COENOSES 10(1): 39-44, 1995

ENVIRONMENT-VEGETATION RELATIONSHIPS IN THE UNDERSTORY OF PYRENNEAN *PINUS SYLVESTRIS* FORESTS: I. AN ORDINATION APPROACH

J. G. Pausas[1,2] & E. Feoli[1]

1 International Institute for Earth, Environment and Marine Sciences and Technologies (IIEM), ICS, via Beirut 7, Trieste, I-34100, Italy
2 Present address: Departament de Biologia Vegetal, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 645, Barcelona, E-08028, Spain

Keywords: Canonical correlation analysis, Correspondence Analysis, Principal Component Analysis, Normalization, *Pinus sylvestris*, Pyrenees, Soil, Understory.

Abstract. Canonical correlation analysis (COR) is applied to test the predictivity of ordination axes obtained by principal component analysis (PCA) and correspondence analysis (CA) of vegetation data (relevés) from Pyrennean *Pinus sylvestris* forests with respect to axes obtained by PCA of some environmental variables. COR shows high correlation between understory composition and the environmental variables. However, higher correlation values are shown by axes obtained by PCA after normalization of relevés. This confirms that PCA has to be seen strictly in connection with data transformation.

Introduction

Vegetation-environment relationships can be studied using empirical data by two main approaches (Feoli 1983): the direct one in which a measure of correlation between a set of environmental variables and a set of species or other biological variables is calculated; and the indirect one, in which a test of separation between groups of relevés obtained by biological data is performed using environmental data (Orlóci 1978). The application of the indirect method will be discussed in another paper based on the same data set (Pausas & Feoli 1995). This paper presents an example of the direct approach by the application of canonical correlation analysis (COR) to two data sets describing the Pyrennean Forest dominated by *Pinus sylvestris* L. One data set consists of a set of relevés (relevés described by species presence and abundance) and the other set consists of a set of environmental variables measured in the same set of relevés. The detailed discussion of results in ecological terms is not aimed here. Instead, we give a further contribution to the analysis of predictivity in plant community studies (Feoli 1983), namely how much a linear ordination of vegetation may be predictive with respect to environmental variables.

In the past 20 years an incredible debate was going on about different ordination methods applied to represent ecological space efficiently by a few number of uncorrelated axes (Feoli & Orlóci 1991). In many papers principal component analysis (PCA), as applied to vegetation data, was indicated to have the worst performance. It is true that when the species exhibit evident unimodal response along gradients (Whittaker 1967), PCA arranges the relevés along a more or less involute curve, but it is also true that PCA gives back perfectly the shape of the objects by offering advantages from many points of view (Feoli & Orlóci 1991). The idea of the "arch effect" and "distorted results" should be definitively forgotten. The arch is not a mathematical artifact (an effect) as was stated in many papers appeared also in outstanding international journals, but the true representation of the shape of the gradient, namely the trajectory of vegetation change in the multidimensional space given by the species. The curve is not a distortion - distortion of what? Of an imaginary straight line on which to order a set of numbers representing scores of environmental variables? It is to be stressed that a set of increasing or decreasing numbers can be written sequentially in any shape. Why should they be placed only along straight lines?!

It is true that the vegetation ordinaton axes showing a horseshoe curve are not "linearly" predictive with respect to environmental variables. However, this is not a weakness of the method, but a weakness of scientists that are pretending something that has not to be pretended. They should know that if they want to test linear correlation with environmental variables the trajectories of vegetation variation should be detrended. Detrending can be performed in many ways and actually it is not a difficult problem (ter Braak 1987, Feoli & Orlóci 1991). If the relevés are scattered everywhere in the ordination scattergram detrending would be a waste of time. Detrending is something that has nothing to do specifically with PCA or CA or other linear and non-linear ordination methods. It is an additional technique which can be applied to any ordination obtained with one of the many available techniques. From this it follows that we cannot say that detrended PCA is "better" than a non-detrended PCA, and that detrended CA is "better" than an ordinary CA. Of course, detrended ordinations will be always "better" if we are delighted to see gradients in straight lines. However, less involute horseshoe ordination curves may be obtained without detrending techniques directly in PCA, i.e. by transforming the original data before the analysis. Transforming the data, e.g. by normalization, standardization, etc., is not a detrending technique, it may be just a way to reduce the curvature of the arch (Dale 1994) as was also shown by Karadzic & Popovic (1994). However, data transformation is always a "manipulation", a sort of "lie". One can say that it is a "necessary lie" for representing gradients in less involute curves. Relevé normalization is one of the most effective transformations for smoothing the curvature not only in PCA, but also in non metric multidimensional scaling (NMDS, Kenkel & Orlóci 1986, Karadzic & Popovic 1994). However, the normalization [$x_{ij} = x_{ij}/(\Sigma_i x_{ij}^{\gamma})^{1/\gamma}$ with i = species, and γ=2 so that $(\Sigma_i x^2)^{1/2} = 1$] of the relevés along a coenocline implies a strong transformation of the response functions: low scores may become higher scores and high scores may become lower scores. If the axes obtained by PCA have to be linearly correlated with environmental axes this transformation would be useful.

Ter Braak (1987) presents a deep overview of the methods of correlation between vegetation and environmental data under the perspective of direct gradient analysis. We would prefer to call direct gradient analysis just the analysis of the response functions of the species or other biological variables against environmental variables in order to describe coenoclines (Whittaker 1967). Measuring correlation is something else. For the correlation analysis ter Braak (1987) supports the use of Canonical Correspondence Analysis (CCA) and Redundancy Analysis (RDA) since they have no limits in the number of variables of the two sets. They do not offer direct, global measures of the correlation between two sets of variables. CCA may be characterized as the "technique that selects the linear combination of environmental variables that maximizes the dispersion of species scores" (ter Braak 1987), while RDA may be characterized as "the technique selecting the linear combination of environmental variables that gives the smallest total residual sum of squares" (ter Braak 1987). The proper method to measure linear correlation between two data matrices, namely the matrix of species and the matrix of environmental variables, is canonical correlation analysis (COR). COR produces a set of ordination axes (canonical variates) for each data matrix in such a way that they are maximally correlated. COR was first suggested for ecological work by Hughes & Lindley (1955) and applied in different ecosystems to exploring vegetation-environment relationships (e.g. Bradfield & Scagel 1984, Carleton 1984, Feoli 1983, Feoli & Ganis 1985, Gittins 1985, Kabzens & Klinka 1987, Courtin et al. 1988, Shaukat & Uddin 1989, Carnevale & Torres 1990, Basnet 1992). One strong limitation of COR is the fact that the number of variables in each of the two sets and in total, must be smaller than the number of objects (Cooley & Lohnes 1971, Gittins 1985, ter Braak 1987). To avoid this drawback it is suggested to summarize the variables in a few orthogonal ones by ordination methods (Orlóci 1978, Feoli & Lagonegro 1985, Carleton 1984, Shaukat & Uddin 1989 and Franklin & Merlin 1992). Bornette et al. (1994) use PCA scores of two matrices in a technique very similar to COR namely the inter-battery factor analysis of Tucker (1958). COR, as applied in the present paper, is useful to measure the predictivity of the indirect ordinations (ordinations based only on species or other biological variables) with respect to environmental data matrices.

The data

Pausas & Fons (1992) have given an overview of the *Pinus* forests in the Pyrenees. For the present work fifty-five relevés from a mature *Pinus sylvestris* forest were selected in the eastern Pyrenees (NE of the Iberian Peninsula), and two data sets were collected:

1) Species data set: species frequency in the understory was recorded as a number of small quadrats (25x25 cm) in which the species occur. In each relevé, 100 systematically placed quadrats were sampled.

2) Environmental data set: a volumetric soil sample of the first 10 cm was collected in each relevé. In each soil sample, pH in water (1:2.5), pH in potassium chloride (1:2.5), stoniness and soil density were measured. Carbonates were analyzed with Bernard calcimeter, and C and N contents with a Carlo Erba elemental analyzer. Thickness of organic F and H layers were measured at 16 points in each relevé and the mean was used for the numerical analysis. Altitude, slope, number of species and moisture index were also included in this data set. Moisture index was calculated based on soil depth (estimated as the mean of 10 points), stoniness, soil texture, and aspect of the relevé, using the method described by Klinca et al. (1984). The number of species of each relevé was considered as a capacity of the environment to host plant diversity. The environmental variables their means and standard deviations are shown in Table 1.

Methods of data analysis

COR was applied in the following ways: all the species were linearly combined by principal component analysis and correspondence analysis (Benzecri et al. 1973, Hill 1973) while all the environmental variables were linearly combined only by PCA. Then, the PCA and CA axes of the two data sets were used in COR. The transformations used in PCA were as follows:

-PCAenv: R-type Principal Component Analysis of standardized environmental data set (environmental ordination).

- PCAsp: Q-type Principal Component Analysis of standardized species frequency.

-PCAspn: Q-type Principal Component Analysis of species frequency using relevé normalization, this is implicit in the chord distance of Orlóci (1978).

The SYN-TAX package (Podani 1993) was used for running these ordination analyses. Wilks' Lambda test with Bartlett's (1947) estimation of chi-squared was used to test

Table 1. Means and standard deviations of environmental variables, and the component correlation of PCAenv. Significant correlations (p<0.01) are indicated by *.

Variable	Mean	SD	Component correlation				
			PCAI	PCAII	PCAIII	PCAIV	PCAV
altitude	1446.182	160.552	.204	.452*	-.582*	-.207	.143
slope	25.709	7.728	-.041	-.357*	-.452*	.026	.346
F thickness	1.707	.766	.594*	-.251	-.215	.577*	.225
H thickness	2.885	1.653	.592*	.052	-.021	.627*	.009
pH	6.257	1.210	.721*	-.294	.235	-.268	-.019
pH_{H2O}-pH_{ClK}	.983	.362	-.687*	.031	-.340	.160	-.165
soil density	1.613	.349	-.345	.246	.237	.502*	-.437*
$CaCO_3$	4.624	9.029	.626*	-.168	.385*	-.286	-.260
stoniness	53.379	16.608	.242	.675*	.208	.160	-.212
C	4.539	2.650	.720*	.499*	-.059	-.000	.198
N	.286	.164	.714*	.288	-.382*	-.068	-.300
C/N	17.017	5.347	-.110	.427*	.507*	-.016	.663*
number of spp	22.364	8.429	.302	-.499*	-.103	-.183	-.204
Moisture index	3.186	.813	.147	-.651*	.206	.328	.116

Table 2. Eigenvalues (λ) and cumulative percentages of variance (%) in the first five ordination components.

		I	II	III	IV	V
PCAenv	λ	.3462E+01	.2217E+01	.1466E+01	.1373E+01	.1145E+01
	%	24.73	40.57	51.04	60.84	69.02
PCAsp	λ	527.63860	422.46500	361.25530	344.95560	301.82100
	%	7.52	13.53	18.68	23.59	27.89
PCAspn	λ	7.52493	5.23685	4.63222	2.72402	1.86357
	%	21.44	36.36	49.56	57.33	62.64
CA	λ	.4306E+00	.3777E+00	.3033E+00	.2795E+00	.2461E+00
	%	8.74	16.40	22.56	28.23	33.22

the significance of each canonical coefficient, as described in Cooley & Lohnes (1971).

Results

Ordinations

The ordinations of the 55 relevés using PCA and CA are shown in Fig. 1. No clear pattern is evident in these ordination scattergrams and no single environmental gradient can be assumed. The eigenvalues and cumulative percentages of variance accounting for each ordination component are indicated in Table 2. PCA with relevé normalization is the ordination which explains most of the variance by its three components (50%). This percentage is similar to that of the environmental ordination (PCAenv). The first three PCAsp and CA axes account respectively for only 18.7 and 22.6% of the variance. This means that there is no overwhelming effect of any single environmental factor. This is evident by the correlation coefficients between PCAenv axes and the original environmental variables as shown in Table 1. The PCAenv axes are significantly correlated with several environmental variables. The first component, explaining 25% of variance, is primarily a function of acidity-basicity and C-N contents, and secondly, of organic layer thickness. The second component is highly correlated, positively with stoniness, altitude and C/N ratio, and negatively with humidity (moisture index), number of species and slope.

Canonical correlation analysis

The canonical correlation coefficients between PCAenv axes and PCAsp, PCAspn and CA axes (vegetational axes) are shown in Table 3. The canonical scattergrams are presented in Fig. 2.

Two significant canonical variates are found between environmental axes, and PCAsp and CA axes. For PCAspn, three significant canonical variates are found. The significant canonical correlation coefficients are R_1= 0.81 and R_2= 0.49 between PCAenv-PCAsp axes; R_1= 0.87 and R_2= 0.60 be-

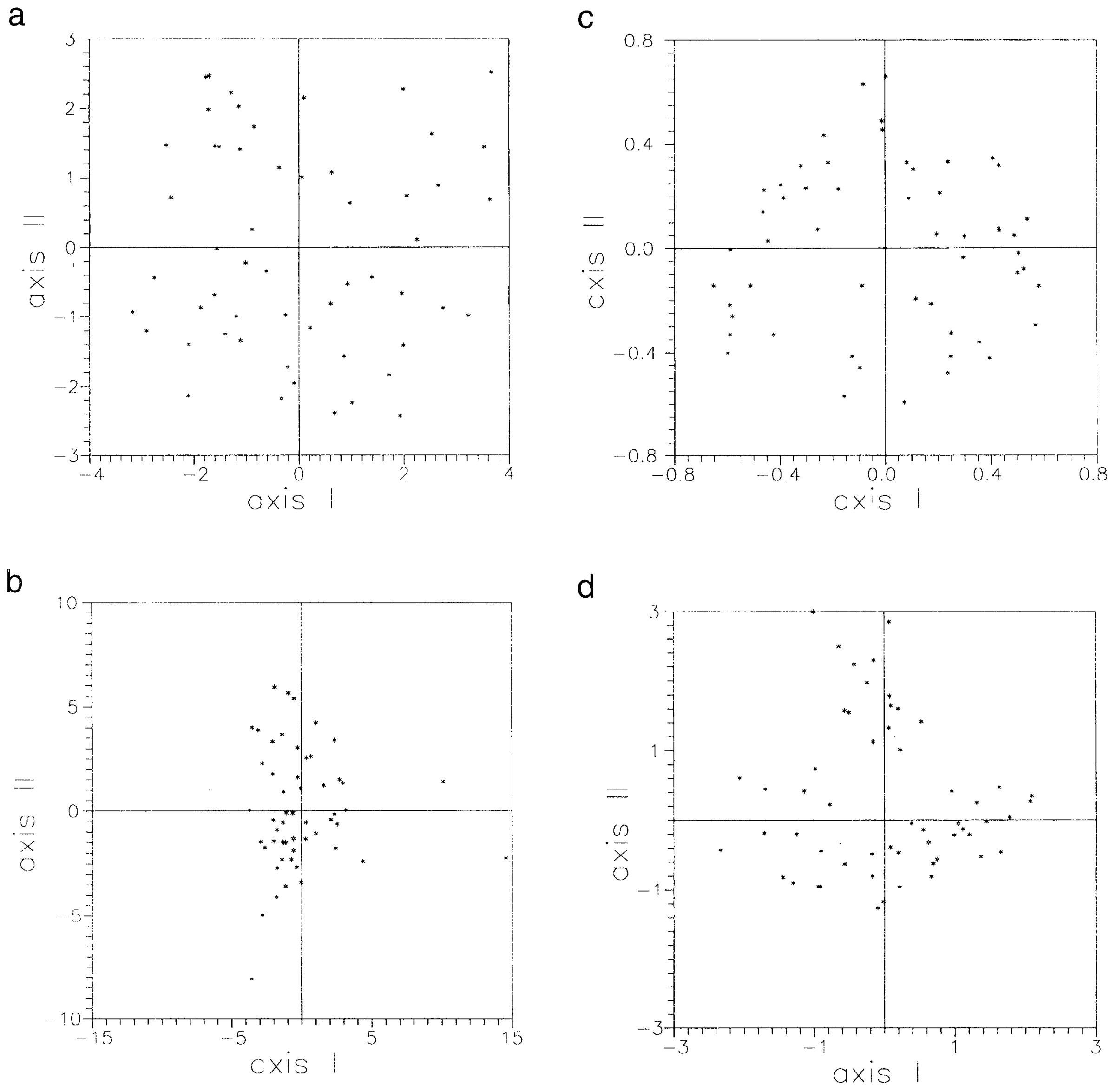

Figure 1. Ordinations of 55 relevés using different methods. a) PCA of standardized environmetal data set; b) PCA of standardized species frequency; c) PCA of species frequency and relevé normalization; and d) CA of species frequency.

tween PCAenv-PCAspn matrices; and R_1= 0.84 and R_2= 0.61 between PCAenv - CA axes. The highest canonical correlation is found between PCAenv and PCAspn axes.

Discussion and conclusions

Based on canonical correlation analysis, we can conclude that there is a high correlation between the measured environmental variables and floristic composition in the understory of *Pinus sylvestris* forests. The results suggest that more than a single environmental variable must be considered to explain the understory vegetation variation. The highest correlation with the environmental ordination (PCAenv) was found for PCA using relevé normalization (PCAspn). PCAspn was also the vegetational ordination method accounting for the highest variance in its components. This confirms that normalization by relevés is very useful when the data matrix has many zero values (very frequent in vegetation data) as was suggested by Orlóci (1978) and Kenkel & Orlóci (1986). In fact, in PCA ordinations relevés with few species in common and lacking many species may be very close to one another, while relevés with many species in common may be very distant. This is a consequence of the fact that Euclidean distance, implicit in the

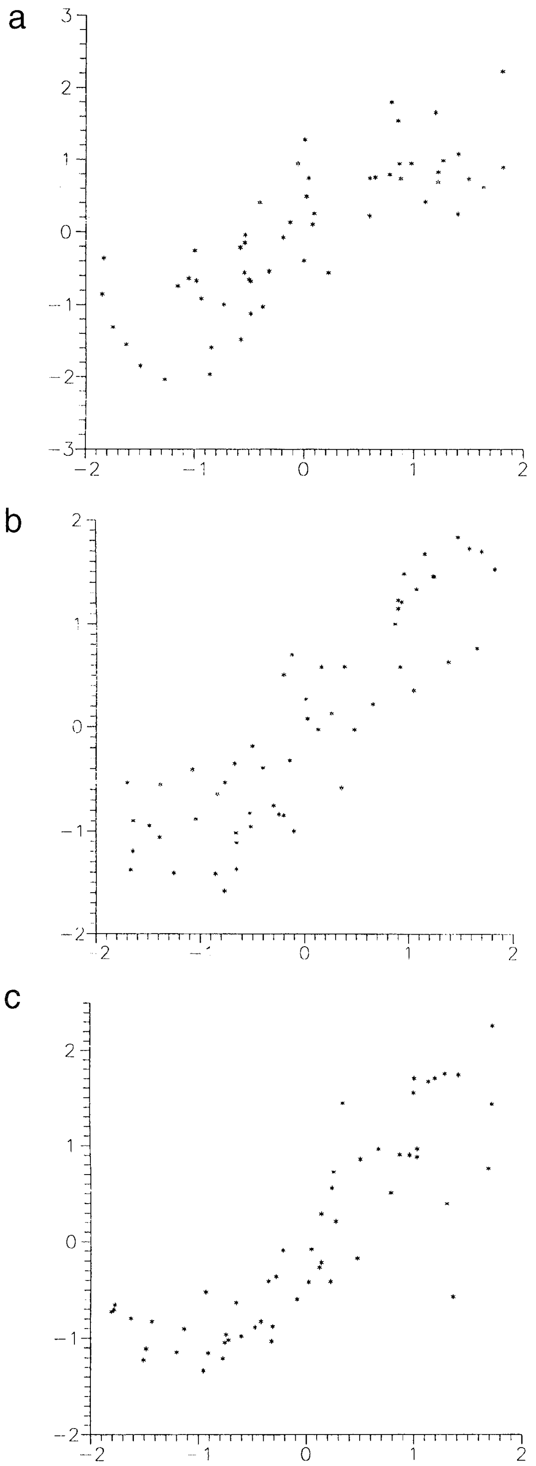

Figure 2. Canonical correlations of the first canonical variates in different analyses: a) between environmental components (PCAenv) and PCAsp; b) between environmental components and PCAspn and c) between environmental components and CA.

Table 3. Eigenvalues (λ), canonical correlation coefficients (R_c), Wilks' Lambdas (Λ), chi-square values (χ^2) and degrees of freedom (df) in the canonical correlation analysis between the first 5 environmental components (PCAenv) and a) 5 components of PCAsp, b) 5 components of PCAspn and c) 5 axes of CA. Significant correlations (p<0.01) are indicated by *.

	R_c	λ	χ^2	df	Λ
a)	.8090*	.655	79.61	25	.1937
	.4948*	.245	28.06	16	.5608
	.4518	.204	14.44	9	.7426
	.2560	.066	3.37	4	.9330
	.0400	.002	.08	1	.9984
b)	.8733*	.763	116.23	25	.0910
	.6013*	.362	46.48	16	.3836
	.4891*	.239	24.71	9	.6008
	.4338	.188	11.45	4	.7897
	.1651	.027	1.34	1	.9727
c)	.8367*	.700	94.56	25	.1423
	.6104*	.373	36.14	16	.4746
	.4727	.223	13.54	9	.7565
	.1600	.026	1.27	4	.9742
	.0154	.000	.01	1	.9998

PCA technique (Orlóci 1978), is very sensitive to the number of species and to their abundance. The results in Table 3 give a further empirical proof that the performance of PCA cannot be judged *a priori* to be worse than the performance of other methods in predicting environmental variation. However, since the species response to environmental factors is usually non-linear (Austin 1987) the application of COR to original data or to their linear transformation may not be suitable for testing the correlation statistically (Austin et al. 1990). Linear methods based on eigenanalysis, such as principal component analysis and correspondence analysis can be used for seeking trends in the analysis of vegetation since they are able to detect non-linearity in orthogonal structures. When a clear curvilinear pattern is shown, unfolding can be necessary before the application of COR. It can be achieved by different techniques such as polynomial regression (Philipps 1978), angular seriation (Feoli & Feoli-Chiapella 1980), detrending (Hill & Gauch 1980) or shortest path adjustment (Lagonegro 1986, Bradfield & Kenkel 1987, Wildi 1992). However, when no clear trend is shown in the ordination scattergram (e.g., more than one environmental gradient in the data which lead to a wide dispersion of points in the scattergrams) unfolding techniques are not necessary.

Acknowledgements. This work was performed at the International Institute for Earth, Environment and Marine Sciences and Technologies (IIEM), ICS, Trieste, Italy, and it was supported by an FPI grant from the Spanish government to the first author. We thank T. Sauras for the valuable collaboration in data collection and soil analysis. The first author would like to thank J. Podani and the Quantitative Ecology Group of Trieste, mainly V. Zuccarello, P. Ganis and Sun C.Y., for their help provided during his stay in Trieste.

References

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

Austin, M.P., A.O. Nicholls & C.R. Margules. 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. Ecol. Monogr. 60: 161-177.

Bartlett, M.S. 1947. Multivariate analysis. J. Roy. Stat. Soc. Series B. 9: 176-197.

Basnet , K. 1992. Effect of topography on the pattern of trees in tabouco (*Dacryodes excelsa*) dominated rain forest of Puerto Rico. Biotropica 24: 31-42.

Benzecri, J.-P. et al. 1973. L'analyse des données. Tome 2, L'analyse des correspondances. Dunod, Paris.

Bornette, G., Amoros, C. & D. Chessel. 1994. Effect of allogenic processes on successional rates in former river channels. J. Veg. Sci. 5: 237-246.

Bradfield, G.E. & N.C. Kenkel. 1987. Nonlinear ordination using flexible shortest path adjustment of ecological distances. Ecology 68: 750-753.

Bradfield, G.E. & A. Scagel. 1984. Correlation among vegetation strata and environmental variables in subalpine spruce-fir forest, southeastern Britsh Columbia. Vegetatio 55: 105-114.

Carleton, T.J. 1984. Residual ordination analysis: a method for exploring vegetation-environment relationship. Ecology 65: 469-477.

Carnevale, N.J. & P.S. Torres. 1990. The relevance of physical factors on species distributions in inland salt marshes (Argentina). Coenoses 5: 113-120.

Cooley, W.W. & P.R. Lohnes. 1971. Multivariate data analysis. Wiley, New York.

Courtin, P.J., K. Klinka, M.C. Feller & J.P. Demaerschalk. 1988. An approach to quantitative classification of nutrient regimes of forest soils. Can. J. Bot. 66: 2640-2653.

Dale, M. 1994. Straightening the horseshoe: a Riemannian resolution? Coenoses 9: 43-53.

Feoli, E. 1983. Predictive use of classification and ordination methods in plant community ecology. In: Ferrari, C., S. Gentile, S. Pignatti and E. Poli Marchese (eds.). Le communità vegetali come indicadori ambientali. pp. 83-108. Studi e Documentazioni 30. Regione Emilia Romagna e Società Italiana di Fitosociologia.

Feoli, E. & L. Feoli Chiapella. 1980. Evaluation of ordination methods through simulated coenoclines: some comments. Vegetatio 42: 35-41.

Feoli, E. & P. Ganis. 1985. Comparison of floristic vegetation types by multiway contingency tables. An example of application of an integrated information system to some beechwood types of the Apennines. Abst. Bot. 9: 1-15.

Feoli, E. & M. Lagonegro. 1985. Analisi multivariati di dati. Libreria Goliardica. Trieste.

Feoli, E. & L. Orlóci. 1991. The properties and interpretation of observations in vegetation study. *In*: Feoli E. & L. Orlóci (eds) Computer assisted vegetation analysis, pp. 3-13. Kluwer, Dordrecht.

Franklin, J. & M. Merlin. 1992. Species-environment patterns of forest vegetation on the uplifted reef limestone of Atiu, Mangaia, Ma'uke and Miti'aro, Cook Islands. J. Veg. Sci. 3: 3-14.

Gittins, R. 1985. Canonical analysis. A review with applications in ecology. Biomathematics 12, Springer-Verlag, Berlin.

Hill, M.O. 1973. Reciprocal averaging: an eigenvector method of ordination. J. Ecol. 61: 237-249.

Hill, M.O. & H.G. Gauch. 1980. Detrended correspondence analysis: an improved ordination technique. Vegetatio 42: 47-58.

Hughes, R.E. & D.V. Lindley. 1955. Application of biometric methods to problems of classification in ecology. Nature 175: 806-807.

Kabzens, R.D. & K. Klinka. 1987. Initial quantitative characterization of soil nutrient regimes. II. Relationships among soil, vegetation, and site index. Can. J. For. Res. 17: 1565-1571.

Karadzic, B. & R. Popovic. 1994. A generalized standardization procedure in ecological ordination: tests with principal component analysis. J. Veg. Sci. 5:259-262.

Kenkel, N.C. & L. Orlóci. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. Ecology 67: 919-927.

Kessel, S.R. & R.H. Whittaker. 1976. Comparison of three ordination techniques. Vegetatio 32: 21-29.

Klinka, K., R.N. Green, P.J. Courtin & F.C. Nuszdorfer. 1984. Site diagnosis, tree species selection, and slashburning guidelines for the Vancouver Forest Region. Ministry of Forests. British Columbia.

Lagonegro, M. 1986. Performances of a proximity index defined on a dendrogram table or a minimum spaning tree graph. GEAD-EQ 4. Trieste.

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

Pausas, J.G. & E. Feoli. 1995 Environment-vegetation relationships in understory of Pyrennean *Pinus sylvestris* forests: II. An approach based on classification.Coenoses (in press) .

Pausas, J.G. & J. Fons. 1992. Estructura i el creixement d'algunes pinedes en diferents condicions topográfiques. Fol. Bot. Misc. 8: 199-213.

Phillips, D.L. 1978. Polynomial ordination: field and computer simulation testing of a new method. Vegetatio 37: 21-39.

Podani, J. 1993. SYN-TAX 5.0: Computer programs for data analysis in ecology and systematics. Abstr. Bot. 17:289-302.

Shaukat, S.S. & M. Uddin. 1989. An application of canonical correlation and principal component analysis to the study of desert environments. Abst. Bot. 13: 17-45.

ter Braak, C.J.F. 1987. Ordination. In: Jongman R.H.G., ter Braak C.J.F. & O.F.R. van Tongeren (eds): Data analysis in community and landscape ecology. Pudoc, Wageningen.

Tucker, L.R. 1958. An inter-battery method of factor analysis. Psychometrika 23:111-136.

Wildi, O. 1992. On the use of Mantel's statistic and flexible shortest path adjustment in the analysis of ecological gradients. Coenoses 7: 91-101.

Whittaker, R.H. 1967. Gradient analysis of vegetation. Biol. Rev. 42: 207-264.

Manuscript received: May 1995