

**Shifted single-peakedness,
Unfolding,
Correspondence analysis,
and
horseshoes**

**Willem J. Heiser
Department of Data Theory
University of Leiden**

Paper presented at the NATO Advanced Research Workshop on Numerical Ecology. Station Marine de Roscoff, Brittany, France. June 3-11, 1986.

Shifted single-peakedness, Unfolding, Correspondence analysis and horseshoes

Willem J. Heiser
Department of Data Theory
University of Leiden

1. Introduction

Ordination and clustering methods all rely on the concept of *distance* and some kind of *reduction* principle in order to facilitate the analysis of structures in data. Usually, this requires the choice of some measure of ecological resemblance as a first step, either between objects (individuals, samples), or between attributes (species, descriptors). Then in ordination the aim is finding a reduced space that preserves distance, i.e. reduction of *dimensionality*, and in cluster analysis the aim is allocating the units of analysis to a reduced number of (possibly hierarchically organised) classes, i.e. reduction of *within-group* distance with respect to *between-group* distance.

This paper will be centered at a third type of method, also based on distance and reduction, but not relying on derived associations or derived dependencies. It is particularly suited for the analysis of species \times samples presence-absence or abundance data; or, perhaps somewhat more generally, for any ecological data matrix that is dimensionally homogeneous (Legendre and Legendre, 1983), and non-negative. In psychology, where its early developments took place in the context of the analysis of individual choice behavior and differential preference strength, the group of methods is called *Unfolding* (Coombs, 1950, 1964). Since the word "unfolding" very plastically describes the major aim of the technique, it will be used as a generic name throughout this paper.

In order to frame the objectives of Unfolding in ecological terms, the first thing to notice is that the basic notion of ecological resemblance need not be confined to distance defined on pairs of units from a single set. If it is assumed that for each species there is a unique combination of the levels or states of the environmental variables that optimizes its possibilities to survive, to be called - perhaps pleonastically - its *ideal niche*, and that the sampling sites approximate these ideal circumstances to a different degree, then species

abundance might be supposed to level off monotonically with the distance of a sampling site from the ideal niche. Here distance could be understood as concrete, geographical distance, or as distance in some abstract space. In the latter case the samples are to be arranged in an orderly fashion, along a *gradient*, reflecting the gradual changes in environmental or community characteristics. Now the Unfolding technique precisely aims at finding those gradients that yield *single-peaked* response functions, i.e. at a reduction to (low-dimensional) *unimodality*. Psychologists study objects called stimuli, want to arrange them along stimulus scales, and one of the major response classes available to them is preference. In these terms, the Unfolding technique aims at finding those stimulus scales that yield single-peaked preference functions.

Coombs developed his form of Unfolding in an attempt to resolve a notorious problem in psychology, i.e. the problem of defining a *psychological unit of measurement* (Coombs, 1950). How can we quantify human judgement without recourse to an arbitrary grade-point system? The ecological equivalent of this issue would be: how can we quantify the differential reactions of species to the environment without capitalizing on the pseudo-exact numerical aspects of abundance? The answer Unfolding has to offer is through the study of *consistency* (or *scalability*) of the behavioral reactions under the condition of single-peakedness.

The first goal of this paper is to convince the reader that the Unfolding technique is the natural general-purpose first candidate for Gradient analysis. However, there exists plenty of room for making more specific assumptions than has been done so far, and hence a lot of slightly different methods are to be considered as a member of the family. Therefore, a second goal is to try to organize the field a little by comparing the various loss functions on which the methods are based, and by showing the interrelations between various special cases. The third goal is to present explicit computational formulas for a convergent Unfolding algorithm, and to sketch a few open problems and lines of development.

2. Non-linearities: a mixed blessing

2.1. Indications for unimodality in ecology and elsewhere

The importance of single-peaked, or unimodal, response curves and surfaces stems from a diversity of scientific areas, ecology being one of the richest sources. Frequently a linear analysis of contingencies showed unexpected nonlinearities, or else regression plots of abundance or coverage against carefully chosen *a priori* gradients were unmistakably bell-shaped. Ihm and van Groenewoud (1984) summarize the early evidence from vegetation studies as follows: "Already Goodall (1954) in one of the first applications of PCA to the

analysis of vegetation data noted the problem caused by the nonlinearity of quantitative species relationships in the interpretation of the principal components. Knowledge about the non-linearity of gradient response was, however, not new. Braun-Blanquet and Jenny (1926) investigated the pH-value of soils in which several species, e.g. *Carex Curvula* (L) and others, were growing in the Swiss Alps and England. They found normal frequency curves for these pH-values. Making the assumption of a uniform distribution of the pH-values - at least in the range of growth of the species studied - one could conclude that also the gradient response was Gaussian. It appears the bell-shaped gradient response curves were first suggested by Igoshina (1927). Gause (1930) studied the abundance of certain species as related to ecological conditions and found that they followed the law of Gauss. The ordination work by Curtis and McIntosh (1951), Bray and Curtis (1957), Cottam and Curtis (1956), Whittaker (1948) and many others all showed the non-linearity of species-site factor relationships. Especially the published examples of gradient responses clearly show the unimodal type of the response curves." (l.c., p.13). For many additional references, cf. Gauch (1982) and Whittaker (1978).

The first articulated unimodal response model in psychology was proposed by Thurstone (1927), building upon nineteenth century work on sensory discrimination. He claimed wider applicability, e.g. as a model for attitude and opinion, but later on abandoned the subject. Hovland, Harvey and Sherif (1957) undertook additional experimental work, and provided convincing evidence for single-peakedness in human evaluative responses. In factor analyses of personality tests one frequently found nonlinearities called - by lack of a full understanding - 'difficulty factors'. Coombs and Smith (1973) and Davison et al. (1980) studied unimodal developmental processes, and a classic example of single-peaked behavior is preference for family compositions in terms of number of children and bias towards boys or girls (e.g., Coxon, 1974). Yet the phenomenon is not very actively studied anymore in psychology, not nearly as much as its special case: monotonicity.

At this point, it might be good to emphasize that it is not unimodality alone, but the fact that the *peaks* of the curves are *shifted* with respect to each other which makes the situation special. For imagine a number of unimodal curves precisely on top of each other, then any transformation of the gradient would provide the same information; thus one could make the curves more skewed, double-peaked, monotonically increasing, any shape at all, by suitable reexpressions of the values against which they are plotted. When the curves are shifted along the gradient, this freedom of simultaneous change of shape is reduced enormously.

The early contributions to ordination by the famous archaeologist Flinders Petrie, source of inspiration for Kendall (1963) and much subsequent work in archaeological seriation (cf.

Hodson et al., 1971), were typically not tailored to the precise shape of the artifact distributions, but primarily to the fact that they should form an overlapping sequence of 'present' counts if the sites were properly ordered (presumably in time). Roberts (1976, section 3.4) has given an interesting graph-theoretical characterization of this ordering problem.

Summarizing, we might say that unimodality is a firmly established empirical phenomenon, that it is only visible when the gradients are carefully chosen, and finally that linear methods like Principal Components Analysis (PCA) will distort expected gradients in a nonlinear fashion (Swan, 1970; Noy-Meir and Austin, 1970). Because these distortions can have widely different forms - depending on such things as the dimensionality of the gradient, the homogeneity of the species and sample variances, and the variability of maximum abundance - it is hazardous to rely on the standard PCA approach, and there is clearly a need for *specialized nonlinear* methods.

2.2 Nonlinear data transformations

If a bivariate distribution of data points is curved, we can straighten it out by transforming one or both of the variables. For instance, if the cloud "accelerates" from left to right, a log transformation of the vertical axis will remove, or mitigate, the acceleration. This is called *linearizing the regression*. If all bivariate distributions among m variables are considered simultaneously, it will generally be necessary to use different transformations to linearize the regression as much as possible *on the average*. This is one of the major objectives in the Gifi system of Nonlinear Multivariate analysis; for a full explanation see de Leeuw (1986a).

Under the assumption of shifted single-peaked response curves and surfaces, we don't expect to find linear bivariate relationships (cf. Greig-Smith, 1983, who has clearly summarized the peculiar shapes one can obtain). Perhaps not too surprisingly, then, the approach using nonlinear data transformations towards linearity turns out to be a move in the wrong direction in this case (Heiser, 1985a), giving more extreme curvature and convolutions than a linear PCA. The previous statement deserves a qualification, because it is only true when the class of transformations, or *admissible reexpressions*, of the variables is defined in the *standard* way. As we shall see later on (section 4.1), there are alternative ways of coding, based on the assumption of shifted single-peakedness, which do give satisfying results.

2.3 The general polynomial model

Instead of bringing in nonlinearity at the data side, it can be introduced in the *functional structure* of the model. McDonald (1962, 1967) and Carroll (1969, 1972) have advocated this general approach. Deviation from linearity - although a heterogeneous phenomenon by its very nature - can always be modelled by a sufficiently rich family of polynomials. Carroll's *polynomial factor analysis* model has the following form:

$$f_{ij} \cong \sum_r a_{ir} z_{rj}, \quad (1)$$

with

$$z_{rj} = \Phi_r(y_{j1}, \dots, y_{js}, \dots, y_{jp}). \quad (2)$$

Here, as in the sequel, f_{ij} denotes the abundance of species i in sample j , or, in the more general terminology of Legendre and Legendre (1983), the value of descriptor i for object j . The symbol \cong is used for approximation in the least squares sense, and the indices run as $i=1, \dots, n$, $j=1, \dots, m$, and $r=1, \dots, q$. So in its full generality, there are p sample gradients, or a p -dimensional space of sample points, with coordinates y_{js} . Then there are q elementary polynomial functions Φ_r that have to be specified on an a priori basis. Thus to obtain a quadratic response surface, for example, one would have to specify:

$$\begin{aligned} \Phi_1(\cdot) &: z_{1j} = 1, \\ \Phi_2(\cdot) &: z_{2j} = y_{j1}, \\ \Phi_3(\cdot) &: z_{3j} = y_{j2}, \\ \Phi_4(\cdot) &: z_{4j} = y_{j1}^2, \\ \Phi_5(\cdot) &: z_{5j} = y_{j2}^2, \\ \Phi_6(\cdot) &: z_{6j} = y_{j1}y_{j2}. \end{aligned}$$

It is easily verified that if only the first three of these are chosen, (1) and (2) reduce to the familiar bilinear form of the PCA model.

Carroll used a steepest descent method for finding optimal values for the parameter sets $\{a_{ir}\}$ and $\{y_{js}\}$. There is little experience with the procedure, however. It is quite heavily loaded with parameters, and does not give a particularly simple parametrization of the species. It has a great many special cases. Perhaps it should better be called a program for research, rather than a model.

When the $\{y_{js}\}$ are fixed to known values, e.g. environmental measurements such as pH-value, soil moisture, elevation and so on, the set-up (1) and (2) becomes formally equivalent to a *multiple regression analysis* problem (Draper and Smith, 1966; Gittins,

1985). Note that although nonlinear predictors are used, the model is now *linear in the parameters*, and can be fitted by standard methods. Also note that in fact we have n independent regression problems, one for each species or row of the data matrix. The last two remarks remain true if the definition of ϕ_r is extended to include logarithmic, exponential or other simple functions. Carroll (1972) has given explicit reparametrizations, constituting the so-called PREFMAP hierarchy of models, to obtain a description of the species response curves or surfaces in terms of the location of the peak, the importance of the relative contributions of the gradient factors, and possibly their interaction.

Fixing the space of sample points or objects and then studying the regression is only one way to simplify the general polynomial model, and is called *direct* Gradient analysis (Whittaker, 1967), or *external* analysis of preferences (Carroll, 1972). These terms are used in contrast to *indirect* Gradient analysis or *internal* analysis of preferences, in which some optimal quantification of the gradient has to be found as well. As we shall see shortly, there is also the possibility of an analysis in between these two extremes, whenever there is *partial* knowledge on the gradient (f.i., a *ranking* of the sites with respect to moisture status, instead of exact numerical measurements). But first a few additional remarks are in order, regarding the reasons for concentrating on unimodal models.

2.4 Strategic reasons for giving priority to shifted single-peakedness

It was remarked earlier: linearity has the virtue of being uniquely defined, but deviation from linearity can have many appearances. From a statistical point of view, it seems wise to progress slowly from very simple to increasingly complex models, and to examine the deviations from the model along the way. In fact, the bilinear model of PCA is already a second type of approximation, the first one being the hypothesis that *all abundances are equal*, up to row- and/or column effects. However ignorant or even indecent this may sound in a field that studies diversity, we may occasionally need to have statistical assurance that we deal with *genuine interaction* between species and sites. If the abundance data are considered to be a contingency table, for instance, the chi-squared test value under the hypothesis of independence should be very large.

The shifted single-peaked model is another second type of approximation, and it has the virtue of having one defining characteristic as well. It is more complex in form than the bilinear model, but not necessarily in terms of number of parameters. The situation is depicted in Figure 1.

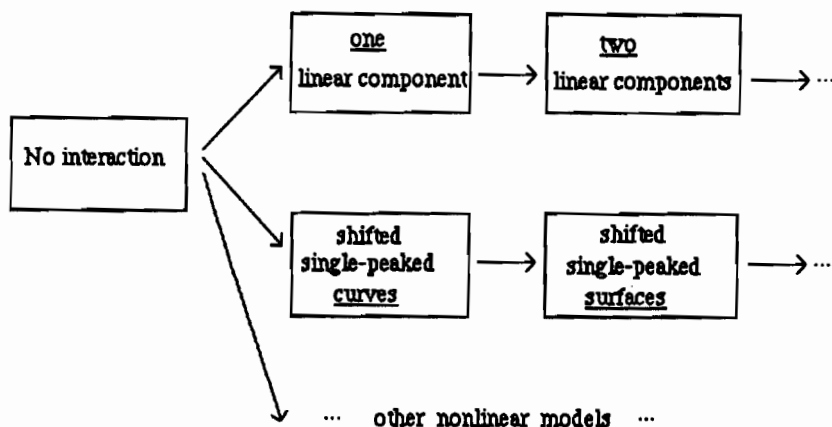


Figure 1. Interaction models can best be partially ordered.

When moving to the right a better fit will always be obtained, but one set of curves might be enough where more components are needed. Of course, other nonlinear models might turn out to be even more appropriate, but in general there is little hope in trying an exhaustive search.

It is difficult to accept that, when two models describe the same data about equally well, one of them is "true" and the other one is "false". Therefore, consider Figure 2. It gives an idealized example of one of those notorious curved chains of sample points from a PCA of abundance data. In addition, however, it gives two directions representing species A and B, respectively. The advantage of making this so-called *joint plot* or *biplot* (Gabriël, 1971) is that it enables the demonstration of a very elementary fact, which is often - if not always - overlooked in the literature. The PCA model implies that, in order to reconstruct the abundances for species A, the sample points should be orthogonally projected onto direction A. If this actually done, and likewise for direction B as well, and *if the curved chain is straightened out, or "unfolded"* into a straight line, locally preserving the distances among the sample points, the projections plotted against the "unfolded" chain get the appearance of Figure 3: shifted single-peaked curves! Any direction in between A and B in Figure 2 would get a curve with its peak in between the peaks of A and B in Figure 3, and more extreme directions (to the left of B, and to the right of A) would get curves with more extremely shifted peaks. This shows that there is no real contradiction between the two ways of representing the data, provided they are interpreted with an open eye. For single-peaked surfaces the PCA representation will be a curved manifold in three dimensions, much less easily recognizable. Under single-peakedness the data themselves already form a curved manifold in m dimensions, which has to be "unfolded" to display its simplicity.

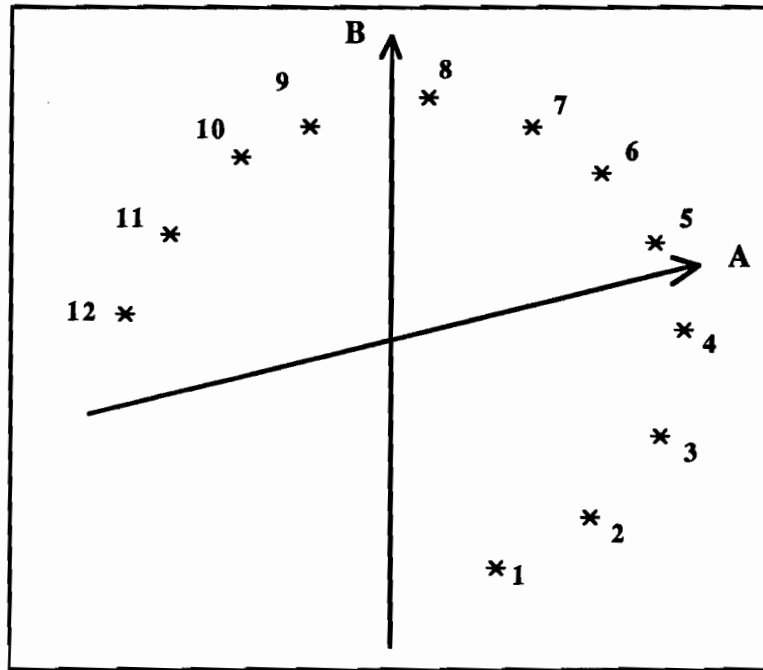


Figure 2. Joint plot of two species (A and B) and a number of sites exhibiting the horseshoe effect.

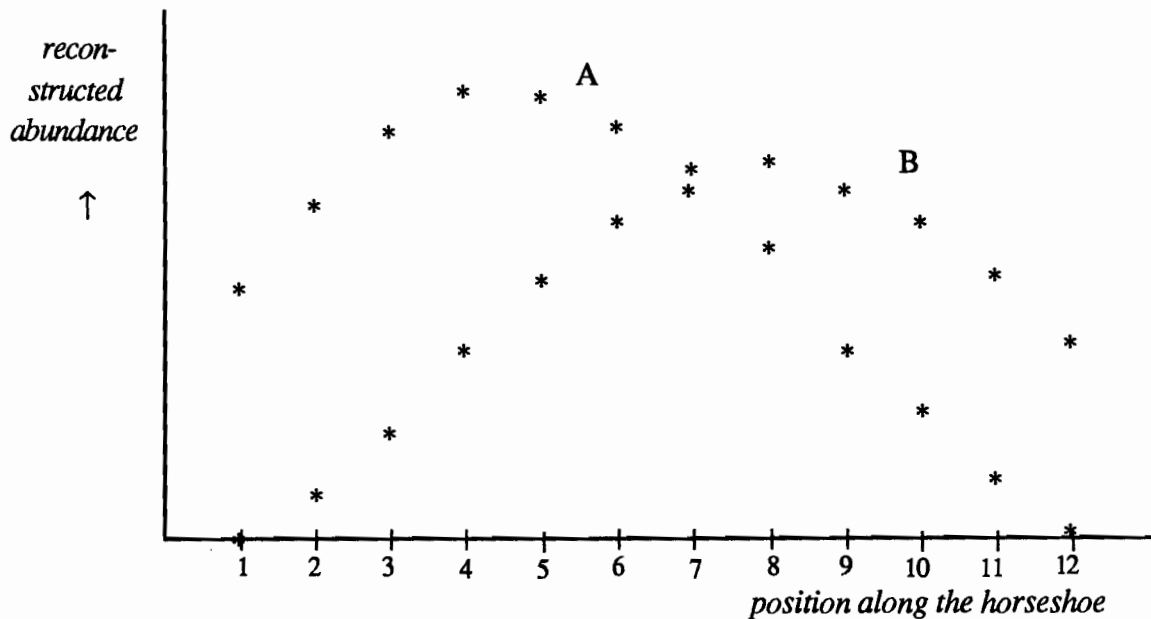


Figure 3 Abundance as a function of position along the horseshoe (Peak A corresponds with direction A of Figure 2, and peak B with direction B).

Of course, these observations are not sufficient for getting a practical method. The occurrence of deviations from the model, including random errors, as well as the possible need to work in high dimensionality, urges us to use and further develop specialized Unfolding methods.

3. A family of loss functions for Unfolding

A curve or surface of any shape could in principle be modelled by means of the general polynomial model. This relatively blind approach implies that many parameters have to be estimated (often repeatedly under different specifications of the model), some of which are not readily interpretable. Under shifted single-peakedness the parametrization can be solely in terms of the *location of the peaks*, and possibly also with respect to remaining aspects of shape: *tolerance* or *species dispersion* (range of the responses along the gradient), *correlated density* in the more-dimensional case, and (lack of) *symmetry*. Any Unfolding method is based on the assumption that abundance is inversely related to the distance of a sample point from the estimated peak location of the species response function, frequently called the *ideal point*. The name "Unfolding" refers to the following metaphor: suppose the model is known, and imagine the sample points painted on a handkerchief. Pick the handkerchief up at the ideal point of species i and *fold* it, for instance by pulling it through a ring. Then observe that the sample spots will appear in the order of the abundances as given in the i 'th row of the data matrix. Because the analysis technique must *construct* the model starting from the data, this process must be reversed; hence the name.

Two major approaches to Unfolding can be discerned: one based on *dissimilarity approximation*, the other on *distance* or *squared distance minimization*. As shall become evident shortly, there is an important sense in which the latter - formally equivalent to Correspondence Analysis - is a special case of the former. The discussion starts with the problem of external Unfolding, where the location of the sample points is fixed in advance, and the ideal points must be determined.

3.1 Locating one set of points with respect to a given set

Suppose the coordinates of m points in p -dimensional space are available in the $m \times p$ matrix Y , the j 'th row of which is denoted with y_j . Now consider n unknown additional points, indexed by i , with coordinates x_i collected in the rows of the $n \times p$ matrix X . The Euclidean distance $d(x_i, y_j)$ is defined by writing its square as:

$$d^2(x_i, y_j) = (x_i - y_j)'(x_i - y_j) . \quad (3)$$

In order to construct a loss function that measures the departure of the model distances from the data, some definition of *dissimilarity* - the empirical counterpart of distance - has to be agreed upon. Just to make a start, suppose this is done in the following way. Since the total

number of occurrences of a species is often of little interest, at least not in the study of species x environment interaction, it is advisable to work with the species-specific proportions

$$p_{ij} = f_{ij} / f_{i+} \quad \text{with} \quad f_{i+} = \sum_k f_{ik} , \quad (4)$$

or some other standardization factor, such as *maximal species abundance*, to make the distributions row-wise comparable. Now the species-sample dissimilarity δ_{ij} and the associated *weights* w_{ij} may be defined as:

$$\delta_{ij} = -\log p_{ij} \quad \text{and} \quad w_{ij} = 1 \quad \text{if} \quad p_{ij} > 0 , \quad (5a)$$

$$\delta_{ij} = 1 \quad \text{and} \quad w_{ij} = 0 \quad \text{if} \quad p_{ij} = 0 . \quad (5b)$$

Other choices will be encountered later. In (5a) and (5b) the weights are merely used to indicate presence or absence; non-occurrence gets an arbitrary unit dissimilarity, and will not cause any increase in loss (because $w_{ij} = 0$). Note that, indeed, dissimilarity is a decreasing function of relative abundance; if p_{ij} approaches zero, then δ_{ij} approaches infinity, and if $p_{ij} = 1$ then $\delta_{ij} = 0$. The interpretation of the latter case depends on the data standardization; under (4) it implies that δ_{ij} only becomes zero if a species occurs in only one sample (in any frequency).

The basic Unfolding loss function is now defined as the weighted least squares criterion

$$\sigma_R^2 = \sum_i \sum_j w_{ij} \{ \delta_{ij} - d(x_i, y_j) \}^2 , \quad (6)$$

the "rectangular" or "off-diagonal" version of Kruskal's so-called *raw STRESS* (Kruskal was the first who explicitly proposed to use least squares distance modelling, in his (1964a, 1964b) papers). Depending on the alterations in the definition of w_{ij} and δ_{ij} , as well as on the choice of *domain* Ω over which σ_R is to be minimized, we get different Unfolding methods.

For the problem of this section Ω is the set of all $n \times p$ matrices, but in addition a provision has to be made for ensuring that δ and d match in scale (assuming that the coordinates of the given set of points are on an arbitrary scale). Because the distance function is *homogeneous*, i.e. $\alpha d(x_i, y_j) = d(\alpha x_i, \alpha y_j)$ for any nonnegative α , adjusting the scale of the coordinates and adjusting the scale of the distances amounts to the same thing. However, we can also adjust the scale of the dissimilarities by just extending their definition as to include an unknown scaling constant:

$$\delta_{ij}(\alpha) = \alpha (-\log p_{ij}) , \quad (7)$$

where the notation $\delta_{ij}(\alpha)$ is used to make the dependence on α fully explicit. Whatever choice is made, the scale adjustment would leave σ_R dependent on the arbitrary scale of the given set of points; this is undesirable, so σ_R has to be *normalized*. As shown by Kruskal and Carroll (1969), various ways of normalization only affect the scale of the loss function, not the argument for which a minimum is attained. De Leeuw and Heiser (1977) have argued that normalization on the distances makes the computational problem considerably more complicated in a number of important special cases. Therefore the external Unfolding problem - as defined here - becomes:

$$\min_{\mathbf{x}_1, \dots, \mathbf{x}_n} \min_{\alpha} \sigma_N^2(\mathbf{x}_1, \dots, \mathbf{x}_n; \alpha), \quad (8a)$$

with

$$\sigma_N^2(\mathbf{x}_1, \dots, \mathbf{x}_n; \alpha) = \sum_i \sum_j w_{ij} \{ \delta_{ij}(\alpha) - d(\mathbf{x}_i, \mathbf{y}_j) \}^2 / \sum_i \sum_j w_{ij} \delta_{ij}^2(\alpha). \quad (8b)$$

This optimization problem has no closed-form solution, it has to be solved iteratively. A convergent algorithm for finding at least a local minimum shall be discussed in some detail now, because it offers the opportunity to illustrate a number of interesting features of these type of algorithms. It is based on the general algorithm model proposed by De Leeuw and Heiser (1977, 1980), called SMACOF.

The minimization of σ_N can be done by repeatedly solving two subproblems. There is a *normalized regression* problem, in this case finding the optimal value of α for fixed distances, and a *relocation* problem, i.e. finding new locations \mathbf{X}^+ starting from some initial guess \mathbf{X} and keeping the rescaled dissimilarities constant at their current values. As to the former, it can be shown that, writing d_{ij} for the fixed distances, the optimal choice of α is

$$\alpha^+ = \sum_i \sum_j w_{ij} d_{ij}^2 / \sum_i \sum_j w_{ij} \delta_{ij} d_{ij}. \quad (9a)$$

The quantities

$$d^+_{ij} = \delta_{ij}(\alpha^+) \quad (9b)$$

(pronounce: the dhats) can be substituted in (8b), thereby reducing it to the basic form (6) with uniformly rescaled weights, due to the normalization factor. This settles the regression part for now.

The relocation part is more difficult. One of the objections to a relatively straight-forward steepest descent method, such as the one used by Kruskal (1964b), is that the partial derivatives of σ_R do not exist at points where $d(\mathbf{x}_i, \mathbf{y}_j)$ becomes zero. In this context it is of

some interest to note that the very same problem emerges in the classic *Fermat* or *generalized Weber problem* (Kuhn, 1967), also called the *location* problem, which is to locate a point x_i among m known points in such a way that

$$\min_{x_i} \sum_j w_{ij} d(x_i, y_j). \quad (10)$$

The SMACOF approach turns out to be closely related to Kuhn's algorithm. It is based on the '*subgradient*', rather than the ordinary '*gradient*' (De Leeuw, 1977).

To elaborate somewhat on the location problem: if w_{ij} is binary and J^i is the index set of the nonzero elements in row i , and if in addition distance is one-dimensional (the ecological gradient is one variable), then (10) reduces to

$$\min_{x_i} \sum_{j \in J^i} |x_i - y_j|, \quad (11)$$

the solution of which is well-known: the *median* of the sample values for which the species is present. This shows that in the case of binary weights (10) is a proper generalization of the median concept to higher dimensions (cf. Austin, 1959). It is also a generalization to the case of differential weights. So it certainly is one perfectly sensible way to estimate the peak of a surface. But in addition it becomes clear that, while (10) could be called a *distance minimization* approach, the external Unfolding problem is different in the sense that (6) aims at *approximation of dissimilarities*. How can this be done?

The interested reader is referred to De Leeuw (1977) and De Leeuw and Heiser (1980) for a general explanation of the SMACOF algorithm model and its rationale. For the Unfolding case also see Heiser (1981). The specific computational steps are as follows. Suppose $d_{ij} = d(x_i, y_j)$ is the distance between the fixed point y_j and some initial estimate x_i of the i 'th point. Then define the matrix A with elements

$$a_{ij} = w_{ij} d^+_{ij} / d_{ij} \quad \text{if } d_{ij} > 0, \quad (12a)$$

$$a_{ij} = 0 \quad \text{if } d_{ij} = 0. \quad (12b)$$

Furthermore, the weights are collected in $W = \{w_{ij}\}$, and the diagonal matrices P and R are defined as:

$$P = \text{diag} (Ae_m), \quad (13a)$$

$$R = \text{diag} (We_m),$$

where e_m denotes an m -vector of ones. The SMACOF algorithm for external Unfolding uses the following two operations:

$$\tilde{X} = P\bar{X} - A\bar{Y}, \quad (14a)$$

$$X^+ = R^{-1}(\tilde{X} + W\bar{Y}). \quad (14b)$$

Here \tilde{X} is a preliminary, unconstrained update, and X^+ is the successor configuration suitable for the present case of fixed column points. Note that in the equally weighted case the last operation (14b) amounts to a uniform rescaling and an adjustment of the centroid. The first operation (14a) carries the burden of the iterative relocation of the species points, because A and P contain information on the size of the current distances d_{ij} , on what they should be (d^+_{ij}), and on how strongly an improvement is desired (w_{ij}). Let's have a closer look by writing (14a) row-wise as a single weighted summation:

$$\tilde{x}_i = \sum_{k \in K} w_{ik} d^+_{ik} (x_i - y_k) / d_{ik}, \quad (15)$$

where K is the subset of the first m integers for which (12a) holds. Thus the preliminary

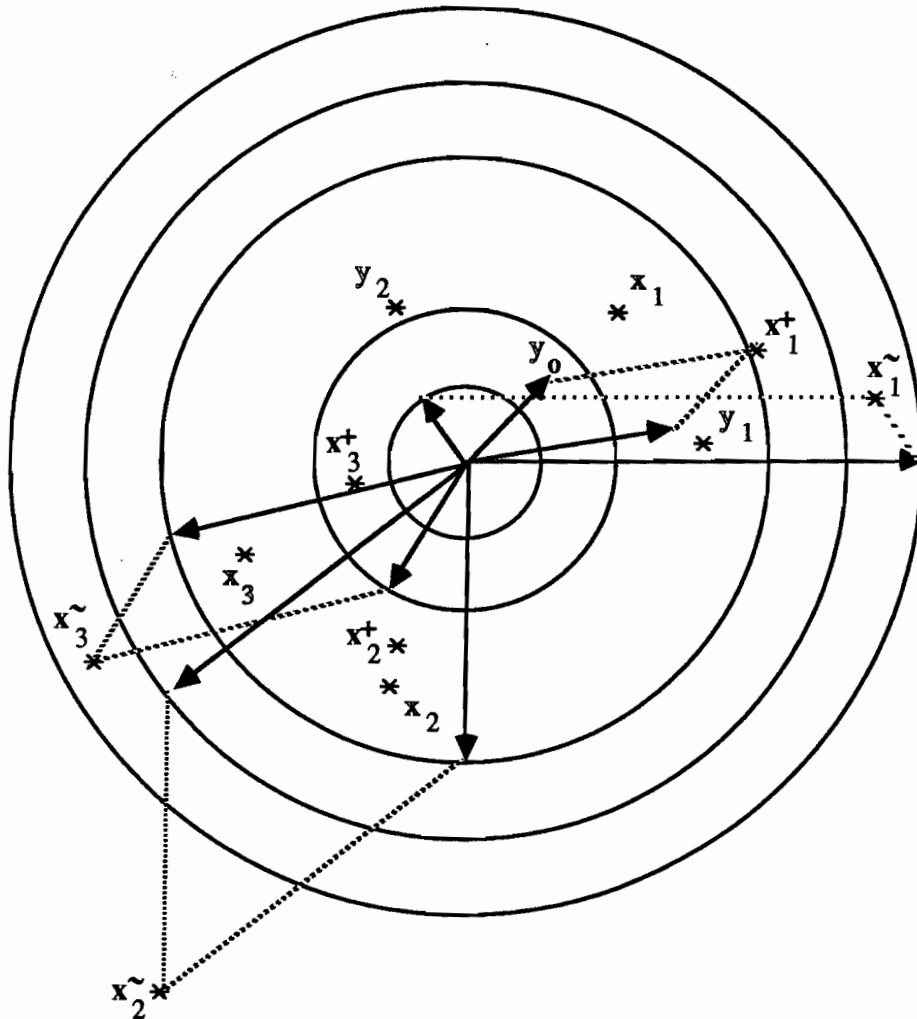


Figure 4. Coordinate-free construction of new species points ($\delta_{11}=1$, $\delta_{12}=6$, $\delta_{21}=5$, $\delta_{22}=4$, $\delta_{31}=4$, $\delta_{32}=2$).

updates are a weighted sum, with weights $w_{ik}d^+_{ik}$, of unit-length difference vectors pointing from the fixed column points towards the current location of i . If the current location of i coincides with a column point, then (12b) comes into effect; the zero difference vector cannot be normalized and is omitted from the summation. Sample sites where species i is absent - or at least where $w_{ij} = 0$, perhaps due to another reason - do not contribute either. The relocation step is illustrated in Figure 4, starting from an arbitrary configuration of three x -points and two y -points, with unit weights and the dissimilarities as given in the Figure caption. Thus there are six difference vectors, and the concentric circles around the origin represent the size of the dissimilarities. The \tilde{x}_i are now simply obtained by vector addition. Next their length has to be divided by 2, the number of y -points, and their origin must be shifted towards y_0 , the centroid of y_1 and y_2 , thus accomplishing (14b). For x^+_1 the latter step is explicitly shown, the other auxiliary lines are omitted for clarity. By visual inspection alone it can be verified that the new distances are closer to the dissimilarities than the old ones. Finally note the fact that each point is relocated independently from the others, in much the same way as there were n independent regression problems under the general polynomial model.

A summary of all steps is given in the next *skeleton algorithm for external Unfolding*:

```

 $\mathbb{X} \leftarrow$  'good guess'
 $\sigma_{\text{OLD}} \leftarrow$  'large'
for iter = 1,..., maxiter do:
  (i) determine  $\mathbb{X}^+$  from (14a) and (14b);
  (ii) calculate  $d(x^+_i, y_j)$  using (3);
  (iii) find  $d^+_{ij}$  from the regression of  $d$  on  $\delta$ ;
  (iv) calculate  $\sigma_{\text{NEW}}$  using (8b);
  (v) if  $(\sigma_{\text{OLD}} - \sigma_{\text{NEW}})$  is not 'small' then
      * set  $\mathbb{X} \leftarrow \mathbb{X}^+$  and  $\sigma_{\text{OLD}} \leftarrow \sigma_{\text{NEW}}$ 
      * go to (i)
  else
      * STOP

```

As a first extension to this scheme we shall now consider the situation in which the sample points are not a priori given, but have to be located as well.

3.2 Reciprocal relocation: internal Unfolding

In internal Unfolding analysis two sets of points have to be located with respect to each other; hence the term 'reciprocal relocation'. As a consequence, the relocations are not independent anymore. It does eliminate the need to rescale the data: the rescaling factor can be absorbed in the unknown coordinates. Therefore, the normalized loss function σ^2_N becomes functionally equivalent to the unnormalized one σ^2_R , i.e. the same up to a constant, and the problem becomes:

$$\min_{\mathbf{x}_1, \dots, \mathbf{x}_n} \min_{\mathbf{y}_1, \dots, \mathbf{y}_m} \sigma^2_R(\mathbf{x}_1, \dots, \mathbf{x}_n; \mathbf{y}_1, \dots, \mathbf{y}_m), \quad (16)$$

The skeleton algorithm of the previous section need not be changed very much. We can skip step (iii) (not for long; it will be reintroduced soon). Only step (i), calculation of the new locations, must really be adjusted. Two additional matrices are required:

$$\mathbf{Q} = \text{diag}(\mathbf{e}_n \mathbf{A}), \quad (17a)$$

$$\mathbf{C} = \text{diag}(\mathbf{e}_n \mathbf{W}), \quad (17b)$$

where \mathbf{e}_n denotes an n -vector of ones. Then, analogous to (14a), a preliminary update for the sample points is found from

$$\mathbf{Y}^{\sim} = \mathbf{Q} \mathbf{Y} - \mathbf{A}' \mathbf{X}. \quad (18)$$

The companion operation (14b) is no longer correct. Instead, the successor configurations \mathbf{X}^+ and \mathbf{Y}^+ must be computed from the system of linear equations:

$$\mathbf{R} \mathbf{X}^+ - \mathbf{W} \mathbf{Y}^+ = \mathbf{X}^{\sim}, \quad (19a)$$

$$\mathbf{C} \mathbf{Y}^+ - \mathbf{W}' \mathbf{X}^+ = \mathbf{Y}^{\sim}. \quad (19b)$$

The interested reader may consult section 3.6 at this point for finding out how these equations come about. How to solve this system most efficiently depends on the size of n and m . Suppose $n > m$ (the other case runs analogously). Then we should first solve

$$(\mathbf{C} - \mathbf{W}' \mathbf{R}^{-1} \mathbf{W}) \mathbf{Y}^+ = \mathbf{Y}^{\sim} + \mathbf{W}' \mathbf{R}^{-1} \mathbf{X}^{\sim}, \quad (20a)$$

which determines \mathbf{Y}^+ up to a shift of origin because the matrix $\mathbf{C} - \mathbf{W}' \mathbf{R}^{-1} \mathbf{W}$ is generally of rank $m-1$ (its null space is the vector \mathbf{e}_m , due to the definition of \mathbf{W} , \mathbf{C} , and \mathbf{R}). Next, any solution of (20a) can be used to determine \mathbf{X}^+ from

$$\mathbf{X}^+ = \mathbf{R}^{-1} (\mathbf{X}^{\sim} + \mathbf{W} \mathbf{Y}^+). \quad (20b)$$

Finally, although this is not really necessary, X^+ and Y^+ can be simultaneously centered so that their joint centroid is in the origin. This settles the relocation part for internal Unfolding.

Now consider a slight generalization in the regression part. Some species might cover a wider range of sites than others, independent of the location of their peaks. If the frequencies are normalized on the sum, this will tend to make the minus log proportions uniformly larger, which might be considered undesirable. This effect can be removed by introducing a scaling parameter for each species as a generalization of (7):

$$\delta_{ij}(\alpha_i) = \alpha_i (- \log p_{ij}) , \quad (21)$$

Note that all that would have to be done for including (21) in the *external* Unfolding algorithm would be to execute it for each species separately, because that would make (9) effectively row-specific, and the row-point movements were done independently anyhow. For internal Unfolding, however, the loss function has to be adjusted explicitly:

$$\sigma_C^2 = \sum_i [\sum_j w_{ij} \{ \delta_{ij}(\alpha_i) - d(x_i, y_j) \}^2 / \sum_j w_{ij} \delta_{ij}^2(\alpha_i)] , \quad (22)$$

where the subscript C in σ_C is used to indicate the conditionality of the regression and normalization (also called "split by rows", cf. Kruskal and Carroll, 1969). Yet the algorithm does not become very much more complicated. Keeping the distances fixed, the normalized regression (9) must simply be done on each row separately, giving α_i^+ . Next new weights can be defined as

$$w_{ij}^+ = w_{ij} / \sum_k w_{ik} \delta_{ik}^2(\alpha_i^+) , \quad (23)$$

which shows that minimizing (22) becomes equivalent to the basic unconditional problem (16), with row-wise rescaled data and row-wise rescaled weights.

Summarizing the steps again in a *skeleton algorithm for row-conditional internal Unfolding* we get:

```

X ← 'good guess'
Y ← 'good guess'
σOLD ← 'large'
for iter = 1, ..., maxiter do:
  (ia) determine X~ from (14a) and Y~ from (18);
  (ib) determine X+ and Y+ from (20a) and (20b);
  (ii) calculate d(x+i, y+j) using (3);
  (iii) for i=1, ..., n do:
        * find d+ij from the regression of the i'th row of {dij}
          on the i'th row of {δij};

```

- (iv) calculate σ_{NEW} using (22);
- (v) if ($\sigma_{OLD} - \sigma_{NEW}$) is not '*small*' then
 - * set $\bar{X} \leftarrow X^+$, $\bar{Y} \leftarrow Y^+$ and $\sigma_{OLD} \leftarrow \sigma_{NEW}$
 - * calculate new weights using (23)
 - * go to (i)
- else
 - * STOP

The algorithm is now illustrated for a classic set of single-peaked abundances.

Example: Internal Unfolding of upland conifer-hardwood forests of northern Wisconsin.

The original data (from Brown and Curtis, 1952) are the "importance values" of seventeen tree species in 55 woodland stands. Importance value is a compound measure of species abundance, it being the sum of relative frequency, relative density, and relative dominance of any species in a given stand. The data were standardized species-wise as indicated in (4),

Table 1. Climax adaptation numbers of tree species used in the analysis of conifer-hardwood data (source: Brown and Curtis, 1952).

Tree species	Climax adaptation number
Pinus banksiana	1
Quercus ellipsoidalis	2
Populus tremuloides	2
Populus grandidentata	2
Pinus resinosa	3
Quercus alba	4
Pinus strobus	5
Betula papyrifera	5
Acer rubrum	6
Quercus rubra	6
Abies balsamea	7
Betula lutea	8
Tsuga canadensis	8
Ulmus americana	8
Tilia americana	8
Ostrya virginiana	9
Acer saccharum	10

with a factor of 105% of the maximum importance values, and coded as (5a) and (5b). This way one obtains small, but non-zero dissimilarity in the maximum abundance cells. To keep the analysis simple, species-specific free scaling parameters were omitted. The discussion in Kershaw and Looney (1985) has served as background; they explain how Brown and Curtis obtained single-peaked importance curves for the species, the way in which a *climax adaptation number* was assigned to each species, and give other details on the original analysis. The species involved here, and their climax adaptation numbers, are given in Table 1. The climax concept implies that the vegetation has developed to a state of equilibrium with the environment, but its intricacies are definitely beyond the scope of the present paper. The adaptation numbers are simply used to label the results of the Unfolding analysis (see Figure 5). Again for reasons of simplicity, the algorithm was executed in two dimensions. Apparently the horizontal axis, ranging from *Pinus banksiana* to *Acer saccharum*, closely resembles the climax number arrangement (product-moment correlation: .97). This is a first, rather strong indication for the validity of the model. But there is plenty of variation to account for in addition to that. For instance, *Pinus resinosa* and *Quercus ellipsoidalis* almost never occur together in the same stand, even though they differ by only

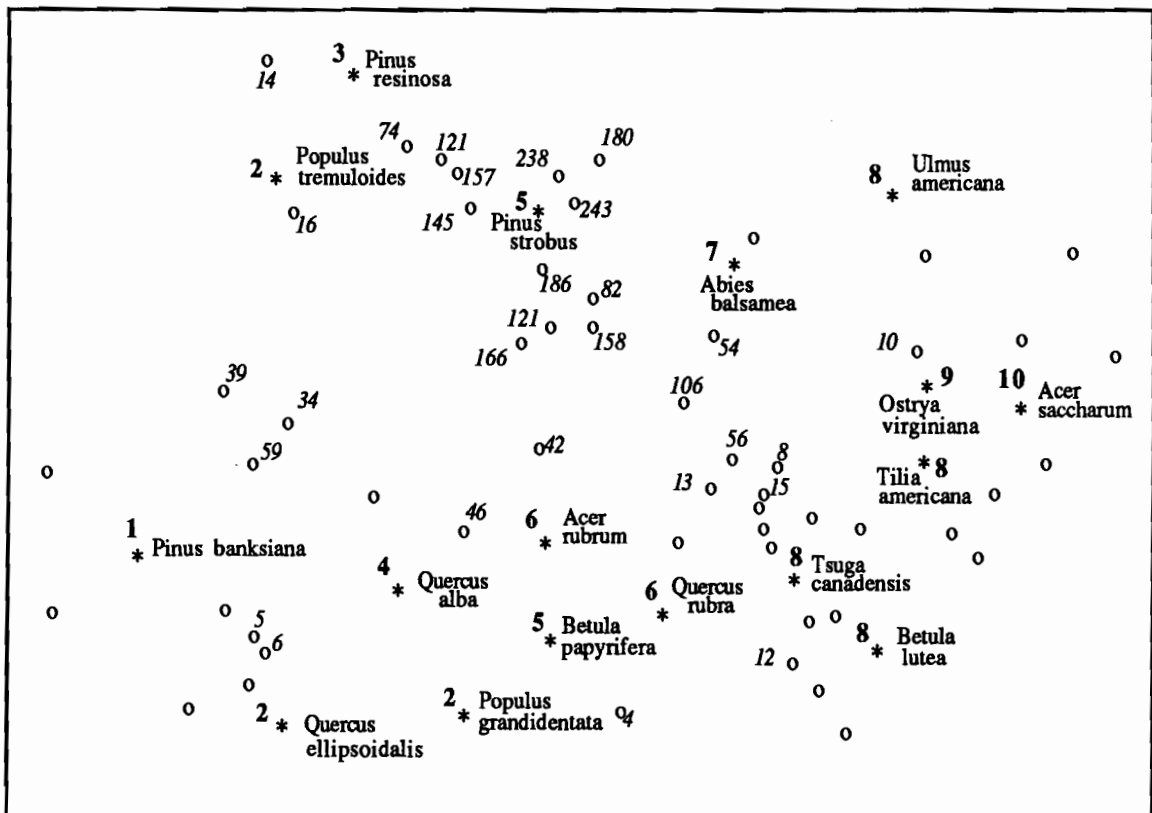


Figure 5. Internal unfolding of conifer-hardwood data (trees labelled with a '*' and their climax adaptation number, sites with a 'o' and the importance values of *Pinus strobus*)

one unit in climax number. The two-dimensional Unfolding analysis shows this by giving them a large separation in the vertical direction, as is also the case for *Betula lutea* and *Ulmus americana*, and, although less strongly, for other pairs.

The model fits the data reasonably ($\sigma_n = .2254$, which is not entirely satisfactory according to the current standards, indicating that a three-dimensional model could be called for, or optimal rescaling of the species profiles). In order to present more concrete evidence for the quality of fit, the stands in Figure 6 are labeled with the original importance values of *Pinus*

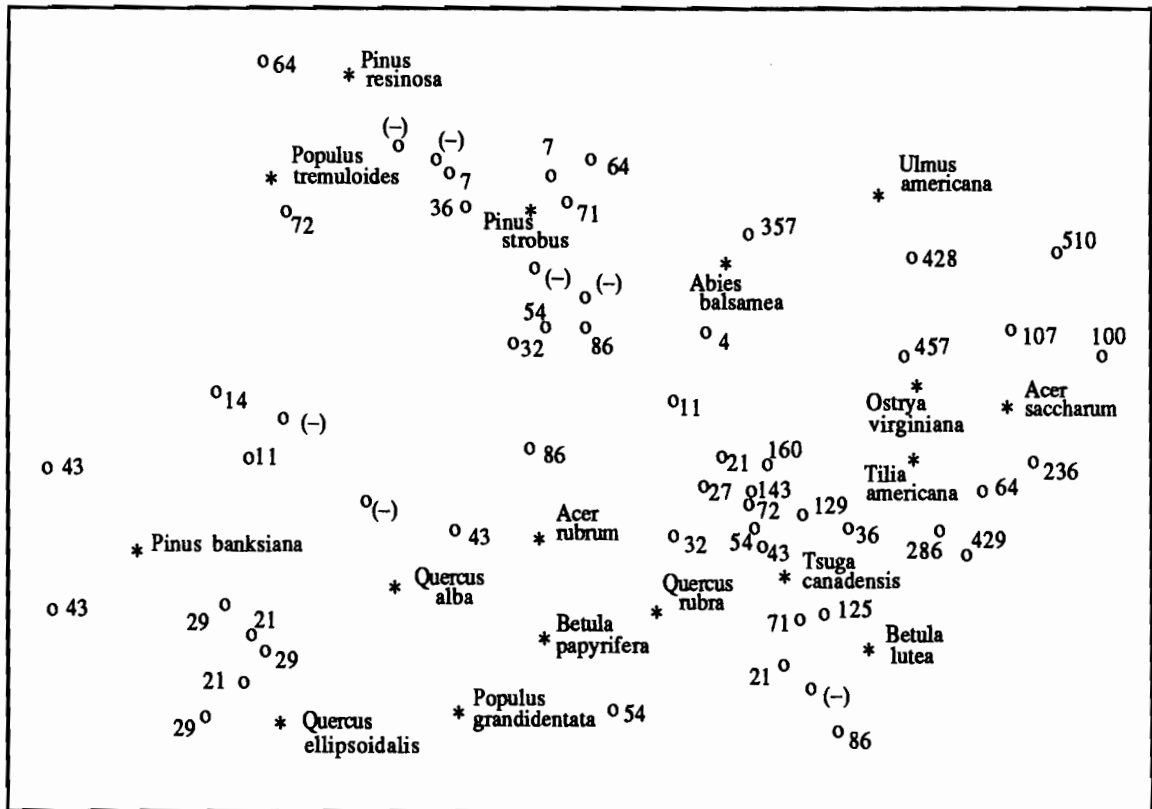


Figure 6. Alternate labelling of the sites: Calcium values (10's lb. per acre).

strobus, which shows the approximate single-peakedness clearly (*Pinus strobus* is absent in the unlabelled sites). Reconstructions of similar quality can be obtained for the other trees.

Since we now have an ordination of the stands along with the optimal tree locations, various stand characteristics can be examined to gain further understanding of the species-environment interaction. In Figure 6 the stands are labelled with their Calcium values. These tend to increase when we move from the lower left to the upper right corner,

although there are important violations against this pattern. It is especially the area around *Ulmus americana* and *Ostrya virginiana* that has characteristically high Calcium values.

3.3 Squared distance minimization: Correspondence analysis

Now that the two basic ways of Unfolding via dissimilarity approximation have been discussed, external when one of the two sets of points is fixed in advance, and internal when both sets are free to vary, let's reconsider the specification of dissimilarities and weights. Suppose that, instead of (5a) and (5b), it is specified that:

$$\delta_{ij} = 0 \quad \text{and} \quad w_{ij} = f_{ij} \quad \text{if} \quad f_{ij} > 0, \quad (24a)$$

$$\delta_{ij} = 1 \quad \text{and} \quad w_{ij} = 0 \quad \text{if} \quad f_{ij} = 0, \quad (24b)$$

where the second one is not really a change, but the first one says that a species point should *coincide* with any site where it occurs, with frequency of occurrence used as weight. When these specifications are substituted in the basic Unfolding loss function (6) one obtains:

$$\sigma^2_{CA} = \sum_i \sum_j f_{ij} d^2(\mathbf{x}_i, \mathbf{y}_j), \quad (25)$$

because the weighted sum of squared dissimilarities and the weighted sum of squared cross products vanish, due to the special structure in (24a) and (24b). The remaining part of the loss function, (25), closely resembles the location problem as defined in (10), but aims at *squared* distance minimization. It is interesting for a number of reasons.

First, note that the SMACOF algorithm breaks down immediately under this specification, because the matrices \mathbf{A} (cf. (12a) and (12b)), and thus \mathbf{P} (13a) and \mathbf{Q} (17a) all vanish. So the specification is at least *incomplete*, it has to be supplemented by a *strong* form of normalization or a radical type of restriction. A good example of the latter is of course the external approach, which now has an easy solution. To see this, it is convenient to write (25) in matrix notation, using the same symbols \mathbf{R} and \mathbf{C} as before (cf. (13b) and (17c)) for the marginal totals of the matrix $\mathbf{F} = \{f_{ij}\}$:

$$\sigma^2_{CA} = \text{tr} \mathbf{X}'\mathbf{R}\mathbf{X} + \text{tr} \mathbf{Y}'\mathbf{C}\mathbf{Y} - 2 \text{tr} \mathbf{X}'\mathbf{F}\mathbf{Y}. \quad (26)$$

For fixed \mathbf{Y} the stationary equations for a minimum of σ^2_{CA} over \mathbf{X} are (setting the partial derivatives with respect to \mathbf{X} equal to zero):

$$\mathbf{X}^+ = \mathbf{R}^{-1}\mathbf{F}\mathbf{Y}, \quad (27a)$$

and, analogously, for fixed \mathbb{X} we obtain

$$\mathbf{Y}^+ = \mathbf{C}^{-1}\mathbf{F}'\mathbb{X} , \quad (27b)$$

Comparing (27a) with the external Unfolding result (14a), it turns out that the solution to squared distance minimization merely involves taking a weighted average of the fixed points, not a transform of some previous estimate such as \mathbf{X}^{\sim} . The best location of a species ideal point now is the centre-of-gravity of the sites it occurs in. For fixed species points, the best location of a site is the centre-of-gravity of the species it is covered with.

The *internal* approach is conceptually somewhat problematical from the present point of view. First, we have to keep away from the trivial solution $\mathbf{X} = \mathbf{Y} = \mathbf{0}$, which certainly would minimize (26). In a one-dimensional analysis, this is usually done by requiring that one of the sets of scores is standardized in the metric of the marginal totals, e.g. $\mathbf{e}_n'\mathbf{R}\mathbf{x} = 0$ and $\mathbf{x}'\mathbf{R}\mathbf{x} = n$ (where the notation \mathbf{x} and \mathbf{y} is used for the vectors of *one*-dimensional species- and site *scores*, whereas x_i and y_j denote the p -dimensional species- and site *points*). The first requirement can be formulated as $\mathbf{J}_R\mathbf{x} = \mathbf{x}$ and inserted in the loss function, here \mathbf{J}_R is the projection operator

$$\mathbf{J}_R = \mathbf{I} - \mathbf{e}_n\mathbf{e}_n'\mathbf{R} / \mathbf{e}_n'\mathbf{R}\mathbf{e}_n \quad (28)$$

that centers all n -vectors, with weights \mathbf{R} . The second one can be handled by introducing a Lagrangean multiplier λ , so that the adjusted minimization problem for the simultaneous estimation of \mathbf{x} and \mathbf{y} becomes

$$\min_{\mathbf{x}} \min_{\mathbf{y}} \{ n + \mathbf{y}'\mathbf{C}\mathbf{y} - 2\mathbf{x}'\mathbf{J}_R'\mathbf{F}\mathbf{y} + \lambda \mathbf{x}'\mathbf{J}_R'\mathbf{R}\mathbf{J}_R\mathbf{x} \} , \quad (29)$$

from which it follows in the usual way that \mathbf{x}^* and \mathbf{y}^* are a solution whenever they satisfy (using the relationships $\mathbf{J}_R'\mathbf{R}\mathbf{J}_R = \mathbf{R}\mathbf{J}_R$ and $\mathbf{R}^{-1}\mathbf{J}_R' = \mathbf{J}_R\mathbf{R}^{-1}$):

$$\mathbf{x}^* = \mathbf{J}_R\mathbf{R}^{-1}\mathbf{F}\mathbf{y}^* \lambda^{-1} , \quad (30a)$$

$$\mathbf{y}^* = \mathbf{C}^{-1}\mathbf{F}'\mathbf{J}_R\mathbf{x}^* . \quad (30b)$$

These are the well-known *reciprocal averaging*, *dual scaling*, or *transition* formulas of Correspondence analysis (e.g., Nishisato, 1980). So under the specifications (24a) and (24b) of trying to minimize the distance between a species and a site in the degree of their abundance Correspondence analysis is a special way of performing internal Unfolding.

In order to obtain a more-dimensional solution, a third normalization condition must be imposed to avoid repetition of the first solution in the columns of \mathbf{X} and \mathbf{Y} (because that would actually give the smallest value of the loss function). How to do this is not free from

arbitrariness *under the present rationale* of the method. *Usually* one requires in addition that the coordinates of the higher dimensions are **R**- or **C**-orthogonal with respect to the earlier ones. This gives the stationary equations of a more-dimensional Correspondence analysis. The formulas are omitted here (but see section 3.6). Healy and Goldstein (1976) have argued that the "usual" normalization conditions are in fact *restrictions*, and they presented an alternative solution based on linear restrictions that can be freshly chosen in any particular application. Whether the freedom gained should be considered an asset or a liability is difficult to say.

Even within the confines of the usual normalization conditions there remains an awkward arbitrariness with regard to the species-site distances in a joint plot. We can just as well normalize y and leave x free, thereby obtaining the same value of the loss function. There is also the possibility to "distribute λ " among x and y . Although in all cases the weighted mean squared distance (25) remains equal, the actual Euclidean distances between species points and site points may change considerably, especially when λ is small. This was one of the reasons for Legendre and Legendre (1983, p. 278) to warn against making biplots; for who can withdraw from considering distances while looking at a configuration of points! Also note that the "folding" interpretation of picking the representation up at a species point i in order to obtain an approximate reconstruction of the i 'th row of the data matrix will give different results under different normalizations.

Finally, we may substitute (30a) in (30b), or vice versa, from which an eigenvalue-eigenvector problem in only one of the sets remains. So in contrast to the general Unfolding problem, Correspondence Analysis "has no memory" for the previous locations of the same set when solved iteratively by alternating between (30a) and (30b); in fact one of the sets of points is superfluous for solving the problem! Therefore the recognition that it is formally a special case of Unfolding has limited value. It is often preferable to view Correspondence Analysis - or, for that matter, Principal Components Analysis - as a way to perform two related, "dual" Multidimensional Scaling problems, in which one tries to fit the so-called chi-squared distances among the rows or columns of the data matrix. This specific viewpoint is more fully explained in Heiser and Meulman (1983a) and Fichet (1986). An up-to-date, comprehensive account of the method was provided by Greenacre (1984), who was also the first who seriously compared Correspondence Analysis with Unfolding in his 1978 dissertation. The use of (24a) and (24b) in connection with the standard Unfolding loss function was suggested by De Leeuw (personal communication) and more fully worked out in Heiser (1981). Hayashi (1952, 1954, 1956, 1974) based his "theory of quantification" almost entirely on (25), and dealt with many of the possible appearances the matrix F can have.

3.4 Approximation with squared distances: Gaussian ordination

In one of his early papers on Multidimensional Scaling, Shepard (1958) adduced evidence for an exponential decay function relating frequency of substitution behaviour to psychological distance. Transferring this idea, we could model expected frequency $E(f_{ij})$ as:

$$E(f_{ij}) = \beta_i e^{-d(\mathbf{x}_i, \mathbf{y}_j)} / \alpha_i, \quad (31)$$

with β_i a positive number representing the *maximum* of the function (attained when the species point \mathbf{x}_i coincides with the site point \mathbf{y}_j), and α_i a positive number representing the *dispersion* or *tolerance* of the species distribution. From (31) it follows that log expected frequency is linear in the distances:

$$\log E(f_{ij}) = \log \beta_i - d(\mathbf{x}_i, \mathbf{y}_j) / \alpha_i. \quad (32)$$

Under this model, then, we could still use the SMACOF algorithm by generalizing the definition of δ_{ij} again a little, writing

$$\delta_{ij}(\alpha_i, \beta_i) = \mu_i - \alpha_i \log f_{ij}, \quad (33)$$

where $\mu_i = \alpha_i \log \beta_i$. In fact, this model inspired the earlier definition of δ_{ij} , (5a), where μ_i could be omitted by fixing β_i equal to one ("to make the curves comparable"). Using (33) instead implies that we no longer have to use a standardization factor like f_{i+} (4) prior to the analysis, but can try to find values that optimize the fit to the data. For the skeleton algorithm it would entail step (iii) to be a linear regression including an *intercept* term. The price is n degrees of freedom and, as experience seems to attest, a less well-behaved algorithm.

Closely related to the exponential decay function is the Gaussian form

$$E(f_{ij}) = \beta_i e^{-d^2(\mathbf{x}_i, \mathbf{y}_j)} / \alpha_i, \quad (34)$$

which was studied in ecology by Ihm and van Groenewoud (1975), Austin (1976), Kooijman (1977), Gauch and Chase (1974), Gauch et al. (1974), and others. Also see Schönemann and Wang (1972). Under the Gaussian decay function it is again the species-site distance that plays the central part. But now log expected frequency is linear in the *squared* distances, and this suggests that we can use (33) in combination with the alternate loss function

$$\sigma_{SD}^2 = \sum_i \sum_j \{ \delta_{ij} - d^2(\mathbf{x}_i, \mathbf{y}_j) \}^2, \quad (35)$$

which is called SSTRESS by Takane et al. (1977), who proposed it as a general MDS loss function, and which was studied in detail for the Unfolding case by Greenacre (1978) and Greenacre and Browne (1986). Here, as in the SMACOF algorithm, δ_{ij} may be a fixed set of dissimilarities, or some function of the original frequencies like (33). The regression principle remains the same. Minimizing (35) would form a feasible and efficient alternative for the Maximum Likelihood methods of Johnson and Goodall (1980) or Ihm and van Groenewoud (1984), or the Least Squares method of Gauch et al. (1974). In the latter methods it is not the data that is transformed, but the distances. The STRESS and SSTRESS methods are based on optimal rescaling to achieve reduction of structural complexity, the same data analytic principle on which the nonlinear Multivariate Analysis and Path Analysis methods are based that are discussed by De Leeuw (1986a, 1986b) in this volume.

It is possible to relate SSTRESS and STRESS in the following way (Heiser and De Leeuw, 1979):

$$\begin{aligned}\sigma_{SD}^2 &= \sum_i \sum_j \{ \sqrt{\delta_{ij}} + d(\mathbf{x}_i, \mathbf{y}_j) \}^2 \{ \sqrt{\delta_{ij}} - d(\mathbf{x}_i, \mathbf{y}_j) \}^2 \\ &\approx 4 \sum_i \sum_j \delta_{ij} \{ \sqrt{\delta_{ij}} - d(\mathbf{x}_i, \mathbf{y}_j) \}^2 ,\end{aligned}\quad (35)$$

the approximation being better if the dissimilarities and distances match well. So we can simulate SSTRESS solutions with the SMACOF algorithm by using an additional square root transformation and choosing the dissimilarities as weights. This way of weighting will tend to de-emphasize good fit of local relationships, in favour of getting the large distances right.

Ihm and van Groenewoud (1984), Ter Braak (1985), and Ter Braak and Barendregt (1986) recently compared Maximum Likelihood estimation under the Gaussian response model with Correspondence Analysis, as we have seen a technique also based on the squared Euclidean distance function. The results are encouraging for CA, especially if the species dispersions are homogeneous.

3.5. Further special cases and extensions

Kershaw (1968) used a *square root* transformation of the abundances to make them less heterogeneous. It is, of course, one of the usual statistical ways to stabilize the variance. Now suppose we take the *inverse square root* as an alternative definition of dissimilarity, and the *frequencies themselves* as weights:

$$\delta_{ij} = 1 / \sqrt{f_{ij}} \quad \text{and} \quad w_{ij} = f_{ij} \quad \text{if} \quad f_{ij} > 0 , \quad (37a)$$

$$\delta_{ij} = 1 \quad \text{and} \quad w_{ij} = 0 \quad \text{if} \quad f_{ij} = 0 , \quad (37b)$$

Then the basic loss function σ^2_R transforms into (P denotes all pairs present, (37a))

$$\begin{aligned}\sigma^2_R &= \sum_{(i,j) \in P} f_{ij} \{1/\sqrt{f_{ij}} - d(\mathbf{x}_i, \mathbf{y}_j)\}^2 \\ &= \sum_{(i,j) \in P} \{1 - \sqrt{f_{ij}} d(\mathbf{x}_i, \mathbf{y}_j)\}^2 = \sum_{(i,j) \in P} \{1 - d(\mathbf{x}_i, \mathbf{y}_j)/\delta_{ij}\}^2,\end{aligned}\quad (38)$$

Thus loss is measured in terms of the ratio of distance and dissimilarity (Underhill, 19??), and we now obviously give more weight to the *small* dissimilarities. It is interesting to compare this weighting structure with yet another loss function, proposed by Ramsay (1977). He similarly argued that dissimilarity measurements in psychology are frequently *lognormally* distributed. The lognormal arises from the *product* of many independent and (nearly) identically distributed random variables. It has been frequently applied as a model for the variation of *nonnegative* quantities (Aitchison and Brown, 1957; Derman et al., 1973), indeed also for abundances (Grundy, 1951). If dissimilarity is assumed to be lognormally distributed we should work with

$$\sigma^2_{MS} = \sum_i \sum_j \{\log \delta_{ij} - \log d(\mathbf{x}_i, \mathbf{y}_j)\}^2,\quad (39)$$

which forms the basis of Ramsay's MULTISCALE algorithm. In order to relate it to the standard loss, we can use the first order approximation

$$\log d(\mathbf{x}_i, \mathbf{y}_j) \approx \log \delta_{ij} + \{\delta_{ij} - d(\mathbf{x}_i, \mathbf{y}_j)\} / \delta_{ij},\quad (40)$$

from which it follows that (De Leeuw and Heiser, 1982):

$$\sigma^2_{MS} \approx \sum_i \sum_j \{1 / \delta_{ij}\}^2 \{\delta_{ij} - d(\mathbf{x}_i, \mathbf{y}_j)\}^2.\quad (41)$$

So Ramsay's loss function can be approximated by using the inverse squared dissimilarities as weights in the standard loss function. The same reasoning is present in (37a), which led to (38).

The choice between so many possible types of transformation of the raw data can be circumvented by defining a radically extended class of transformations as

$$\delta_{ij} \geq \delta_{ik} \quad \text{if} \quad f_{ij} < f_{ik}.\quad (42)$$

So dissimilarity should increase whenever abundance decreases, for each species separately. This specification would form the basis of a row-conditional, *nonmetric* Unfolding algorithm. The idea to pose merely monotonicity (42) as the basis of the technique is due to Coombs (1950). He did not provide a working algorithm, however; it was not until the sixties that Shepard, Kruskal, Guttman and others developed general

nonmetric MDS algorithms (Kruskal, 1977; De Leeuw and Heiser, 1982). Technically, our skeleton algorithm only needs alteration in step (iii), where the type of regression performed should be of the *monotonic*, or *isotonic*, variety (Kruskal, 1964a, 1964b; Barlow et al., 1972). Yet the nonmetric Unfolding case always remained something of a problem, due to a phenomenon called *degeneration*: a tendency to collapse many points, or, anyhow, to make all distances equal (cf section 4.3). These problems, and proposals to resolve them (although not fully satisfactorily), are explained in Kruskal and Carroll (1969) and in Heiser (1981), who argued that it is necessary to put *bounds* on the regression. Subsequently Heiser (1985, 1986) proposed a *smoothed* form of monotonic regression in order to obtain a better behaving algorithm, and this refinement might make standard application of nonmetric Unfolding feasible.

The one-dimensional case of any STRESS minimizing algorithm deserves special care. Guttman (1968) already pointed out its special status, and De Leeuw and Heiser (1977), also see Heiser (1981), showed that the SMACOF algorithm does not really resolve the combinatorial complications that arise in this case. Quite independently, Wilkinson (1971) made some insightful observations on a form of one-dimensional Unfolding, and showed the connection with the so-called travelling salesman problem. Poole (1984) analysed the situation along the lines of the graphical version of the algorithm in Figure 4, and proposed an improvement for the one-dimensional case. Fortunately we now also have Hubert and Arabie (1986), who provided a *globally convergent*, dynamic programming algorithm for one-dimensional MDS, extending the work of Defays (1978). Little is known about its performance in the Unfolding situation, but it surely marks an exciting step forward.

3.6 Restrictions on the locations

In this section the major tools are described for restricting the locations of either the species points, or the site points, or both. This is done first for the SMACOF algorithm, next for Correspondence Analysis. Remember the SMACOF algorithm always starts with the preliminary updates \tilde{X} and \tilde{Y} , as defined in (14a) and (18). These provide the basic corrections necessary to obtain a better fit to the dissimilarities. From the general results of De Leeuw and Heiser (1980) it then follows that the remaining task is to find

$$\min_{(\mathbf{X}, \mathbf{Y}) \in \Omega} \text{tr} \{ \mathbf{X}'\mathbf{R}\mathbf{X} + \mathbf{Y}'\mathbf{C}\mathbf{Y} - 2\mathbf{X}'\mathbf{W}\mathbf{Y} - 2\mathbf{X}'\tilde{\mathbf{X}} - 2\mathbf{Y}'\tilde{\mathbf{Y}} \}, \quad (43)$$

where Ω is the domain of minimization, or *feasible region*. When \mathbf{X} and \mathbf{Y} are completely free, Ω is the set of all (combined) $n \times p$ and $m \times p$ matrices, and from equating the partial derivatives to zero one obtains the system of linear equations (19a) and (19b) for the unrestricted internal Unfolding problem. In De Leeuw and Heiser (1980) it is also shown that it is not at all necessary to solve problem (43) *completely*; it suffices to move from a feasible point into the right direction for minimizing it. The algorithm will still converge to at least a local minimum. This important fact opens the possibility to use *Alternating Least Squares*, i.e. to split the parameter set into subsets, and to alternate among the subset minimizations. The obvious candidate for a first split in the Unfolding situation is into \mathbf{X} and \mathbf{Y} , and accordingly (43) can be split into two subproblems (again writing \mathbb{X} and \mathbb{Y} for fixed matrices, and after some rearrangement of terms):

$$\min_{\mathbf{X} \in \Omega_{\mathbf{X}}} \text{tr} \{ \mathbf{X} - \mathbf{R}^{-1}(\tilde{\mathbf{X}} + \mathbf{W}\mathbb{Y}) \}' \mathbf{R} \{ \mathbf{X} - \mathbf{R}^{-1}(\tilde{\mathbf{X}} + \mathbf{W}\mathbb{Y}) \} + \text{constant}, \quad (44a)$$

$$\min_{\mathbf{Y} \in \Omega_{\mathbf{Y}}} \text{tr} \{ \mathbf{Y} - \mathbf{C}^{-1}(\tilde{\mathbf{Y}} + \mathbf{W}'\mathbb{X}) \}' \mathbf{C} \{ \mathbf{Y} - \mathbf{C}^{-1}(\tilde{\mathbf{Y}} + \mathbf{W}'\mathbb{X}) \} + \text{constant}. \quad (44b)$$

These are two *projection* problems, one in the metric \mathbf{R} and the other in the metric \mathbf{C} . The former immediately gives (14b), the solution to the external Unfolding problem when \mathbf{Y} is fixed. It will be evident that there is a variety of possibilities now *in between* the internal and the external approach (in between indirect and direct Gradient Analysis).

In Heiser (1981, chapter 8) two examples of restricted Unfolding were studied in detail. For preference data with respect to family compositions, i.e. combinations of number of sons and number of daughters, *equality constraints* were used in such a way that the family points would always form a rectangular grid in two dimensions. So personal preference was supposed to be single-peaked with respect to the grid, of which the spacings were left

free to vary. The resulting value of STRESS turned out to be only slightly higher than in the unrestricted case, thus confirming the validity of supposing lack of interaction. The second example concerned preferences of 137 Members of the Dutch Parliament for nine political parties, and it used their stands on seven controversial issues as *inequality constraints*. Note that (44a) and (44b) can be split further down into dimension-wise components, and this way each axis was associated with a single issue; the subproblems become weighted monotone regression problems. For further examples and refinements, as well as references to other work on restricted MDS, see De Leeuw and Heiser (1980), Heiser and Meulman (1983a, 1983b), and Meulman and Heiser (1984). Heiser (1981, chapter 6) also discusses the possibility to impose *centroid constraints*, implying that each species should be located in the centre-of-gravity of the sites in which it is dominant.

This brings us back to Correspondence Analysis. Remember the basic averaging formulas (27a) and (27b). In order to incorporate restrictions on \mathbf{X} and \mathbf{Y} , these weighted averages must now be regarded as the *preliminary* updates. Suppose we normalize \mathbf{Y} and keep it fixed, and want to restrict $\mathbf{X} \in \Omega_{\mathbf{X}}$. If we write

$$\mathbf{X} = \mathbf{X}^{\approx} + (\mathbf{X} - \mathbf{X}^{\approx}) \quad \text{with} \quad \mathbf{X}^{\approx} = \mathbf{R}^{-1}\mathbf{F}\mathbf{Y}, \quad (45)$$

then it may be verified that the CA loss function transform into

$$\sigma_{\text{CA}}^2 = \text{tr} (\mathbf{X} - \mathbf{X}^{\approx})'\mathbf{R}(\mathbf{X} - \mathbf{X}^{\approx}) + \text{tr} \mathbf{Y}'(\mathbf{C} - \mathbf{F}'\mathbf{R}^{-1}\mathbf{F})\mathbf{Y}. \quad (46)$$

The second term on the right-hand side of (46) is constant, so we again end up with a projection problem in the metric \mathbf{R} , in which \mathbf{X}^{\approx} rather than $\mathbf{R}^{-1}(\mathbf{X}^{\sim} + \mathbf{W}\mathbf{Y})$ must be projected onto the feasible region. All the possibilities of restrictions mentioned for the SMACOF algorithm are now open to us for Correspondence Analysis. Historically, it is not quite fair to say this, because a lot of them were used earlier in the developing Gifi system (cf. Gifi, 1981). Still, the formulation presented here is new, and especially putting together (44a) and (46) clarifies the similarities and differences between Unfolding and Correspondence Analysis a great deal. Ter Braak (1986a, 1986b) has further developed the case in which the site locations are linear combinations of environmental variables, under the name "Canonical Correspondence Analysis".

A special example of restrictions in CA is Hill and Gauch' (1980) method of *Detrended Correspondence Analysis*. They don't compute all dimensions simultaneously, but work successively. Their aim is to remove the horseshoe effect, and other nonlinearities in higher dimensions. To bring it in the present formulation, suppose \mathbf{x}_1 is the first set of CA scores, satisfying - as explained in section 3.3 - $\mathbf{J}_{\mathbf{R}}\mathbf{x}_1 = \mathbf{x}_1$ and $\mathbf{x}_1'\mathbf{R}\mathbf{x}_1 = n$. Then, instead of requiring \mathbf{R} -orthogonality of \mathbf{x}_2 , i.e. $\mathbf{x}_2'\mathbf{R}\mathbf{x}_1 = 0$, the idea is to have \mathbf{x}_2 *locally centered*.

To do this, an $n \times k_G$ matrix G can be formed on the basis of x_1 , indicating a partitioning into k_G blocks of species that are close together on x_1 . Thus G is binary and $G'G$ is diagonal. The projection matrix

$$J_G = I - G(G'G)^{-1}G' \quad (47)$$

is the required block-wise centering operator, and the new requirement becomes $J_G x_2 = x_2$. This can be inserted in (46), which shows that we have to solve

$$\min_{x_2} (J_G x_2 - x_2^*)' R (J_G x_2 - x_2^*) . \quad (48)$$

The weak point in this method is that it does not provide a unique, convincing definition of G , as a result of which it may sometimes detrend too much, sometimes too little. This objection is comparable to the earlier remark regarding the restrictions of Healy and Goldstein (1976), if understood as a general way to proceed.

4. Miscellaneous issues and discussion

4.1 Homogeneity analysis

Homogeneity Analysis is the key method of the Gifi system of Nonlinear Multivariate Analysis (De Leeuw, 1984; 1986a). It employs indicator matrices as a basis for all nonlinear transformations of a given set of variables, and selects precisely those transformations that are as homogeneous as possible. If the data matrix F in Correspondence Analysis is chosen as the set of concatenated indicator matrices we obtain CA solutions that are essentially equivalent to those of Homogeneity Analysis. An extended discussion on the details of this connection can be found in Heiser (1981, chapter 4). There, as well as in Heiser (1985), it was argued that in the case of shifted single-peaked variables the homogeneity approach should not be followed without restraint. If we think it is characteristic for species to have distributions that are shifted with respect to each other, we should not center them (which is part of making them as homogeneous as possible). If, moreover, the variables are thought to give an *asymmetrical* type of information, i.e. high abundance indicates *similarity* of sites and low abundance *dissimilarity*, then we should not try to give equally dissimilar sites as much as possible the same quantification.

Homogeneity analysis can still be used provided the right kind of *change of variables*, or variable *coding*, is chosen. One possibility is to use *conjoint coding* (Heiser, 1981, p.123), which associates a nested sequence of sites to each species. This method was proposed

earlier by Wilkinson (1971) and, independently, by Hill et al. (1975), who called it the "method of pseudo-species" (also see Hill, 1977). A second possibility is to use *convex coding* (Heiser, 1981, section 5.3), which is especially tailored to the situation that there are more species or individuals than sites, because it uses the geometrical property that the site space can be partitioned into so-called isotonic regions. Since these alternative ways of coding have not yet been used a great deal, their data analytic value is unsure.

4.2 Optimal rearrangement

It is well-known that both Correspondence Analysis and Homogeneity Analysis have a remarkable rearrangement property: if the rows of the table can be reordered in such a way that all columns become single-peaked, or have the so-called *consecutive ones* property, then both techniques will find the correct ordering as their first dimension (see Guttman (1950) and Hill (1974) for somewhat less general statements; Heiser (1981, section 3.2) proved the proposition in the form stated here; see Schriever (1985) for a comprehensive discussion of such ordering properties).

One would of course like to be able to say that *each* Unfolding method shares this property, but it is an open question under what conditions anyone Unfolding technique can be said to achieve an optimal rearrangement in the above sense. Perhaps it is necessary to assume *symmetry* of the single-peaked functions. A second open question is how to devise an efficient method that directly optimizes the single-peakedness condition. Wilkinson (1971) proposed a combinatorial method to find a permutation yielding consecutive ones, but little is known about its effectiveness.

4.3 Horseshoes

It is important to discern at least *four* different situations in which a curved configuration of points can arise from a more-dimensional analysis. All of them have occasionally been indicated with the term "horseshoe".

In the first place there is the polynomial curvature emerging in Correspondence Analysis and Homogeneity analysis when the first dimension is strongly dominant. This could best be called the *Guttman effect*, as is usually done in France, because it gives the right credit to Guttman (1950). The background of this phenomenon was discussed recently in greater detail by De Leeuw (1982) and Van Rijkevorsel (1986).

In the second place there is the more strongly curved, sometimes even convoluted case obtained when the principal components of single-peaked data are studied directly (i.e.,

without the normalizations, centering, and weighting involved in CA). Here the points are frequently distributed along the greater part of circles, ellipses and ellipsoids; it is much more difficult to recognize such regularities in practice. Therefore, in evaluation studies of ordination techniques, such as Whittaker and Gauch (1978), CA is usually considered to be the more satisfactory technique.

In as much the data are reasonably single-peaked, STRESS minimizing Unfolding techniques will not produce any curvature at all. Yet in some circumstances a horseshoe effect can be encountered as well. Again if the data are close to being one-dimensional, this time *in terms of the distances*, both MDS and Unfolding tend to produce C- or S-shaped configurations. Shepard (1974, p.386) characterized the situation as follows: "Evidently, by bending away from a one-dimensional straight line, the configuration is able to take advantage of the extra degrees of freedom provided by additional dimensions to achieve a better fit to the random fluctuations in the similarity data. In some published applications, moreover, the possibility of the more desirable one-dimensional result was mistakenly dismissed because the undetected occurrence of merely a local minimum (which is especially likely in one-dimension) made the one-dimensional solution appear to yield an unacceptably poor monotone fit and/or substantive interpretation." Meanwhile there has been considerable technical progress for the one-dimensional case (cf. section 3.5). Also, it seems likely that the MDS-horseshoe frequently arises from the occurrence of large tie-blocks of large dissimilarities, for instance when they are derived from presence-absence data. In such cases it is advisable to *down-weight the large distances*, which also forms the basis of the so-called *parametric mapping* technique (Shepard and Carroll, 1966). In many of the specifications in the previous sections down-weighting was used as well.

Finally, there is a typical horseshoe effect for Unfolding, due to *regression to the mean*. If the regression part in the Unfolding algorithm is not selected carefully, for instance a straight-forward monotone regression is inserted, then the technique capitalizes on a general property of many kinds of regression to yield regressed values that are more homogeneous than the regressants. The Unfolding technique is attracted to the extreme case of (nearly) *equal dhats*, because it can so easily find a configuration with equal distances: all points of one set collapsed at a single location, all points of the other set distributed on part of a circle or sphere around it. Linear or polynomial regression *without an intercept*, and restricted forms of monotone regression seem to provide the best safeguards against this type of *degeneration* (cf. section 3.5).

4.4 Q and R

Noy-Meir and Whittaker (1978) remark on this issue: "Another division among ordinations was between R -methods, which derive axes from a matrix of similarities between species (e.g. factor analysis as normally used), and Q -methods, which derive them from a matrix of similarities between samples (...). The issue was confused because it was not always clear whether R/Q referred to the matrix used (between species / between sites), to the standardization applied (by species / by sites), or to the ordination obtained (of species / of sites)." (l.c., p.342). A few pages later they state that the R/Q problem can be resolved by recognition of the dual character of many multivariate methods, as explained so clearly, indeed, in Gower (1966).

To a psychometrician, it is very remarkable in the above quotation that the case of similarities between species is called R , because the comparable case of similarities between people (or between psychological types, or between aggregate units like schools or political parties) would definitely be called Q by him. Is R what is commonly done, and Q its complement?

Burt and Stephenson had quite an extensive controversy on the issue in the late thirties and forties, in which Burt put forward, and rudimentally proved, the duality properties of Principal Components Analysis. A joint statement of the things they agreed upon and on their differences appeared in Burt and Stephenson (1939). Stephenson did not deny that one can analyse the same data matrix *as is* or *turned on its side*, with comparable or - at least in principle - transferable results. But he claimed that R and Q are *entirely different methodologies*, with different aims at both the theoretical and the operational level. These were further explained in a remarkable book, Stephenson (1953), which appears to have been misunderstood by many psychometricians at the time. His claims of putting forward a whole new perspective might not surprise ecologists, since they seem to have been using it already quite commonly. The most salient aspects of the R/Q distinction according to Stephenson are that R studies interdependencies among "objective" descriptors, while Q studies dependencies of behavioral reactions upon preset conditions. In R a "factor" is a linear combination of measured variables, in Q the term "factor" is more frequently, although not unequivocally, used in the sense of Fisher's design of experiments. R aims at forming compounds of quantitative individual differences between organisms (perhaps temporarily only qualitatively measured), while Q tries to assess *qualities of performance* of any one organism (perhaps eventually grouped with others into classes, or quantified along a scale).

Applying the distinction to the example of the conifer-hardwood forests (section 3.2): it would be a matter of R to combine - within one stand, repeated over similar stands - relative

frequency, relative density, relative dominance, and possibly other indicators of population strength into one importance measure. A Principal Components Analysis could then show that two compound measures of importance are needed, or even more, and could in any case replace simple addition. On the other hand, it is quite Q to assign climax adaptation numbers to the species, or to scale them - as we have done - on the basis of similar reactions to the environment. Stephenson would have used PCA for this, with species as variables. In addition, it is also very Q to relate the classes, or the scale of species, to systematic stand characteristics. Undoubtedly, Stephenson would have preferred to select the stands on the basis of, e.g., "factorial" combinations of Calcium value and moisture status, rather than measuring these as covariates. He thought that Fisher's randomization principle, in combination with systematic variation, should be applied in non-experimental contexts too.

In conclusion, we can make R and Q observations, and can have R and Q datamatrices, either of which can be useful. There is not a classification of analysis methods involved, although it seems to be true that MDS and Unfolding are eminently suited for Q purposes.

5. References

- Aitchison, J. and Brown, J.A.C. (1957). **The Lognormal Distribution**. New York: Cambridge University Press.
- Austin, M.P. (1976). On non-linear species response models in ordination. **Vegetatio**, **33**, 33-41.
- Austin, T.L.jr. (1959). An approximation to the point of minimum aggregate distance. **Metron**, **19**, 10-21.
- Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). **Statistical Inference under Order Restrictions**. New York: Wiley.
- Braun-Blanquet, J. and Jenny, H. (1926). Vegetationsentwicklung und Bodenbildung in der alpinen Stufe der Zentralalpen. **Neue Denkschr. Schweiz. Naturforsch. Ges.**, **63**, 175-349.
- Bray, R.J. and Curtis, J.T. (1957). An ordination of the upland forest communities of Southern Wisconsin. **Ecol. Monogr.**, **27**, 325-249.
- Brown, R.T. and Curtis, T.T. (1952). The upland conifer-hardwood forests of northern Wisconsin. **Ecol. Monogr.**, **22**, 217-234.
- Burt, C. and Stephenson, W. (1939). Alternative views on correlations between persons. **Psychometrika**, **4**, 269-281.
- Carroll, J.D. (1969). Polynomial factor analysis. **Proc. 77'th Annual Convention of the APA**, **4**, 103-104.
- Carroll, J.D. (1972). Individual differences and multidimensional scaling. In R.N. Shepard et al., **Multidimensional Scaling, Vol I: Theory**. New York: Seminar Press, 105-155.
- Coombs, C.H. (1950). Psychological Scaling without a unit of measurement. **Psych. Rev.**, **57**, 148-158.
- Coombs, C.H. (1964). **A Theory of Data**. New York: Wiley.
- Coombs, C.H. and Smith, J.E.K. (1973). On the detection of structure in attitudes and developmental processes. **Psych. Rev.**, **80**, 337-351.

- Cottam, G. and Curtis, J.T. (1956). The use of distance measures in phytosociological sampling. *Ecology*, **37**, 451-460.
- Coxon, A.P.M. (1974). The mapping of family-composition preferences: A scaling analysis. *Social Science Research*, **3**, 191-210.
- Curtis, J.T. and McIntosh, R.P. (1951). An upland continuum in the prairie-forest border region of Wisconsin. *Ecology*, **32**, 476-496.
- Davison, M.L., King, P.M., Kitchener, K.S. and Parker, C.A. (1980). The stage sequence concept in cognitive and social development. *Developm. Psych.*, **16**, 121-131.
- De Leeuw, J. (1977). Applications of convex analysis to multidimensional scaling. In: J.R. Barra et al. (Eds.), **Recent Developments in Statistics**. Amsterdam: North-Holland, 133-145.
- De Leeuw, J. (1982). Nonlinear Principal Component Analysis. In: H. Caussinus et al. (Eds.), **COMPSTAT 1982**. Vienna: Physica Verlag, 77-89.
- De Leeuw, J. (1984). The Gifi system of Nonlinear Multivariate Analysis. In: E. Diday et al. (Eds.), **Data Analysis and Informatics, III**. Amsterdam: North-Holland, 415-424.
- De Leeuw, J. (1986a). Nonlinear Multivariate Analysis with Optimal Scaling. In: these **Proceedings**.
- De Leeuw, J. (1986b). Nonlinear Path Analysis with Optimal Scaling. In: these **Proceedings**.
- De Leeuw, J. and Heiser, W.J. (1977). Convergence of correction matrix algorithms for multidimensional scaling. In: J. Lingoes (Ed.), **Geometric representations of relational data**. Ann Arbor: Mathesis Press, 735-752.
- De Leeuw, J. and Heiser, W.J. (1980). Multidimensional scaling with restrictions on the configuration. In: P.R. Krishnaiah (Ed.), **Multivariate Analysis, Vol V**. Amsterdam: North-Holland, 501-522.
- De Leeuw, J. and Heiser, W.J. (1982). Theory of Multidimensional Scaling. In: P.R. Krishnaiah and L.N. Kanal (Eds.), **Handbook of Statistics, Vol 2**. Amsterdam: North-Holland, 285-316.
- Defays, D. (1978). A short note on a method of seriation. *Brit. J. Math. Stat. Psych.*, **31**, 49-53.
- Derman, C., Gleser, L.J. and Olkin, I. (1973). **A Guide to Probability Theory and Application**. New York: Holt, Rinehart and Winston.
- Draper, N.R. and Smith, H. (1966). **Applied Regression Analysis**. New York: Wiley.
- Fichet, B. (1986). Distances and Euclidean distances for presence-absence characters and their application to factor analysis. In: J. de Leeuw et al. (Eds.), **Multidimensional Data Analysis**. Leiden: DSWO Press, in press.
- Gabriël, K.R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, **58**, 453-467.
- Gauch, H.G. (1982). **Multivariate analysis in community ecology**. Cambridge: Cambridge University Press.
- Gauch, H.G. and Chase, G.B. (1974). Fitting the Gaussian curve to ecological data. *Ecology*, **55**, 1377-1381.
- Gauch, H.G., Chase, G.B. and Whittaker, R.H. (1974). Ordination of vegetation samples by Gaussian species distributions. *Ecology*, **55**, 1382-1390.
- Gause, C.F. (1930). Studies of the ecology of the orthoptera. *Ecology*, **11**, 307-325.
- Gifi, A. (1981). **Nonlinear Multivariate Analysis**. Leiden: Department of Data Theory, University of Leiden.
- Gittins, R. (1985). **Canonical Analysis: A Review with Applications in Ecology**. Berlin: Physica Verlag.
- 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.
- Gower, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, **53**, 325-338.

- Greenacre, M.J. (1978). **Some objective methods of graphical display of a data matrix**. Special Report, Dept. of Statistics and Operations Research, University of South-Africa, Pretoria.
- Greenacre, M.J. (1984). **Theory and Applications of Correspondence Analysis**. London: Academic Press.
- Greenacre, M.J. and Browne, M.W. (1986). An efficient alternating least squares algorithm to perform multidimensional Unfolding. *Psychometrika*, **51**, in press.
- Greig-Smith, P. (1983). **Quantitative Plant Ecology**, 3rd Ed. London: Blackwell Scient. Publ.
- Grundy, P.M. (1951). The expected frequencies in a sample of an animal population in which the abundances of species are lognormally distribute. I. *Biometrika*, **38**, 427-434.
- Guttman, L. (1950). The Principal Components of scale analysis. In: S.A. Stouffer et al., **Measurement and Prediction**. Princeton: Princeton University Press.
- Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, **33**, 469-506.
- Hayashi, C. (1952). On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Ann. Inst. Statist. Math.*, **2**, 93-96.
- Hayashi, C. (1954). Multidimensional quantification - with applications to analysis of social phenomena. *Ann. Inst. Stat. Math.*, **5**, 121-143.
- Hayashi, C. (1956). Theory and example of quantification, II. *Proc. Inst. Stat. Math.*, **4**, 19-30.
- Hayashi, C. (1974). Minimum dimension analysis MDA. *Behaviormetrika*, **1**, 1-24.
- Healy, M.J.R. and Goldstein, H. (1976). An approach to the scaling of categorised attributes. *Biometrika*, **63**, 219-229.
- Heiser, W.J. (1981). **Unfolding Analysis of Proximity Data**. Ph.D.Thesis, University of Leiden, The Netherlands.
- Heiser, W.J. (1985a). Undesired nonlinearities in Nonlinear Multivariate Analysis. In: E. Diday, **Data Analysis and Informatics, IV**. Amsterdam: North-Holland, in press.
- Heiser, W.J. (1985b). Multidimensional Scaling by optimizing goodness-of-fit to a smooth hypothesis. **Internal Report RR-85-07**, Dept. of Data Theory, University of Leiden.
- Heiser, W.J. (1986). Order invariant Unfolding analysis under smoothness restrictions, **Internal Report RR-86-07**, Dept. of Data Theory, University of Leiden.
- Heiser, W.J. and De Leeuw, J. (1979). How to use SMACOF-I (2nd edition). **Internal Report**, Dept. of Data Theory, University of Leiden.
- Heiser, W.J. and Meulman, J. (1983a). Analyzing rectangular tables by joint and constrained Multidimensional Scaling. *J. Econometrics*, **22**, 139-167.
- Heiser, W.J. and Meulman, J. (1983b). Constrained Multidimensional Scaling, including confirmation. *Applied Psych. Meas.*, **22**, 139-167.
- Hill, M.O. (1974). Correspondence Analysis: a neglected multivariate method. *Applied Statistics*, **23**, 340-354.
- Hill, M.O. (1977). Use of simple discriminant functions to classify quantitative phytosociological data. In: E. Diday et al. (Eds.), **Data Analysis and Informatics, I**. Le Chesnay, France: INRIA, 181-199.
- Hill, M.O., Bunce, R.G.H. and Shaw, M.W. (1975). Indicator species analysis, a divisive polythetic method of classification, and its application to a survey of native pinewoods in Scotland. *J. Ecol.*, **63**, 597-613.
- Hill, M.O. and Gauch, H.G. (1980). Detrended Correspondence analysis: an improved ordination technique. *Vegetatio*, **42**, 47-58.
- Hodson, F.R. et al. (Eds.) (1971). **Mathematics in the Archaeological and Historical Sciences**. Edinburgh: Edinburgh University Press.
- Hovland, C.I., Harvey, O.J., and Sherif, M. (1957). Assimilation and contrast effects in reactions to communication and attitude change. *J. Abnorm. Soc. Psych.*, **55**, 244-252.

- Hubert, L. and Arabie, Ph. (1986). Unidimensional scaling and combinatorial optimization. In: J. de Leeuw et al. (Eds.), **Multidimensional Data Analysis**. Leiden: DSWO Press (in press).
- Igoshina, K.N. (1927). Die Pflanzengesellschaften der Alluvionen der Flüsse Kama und Tschussowaja (in Russian with German summary). **Trav. de l'Inst. Biol. à l'Univ. de Perm**, **1**, 1-117.
- Ihm, P. and Van Groenewoud, H. (1975). A multivariate ordering of vegetation data based on Gaussian type gradient response curves. **J. Ecol.**, **63**, 767-777.
- Ihm, P. and Van Groenewoud, H. (1984). Correspondence Analysis and Gaussian Ordination. **COMPSTAT Lectures 3**. Vienna: Physica Verlag, 5-60.
- Johnson, R.W. and Goodall, D.W. (1980). A maximum likelihood approach to non-linear ordination. **Vegetation**, **41**, 133-142.
- Kendall, D.G. (1963). A statistical approach to Flinders Petrie's sequence dating. **Bull. Inst. Statist. Inst.**, **40**, 657-680.
- Kershaw, K.A. (1968). Classification and ordination of Nigerian savanna vegetation. **J. Ecol.**, **56**, 467-482.
- Kershaw, K.A. and Looney, J.H.H. (1985). **Quantitative and Dynamic Plant Ecology**, 3rd Ed. London: Edward Arnold Publ.
- Kooijman, S.A.L.M. (1977). Species abundance with optimum relations to environmental factors. **Ann. Systems Res.**, **6**, 123-138.
- Kruskal, J.B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. **Psychometrika**, **29**, 1-28.
- Kruskal, J.B. (1964b). Nonmetric multidimensional scaling: a numerical method. **Psychometrika**, **29**, 115-129.
- Kruskal, J.B. (1977). Multidimensional Scaling and Other Methods for Discovering Structure. In: Enslein, K. Ralston, A. and Wilf, H.S. (Eds.), **Statistical Methods for Digital Computers**, Vol III. New York: Wiley, 296-339.
- Kruskal, J.B. and Carroll, J.D. (1969). Geometrical models and badness-of-fit functions. In: P.R. Krishnaiah (Ed.), **Multivariate Analysis II**. New York: Academic Press, 639-671.
- Kuhn, H.W. (1967). On a pair of dual nonlinear programs. In: J. Abadie (Ed.), **Methods of nonlinear programming**. Amsterdam: North-Holland, 38-54.
- Legendre, L. and Legendre, P. (1983). **Numerical Ecology**. Amsterdam: Elsevier Scient. Publ.
- McDonald, R.P. (1962). A general approach to nonlinear factor analysis. **Psychometrika**, **27**, 397-415.
- McDonald, R.P. (1967). Nonlinear factor analysis. **Psychometric Monograph**, **15**.
- Meulman, J. and Heiser, W.J. (1984). Constrained multidimensional scaling: more directions than dimensions. In: T. Havránek et al. (Eds.), **COMPSTAT 1984, Proceedings in Computational Statistics**. Vienna: Physica Verlag, 137-142.
- Nishisato, S. (1980). **Analysis of categorical data: dual scaling and its applications**. Toronto: University of Toronto Press.
- Noy-Meir, I. and Austin, M.P. (1970). Principal component ordination and simulated vegetational data. **Ecology**, **51**, 551-552.
- Noy-Meir, I. and Whittaker, R.H. (1978). Recent developments in continuous multivariate techniques. In: R.H. Whittaker (Ed.), **Ordination of plant communities**. The Hague: Dr. Junk Publ., 337-378.
- Poole, K.T. (1984). Least squares metric, unidimensional unfolding. **Psychometrika**, **49**, 311-323.
- Ramsay, J.O. (1977). Maximum likelihood estimation in multidimensional scaling. **Psychometrika**, **42**, 241-266.
- Roberts, F.S. (1976). **Discrete mathematical models**. Englewood Cliffs, NJ: Prentice Hall.
- Schriever, B.F. (1985). **Order Dependence**. Ph.D. Thesis, Amsterdam: Mathematical Centre.

- Shepard, R.N. (1958). Stimulus and response generalization: deduction of the generalization gradient from a trace model. *Psych. Rev.*, **65**, 242-256.
- Shepard, R.N. (1974). Representation of structure in similarity data: problems and prospects. *Psychometrika*, **39**, 373-421.
- Shepard, R.N. and Carroll, J.D. (1966). Parametric representation of nonlinear data structures. In: P.R. Krishnaiah (Ed.), *Multivariate Analysis, Vol. I*. New York: Academic Press, 561-592.
- Stephenson, W. (1953). *The Study of Behavior: Q-technique and its methodology*. Chicago: University of Chicago Press.
- Swan, J.M.A. (1970). An examination of some ordination problem by use of simulated vegetation data. *Ecology*, **51**, 89-102.
- Takane, Y., Young, F.W., and De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika*, **42**, 7-67.
- Ter Braak, C.J.F. (1985). Correspondence Analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics*, **41**, 859-873.
- Ter Braak, C.J.F. (1986a). Canonical Correspondence Analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**, in press.
- Ter Braak, C.J.F. (1986b). The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio*, **65**, in press.
- Ter Braak, C.J.F. and Barendregt, L.G. (1986). Weighted averaging of species indicator values: its efficiency in environmental calibration. *Math. Biosciences*, **78**, 57-72.
- Thurstone, L.L. (1927). A law of comparative judgment. *Psych. Rev.*, **34**, 278-286.
- Van Rijckevorsel, J.L.A. (1986). About horseshoes in multiple correspondence analysis. In: W. Gaul and M. Schader (Eds.), *Classification as a tool of research*. Amsterdam: North-Holland, 377-388.
- Whittaker, R.H. (1948). *A vegetation analysis of the Great Smokey Mountains*. Ph.D. Thesis, University of Illinois, Urbana.
- Whittaker, R.H. (1967). Gradient analysis of vegetation. *Biol. Rev.*, **42**, 207-264.
- Whittaker, R.H. (1978). *Ordination of Plant Communities*. The Hague: Dr. W. Junk Publ.
- Whittaker, R.H. and Gauch, H.G. (1978). Evaluation of ordination techniques. In: R.H. Whittaker (Ed.), *Ordination of Plant Communities*. The Hague: Dr. W. Junk Publ., 277-336.
- Wilkinson, E.M. (1971). Archaeological seriation and the travelling salesman problem. In: F.R. Hodson et al. (Eds.), *Mathematics in the Archaeological and Historical Sciences*. Edinburgh: Edinburgh University Press.