental/biotic components.

Other developments in CCA concern its relationship to other methods; first to maximum-likelihood canonical Gaussian ordination (CGO) (Johnson and Altman, 1999). Yee (2004) developed a new class of quadratic reduced-rank vector generalized linear models for maximum-likelihood CGO. Second, Zhu *et al.* (2005) showed the equivalence between CCA and the method of linear discriminant analysis (LDA). A probabilistic model for LDA in the context of ecological ordination is presented which sheds light on the CCA algorithm and which is generalized to perform constrained ordination with flexible response functions.

A novel approach to comparing two ordinations from two communities (eg, plants and animals) sampled from the same locations entitled 'Co-correspondence analysis' was proposed by ter Braak and Schaffers (2004). Rather than the previous attempts where simple indirect correspondence analysis or reciprocal averaging was applied to both data sets and the resulting ordination axes were then correlated (Hájek *et al.*, 2002), the method maximizes the weighted covariance between weighted averaged species scores of one community and the weighted average species scores of the other community. Methods are linked to techniques of co-inertia analysis (Dolédec and Chessel, 1994; Dray *et al.*, 2004). The result is the identification of patterns that are common to both communities.

Finally, Anderson and Willis (2003) argued that unconstrained (indirect) ordinations, including methods like MDS, that are generally considered robust, may still mask some meaningful patterns and suggested a new framework for deductive situations where there is reason to construct a priori hypotheses concerning species-environment relationships. They demonstrate use of a type of constrained ordination based on use of principal coordinates analysis (PCO) followed by either canonical discriminant analysis (CDA) or a canonical correlation analysis (CCorA).

VI Conclusion

The application of methods of numerical classification and ordination in biogeography and ecology is at an interesting crossroads. The major techniques in use in both areas at the present time have been around for 15–25 years. During that time, use of particular methods has tended to be based on tradition in different branches of the two subjects, rather than a full appreciation of the merits and demerits of particular techniques. Almost all methods have now been shown to have their limitations and, despite quite serious criticisms, such as those of McCune and Grace (2002), the range of techniques discussed above all still appear to enjoy widespread application. However, the extent to which users are fully aware of the problems with any given method is often an open question.

The major development of recent years has been the greatly increased understanding of the importance of taking account of both scale and spatial structure (spatial autocorrelation and dependence) between and among samples in the application of these multivariate methods. Surprisingly, virtually all of these developments have come from biologists and ecologists (eg, Wagner, 2003; 2004; Fortin and Dale; 2005; Wagner and Fortin, 2005), rather than from geographers and biogeographers. There is a certain irony in the fact that biogeographers are now having to take lessons in the importance of spatial analysis from biologists and ecologists!

References

Anderson, M.J. and **Gribble, N.A.** 1998: Partitioning the variation among spatial, temporal and environmental components in a multivariate dataset. *Australian Journal of Ecology* 23, 158–67.

Anderson, M.J. and **Willis, T.J.** 2003: Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology* 84, 511–25.

Beals E.W. 1984: Bray-Curtis ordination: an effective strategy for analysis of multivariate ecological data. *Advances in Ecological Research* 14, 1–55.

Belbin, L. and **McDonald, C.** 1993: Comparing three classification strategies for use in ecology. *Journal of Vegetation Science* 4, 341–48.

Bezdek, J.C. 1974: Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology* 1, 57–71.

— 1981: *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.

Borcard, D. and **Legendre, P.** 1994: Environmental control and spatial structure in ecological communities: an example using orbatid mites (Acari, Oribatei). *Environmental and Ecological Statistics* 1, 37–53.

— 2002: All-scale spatial analysis of ecological data by means of principal coordinates on neighbour matrices. *Ecological Modelling* 153, 51–68.

Borcard, D., **Legendre, P.**, **Avois-Jacquet, C.** and **Tuomisto, H.** 2004: Dissecting the spatial structure of ecological data at multiple scales. *Ecology* 85, 1826–32.

Borcard, D., **Legendre, P.** and **Drapeau, P.** 1992: Partialling out the spatial component of ecological variation. *Ecology* 73, 1045–55.

Breiman, L., **Friedman, J.H.**, **Olshen, R.A.** and **Stone, C.J.** 1984: *Classification and regression trees*. New York: Chapman and Hall.

Clarke, K.R. 1993: Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* 18, 117–43.

Clarke, K.R. and **Ainsworth, M.** 1993: A method of linking multivariate community structure to environmental variables. *Marine Ecology Progress Series* 92, 205–19.

Clarke, K.R. and **Gorley, R.N.** 2001: *PRIMER v5: User Manual/Tutorial*. Plymouth: PRIMER-E.

Clarke, K.R. and **Green, R.H.** 1988: Statistical design and analysis for a 'biological effects' study. *Marine Ecology Progress Series* 46, 213–26.

Clarke, K.R. and **Warwick, R.M.** 1994: Similarity-based testing for community pattern: the 2-way layout with no replication. *Marine Biology* 118, 167–76.

Corney, P.M., **Le Duc, M.G.**, **Smart, S.M.**, **Kirby, K.J.**, **Bunce, R.G.H.** and **Marrs, R.H.** 2004: The effect of landscape-scale environmental drivers on the vegetation composition of British woodlands. *Biological Conservation* 120, 491–505.

Couteron, P. and **Ollier, S.** 2005: A generalized, variogram-based framework for multi-scale ordination. *Ecology* 86, 828–34.

Cox, T.F. and **Cox, M.A.A.** 1996: *Multidimensional scaling*. London: Chapman and Hall.

Cushman, S.A. and **McGarigal, K.** 2003: Hierarchical, multi-scale decomposition of species-environment relationships. *Landscape Ecology* 17, 637–46.

Cushman, S.A. and **Wallin, D.O.** 2002: Separating the effects of environmental, spatial and disturbance factors on forest community structure in the Russian Far East. *Forest Ecology and Management* 168, 201–15.

Dale, M.R.T. and **Fortin, M.-J.** 2002: Spatial autocorrelation and statistical tests in ecology. *Ecoscience* 9, 162–67.

Diniz-Filho, J.A.F., **Bini, L.M.** and **Hawkins, B.** 2003: Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography* 12, 53–64.

Dolédec, S. and **Chessel, D.** 1994: Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology* 31, 277–94.

Dray, S., **Chessel, D.** and **Thioulouse, J.** 2004: Co-inertia analysis and the linking of ecological data tables. *Ecology* 84, 3078–89.

Equihua, M. 1991: Fuzzy clustering of ecological data. *Journal of Ecology* 78, 519–34.

Everitt, B.S., **Landau, S.** and **Leese, M.** 2001: *Cluster analysis* (fourth edition). London: Arnold.

Faith, D.P., **Minchin, P.R.** and **Belbin, L.** 1987: Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69, 57–68.

Fortin, M.-J. and **Dale, M.** 2005: *Spatial analysis: a guide for ecologists*. Cambridge: Cambridge University Press.

Fortin, M.-J. and **Gurevitch, J.** 2001: Mantel tests: spatial structure in field experiments. In Scheiner, S.M. and Gurevitch, J., editors, *Design and analysis of ecological experiments* (second edition), Oxford: Oxford University Press, 308–26.

Gauch, H.G. 1982: *Multivariate analysis for community ecology*. New York, Cambridge University Press.

Gittins, R. 1969: The application of ordination techniques. In Rorison, I.H., editor, *Ecological aspects of the mineral nutrition of plants*, Oxford: Blackwell Scientific, 37–66.

Goodall, D.W. 2002: Probabilistic classification and its application to vegetation science. *Community Ecology* 3, 147–57.

Graham, M.H. 2003: Confronting multicollinearity in ecological multiple regression. *Ecology* 84, 2809–15.

Greig-Smith, P. 1983: *Quantitative plant ecology* (third edition). London: Butterworths.

Hájek, M., **Hekera, P.** and **Hájková, P.** 2002: Spring fen vegetation and water chemistry in the Western Carpathien flysch zone. *Folia Geobotanica* 37, 205–24.

Hervada-Sala, C. and **Jarauta-Bragulat, E.** 2004: A program to perform Ward's clustering method on several regionalised variables. *Computers and Geosciences* 30, 881–86.

Hill, M.O. 1979a: *TWINSPAN - a FORTRAN program for arranging multivariate data in an ordered two way table by classification of the individuals and the attributes*. Ithaca, NY: Cornell University, Department of Ecology and Systematics.

— 1979b: *DECORANA – a FORTRAN program for detrended correspondence analysis and reciprocal averaging*. Ithaca, NY: Cornell University, Department of Ecology and Systematics.

Hill, M.O. and **Šmilauer, P.** 2005: *TWINSPAN for Windows version 2.3*. Huntingdon and České Budějovice, Centre for Ecology and Hydrology and University of South Bohemia.

Jackson, D.A. and **Somers, K.M.** 1991: Putting things in order: the ups and downs of detrended correspondence analysis. *The American Naturalist* 117, 704–12.

Jetz, W., Rahbek, C. and **Lichstein, J.W.** 2005: Local and global approaches to spatial data analysis in ecology. *Global Ecology and Biogeography* 14, 97–98.

Jiménez-Gonzáles, S. 1999: Agrupación de datos con dependencia especial. El método de Ward adapto y la estimación iterative de semivariograms (Clustering of spatially dependent data. Adapted Ward's method and iterative variogram estimation). Final Engineering Project, Universitat Politècnica de Catalunya, Barcelona, Spain, 151 pp.

Jiménez-Gonzáles, S., Pawlowsky-Glahn, V., Egozcue, J.J. and **Ramirez-Llorente, R.** 1998: Ward's method adapted to spatially dependent data. In *Proceedings of the Fourth Meeting of the International Association for Mathematical Geology, Barcelona, Spain*, 445–50.

Johnson, K.W. and **Altman, N.S.** 1999: Canonical correspondence analysis as an approximation to Gaussian ordination. *Environmetrics* 10, 39–52.

Jongman, R.H.G., ter Braak, C.J.G. and **van Tongeren**, **O.F.R.** 1995: *Data analysis in community and landscape ecology* (second edition). Cambridge: Cambridge University Press.

Kent, M. and **Ballard, J.** 1988: Trends and problems in the application of classification and ordination methods in plant ecology. *Vegetatio* 78, 109–24.

Kent, M. and **Coker, P.** 1992: *Vegetation description and analysis a practical approach*. Chichester: Wiley.

Kent, M., Moyeed, R.A., Reid, C.L., Pakeman, R. and **Weaver, R.** 2006: Geostatistics, spatial rate of change analysis and boundary detection in plant ecology and biogeography. *Progress in Physical Geography* 30, 201–31.

Kershaw, K.A. 1973: *Quantitative and dynamic plant ecology*. London: Arnold.

Kershaw, K.A. and **Looney, J.H.H.** 1985: *Quantitative and dynamic plant ecology* (third edition). London: Arnold.

Kruskal, J.B. 1964a: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29, 1–27.

— 1964b: Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29, 115–29.

Lee, P.Y. and **Rotenberry, J.T.** 2005: Relationships between bird species and tree species assemblages in forested habitats of eastern North America. *Journal of Biogeography* 32, 1139–50.

Legendre, P. 1993: Spatial autocorrelation: trouble or new paradigm? *Ecology* 74, 1659–73.

Legendre, P. and **Legendre, L.** 1998: *Numerical ecology*. Amsterdam: Elsevier.

Legendre, P., Dale, M.R.T., Fortin, M.-J., Gurevitch, J., Hohn, M. and **Myers, D.** 2002: The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* 25, 601–15.

Lennon, J.J. 2000: Red-shifts and red herrings in geographical ecology. *Ecography* 23, 101–13.

Lepš, J. and **Šmilauer, P.** 2003: *Multivariate analysis of ecological data using CANOCO*. Cambridge: Cambridge University Press.

Makarenkov, V. and **Legendre, P.** 2002: Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. *Ecology* 83, 1146–61.

McArdle, B.H. and **Anderson, M.J.** 2001: Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82, 290–97.

McCune, B. 2004: Nonparametric multiplicative regression for habitat modeling. Retrieved 1 October 2005 from http://www.pcord.com/NPMRintro.pdf

McCune, B. and **Grace, J.B.** 2002: *Analysis of ecological communities*. Gleneden Beach, OR: MjM Software.

McCune, B. and **Mefford, M.J.** 1999: *PC-ORD, Multivariate analysis of ecological data, version 4*. Gleneden Beach, OR: MjM Software.

— 2004: *Nonparametric multiplicative habitat modelling, version 1.0*. Gleneden Beach, OR: MjM Software.

Méot, A., Legendre, P. and **Borcard, D.** 1998: Partialling out the spatial component of ecological variations: questions and propositions in the linear modelling framework. *Environmental and Ecological Statistics* 5, 1–27.

Mielke, P.W. Jr, Berry, K.J. and **Brier, G.W.** 1981: Application of multi-response permutation procedures for examining seasonal changes in monthly sea-level pressure patterns. *Monthly Weather Review* 109, 120–26.

Mielke, P.W. Jr, Berry, K.J. and **Johnson, E.S.** 1976: Multiresponse permutation procedures for *a priori* classifications. *Communications in Statistics* A5, 1409–24.

Millar, R.B., Anderson, M.J. and **Zunun, G.** 2005: Fitting nonlinear environmental gradients to community data: a general distance-based approach. *Ecology* 86, 2245–51.

Minchin, P.R. 1987: Simulation of multidimensional community patterns: towards a comprehensive model. *Vegetatio* 71, 145–56.

Økland, R.H. 1996: Are ordination and constrained ordination alternative or complementary strategies in general ecological studies? *Journal of Vegetation Science* 7, 289–92.

Økland, R.H. and **Eilertsen, O.** 1994: Canonical correspondence analysis with variance partitioning: some comments and an application. *Journal of Vegetation Science* 5, 117–26.

Oksanen, J. and **Minchin, P.R.** 1997: Instability of ordination results under changes in input data order: explanations and remedies. *Journal of Vegetation Science* 8, 447–54.

Palmer, M.W.1993: Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* 74, 2215–30.

Pawlowsky-Glahn, V., Simarro-Grande, G. and **Martin-Fernandez, J.A.** 1997: Spatial cluster analysis using a generalised Mahalanobis distance. In *Proceedings of the Third Meeting of the International Association for Mathematical Geology, Barcelona, Spain*, 175–80.

Pisces Conservation Ltd. 2004: *Fuzzy grouping for clustering objects while accounting for uncertainty*. Lymington: Pisces Conservation Ltd (www.pisces-conservation.com).

Quinn, G.P. and **Keough, M.J.** 2002: *Experimental design and data analysis for biologists*. Cambridge: Cambridge University Press.

Roberts, D.W. 1989: Fuzzy-systems vegetation theory. *Vegetatio* 83, 71–80.

Rogerson, P.A. 2005: A set of associated statistical tests for spatial clustering. *Environmental and Ecological Statistics* 12, 275–88.

Scull, P., Franklin, J. and **Chadwick, O.A.** 2004: The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological Modelling* 181, 1–15.

Shepard, R.N. 1962: The analysis of proximities: multidimensional scaling with an unknown distance function. *Psychometrika* 27, 125–39, 219–46.

Svenning, J.C., Kinner, D.A., Stallard, R.F., Engelbrecht, B.M.J. and **Wright, S.J.** 2004: Ecological determinism in plant community structure across a tropical forest landscape. *Ecology* 85, 2526–38.

Tausch, R.J., Charlet, D.A., Weixelman, D.A. and **Zamudio, D.C.** 1995: Patterns of ordination and classification instability resulting from changes in input order. *Journal of Vegetation Science* 6, 897–902.

Telford, R.J. and **Birks, H.J.B.** 2005: The secret assumption of transfer functions: problems with spatial autocorrelation in evaluating model performance. *Quaternary Science Reviews* 24, 2173–79.

ter Braak, C.J.F. 1986: Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67, 1167–79.

— 1987: The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio* 69, 69–77.

— 1988a: CANOCO – an extension of DECORANA to analyse species-environmental relationships. *Vegetatio* 75, 150–60.

— 1988b: Partial canonical correspondence analysis. In Block, H.H., editor, *Classification and related methods of data analysis*, Amsterdam: North Holland Press, 551–58.

ter Braak, C.J. F. and **Schaffers, A.P.** 2004: Co-correspondence analysis: a new ordination method to relate two community compositions. *Ecology* 85, 834–46.

ter Braak, C.J.F. and **Šmilauer, P.** 2003: *CANOCO reference manual and user's guide to Canoco for Windows: software for canonical community ordination (version 4.5)*. Ithaca, NY: Microcomputer Power.

ter Braak, C.J.F., Hoijtink, H., Akkermans, W. and **Verdonschot, P.F.** 2003: Bayesian model-based cluster analysis for predicting macrofaunal communities. *Ecological Modelling* 160, 235–48

Titeux, N., Dufrêne, M., Jacob, J.-P., Paquay, M. and **Defourny, P.** 2004: Multivariate analysis of a fine-scale breeding bird atlas using a geographic information system and partial correspondence analysis: environmental and spatial effects. *Journal of Biogeography* 31, 1841–56.

Urban, D.L. 2002: Classification and regression trees. In McCune, B. and Grace, J.B., editors, *Analysis of ecological communities*, Gleneden Beach, OR: MjM Software, Chapter 29.

van Groenewoud, H. 1992: The robustness of correspondence, detrended correspondence and TWINSPAN analysis. *Journal of Vegetation Science* 3, 239–46.

Vayssieres, M.P., Plant, R.E. and **Allen-Diaz, B.H.** 2000: Classification trees: an alternative approach for predicting species distributions. *Journal of Vegetation Science* 11, 679–94.

— 2003: Spatial covariance in plant communities: integrating ordination, geostatistics and variance testing. *Ecology* 84, 1045–57.

Wagner, H.H. 2004: Direct multi-scale ordination with canonical correspondence analysis. *Ecology* 85, 342–51.

Wagner, H.H. and **Fortin, M.-J.** 2005: Spatial analysis of landscapes: concepts and statistics. *Ecology* 86, 1975–87.

Waite, S. 2000: *Statistical ecology in practice: a guide to analysing environmental and ecological field data*. Harlow: Pearson-Prentice-Hall.

Ward, J.H. 1963: Hierarchical grouping to optimize an objective function. *American Statistical Association Journal* 58, 236–44.

Williams, W.T., Lambert, J.M. and **Lance, G.N.** 1966: Multivariate methods in plant ecology. V. Similarity analysis and information analysis. *Journal of Ecology* 54, 427–45.

Yee, T.W. 2004: A new technique for maximum-liklihood canonical Gaussian ordination. *Ecological Monographs* 74, 685–701.

Zhu, M., Hastie, T.J. and **Walther, G.** 2005: Constrained ordination analysis with flexible response functions. *Ecological Modelling* 187, 254–36.