COENOSES 6 (1): 3-13, 1991

REFLECTIONS ON SPACES AND RELATIONSHIPS IN ECOLOGICAL DATA ANALYSIS: EFFECTS, PROBLEMS, POSSIBLE SOLUTIONS

S. Camiz, University of Roma, Via S. Francesco a Ripa, 57, I-00153 Roma, Italy

Keywords: Ecological spaces, Data analysis, Nonlinearity, Generalised co-ordinates

Abstract. The horse-shoe effect, that results when linear analysis methods are used with nonlinear relationships between species and environmental factors, is considered. Limits of the classical eigenmethod are examined and their use is suggested for exploratory purposes. The affine space model, that underlies eigenmethods, is considered unsuitable for indirect gradient analysis and a better mathematical model is suggested, based on generalised co-ordinates. The technical difficulties of this approach are considered. Suggestions are made to investigate experimentally the actual relationships of characteristic species to environmental factors.

Introduction

The analysis of ecological data usually deals with matrices containing, for each sampled relevé, either species abundance or presence/absence, ecological measures or some combination of these. The aims of the analysis are both classification of relevés, in order to arrange them according to a syntaxonomy, already known or yet to be defined, and ordination along ecological gradients, considered causal for β-diversity (Whittaker 1967).

The increasing diffusion of data analytical methods in conjunction with the spread of modern computing equipment, which made easier data analysis through well-known statistical methods (see, e.g., Dagnelie 1965 a, b, Goodall 1970), raised questions about the adequacy of these methods in applications to ecological data. The initial ordination methods introduced in the 1950's were principal components analysis (PCA) (Goodall, 1954), an eigenmethod, and Bray and Curtis (1957) ordination (BCO), an ordering of relevés on a probable ecological gradient, through a very simple algorithm. These methods represent a first approximation to indirect gradient analysis (Whittaker, 1967), PCA more analytical and BCO more intuitive. Unfortunately, indirect gradient analysis, performed with eigenanalysis tools, such as a PCA, proved to be inadequate, because the underlying mathematical model does not fit the actual species/environmental factor relationships. The recognition of this inadequacy led to comparative studies (e.g., Austin and Orlóci 1966, Orlóci 1966, van der Maarel 1969, Austin and Noy-Meir 1971, Gauch and Whittaker 1972, Kessel and Whittaker 1976, Austin 1976b, Gauch, Whittaker and Wentworth 1977), in order to discover which method, PCA or BCO, gives better results.

From the very beginning, suspicions that PCA may not be an effective method were raised (Goodall 1963, van Groenewoud 1965, see also Guttman 1953), confirmed later by many workers who dealt with simulated data and brought into focus the 'horse-shoe' effect (Austin and Noy-Meir 1971). This effect notwithstanding, Noy-Meir (1971, 1974) points out that PCA can be used successfully when the analysis is limited to the identification of phytosociological entities, since each such entity can be delineated through a linear combination of variables, and the multilinear model that underlies most factor analyses holds. The inadequacy rises when the researcher's interest turns toward indirect gradient analysis, since the environmental effects should not be considered as having linear constraint on species performance.

The horse-shoe effect can then be explained as the result of the representation into linear space of the nonlinear relationship between species performance and environmental factors (van Groenewoud 1965, Orlóci 1980, Fewster and Orlóci 1983, see also Heiser 1987). The recognition of this led to attempt methodological improvements to increase reliability, if evidence of environmental factors through indirect gradient analysis is required. Whereas the development of some early intuitive simple methods, such as BCO, and its various modifications (Orlóci 1966, Goff and Cottam 1967), even if still supported by some authors (Beals 1984), seems now to have reached the end of their utility, the investigation on more suitable eigentechniques followed several directions:

1. improving existing methods, through data transformations prior to eigenanalysis (Benzécri 1973-82, Ihm and van Groenewoud 1975);

2. looking for ways to reduce the horse-shoe effect following the analysis (Hill and Gauch, 1980);

3. adapting nonmetric multidimensional scaling (NMMDS) methods to weaker underlying hypotheses (Fewster and Orlóci 1983);

4. developing methods which better conform with

the response/factor relationship model (Gauch and Chase 1974, Gauch, Chase and Whittaker 1974).

The introduction of canonical contingency table analysis, known broadly as correspondence analysis, in ecological analyses (Hill 1973, Feoli and Orlóci 1979, Orlóci 1981, ter Braak 1984, 1985, Orlóci and Orlóci 1988), seems a best choice among the methods based on the multilinear models. However, as a double standardisation is involved, a complex nonlinear transformation is introduced. Analysis of simulated data seem to confirm a reduction of the horse-shoe effect (Kenkel and Orlóci 1986), although depending on the β-diversity along the considered gradients. A suggestion for detrending *a posteriori* comes from Hill and Gauch (1980) and ter Braak (1984 and 1985), in order to completely remove any nonlinear trend considered as a product of the method of analysis.

In another direction, based on the evidence that the Gaussian model could be a reasonable approximation for a nonlinear response type, methods were developed known as Gaussian ordination (Gauch and Chase 1974, Gauch, Chase and Whittaker 1974, Johnson and Goodall 1980, Fewster and Orlóci 1983). Some criticised the Gaussian choice (Austin, 1976a, b), arguing that nothing inherent dictates a Gaussian type response and in fact species can follow several different response models which may be skewed or even bimodal, etc. Furthermore experiments followed, dropping the eigentechniques in favour of Nonmetric Multidimensional Scaling (NMMDS). They are based on iterative procedures aiming at recovering the data pattern in a reduced dimensional space, without any underlying hypothesis, and seem to be suitable for the purpose of representing relevés in environmental factors' spaces (Fewster and Orlóci 1973, Gauch, Whittaker and Singer 1981, Orlóci, Kenkel and Fewster 1984, Minchin 1985, Kenkel and Orlóci 1986, Bradfield and Kenkel 1987). The recent developments are summarised in Table 1; for further reviews and discussions, see, e.g., Noy-Meir and Whittaker (1977), De Leeuw (1987), and Heiser (1987).

A review of the work performed up to present on the matter, an effort of investigation is apparent, aiming to represent relevés in environmental factor spaces, through species performance values. Interestingly, some (Feoli and Feoli Chiapella 1980) find the horse-shoe effect ecologically informative for revealing non linear trends. I will try then to organise the matter in this paper and to draw some conclusions.

The geometrical model of factor analyses

We are considering an n-dimensional affine space **E**, which is a couple formed by a set **S** of elements, called points (see, e.g., Bourbaki 1962, Godement 1965), and an associated n-dimensional vector space **V**, such that, given any point O belonging to **S**, a 1 to 1 correspondence exists between the points of **S** and the vectors of **V**, on the condition that:

1) O of **S** corresponds to the null vector of **V**;

2) **E** is invariant under linear transformations and origin translation.

With this model, if linear relationship assumptions could be made, we could imagine to represent the relevés as points in an affine space, sustained by different vector spaces, let they be spanned by species,

Table 1. Proposed solutions to indirect gradient analysis, according whether the functional from of the response is assumed or not and, in this case, to the type of adjustment.

Functional form of response		
Assumed	not assumed	
	Double adjustment	Other
McDonald (1962, 1967) polynomial factor analysis	Hill (1973) reciprocal averaging	Orlóci, Kenkel and Fewster (1984), Kenkel and Orlóci (1986) Multidimensional scaling, chord as external distance
Ihm and Van Groenewoud (1965, 1984) eigenanalysis	Feoli and Orlóci (1979) concentration analysis	Bradfield and Kenkel (1987) Minimum spanning tree distance
Gauch and Chase (1974) Gaussian model	Hill and Gauch (1980) detrended correspondence analysis	Von Rijkevorsel (1987) Homogeneity analysis and fuzzy coding
Phillips (1978) polynomial ordination		
Johnson and Goodall (1980) curve fitting		
Fewster and Orlóci (1983) Nonmetric multidimensional scaling		

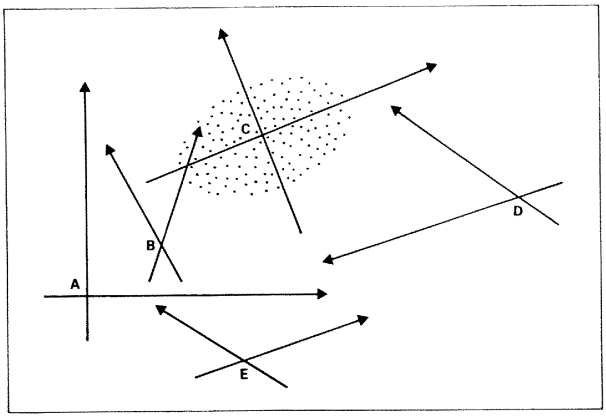

Fig. 1. Five different bases of vector spaces, having different origin, but all sustaining the 2-dimensional affine space corresponding to this page. The set C, resulting from Principal Components Analysis of the dotted points, aiming at optimising the representation of the points' inertia, is considered more suitable for their representation, in that each vector can be interpreted as a factor influencing the points' scattering.

environmental factors, life forms, or other collections of objects (Fig. 1). While this idea frees the model from the constraint that all spaces have the same zero value, it also implies that the spaces are isomorphic, have the same dimensions, as the true dimensions of the affine space. The classical analysis proceeds in the following way:

1) consider the species as generating a vector space sustaining the affine space of relevés;

2) define a scalar product, in order to transform the affine space into a Euclidean space;

3) extract an orthogonal base for the vector space, via an eigenanalysis technique.

As a result, the found dimension of the base formed by the eigenvectors should correspond to the dimension of any other space sustaining the affine space, and consequently there should be a correspondence between the reduced dimension subspaces spanned by the first eigenvectors of the analysed space and subspaces of the other vector spaces. This is, in practice, what we mean when we "interpret" the extracted factors.

It must be pointed out that the use of PCA as confirmatory analysis or as a modelling tool, can raise serious questions. In fact, it is to be noted that the extracted principal component axes are unique to the data and should not be interpreted as real factors. Being optimal for the task of representing data in a reduced dimensional space, the axes are only *suggestive* of the factors (or more likely their combinations) that actually influence the sampled phenomenon under study, and not the factors themselves. On the other hand, it is evident that, although the factorial axes are linear correlations of the original variables, any non linear trend that can be revealed by inspection of the pattern of points, can be interpreted as a nonlinear factor that is likely to be causal for the phenomenon under investigation. Therefore, the use of PCA should be limited to exploratory analyses, and in this context it is highly appropriate.

In order to confirm what the analysis has actually shown, or to formalise it in a mathematical model, one must perform more specific analyses, such as analysis of variance, regression or discriminant analysis, using data collected for the purpose, checking conditions otherwise hard to substantiate, such as random choice in sampling, random distribution of residuals, multinormality of distributions, or, in case of eigen analyses, linear relationships of variables. In addition, if nonlinear trends are to be investigated, the functional form of the trend must be determined empirically, through nonlinear regression or curve fitting, and theoretically justified. These aspects must be stressed since the use of an exploratory and descriptive tool, such as PCA, is far from being an ideal tool for the purpose.

Environmental data

Let us consider a table **Y**, where each element y_{ij} is the measure of the level of an ecological factor j in relevé i, i = 1, ... r, j = 1, ... f. Each matrix row $\mathbf{Y}_i$ is a vector in the f-dimensional space of ecological factors F and each column $\mathbf{Y}_j$ is a vector in the r-dimensional relevé space, both of being subspaces of the spaces of *all* the ecological factors **F** and *all the possible* relevés **R**, respectively.

The factors extracted in an eigenanalysis are linear combinations of the environmental variables measured, so that they can give predictions of non-measured environmental factors, linearly related to the measured

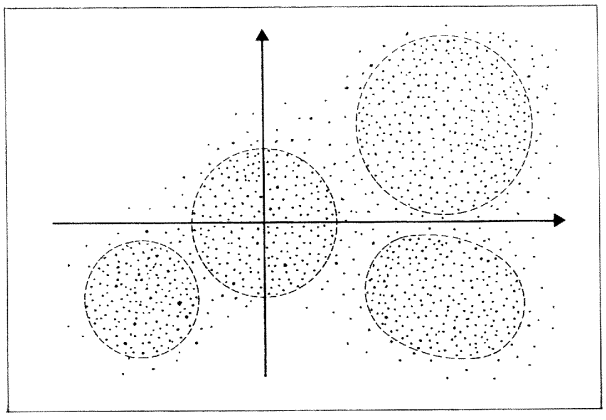

Fig. 2. Factors are extracted from environmental variables and relevés are scattered in the whole factor space. Relevé groups, within dotted lines, are likely to correspond to phytosociological entities. No horseshoe effect is present.

ones (Fig. 2). In addition, the examination of covariance or correlation matrices, and the pattern of variables in the factor spanned space provide reliable information of the correlation among the measured factors, that can be interpreted as a non-random sample correlation among the factors under investigation.

In this case, the aim of the analysis is to order the relevés along factors, and to identify non measured underlying factors via eigentechniques. Data transformations, ordination, and classification are usually done. Discriminant analysis is used when a reference classification exists based on external criteria. However, unless special care is taken both in sampling and in analyses assumptions, it must be stressed that the analysis results are not to be inferred to a population, and not even the axes may be considered as the true factors.

Vegetation data

Let us consider a table **X**, where each element x_{ij} of which is a transformation of the abundance of species j in the relevé i, i = 1, ... r, j = 1, ... s. Such transformation can mean simply 1 for species presence and 0 for absence (PA), the Braun-Blanquet (Westhoff and Van der Maarel 1973) code 0, r, +, 1, ..., 5 (BB), species density, i.e., the number of species individuals (SI), or percentage species cover (CP). Such a table can be considered as a contingency table, treated as a data matrix. In addition, each matrix row $\mathbf{X}_i$ is a vector in the s-dimensional species space **S** and each column $\mathbf{X}_j$ is a vector in the r-dimensional relevé space, both being a subspace of the spaces of *all* the species populations **S** and *all* the possible relevés **R**, respectively.

The aim of the analysis of such a table is two-fold: on the one hand the matrix is rearranged into blocks, such that the relevés of similar floristic composition are grouped together and the co-occurring species are placed side by side: for this a suitable clustering technique is needed. On the other hand, an understanding of the ecological reasons of the variation among the table blocks is attempted, namely to identify the ecological gradients along which the species populations undergo trended variation, and along which the relevés can be arranged in a natural order. For this aim, the idea of factor extraction via an eigenanalysis is seducing. Indeed, a most coherent procedure could be a combination of these methodologies, so that a successful analysis should give both an ordination of relevés along understood gradients and a syntaxonomical classification (Camiz 1988). Translating these aims into the language of affine spaces, it would mean to look for an orthogonal base of the subspace of species space, corresponding to orthogonal factors in the corresponding subspace of environmental factors: in this latter space, species should be represented by vectors and relevés by points, and the clusters should correspond to density phases in the cloud of points.

In fact, PCA gives co-ordinates to relevés that are correlated to how each *extracted* factor influenced the relevé diversity. Whereas in an analysis of environmental factors this could mean the exact position of a relevé according to extracted factors, that can be considered as compounds of environmental factors, in this case, the relevé scores should represent the share of a phytosociological entity in the relevé. This is unlikely to be distributed at random on the factor space (as it should be expected, if no other underlying cause would exist), because 1) different entities are likely to be mutually exclusive in nature, 2) the sampling of relevés, for phytosociological purposes, far from being a random sampling, is performed according to the homogeneity of vegetation (for further comments, see Goodall and Feoli 1988), so that it is likely to strengthen the group pattern of the data table, forcing to mutual exclusiveness of the entities, and 3) the relevés are very often fully attributable to different phytosociological entities (unless fuzzy attributions are allowed, see Roberts 1986, Feoli and Zuccarello 1986, 1988 and Dale 1988).

The arranged position on the factor space of phytosociological entities reflects these matters, and PCA usually places entities' centroids close to the factors, and on opposite sides of a factor those which are mutually exclusive. The procedure is such that, given three entities A, B, C, A and B being mutually exclusive and C partially co-occurrent with A and B, in a PCA scatter diagram on the first two axes, A and B will be likely placed at the opposite extremes of the first axis, and C at the extreme of the second, centred on the first axis' origin. A clustering technique, based on the first set of factor scores, is often able to partition the relevés into groups that can be identified as phytosociological associations. If intermediate relevés are considered, the pattern has a horse-shoe shape, that can be interpreted as an indirect manifestation of a nonlinear environmental factor, as far as the diversity of the phytosociological entities can be interpreted in such terms. If such a factor exists, phytosociological entities, as well as species, are supposed to occupy a niche, described by a nonlinear function (Fig. 3).

It is therefore evident that through PCA, the aim to get principal axes corresponding to environmental factors is far from being reached, owing to the complex co-occurrence pattern and correlation characteristics of the ecological factors that normally influence plant growth. Of course, the analysis is complicated by both the fuzzy and irregular border that separates different syntaxonomic units in nature and by the noise that is always present in data. Naturally, an analysis simply aiming at relevé assignments to syntaxonomical units, can be easily developed, either through the identification of characteristic species of known units, or through cluster analysis and *a priori* or *a posteriori* original data transformations, which modify the relative weight of

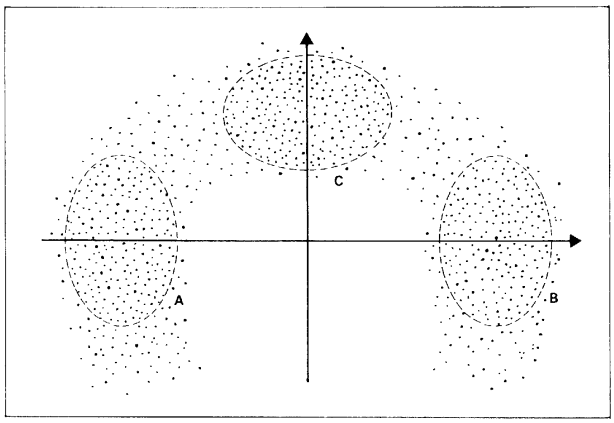

Fig. 3. PCA scatter diagram of relevés sampled along a major ecological factor. Groups A, B, and C may represent mutually exclusive phytosociological entities. The other points may be transitional relevés. The principal axes do not represent ecological gradients. The horseshoe effect present can be interpreted as the main ecological factor.

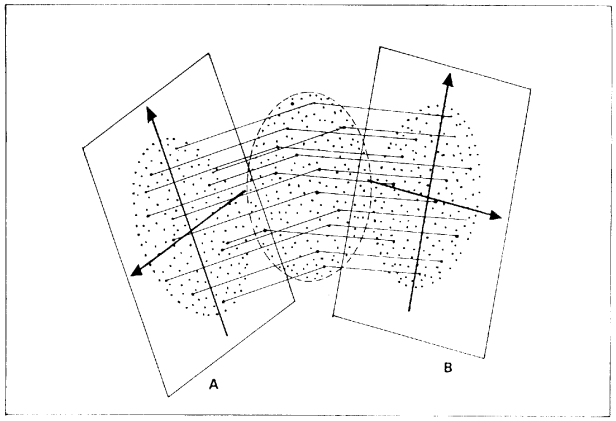

Fig. 4. Canonical correlation analysis extracts couples of orthogonal axes from two sets of variables. Each couple must be as correlated as possible. The correlation coefficient is a measure of the goodness of the linear representation by the two sets.

the species. In this regard, PCA (as well as CA, or NMMDS, that will be discussed later), resulting in a dimension reduction, can be used successfully. Compared to mere identification and cluster analysis, the latter procedures have the advantage of allowing a dramatic graphical representation of the relevés' scattering and the species' correlation pattern. The interpretation of the axes through species and environmental, (if applicable) correlations can result in the recognition of syntaxonomical units, suggesting a vegetation environmental basis for the species space.

In any case, it must be kept in mind that *all* methods of analysis can recover most of the information present in data set, but not information *outside*. This means that any analysis can synthetise information, but cannot equate the computed gradients or syntaxa (i.e., clusters) with true gradients or syntaxa in nature. For this reason, interpretation of results is subjective, left entirely to the researcher, at least until an ecologist's knowledge is imparted through professional programming to an expert system (Camiz 1988).

Canonical correlation analysis: combining the analysis of vegetation and environmental data

Canonical correlation analysis (see, e.g., Gittins 1985) aims at defining the best linear relationships between two sets of variables, and vegetation or environmental. Eigenanalysis techniques of projection spaces are the usual departure points. The canonical co-ordinates can be interpreted as the best representation of the synthesis of one set performed by the other (Fig. 4). Whereas this method can be used successfully among different sets of environmental variables, it could be misleading if used for understanding relationships between species distributions and ecological factors, because of the usual nonlinear relationship between the two. In fact, when using canonical analysis, we should keep in mind that each factor belongs to a subspace spanned by species performances, and each species performance belongs to a subspace of environmental factors, both being a linear combination of the other space vectors. If this is not the case, the two spaces **A** and **B** are not subspaces of the same vector space and we cannot consider interaction between species and factors as linear correlations.

It is interesting to note that, limiting the analysis to a relevé by species table, a particular canonical analysis can be performed. Correspondence analysis (CA), in fact, can be considered as a canonical analysis, suitable for the treatment of contingency tables, as no cause/effect relationship is expected to exist between rows and columns. Here, the vector spaces involved are both species and relevé spaces, considered as subspaces of the space of all the existing plants. Actually, any sampled plant is assigned to a species, as well as to a quadrat where it was found. The analysis then tends to explain if and how the same phenomenon, namely the table structure, can be similarly described by the two subspaces, that is, what the named subspaces have in common. As relevés can be considered environmental factor dependent, we are back to the relationship between species and environmental gradients, although the underlying geometry is not so evident, as a double standardisation of the table is performed prior to the computation of distances.

It could be due to this obscure aspect that the horseshoe was not considered by Hill and Gauch (1980) a manifestation of nonlinear trends, but an aberration of the method when they proposed Detrended Correspondence Analysis (DCA). While Hill and Gauch (1980) remove the houveshoe analytically (Fig. 5), ter Braak

(1984 and 1985) forces the detrended factors to be maximally correlated with introduced environmental factors, through regression techniques.

Actually, their aim is to obtain a representation of relevés in environmental factor space, through the known position of relevés in species space, via subsequent adjustments of the relevé factor scores, in order to remove nonlinear correlations among the extracted axes. Unfortunately, this heavy manipulation is likely to add uncontrolled artificial distortion (Greenacre 1984), maybe in the opposite sense of that introduced by previous data manipulations. It is evident that investigation of the actual amount of distortion affecting the results, attributable to each of the manipulations of data during the analyses, should be considered in the light of: 1) double standardisation prior the analysis, 2) eigenanalysis, and 3) detrending.

The generalised co-ordinate spaces: a possible solution

As a cause/effect relationship exists between environmental factors and species performances, if we want to build a geometrical model we must represent the species performances in environmental factor space, considered as the reference space; in this space, species performance with respect to each factor is an uncertain curve that is usually approximated by a Gaussian curve. In this model, distributions of species with respect to an ecological factor have the form of a series of asymmetrical curves which overlap along the gradient and about which the individual performances are randomly distributed. Dealing with two or more factors, the species distributions appear as a set of curved surfaces overlapping on the whole plane.

Such a situation, could be viewed in the frame of generalised co-ordinates such as latitude and longitude on the earth's surface, or even ordinary cartesian co-ordinates in 3-dimensional space, if considered as intersections of orthogonal planes (Fig. 6) as suggested by Feoli, Lagonegro and Zampar (1982) for which they present the computer program PUPOLI (Lagonegro and Feoli 1984). Let E be an n-dimensional space, each point i corresponding to an n-tuple (e_1, ..., e_n) of co-ordinates. Let us consider now n families H_j of (n-1)-dimensional hypersurfaces in that space, each hypersurface h_i in a family being described by the variation in a parameter p_{ij}, $j = 1, \ldots, n$, such that

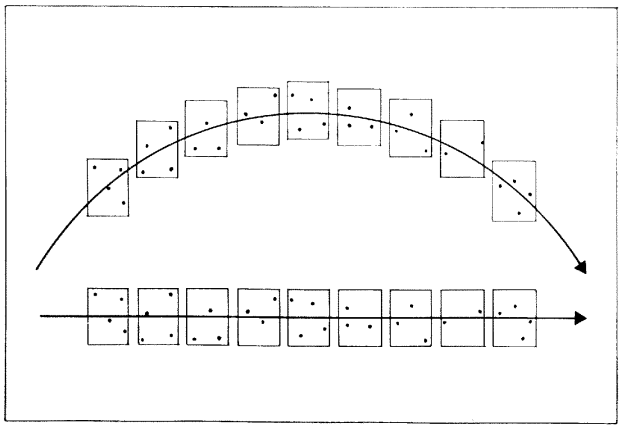

Fig. 5. The procedure of detrending as described by Hill and Gauch (1980): relevés are divided into groups and each group is detrended by projection. Relevés of two different groups may obtain distorted reciprocal position.

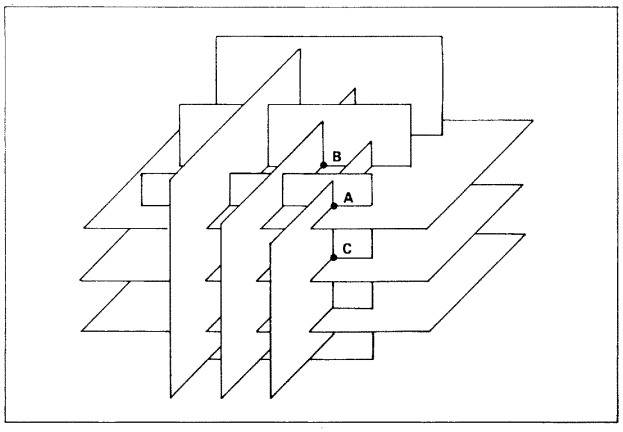

Fig. 6. Cartesian 3-dimensional space represented as having three generalised coordinate sets. Each point is the intersection of three planes, each parallel to coordinate planes (x, y), (x, z), and (y, z) respectively.

1) each point of the space belongs to at least one hypersurface of each family;

2) the intersection of n hypersurfaces, each belonging to a different family, is a point (A, B, C, Fig. 6);

3) two hypersurfaces of the same family do not intersect;

4) in all points the equations of the family of hypersurfaces may be written in explicit form;

5) exceptions to previous conditions are acceptable only in a number of cases that remain undetermined, or determined conventionally.

Under these conditions, for each point i, the value p_{ij} of the parameter identifying the hypersurface of the j-th family where point i lies, can be used as a generalised co-ordinate of the space. In fact, each n-tuple ($p_{h1} \cdots$, p_{hn}) corresponds to a particular point and each point corresponds to an n-tuple. The exceptions are only the points where two hypersurfaces, belonging to the same family intersect, as they can have the co-ordinates of both the intersecting hypersurfaces. As an example, each point of the physical space, on the earth's surface is defined by its altitude, latitude and longitude: altitude is conventionally referenced from sea (zero) level, so we can consider it as a family of spherical concentric surfaces; latitude is an arch distance from the

equator line, so it can be represented as a family of semicones having vertices on the centre and axis along the z-axis; longitude, an arch distance from Greenwich meridian, is a family of half planes all bordered by the z-axis. If we consider now an orthogonal system of co-ordinates, centred on the earth's centre, with the x-axis connecting the centre with the intersection on earth surface of the equator with the Greenwich meridian, the y-axis connecting it with the intersection of equator with the 90° meridian East, and the z-axis connecting the centre with the North pole, we obtain the following equations for the three hypersurface families (that we now approximate, considering the earth's zero level as a perfect sphere):

$$(x^2 + y^2 + z^2) - h^2 = 0, \quad (1)$$

$$(z / \text{sqrt}(x^2 + y^2)) - \text{tg}(\text{lat}) = 0, \quad (2)$$

$$-90° \leq \text{lat} \leq +90°,$$

$$(y / x) - \text{tg}(\text{long}) = 0, \quad (3)$$

$$-180° \leq \text{long} \leq +180°,$$

where all the hypersurfaces of each family are described by the variation of the parameters h, lat, and long respectively (Fig. 7). It is evident that equation (2) becomes indeterminate in the earth's centre ($x = y = z = 0$), as does equation (3) along the z-axis ($x = y = 0$). We can conventionally set to zero the undetermined parameter in those cases. Based on equations (1), (2), and (3), we can derive the new co-ordinates, by expliciting the parameters (condition 4 must hold, then), so that they become,

$$\text{altitude: } h = \text{sqrt}(x^2 + y^2 + z^2) - \text{earth radius} \quad (4)$$

$$\text{latitude: lat} = \text{arctg}(z / \text{sqrt}(x^2 + y^2)) \quad (5)$$

where lat > 0 = North; lat < 0 = South

$$\text{longitude: long} = \text{arctg}(y / x) \quad (6)$$

where long > 0 = East; long < 0 = West

Cartesian co-ordinates can be derived from these equations, by solving the system in x, y, and z; in the same way they can be considered as families of planes parallel to y-z-plane, x-z-plane, and x-y-plane, respectively. The system (1), (2), (3) must be solvable with respect to x, y, z:

$$x = (h + \text{earth radius}) * \cos(\text{lat}) * \cos(\text{long}) \quad (7)$$

$$y = (h + \text{earth radius}) * \cos(\text{lat}) * \sin(\text{long}) \quad (8)$$

$$z = (h + \text{earth radius}) + \sin(\text{lat}) \quad (9)$$

Here, each equation is to be considered as the equation of a set of parallel planes in the earth geographical co-ordinates, with varying parameters x, y, z, respectively.

Looking at the systems (1), (2), (3), and (7), (8), (9), we can state that it is much easier, even in this simple example, to write equations of curve surfaces in cartesian co-ordinates, than those of linear manifolds based on nonlinear coordinates. Nevertheless, transformations of co-ordinate systems are possible in this context, provided that we know the equation of the families of hypersurfaces as defined by the other set; we may also consider the inverse problem, that is having the co-ordinates of points in a set transformed them into co-ordinates of the other set, provided that we have inverse transformation equations.

Such a frame of reference applies in the case of relevés: in the space of ecological factors **F**, each relevé corresponds to an f-tuple of measures taken from Y, as ordinary co-ordinates, as well as an s-tuple of species cover/abundance values, to be considered as generalised co-ordinates, in the sense that the set of points where a given species has a given cover value is a hypersurface (Fig. 8).

If we knew the equations that define the species hypersurfaces as functions of environmental parameters (Johnson and Goodall 1980, Fewster and Orlóci 1983),

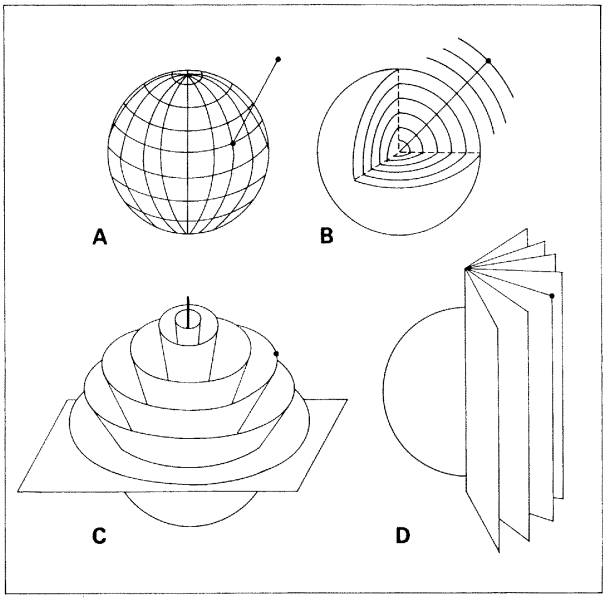

Fig. 7. The three coordinates for measuring positions of a satellite (A): altitude is represented by concentric spheres (B), latitude by cones having the polar axis as common axis (C), and longitude by semiplanes intersecting in the polar axis (D).

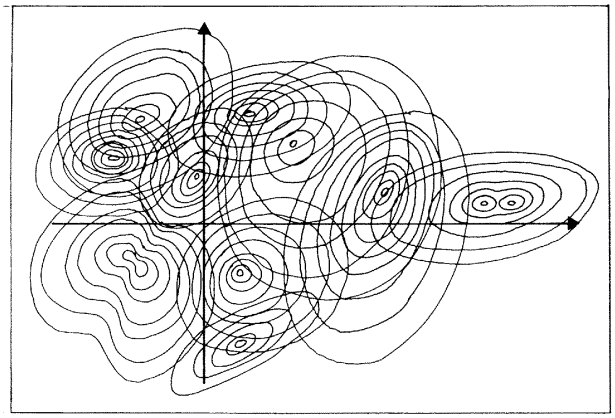

Fig. 8. In the space of two environmental factors, each species can be represented by its niche, here represented as a set of isospecies. Isospecies are here concentric curves whose points correspond to relevés with the same species cover values.

the direct problem of defining coenoclines and coenoplanes would be solved. In order to solve the inverse problem, the search for ecological gradients, based on species abundance data, it is strictly necessary to know the species response curve or, at least, some model should be hypothesised, in order to allow parameter estimation.

In the vector space spanned by f ecological factors, that we can consider as being orthogonal to each other, the set of points having the same cover value for a species s_i can be approximated as f—1 dimensional regular hypersurfaces depending on the species value itself. It should be expected that f species would be sufficient to describe the same ecological space. In fact, since as a species niche is but a part of the full ecological space, the number of species needed is higher (and the number of species in nature confirms this idea), in order to cover the space.

In such an environmental factors space, if we admit that the distribution of each species is Gaussian for each environmental factor, both species and factors are normalised, and the distribution parameters are always the same, the model is an orthogonal vector space, spanned by factors, where species co-ordinates are hypercircles of different centres but with radius proportional to the cover value; as a result, the distance of a point from a centre of a species distribution is the measure of that species' cover.

Let us imagine now that our data table is noiseless, that is, random variation is zero, and that we know already the position of each species mode. We can place a relevé in the intersection of the hypercircle of species s_1 corresponding to value r_{j1}, with the hypercircle of species s_2 corresponding to value r_{j2}, ..., with the hypercircle of species s_{i2} corresponding to value r_{ji}, etc. j = 1, ..., r. In this way, we can find positions of all relevés that are meaningful in terms of the environmental factors. The same can be accomplished if the metrics along each factor are not the same or factors are not orthogonal (that would mean the existence of a covariance pattern among factors), or the isospecies (that is the considered hypersurfaces) are not hypercircles, but hyperelliptycal or any other known hypersurfaces. The only thing needed is to know the exact functional form of both the isospecies and its inverse transformation (Fig. 9).

The first problem that must be considered in this context is the problem of noise in the data. We can estimate the actual position of relevés in the same way we estimate points in linear space, i.e., via regression (least squares) techniques, once we assume that the noise has random distribution. A second serious problem arises when we drop the requirement of knowing the parameters of the Gaussian distribution of each species in the environmental factor space. Gaussian ordination (GO) (Gauch and Chase 1974, Gauch, Chase and Whittaker 1974) defines both the position and parameters of a species. A third serious difficulty arises if we drop the last requirement, that is, knowing the functional form of response to environmental factors. In this last case, no parametric model can be considered, except indirectly inferred, such as the horseshoe effect in a multilinear space, or other effects, which may serve as the signature by which some species response curves may be identified.

Non parametric methods: possible exploratory solutions?

In the case that estimation of species cover/abundan-

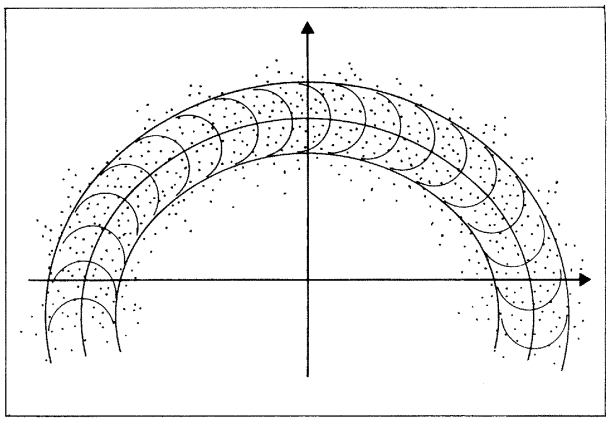

Fig. 9. A probable representation of the main ecological factor influencing the scatter of relevés in Fig. 3. A second minor factor may be considered affecting the scatter around the first one. If the equation of the curves representing the factors is be known, a generalised coordinate system is defined operationally.

ce accords with the Braun-Blanquet code, we may consider Multiple Correspondence Analysis (Benzécri 1973-82, Lebart, Morineau et Tabard 1977) a suitable exploratory tool, although seldom used in ecology studies (see Romane 1972). MCA is a Generalised (to several sets of variables) Canonical Analysis (Carroll 1968, Kettenring 1971, Casin and Turlot 1986) specially designed for the analysis of a contingency table (Burt's table, Burt 1950). The subsequent eigenanalysis is supposed to be particularly robust, and it is likely that the distribution of both relevé and species scores could contribute understanding about how species co-occur and how they are aligned to environmental gradients.

NMMDS techniques (Kruskal 1964a and b) are another possibility that seem promising, once the requirement for a specific underlying model is dropped. In fact, once a distance is defined among relevés, the aim is to recover the same distances between relevés in a reduced space, through iterative computations, aiming at reducing a defined stress value. This can be viewed as an optimisation problem, depending on the definition of distance and stress. Although these methods were already tested in ecological data analysis on real and simulated data (Fewster and Orlóci 1973, Orlóci, Kenkel and Fewster 1984, Minchin 1985, Kenkel and Orlóci 1986, Bradfield and Kenkel 1987), no theory can help, up to now, in defining the number of dimensions most suitable to represent the data, that is to infer the actual number of environmental gradients. Furthermore, special care must be exercised in computation of distances, in order to avoid overweighting of the co-occurrent species and underweighting the rare ones.

Characteristic species: a suggestion

Most of the problems derive as the researcher wishes to store species cover data taken in the field in a computer's memory, to obtain automatically, by shaking the computer, as it were, a scatter diagram where each relevé is set in its exact position according to the environmental underlying factors (if possible, labelled with their names). Of course, the researcher may come close to this goal, through the use of exploratory techniques sufficiently specialised for this type of data, if assisted by a suitable expert system (Camiz, 1988). I think that presently existing tools can provide reliable information about phytosociological pattern, as revealed by data, and suggestions of the actual number of environmental gradients influencing the pattern, even if horseshoe or other arch effects are present (Feoli and Orlóci 1976, Orlóci 1988). The trained researcher can extract understanding from these data patterns and derive conclusions. If the researcher wants to upgrade to a more precise definition of the gradients themselves and the relationships that species have with the gradients, he must consider more details about the actual relationships that can only be derived from field surveys or experiments.

In order to reduce the enormous workload, it would be sufficient to limit the study, at least at the beginning, to the species considered phytosociologically characteristic. Once their relationships with environmental gradients are identified, equations could be derived that could well be used in building a system of generalised co-ordinates that could cover the entire ecological space. Each relevé, in turn, could be placed at the appropriate point in this space, based only on the performance data of characteristic species.

Acknowledgement. I am grateful to Prof. L. Orlóci for the encouragement during my visit at the University of Western Ontario. The work was partially supported by a grant of Italian Consiglio Nazionale delle Ricerche (C.N.R.).

REFERENCES

Austin, M.P. 1976a. On non-linear species response models in ordination. Vegetatio. 33: 33-41.

Austin, M.P. 1976b. Performance of four ordination techniques assuming three different non-linear species response models. Vegetatio. 33: 43-49.

Austin, M.P. and I. Noy-Meir. 1971. The problem of non linearity in ordination: experiments with two-gradient models. J. Ecol. 59: 763-773.

Austin, M.P. and L. Orlóci. 1966. Geometric models in ecology. II. An evalutation of some ordination techniques. J. Ecol. 54: 217-227.

Beals, E.W. 1984. Bray-Curtis Ordination: An Effective Strategy for Analysis of Multivariate Ecological Data. Advances in Ecological Research 14: 1-55.

Benzecri, J.P. 1973-82. L'Analyse des données. Tome II: L'Analyse des correspondances. Dunod, Paris.

Bourbaki, N. 1962. Eléments de Mathématiques. Livre II, Algèbre, chap. 2, Algèbre lineaire. Hermann, Paris, ASI 1236.

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

Bray, J.R. and J. T. Curtis. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. 27: 325-349.

Burt, C. 1950. The factorial analysis of qualitative data. Br. J. Psycol. (statistical section) 3: 166-185.

Camiz, S. 1988. Expert systems: utility in community studies and examples. Coenoses, 3: 33-40.

Carroll, J.D. 1968. A generalisation of canonical correlation analysis to three or more sets of variables. Proc. 76th Amer. Psych. Assoc.: 227-228.

Casin, P. and J.C. Turlot. 1986. Une présentation de l'analyse canonique generalisée dans l'éspace des individus. Revue de Statistiques Appliquées, 35: 65-75.

Dagnelie, P. 1965a. L'étude des communautés végétales par l'analyse statistique des liaisons entre les espèces et les variables écologiques: principes fondamentaux. Biometrics 21: 345-361.

Dagnelie, P. 1965b. L'étude des communautés végétales par l'analyse statistique des liaisons entre les espèces et les variables écologiques: un example. Biometrics 21: 890-907.

Dale, M.B. 1988. Some fuzzy approaches to phytosociology:

ideals and instances. Folia. Geob. Phytotax. 23: 239-274.

De Leeuw, J. 1987. Nonlinear Multivariate Analysis with Optimal Scaling. In: Legendre P. and L. Legendre (eds.), Developments in numerical ecology. Springer Verlag. Berlin. pp. 157-188.

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

Feoli, E. and L. Orlóci. 1979. Analysis of concentration and detection of underlying factors in structured tables. Vegetatio, 40: 49-54.

Feoli, E. and V. Zuccarello. 1986. Ordination based on classification: yet another solution? Abstracta Botanica 10: 203-219.

Feoli, E. and V. Zuccarello. 1988. Syntaxonomy: a source of useful fuzzy sets for environmental analysis? Coenoses 3: 141-147.

Feoli, E., M. Lagonegro and A. Zampar. 1982. Classificazione e ordinamento della vegetazione. Metodi e programma di calcolo. AQ 35 CNR, Roma.

Fewster, P.H. and L. Orlóci. 1983. On choosing a resemblance measure for non-linear predictive ordination. Vegetatio 54: 27-35.

Gauch, H.G. JR. and G.B. Chase. 1974. Fitting the gaussian curve to ecological data. Ecology 55: 1377-1381.

Gauch, H.G. JR., G.B. Chase and R.H. Whittaker. 1974. Ordination of vegetation samples by Gaussian species distributions. Ecology 55: 1382-1390.

Gauch, H.G. JR. and R.H. Whittaker. 1972. Comparison of ordination techniques. Ecology 53: 868-875.

Gauch, H.G. JR., R.H. Whittaker and S.B. Singer. 1981. A comparative study of nonmetric ordinations. Journal of Ecology 69: 135-152.

Gauch, H.G. JR., R.H. Whittaker and T.R. Wentworth. 1977. A comparative study of reciprocal averaging and other ordination techniques. J. Ecol. 65: 157-174.

Gittins, R. 1985. Canonical Analysis. A review with applications in Ecology. Springer Verlag, Berlin.

Godement, R. 1966. Cours d'algèbre. Hermann, Paris.

Goff, F.G. and G. Cottam. 1967. Gradient analysis: the use of species and synthetic indices. Ecology 48: 793-801.

Goodall, D.W. 1954. Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. Aust. J. Bot. 2: 304-324.

Goodall, D.W. 1963. The continuum and the individualistic association. Vegetatio 11: 297-316.

Goodall, D.W. 1970. Statistical plant ecology. Annual Review of Ecology and Systematics 1: 99-124.

Goodall, D.W. and E. Feoli. 1988. Application of probabilistic Methods in the Analysis of Phytosoċiological Data. *Coenoses* 3: 1-9.

Greenacre, M.J. 1984. Theory and applications of Correspondence Analysis. Academic Press, New York.

Guttman, L. 1953. A note on Sir Cyril Burt's factorial analysis of qualitative data. Br. J. of Stat. Psychol. 6: 1-4.

Heiser, W.J. 1987. Joint ordination of species and sites. The Unfolding technique. In: Legendre P. and L. Legendre (eds.), Developments in Numerical Ecology, pp. 189-225. Springer Verlag, Berlin.

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

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

Ihm, P. and H. Van Groenewoud. 1975. A multivariate ordering of vegetation data based on Gaussain type gradient response curves. J. Ecol. 63: 767-777.

Ihm, P. and H. Van Groenewoud. 1984. Correspondence Analysis and Gaussian Ordination. Compstat lectures, Physica Verlag 3: 5-60.

Johnson, R.W. and D.W. Goodall. 1980. A maximum likelihood approach to non-linear ordination. Vegetatio 41: 133-142.

Kenkel, N.C. and L. Orlóci. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. Ecology 67: 919-928.

Kessel, R. and R.H. Whittaker. 1976. Comparison of three ordination techniques. Vegetatio 33: 21-29.

Kettenring, R.J. 1971. Canonical analysis of several sets of variables. Biometrika 58:

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

Kruskal, J.B. 1964b. Nonmetric multidimensional scaling: a numerical method. Psychometrika 29: 115-129.

Lagonegro, M. and E. Feoli. 1984. Three-Packages for classification and ordination of multivariate data. Libreria Goliardica, Trieste.

Lebart, L., A. Morineau and N. Tabard. 1977. Techniques de la description statistique. Dunod, Paris.

McDonald, R.P. 1962. A general approach to nonlinear factor analysis. Psychometrika 27: 397-415.

McDonald, R.P. 1967. Numerical Methods for Polynomial Models in Nonlinear Factor Analysis. Psychometrika 32: 77-112.

Minchin, P.R. 1985. A comparative evaluation of ordination techniques. Symposium on Theory and Models in Vegetation Science. IAVS. Uppsala.

Noy-Meir, I. 1971. Multivariate analysis of the semi-arid vegetation in south-eastern Australia: nodal ordination by component analysis. Proc. Ecol. Soc. Aust. 6: 159-193.

Noy-Meir, I. 1974. Catenation: quantitative methods for the definition of coenoclines. Vegetatio 29: 89-99.

Noy-Meir I. and R.H. Whittaker. 1977. Continuous multivariate methods in community analysis: some problems and developments. Vegetatio 33: 79-98.

Orlóci, L. 1966. Geometric models in ecology. I. The theory and application of some ordination methods. J. Ecol. 54: 193-215.

Orlóci, L. 1981. Probing time series vegetation data for evidence of succession. Vegetatio 46: 31-35.

Orlóci, L. 1980. Non-linear data stuctures and their description. In: L. Orlóci, C.R. Rao and W.M. Stiteler (eds.), Multivariate Methods in Ecological Work, pp. 191-202. Statistical Ecology Series Vol. 7. ICPH, Burtonsville, MD.

Orlóci, L. 1988. Community Organisation: Recent Advances in Numerical Methods. Canadian Journal of Botany 66:

Orlóci, L., N.C. Kenkel and P.H. Fewster. 1984. Probing simulated vegetation data for complex trends by linear and nonlinear ordination methods. Abstr. Bot., 8: 163-172.

Orlóci, L. and M. Orlóci. 1988. On recovery, Markov chains and Canonical Analysis. Ecology 69: 1260-1265.

Phillips, D.L. 1978. Polynomial ordination: Field and computer simulation testing of a new method. Vegetatio 37: 129-140.

ROBERTS, D.W. 1986. Ordination on the basis of fuzzy sets theory. Vegetatio 66: 123-131.

ROMANE, F. 1972. Utilisation de l'analyse multivariable en phytoécologie. Investigación Pesquera 36: 131-139.

TER BRAAK, C.J.F. 1984. Correspondence analysis of Incidence and Abundance Data: Properties of a Unimodal response model. IWIS-TNO Wageningen (NL), A85ST11. 31 pp.

TER BRAAK, C.J.F. 1985. Canonical Correspondence Analysis: a new Eigenvector Technique for Multivariate Direct Gradient Analysis. IWIS-TNO Wageningen (NL), A85ST17. 24 pp.

VAN DER MAAREL, E. 1969. On the use of ordination models in phytosociology. Vegetatio 19: 21-46.

VAN GROENEWOUD, H. 1965. Ordination and classification of Swiss and Canadian forests by various biometric and other methods. Ber. Geobot. Inst. ETH Stiftung Rübel, Zürich; 36: 28-102.

VAN RIJCKEVORSEL, J. 1987. The application of fuzzy coding and horse shoes in multiple correspondence analysis. DSWO Press. Leiden.

WESTHOFF, V. and E. VAN DER MAAREL. 1973. The Braun-Blanquet approach. In: R.H. Whittaker (ed.), Handbook of Vegetation Science, pp. 617-725. Part V: Ordination and Classification of Vegetation. Junk. The Hague.

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

Manuscript received: September 1988