COENOSES 1: 53-56, 1986

AUTOCORRELATION FOR MEASURING PREDICTIVITY IN COMMUNITY ECOLOGY: AN EXAMPLE WITH STRUCTURAL AND CHOROLOGICAL DATA FROM MIXED FOREST TYPES OF NE ITALY

Enrico Feoli and Paola Ganis

Dipartimento di Biologia, Università degli Studi di Trieste, 34100 Italia.

Keywords: Autocorrelation, Ecology, Predictivity, Vegetation.

Abstract: The autocorrelation method, widely used in spatial pattern analysis mainly by geographers, is tested and proposed to measure the predictive value of different variables describing the communities. An example of application is based on three matrices of 20 vegetation types from the mixed woods of Friuli Venezia-Giulia (NE Italy). The predictivity has been measured within and between three sets of variables: a) 8 environmental variables measured by ecological indicator values; b) 15 life form-growth form variables, c) 23 chorological variables. The method, being freer from restrictive assumptions about the relationships between the variables, appears to complement other methods well, such as canonical correlation analysis and discriminant analysis.

Introduction

The analysis of correlations between different ecological variables, describing the communities, plays an important role in ecology. The methods may follow two main approaches, a direct one and an indirect one. The direct approach is based on direct computations of correlation coefficient between variables, or sets of variables; multiple regression analysis and canonical correlation analysis are cases in point. The indirect approach is based on measures of separation between classes of objects, discriminant analysis for continuous data and contingency table analysis for discrete data, are the typical methods. In the latter two cases, the separation and association of the classes, defined by a set of variables, is measured by considering another set: the sharper the relationship, the higher the predictivity. Descriptions of such methods are given in standard texts of multivariate analysis (see Orlóci 1978; Orlóci & Kenkel 1985). Short reviews with examples from vegetation ecology are found elsewhere (e.g. Feoli 1983 and Feoli & Scimone 1984).

In the present paper a method for measuring predictivity between different sets of ecological variables is proposed. The method uses an autocorrelation test of non-randomness of the distribution of values of single or composite variables in one set in the hyperspace defined by the variables of another set. Such an autocorrelation is widely used by geographers in spatial pattern analyses (Ebdon 1977) as described in detail by Cliff and Ord (1973, 1981).

The method

Let us consider some sets of variables such as X, Y, Z, which describe a set of n communities (C1, C2, C3, ..., Cn). Each set of variables defines a space (SX, SY, SZ) in which the communities are located as points (see Feoli 1984, for a discussion on different vegetation spaces). In such spaces we can compute $n(n-1)/2$ distances (d_{ij}) between the points. The formulas for computing the distances are several (see Orlóci & Kenkel 1985 and reference therein). We let the choice up to the user. In the space SX, SY, SZ, the values of any variable, may be arranged in a random or in an ordered way. Let us consider:

a) the distribution of the values of the variables of one set, e.g. X, in the space defined by the same set (SX),

b) the distribution of the values of variables of one set, e.g. X, in the space defined by another set, e.g. SY or SZ.

In the case a) we will measure the internal predictivity, in the case b) the external predictivity. If a variable shows a random distribution in a space, the ordination of the points in the space will not be predictive with respect to that variable. For measuring predictivity we propose a revised version of Moran's I (symbolism follows Ebdon, 1977):

$$I= [n \Sigma_{(p)} w_{ij} \ (x_i-\bar{x}) \ (x_j-\bar{x})]/(\Sigma_{(p)} w_{ij} \Sigma (x-\bar{x})^2)$$

where x_i and x_j are the values of variable x respectively in relevés or vegetation types i and j, $w = 1/d_{ij}$ is the weight given to the relationship between i and j, where d_{ij} is the distance between i and j, p is the summation index going from 1 to $n(n-1)/2$. $\bar{x}$ is the grand mean.

Two forms of a null hypothesis apply for the significance of I: the hypothesis of normality and the hypothesis of randomization. The first involves the assumption that the values of the variables constitute a random sample of values drawn from a normally distributed population, the second is free from any assumption. In both cases the test is based on the variable:

$$z= (I-EI)/sI$$

where EI is the expected value of I ($EI= -1/(n-1)$) and sI is the standard deviation of I. This is calculated in different ways for the two null hypotheis. The formula of sI, under the hypothesis of randomization, which in general is the 'safer' choice (see Ebdon 1977), is the following:

$$sI\ \sqrt{\frac{n(n^2+3-3n)A+3B^2-nC]-\bar{K}(n^2-n)A+6B^2-2nC]}{(n-1)\ (n-2)\ (n-3)B^2}}$$

where $A=\Sigma_{(p)}w_{ij}^2$, $B=\Sigma_{(p)}w_{ij}$, $c=\Sigma_i(\Sigma_j w_{ij})^2$ and k is the kurtosis of x. The two-tailed critical value of 2 at the significance level of 0.05 is 1.960.

The computation of d_{ij} and the use of x as a single or a composite variable, as for instance principal components, eigenvectors, and/or different kinds of ordination axes, may characterize the application of the method.

Example

Data

The data used for the example are presented in Tables 1,2 and 3. The columns of these tables represent 20 vegetation types of mixed broad-leaved forests of Friuli Venezia-Giulia described by Poldini (1982). In Table 1, the types are described by average indicator values (Landolt 1977), computed using the 5 frequency classes of species (Braun-Blanquet 1964) as weights. The use of average ecological indicator values proved to be useful in several trials

Table 1. Description of 20 forest vegetation types of NE Italy on the basis of average ecological indicator values.

Humidity	1.76	2.07	1.90	2.30	1.96	2.45	2.44	2.34	2.47	2.61	2.69	2.64	2.77	2.82	2.92	2.65	2.68	2.40	2.13	2.03
pH	3.84	3.72	3.91	3.57	3.68	3.42	3.20	3.60	3.52	3.33	3.38	3.10	3.43	3.32	3.37	3.50	3.41	3.64	3.56	3.68
Nutrients	2.28	2.55	2.22	2.46	2.32	2.54	2.45	2.45	2.54	2.62	2.69	2.62	2.87	2.91	3.01	2.64	2.73	2.53	2.27	2.19
Humus	2.90	3.10	2.94	3.10	3.07	3.27	3.28	3.28	3.26	3.37	3.38	3.39	3.44	3.46	3.46	3.32	3.40	3.21	3.05	2.99
Dispersion	3.12	2.95	2.81	3.37	3.17	3.49	3.54	3.32	3.53	3.58	3.61	3.63	3.68	3.66	3.57	3.45	3.49	3.25	3.04	2.83
Light	3.06	2.98	3.16	2.96	3.05	2.82	2.89	2.83	2.80	2.66	2.65	2.66	2.46	2.41	2.35	2.72	2.53	2.86	3.06	3.17
Temperature	4.71	4.46	4.02	4.16	4.29	3.89	3.77	3.84	3.78	3.73	3.87	3.77	3.80	3.64	3.54	3.51	3.50	3.58	3.71	3.62
Continentality	2.60	2.70	3.25	2.85	2.95	2.84	2.83	3.04	2.90	2.84	2.67	2.65	2.60	2.68	2.70	2.94	2.87	3.09	3.19	3.24

Table 2. Description of the 20 vegetation types of Table 1 based on frequencies of the species in the life-growth form. The nomenclature follows Ellenberg and Mueller-Dombois (1965/66). Abbreviations: P = Phanerophytes; Ch = Chamaephytes; T = Therophytes; G = Geophytes; H = Hemicryptophytes; lian = lianas; rept = reptant; suffr = suffruticose; rad = radicigemma (root-budding); bulb = bulbous; rhiz = rhizome; ros = rosulate.

P caesp	37	34	33	39	44	42	28	53	60	51	55	22	43	47	40	46	33	27	40	36
P scap	10	5	5	13	11	20	22	12	26	26	20	22	24	29	28	13	23	10	18	18
P lian	11	6	3	11	10	6	6	5	8	8	7	7	9	10	7	4	8	4	3	0
Ch rept	0	0	0	0	0	0	0	1	2	1	1	3	5	2	3	0	1	0	1	0
Ch suffr	0	0	10	70	71	30	50	80	70	70	30	0	40	20	0	0	20	10	22	2
Ch frut	0	0	0	0	0	1	4	0	0	1	0	1	0	0	0	4	1	2	5	5
T scap	0	3	0	2	2	3	5	0	2	2	3	5	0	0	0	0	0	0	3	1
G rad	0	0	0	3	1	1	0	3	4	3	5	2	3	2	1	0	0	3	0	0
G bulb	2	4	3	0	1	5	1	9	8	5	13	7	15	13	8	4	6	7	6	8
G rhiz	6	7	16	17	25	42	23	47	36	32	43	37	44	57	62	35	61	11	20	21
H caesp	4	7	15	17	16	16	32	24	25	25	24	21	18	15	12	17	19	20	22	26
H rept	0	0	0	1	0	3	2	2	2	5	3	4	6	6	3	1	4	0	1	2
H scap	3	1	42	37	25	60	74	55	58	58	45	60	51	67	57	65	77	41	48	48
H ros	1	0	6	10	6	15	15	10	8	8	11	9	15	12	15	2	11	1	5	5
NP	16	15	10	14	16	9	4	14	18	11	11	7	15	8	8	14	10	10	15	14

Table 3. Description of the 20 vegetation types of Table 1 based on the frequencey of the species in chorological variables. The nomenclature follows Pignatti (1982).

Cosmopolitan	0	0	0	2	0	3	7	2	4	2	4	5	3	10	10	5	7	0	1	2
Circumboreal	8	12	13	10	13	22	20	18	18	21	11	15	22	21	25	15	21	13	13	17
Eurasiatic	7	3	12	17	16	30	34	33	43	38	36	37	36	43	36	33	40	25	27	23
Eurosyberian	0	0	1	15	7	20	20	20	18	23	27	30	19	24	21	24	28	14	20	15
Euopean	10	10	16	44	33	54	58	53	70	69	61	60	74	87	80	48	67	29	38	34
Arctic-Alpin	0	0	0	0	0	0	0	2	0	1	0	0	0	0	1	0	1	1	1	0
Pontic	11	12	21	12	21	20	10	24	25	16	21	10	13	13	5	8	15	2	14	13
Medit.-Pontic	6	6	5	6	6	6	5	10	9	6	10	7	8	2	5	9	6	7	8	7
Eurymedit.	12	10	4	22	20	14	10	12	10	7	19	11	17	12	6	1	5	6	5	0
Stenomedit.	20	4	1	4	6	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Medit.-Atlantic	2	4	6	5	5	6	5	7	11	7	8	8	11	9	7	4	5	2	3	6
Subatlantic	0	2	3	4	4	2	4	3	0	0	0	0	0	0	0	1	1	3	3	4
Medit.-Mont.	2	4	21	1	6	6	8	26	23	15	12	6	13	17	23	34	31	27	45	47
S-Illyrian	3	5	15	1	7	2	2	1	0	0	0	0	0	0	0	2	1	1	2	9
N-Illyrian	0	0	0	0	0	0	0	1	0	0	0	0	4	2	7	1	1	0	4	5
S-SE-European	3	5	9	13	12	18	20	9	10	12	12	8	8	13	6	1	5	1	4	0
Alpine	0	0	0	0	0	2	2	1	1	0	0	2	0	1	2	7	4	0	1	5
E-Alpine	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
Alpine-Carp.	0	0	0	0	0	1	0	0	0	0	0	1	4	0	0	0	0	0	0	0
Endemic	0	0	0	0	0	0	0	3	3	3	3	3	5	4	3	4	4	2	5	6
Paleotemperate	3	6	10	9	8	16	12	16	18	18	12	6	12	10	7	10	13	5	10	11
Paleosubtrop.	3	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
European-Caucasic	0	0	0	6	4	5	7	0	0	1	5	2	1	0	0	0	0	0	1	0

Table 4. Values of z (see the text) for the variables in Tables 1, 2 and 3, within the space which they themselves define (internal) and within the space defined by the other variables (external). Ecol.= environmental variables; Biol.= life-growth forms; Chorol.= chorological variables.

	Internal	External	
	Ecol.	Biol.	Chorol.
Humidity	4.6888	–	3.1928
pH	3.5786	–	2.9631
Nutrients	4.3848	–	2.5797
Humus	4.7042	–	3.3895
Dispersion	4.6452	–	3.7367
Light	4.7940	–	3.0718
Temperature	2.5825	2.3039	2.4242
Continentality	2.9398	–	2.3116
	Biol.	Ecol.	Chorol.
P caesp	1.9639	–	–
P scap	–	3.2603	3.0303
P lian	–	–	–
Ch rept	–	2.1550	–
Ch suffr	–	–	–
Ch frut	–	–	–
T scap	–	–	–
G rad	–	–	–
G bulb	–	2.0988	–
G rhiz	2.2047	3.0811	2.3471
H caesp	–	–	–
H rept	–	3.6947	3.0313
H scap	2.3092	2.9001	2.4065
H ros	–	2.0563	2.6770
NP	–	–	–
	Chorol.	Ecol.	Biol.
Cosmopolitan	1.9958	2.9843	–
Circumboreal	2.1221	2.0291	–
Eurasiatic	3.3652	3.9030	2.1266
Eurosiberian	2.6684	3.4748	–
European	3.6948	4.0472	2.2385
Arctic-Alpine	–	–	–
Pontic	–	–	–
Medit.-Pontic	–	–	–
Eurymedit.	2.0661	–	–
Stenomedit.	–	–	–
Medit.-Atlantic	2.3645	–	–
Subatlantic	–	2.9512	–
Medit.-Mont.	2.7185	–	–
S-Illyrian	2.1050	2.8444	–
N-Illyrian	–	–	–
S-SE-European	2.7068	–	–
Alpine	–	–	–
E-Alpine	–	–	–
Alpine-Carp.	–	–	–
Endemic	–	–	–
Paleotemperate	–	–	2.4744
Paleosubtrop.	–	–	–
European-Caucasic	–	–	–

(Persson 1981; Feoli & Lagonegro 1982; Jurko 1983). Interesting to note, that linear relationships were observed by several users between the indicator values and actual instrument readings (Feoli, Nasimbeni & Cicigoi 1982; Bezzi, Feoli, & Orlandi 1984; Lagonegro & Feoli 1985). In Tables 2 and 3, the types are described respectively by the sum of frequency classes in the Raunkiaer's life-growth forms (Mueller-Dombois & Ellenberg 1974) and in the chorological variables (nomenclature follows Pignatti 1982). Tables 2 and 3 are respectively, structural and chorological descriptions of the vegetation types. Such descriptions are widely used to interpret vegetational variation in terms of dynamics and hystorical facts.

Computations

The index I has been computed for each variable of the three matrices in Tables 1, 2 and 3. The d_{ij} values have been calculated by the formula of chord distance (Orlóci 1978) applied to the raw data. If, instead of raw data, the scores of ordination axes would be used, the index I will measure the predictivity of the axes. The Lagonegro and Feoli (1985) program package have been used for calculating the matrices of chord distances. The computer program for I and its significance test is part of a broader package for spatial pattern analysis written by Paola Ganis. The listings are available from her on request free of charge.

Results

The results of the analysis are presented in Table 4. Only the significant values of z are considered further. The internal predictivity values are given in the first column, the external in the other two columns.

The internal predictivity is very high for the first set of variables, they are all arranged in a non-random way in the space which they themselves define. The predictive value is much lower for the chorological variables and very low for life-growth forms. As far as the external predictivity is concerned, the life-growth forms have very low predictivity for the environmental variables. In the biological space only the values of temperature follow a non-random pattern of dispersion.

Contrasting with this, in the space of chorological variables all the environmental variables have a non-random dispersion. On the basis of Table 4 we can conclude that the environmental variables are more predictive with respect to chorology and structure of vegetation, than vice versa. The environmental variables are also more predictive than chorology with respect to structure, and structure with respect to chorology. In the considered vegetation system, chorology is more predictive than structure. This means that the differences between the types are not sharp on the basis of the structural characters used, i.e. life-growth forms.

Conclusions

The method of autocorrelation described in this paper is efficient to detect significant relationships between sets of variables. It is freer from restrictive assumptions which so heavily burden discriminant and canonical correlation analysis. For this reason the autocorrelation method can be considered as a complement of such methods in the explorative phase of data analysis. The variable z can be considered as a measure of intraset or interset correlation, depending on whether it is computed for internal or external predictivity.

The method can be used with ordination axes that may be linear or non-linear combinations of variables, in which case z will measure the pattern of composite variables in different spaces. At this point it is evident that the possible applications of the method are several. We believe it will be useful especially in interpreting ordinations, when the values of one variable, simple or composite, is superimposed onto the points of a scattergram.

Acknowledgements. The work has been supported by a grant of the Italian CNR within the project «Structure and dynamics of forest». We thank D.W. Goodall and L. Orlóci for reading and commenting on the text.

REFERENCES

BEZZI, A., FEOLI, E. and ORLANDI, D. 1984. Sintesi sulla vegetazione degli alpeggi della Val Rendena (Trento). Ann. Istituto Sperim. Assest. For. e Alp. 8: 237-287.

BRAUN-BLANQUET, J. 1964. *Pflanzensoziologie. Grundzuge der Vegetationskunde.* 3 Aufl. Springer-Verlag, Berlin.

CLIFF, A.D. and Ord, J.K. 1973. *Spatial autocorrelation.* Pion, London.

CLIFF, A.D. and Ord, J.K. 1981. *Spatial processes. Models and applications.* Pion, London.

EBDON, D. 1977. *Statistics in Geography,* Basil Blackwell, Oxford.

ELLENBERG, H. and MUELLER-DOMBOIS D. 1965/66. A key to Raunkiaer plant life forms with revised subdivisions. Ber. Geobot. Inst. ETH 37: 56-72.

FEOLI, E. 1983. Predictive use of classification and ordination methods in plant community ecology. A summary with examples. In: FERRARI C. (ed.), *Le comunità vegetali come indicatori ambientali.* Studi e Documentazioni Regione Emilia-Romagna, Bologna.

FEOLI, E. 1984. Some aspects of classification and ordination of vegetation data in perspective. Studia Geobotanica 4: 7-21.

FEOLI, E. and SCIMONE, M. 1984. A quantitative view of textural analysis of vegetation and examples of application of some methods. Archivio Botanico e Biogeografico Italiano 60: 73-94.

FEOLI, E. and LAGONEGRO M. 1982. Syntaxonomical analysis of beech woods of Apennines based on the program package IAHOPA. Vegetatio 50: 129-174.

FEOLI, E., NASSIMBENI, P.L. and CICIGOI, G. 1982. I prati falciabili delle Valli del Natisone. II. Correlazione tra suolo e vegetazione. Centro Regionale per la Sperimentazione Agraria per il Friuli Venezia-Giulia, Udine.

JURKO, J. 1983. Ecomorphological survey of the leaf characteristics in selected Central European plant communities. Folia Geobot. Phytotax. 18/3: 287-300.

LAGONEGRO, M. and FEOLI, E. 1985. *Analisi multivariata di dati. Manuale d'uso di programmi BASIC per personal computers.* Libreria Goliardica, Trieste.

LANDOLT, E. 1977. Okologische Zeigerwerte zur Schweizer Flora. Ber. Geobot. Inst. ETH, 64: 64-207.

MUELLER-DOMBOIS, D. and ELLENBERG. 1974. *Aims and methods of vegetation ecology.* J. Wiley, New York.

ORLÓCI, L. 1978. *Multivariate analysis in vegetation research.* Junk, The Hague.

ORLÓCI, L. and KENKEL, N. 1985. *Introduction to data analysis with examples from population and community ecology.* Statistical Ecology Monographs, vol. 1. International Cooperative Publishing House, Maryland.

PERSSON, S. 1981. Ecological indicator values as an aid in the interpretation of ordination diagrams. J. Ecol. 69: 71-84.

PIGNATTI, S. 1982. *Flora d'Italia.* Calderini, Bologna.

POLDINI, L. 1982. Ostrya carpinifolia-rich woods and bushes of Friuli Venezia-Giulia. Studia Geobotanica 2: 69-122.

Manuscript received: October 1985