[TITLE]

[authors]

ter Braak, C. J. F. (1988). Partial canonical correspondence analysis. In *Classification and related methods of data analysis,* (ed. H. H. Bock), pp. 551-558. Amsterdam: Elsevier (North-Holland).

Classification and Related Methods of Data Analysis
H.H. Bock (Editor)

PARTIAL CANONICAL CORRESPONDENCE ANALYSIS

Cajo J.F. TER BRAAK

TNO Institute of Applied Computer Science, Box 100, 6700 AC
Wageningen, The Netherlands

Canonical correspondence analysis is (multiple) correspondence analysis in which the ordination axes are constrained to be linear combinations of external, explanatory variables. We consider the case where the set of explanatory variables is subdivided in two sets, a set of covariables and a set of variables-of-interest. This leads to partial canonical correspondence analysis. Its ordination diagram displays the unimodal relationships between a set of response variables and the variables-of-interest after the effects of the covariables have been partialed out. The derivation shows that the response data can be incidence data, count data, compositional data or nominal data.

1. INTRODUCTION

Canonical correspondence analysis is a multivariate analysis technique to display unimodal relationships between a set of response variables and a set of explanatory variables in a low-dimensional space, called an ordination diagram [20,21]. Canonical correspondence analysis has been used in ecology as a simple form of constrained multidimensional unfolding [4,10,12] to relate the occurrences or abundances of a number of species to environmental variables [22]. Applied to nominal variables, canonical correspondence analysis is identical to redundancy analysis of qualitative variables [14] used, for example, to relate nominal welfare variables to social background variables. Here we consider the case where the set of explanatory variables is subdivided in two sets, a set of p covariables and a set of q variables in the effects of which one is particularly interested. Stated informally, we want an ordination diagram of the unimodal relationships between the response variables and the q variables of interest after eliminating the effects of the p covariables. The object is thus to partial out the effects of the covariables, hence the name partial canonical correspondence analysis. Ter Braak [20] derived canonical correspondence analysis as an approximation to canonical Gaussian ordination. Here we define partial canonical Gaussian ordination, derive partial canonical correspondence analysis as an approximation and give an example. Our derivation starts from a constrained generalized linear model and shows that the technique can be applied to nominal data (multi-way contingency data), compositional data, count data and incidence data, with quantitative or qualitative explanatory variables. Related work on partial analysis is given in [3,15,25].

2. THEORY

Let Y and Z be real matrices of order $n \times m$ and $n \times (p+q)$, containing n observations of m nonnegative response variables and p+q explanatory variables, respectively. The p+q explanatory variables are subdivided in p covariables (including the vector $\underline{1}_n$) and q variables of interest and $Z = (Z_1, Z_2)$ is partitioned accordingly. The response variables can be incidences (1/0) or counts of animals or plants in regions, or fractions of constituents in a composition. For nominal variables, Y is a multivariate indicator matrix [7,9] with as many columns as categories. The elements of a matrix B are denoted by b_{ij}, the j-th column of B by $\underline{b}_j$ and the i-th row of

B by $\underset{\sim}{b}_{(i)}$, a column vector, and a generalized inverse of B by B^-. The symbol E denotes expectation.

We now define partial canonical Gaussian ordination as a constrained, generalized linear model.

Definition: For any integer $r<q$, the model of partial canonical Gaussian ordination is

$$\text{link } (Ey_{ik}) = \phi_i + a_k - \frac{1}{2} (\underset{\sim}{z}_{(i)} - \underset{\sim}{u}^*_{(k)})' M (\underset{\sim}{z}_{(i)} - \underset{\sim}{u}^*_{(k)}) \qquad (1)$$

where link is a natural link function (Table 1) [16:p.24] and M is constrained to

$$M = DD' \text{ with } D = \begin{pmatrix} F & G \\ 0 & C \end{pmatrix} \qquad (2)$$

with F, G and C parameter matrices of order $p \times p$, $p \times r$ and $q \times r$, respectively, and 0 is a matrix of order $q \times p$ with zeroes; $\underset{\sim}{u}^*_{(k)}$ is a $(p+q)$-vector representing the optimum of response variable k, a_k is a scalar related to the maximum expected response, and ϕ_i is an incidental parameter for sampling unit i, which takes care of the constant-sum constraint, if present [16:p. 106, p. 142].

Table 1 shows for various types of data the appropriate link function, error distribution and ϕ_i. A statistical interpretation of partial canonical Gaussian ordination is that the m response variables (in Y) are explained by two sets of explanatory variables (in $Z = (Z_1, Z_2)$) by a generalized linear model (GLM) [16] with as predictor a quadratic form in the explanatory variables. It is a unimodal regression model (Fig. 1) with constraints. The difference with standard GLM, which is applied to each response variable separately, is that the parameter matrix M is identical for all response variables and that M is constrained to be positive semi-definite of rank at most p+r, so as to allow an r-dimensional representation of the partial effects of the q variables of interest on the response variables. This becomes clear by writing the model as a constrained ordination model. By setting $\underset{\sim}{x}_{(i)} = D'\underset{\sim}{z}_{(i)}$ and $\underset{\sim}{u}_{(k)} = D'\underset{\sim}{u}^*_{(k)}$, the model is transformed to the canonical form (Fig. 1)

$$\text{link } (Ey_{ik}) = \phi_i + a_k - \frac{1}{2} (\underset{\sim}{x}_{(i)} - \underset{\sim}{u}_{(k)})'(\underset{\sim}{x}_{(i)} - \underset{\sim}{u}_{(k)}) \ . \qquad (3)$$

By this transformation, the $n \times (p+q)$ matrix Z is transformed to a $n \times (p+r)$ matrix X, whose i-th row is $\underset{\sim}{x}_{(i)}$. In terms of variables (the columns of Z and X), the p+q explanatory variables are transformed to p+r axes of a new coordinate system, called ordination axes, by

$$\underset{\sim}{x}_s = Z_1 \underset{\sim}{f}_s \qquad (1 \leq s \leq p) \qquad (4a)$$

$$\underset{\sim}{x}_s = Z_1 \underset{\sim}{g}_s + Z_2 \underset{\sim}{c}_s \qquad (p < s \leq p+r) \qquad (4b)$$

Table 1. Types of response which can be analysed by model (1) which is the basis of partial canonical correspondence analysis (ϕ = incidental parameter, l = index of the L nominal variables with, in total m, categories, ref = references for related models).

type of response	example	link	error*	ϕ	ref
incidence	artifacts in graves pick-any-out-of-m data	logit	Bernoulli	0	13,19 2
abundance	species in regions	log	Poisson	0	13,19
compositions	pollen data electrophoresis data	log	multinomial	ϕ_i	13,16 23
nominal	multiple-choice data	log	multinomial	ϕ_{il}	1,8

*) including extensions to quasi-likelihood models [16].

i.e. the first p ordination axes are a linear combination of the p covariables and the last r axes are a linear combination of all p+q explanatory variables. Model (3) without the constraints in (4) is the Gaussian ordination model [6,11,19] and contains Ihm & Van Groenewoud's [13] generalized logit model. If $a_k = a$ and $\phi_i = 0$, the model shows shifted single-peakedness [11].

In the sequel we focus on the estimation of a basis for the column space of the matrix D and on the estimation of the optima after transformation $\underline{u}_{(k)} = (u_{k1}, u_{k2}, \ldots, u_{k(p+r)})'$. Under the assumption that the $\{y_{ik}\}$ are either independent Bernoulli variables when link(.) = logit(.), or independent Poisson or multinomial variables when link(.) = log(.) (Table 1), with expectations defined by (1), the maximum likelihood equations for u_{ks} and the elements of D become, after some rearrangement [20]

$$u_{ks} = \Sigma_i\, y_{ik}x_{is}/y_{+k} - [\Sigma_i\, (x_{is}-u_{ks})(Ey_{ik})/y_{+k}] \tag{5}$$

$$\Sigma_i\, z_{ij}[\Sigma_k\, y_{ik}(x_{is}-u_{ks})] = \Sigma_i\, z_{ij}[\Sigma_k\, (x_{is}-u_{ks})Ey_{ik}] \tag{6}$$

for k=1, ..., m; j=1, ..., p+q and s=1, ..., p+r, where $y_{+k} = \Sigma_i\, y_{ik}$.

We now derive partial canonical correspondence analysis as an approximation to Eqs (5)-(6) under the following simplifying conditions:

C1. the maxima are equal (a_k=a, k=1, ..., m), or random and independent of the optima $\underline{u}_{(k)}$,

C2. the optima $\underline{u}_{(k)}$ are uniformly distributed over a hypercube A with sides parallel to the ordination axes and of length much larger than 1,

C3. the sampling points $\underline{x}_{(i)}$ are uniformly distributed over a 'large' hypercube B that is contained in A and that has the origin as centroid,

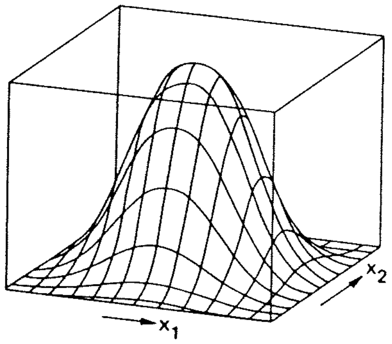
C4. m and n are large so that the optima and sampling points are densely spaced. For nominal variables, the number of classes per variable should be large.

Under these conditions, Ey_{ik} is approximately symmetric about x_{is} and about u_{ks} for each s [19], so that we may use the approximations

$$\Sigma_k\, (x_{is}-u_{ks})Ey_{ik} \approx 0 \tag{7}$$

$$\Sigma_i\, (x_{is}-u_{ks})Ey_{ik} \approx -\lambda_s^* u_{ks}y_{+k}\ . \tag{8}$$

Fig. 1. A unimodal relationship between a response variable y and two regressors (Eq. (3) with p+r = 2 and link = log).



The proportionality constant λ_s^* comes in because the unimodal response surfaces are the more truncated the more their optima lie towards or beyond the edge of the sampling region [19,20]. Using Approximation (8) and the equation $\lambda_s = 1-\lambda_s^*$, we obtain from (5)

$$\lambda_s u_{ks} = \Sigma_{i=1}^{n} y_{ik} x_{is}/y_{+k} \quad . \tag{9}$$

By inserting (4) in (6) and using Approximation (7) we obtain

$$(Z_1'RZ_1)\mathbf{f}_s = Z_1'R\mathbf{x}_s^* \qquad (1 \leq s \leq p) \tag{10}$$

$$(Z'RZ)\mathbf{d}_s = Z'R\mathbf{x}_s^* \qquad (p < s \leq p+r) \tag{11}$$

where $R = \text{diag}\ (y_{i+})$ with $y_{i+} = \sum_i y_{ik}$ and $\mathbf{x}_s^* = (x_{1s}^*, \ldots, x_{is}^*, \ldots, x_{ns}^*)'$ with

$$x_{is}^* = \Sigma_{k=1}^{m} y_{ik} u_{ks}/y_{i+} \tag{12}$$

Equations (4) and (9)-(12) can be solved in a similar way as the transition formulae of canonical correspondence analysis [20].

Because Z_2 contains the variables of interest, it would be convenient to solve for the last r ordination axes without having to extract the first p ordination axes. Fortunately, this can be achieved by making the partitioning of Z in Z_1 and Z_2 explicit. By solving (11) for the component $\mathbf{c}_s$ in $\mathbf{d}_s' = (\mathbf{g}_s', \mathbf{c}_s')$ and using the standard formula for the inverse of a partitioned matrix [18,p.33], we obtain for $s > p$

$$\mathbf{c}_s = (\tilde{Z}_2'R\tilde{Z}_2)^- \ \tilde{Z}_2'R\mathbf{x}_s^*, \text{ where} \tag{13}$$

$$\tilde{Z}_2 = (I-Z_1^\circ)Z_2 \tag{14}$$

and where the notation B° is used to denote $B(B'RB)^-B'R$, the projection operator on V(B), the column space of B, in the metric defined by R. Further, $\mathbf{x}_s$ is the projection of $\mathbf{x}_s^*$ on $Z = (Z_1,Z_2)$ as follows from (4b) and (11), so that

$$\mathbf{x}_s = Z_1^\circ \mathbf{x}_s^* + \tilde{Z}_2^\circ \mathbf{x}_s^* \quad . \tag{15}$$

But, in canonical correspondence analysis the last r ordination axes are required to be orthogonal to the first p ordination axes [20], so that

$$Z_1^\circ \mathbf{x}_s = Z_1^\circ \mathbf{x}_s^* = \mathbf{0} \qquad (p < s \leq p+r) \tag{16}$$

because $V(Z_1) = V(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p)$. Therefore,

$$\mathbf{x}_s = \tilde{Z}_2^\circ \mathbf{x}_s^* = \tilde{Z}_2(\tilde{Z}_2'R\tilde{Z}_2)^-\tilde{Z}_2'R\mathbf{x}_s^* = \tilde{Z}_2\mathbf{c}_s \qquad (p < s \leq p+r) \ . \tag{17}$$

The last r ordination axes can thus be obtained from (9), (12), (13) and (17). These equations form the transition formulae of partial canonical correspondence analysis and define an eigenvalue problem akin to that of canonical correspondence analysis [20]. This can be verified by inserting consecutively in (13) the equations (12), (9) and (17), giving

$$(S_{21}K^{-1}S_{12} - \lambda S_{22})\underline{c}_s = 0 \tag{18}$$

where $S_{21} = \tilde{Z}_2'Y$, $S_{12} = Y'\tilde{Z}_2$, $S_{22} = \tilde{Z}_2'R\tilde{Z}_2$ and $K = \mathrm{diag}(y_{+k})$.

In summary, partial canonical correspondence is a canonical correspondence analysis technique whereby p+r orthogonal axes are constructed. The first p axes are linear combinations of the p covariables only and the subsequent r axes are linear combinations of the p covariables and the q variables of interest. As the covariables are of less interest in the analysis, the first p axes are usually ignored. The subsequent r axes are considered as the first r ordination axes of partial canonical correspondence analysis. They give a low-dimensional representation of the unimodal relationships according to model (1) with constraint (2) between the variables of interest and the response variables after partialing out the effects of the covariables. Technically, the only difference with canonical correspondence analysis is that the matrix of explanatory variables is replaced by the matrix $\tilde{Z}_2$ of residuals of a multivariate multiple regression of Z_2 on Z_1 (14).

Special cases of partial canonical correspondence analysis are:

1. Canonical correspondence analysis [14,20] if Z_1 is a n×1 matrix of 1's (a single trivial covariable only).
2. Partial correspondence analysis if Z_2 is a n×n identity matrix (no variables of interest) or any arbitrary n×(n-1) matrix of rank n-1 (too many variables of interest [21]).
3. Multiple correspondence analysis [7,9] if Z_1 and Z_2 are as specified in 1 and 2 above (no explanatory variables or too many of them).
4. Weighted averaging ordination [6,21] if $p = 0$ and $q = 1$ (a single variable of interest).

Our definition of partial correspondence analysis differs from that by Yanai [25].

3. ORDINATION DIAGRAM

As in correspondence analysis, the results can be presented in an ordination diagram in which the rows and columns of Y are represented by points at locations $\underline{x}_{(i)}$ and $\underline{u}_{(k)}$. To the extent that the analysis approximates the fitting of Gaussian surfaces (1), the points for response variables are approximately the optima of these surfaces; hence, the estimated value of Ey_{ik} decreases with the distance between the points of sampling unit i and response variable k (Fig. 1). The estimated values are, of course, conditional on the values of the covariables.

In partial canonical correspondence analysis the ordination diagram can be supplemented with arrows for the variables of interest (Fig. 2). This is done in such a way that, in conjunction with the points for response variables, the arrows give a weighted least squares approximation of the elements of the m×q matrix $W = K^{-1}Y'\tilde{Z}_2$. The (k, j)-th element of W is the weighted average of response variable k with respect to variable of interest j, after this variable is adjusted for the covariables. In a unimodal model, the weighted average indicates the centre of a response curve. So the matrix W summarizes unimodal relationships, like a matrix of partial correlation coefficients summarizes linear relationships. In the approximation of W, response variables are given weights proportional to their total y_{+k}. The coordinates of the supplementary arrows can be obtained by a multivariate regession of W on $U = \{u_{ks}\}$, i.e. by

$$C_r = W'KU(U'KU)^{-1}. \tag{19}$$

The approximation to W is then given by the bilinear model UC_r'. The plot of points for response variables and arrows for variables of interest is thus a biplot [5], termed the species-environment biplot in [20]. This plot is not just supplementary, as it can be made central to (partial) canonical correspondence analysis [22].

4. EXAMPLE

The example is taken from H. Smit (in prep.). Smit studied the abundances of diatom species in dykes in the province of Zuid Holland (The Netherlands), with special reference to the effects of water pollution. A sample of 402 dykes was taken, which contained in total 330 species. Variables that indicate pollution were compounds with phosphorus (P) and nitrogen (N), and biological oxygen demand (BOD). Apart from variation in pollution, the sample showed strong natural variation due to the season of sampling and due to a gradient from fresh to brackish water. This natural variation was partialed out by specifying a season indicator variable and the chloride concentration (Cl) as covariables. Partial canonical correspondence analysis on diatom species with 24 variables-of-interest showed a first axis ($\lambda_1 = 0.10$) that was a clear pollution gradient as indicated by the arrows for P, BOD and N in the ordination diagram (Fig. 2). The second axis ($\lambda_2 = 0.05$) revealed the importance of other natural variation, notably soil type and dyke width. Species of polluted waters are represented on the right hand side of the diagram (Fig. 2), e.g. Navicula accomoda and N. subminuscula, whereas species of unpolluted waters lie on the left hand side, e.g. Eunotia pectinalis. Species in the middle have their optimum at intermediate pollution levels or are indifferent [20]. Which possibility is most likely can be decided upon by plotting the abundance values on the ordination diagram. Despite their occurrence at high values of P and BOD, two species of brackish waters, Melosira jürgensii and Navicula diserta, are displayed on the left hand side of the diagram, because brackish waters naturally have high P- and BOD-values. This illustrates that Fig. 2 displays partial effects.

5. DISCUSSION

In this paper partial canonical correspondence analysis is derived as an approximation to maximum likelihood estimation of a particular unimodal model. But it does not maximize a likelihood. What is being maximized is the least-squares criterion of multiple correspondence analysis [7,11,12,21,24] with the additional constraints (a) that the axes are linear combinations of all explanatory variables and (b) that the axes are orthogonal to the covariables. We note that the orthogonality constraints do not follow necessarily from the maximum likelihood approach (see below Eq. (15)). They are not sufficient either; we conjecture that when the Guttman effect [9,19] crops up, the transition formulae have solutions close to solutions of the maximum likelihood equations that correspond to local maxima. Such solutions can be excluded by "detrending" [6,19,23] or by deleting explanatory variables [21]. Other loss-functions are considered in [10,12,17].

In the dual scaling approach to correspondence analysis [9], category scores form the optimal quantification [7] of the corresponding nominal variables. This paper gives reason to interpret category scores as optima of underlying response curves (termed trace lines in [24]). The properties of correspondence analysis in terms of a unimodal model were explored earlier by Torgerson [24: point items], Heiser [10,11] and Ihm & Van Groenewoud [13].

For data with a constant-sum constraint ($\phi_i \neq 0$ in Table 1), model (1) can be rewritten as

$$\log\,(Ey_{ik}) = \phi_i^* + a_k^* + \underline{z}'_{(i)}\, M\, \underline{u}^*_{(k)} \qquad (20)$$

where ϕ_i^* and a_k^* have absorbed the quadratic forms in $\underline{z}_{(i)}$ and $\underline{u}^*_{(k)}$ in Eq. (1), respectively. Model (20) with p=1 and r=1 is the qualitative logistic regression model, from which Anderson [1] developed his regression method for ordinal response variables (cf. [8]). The results of this paper can be used to show that his method can be approximated by canonical correspondence analysis with ordinal constraints, as in Gifi [7], on the category scores. It is surprising that for nominal and compositional data the unimodal unfolding model (1) can be reexpressed as a generalized bilinear model (20)!

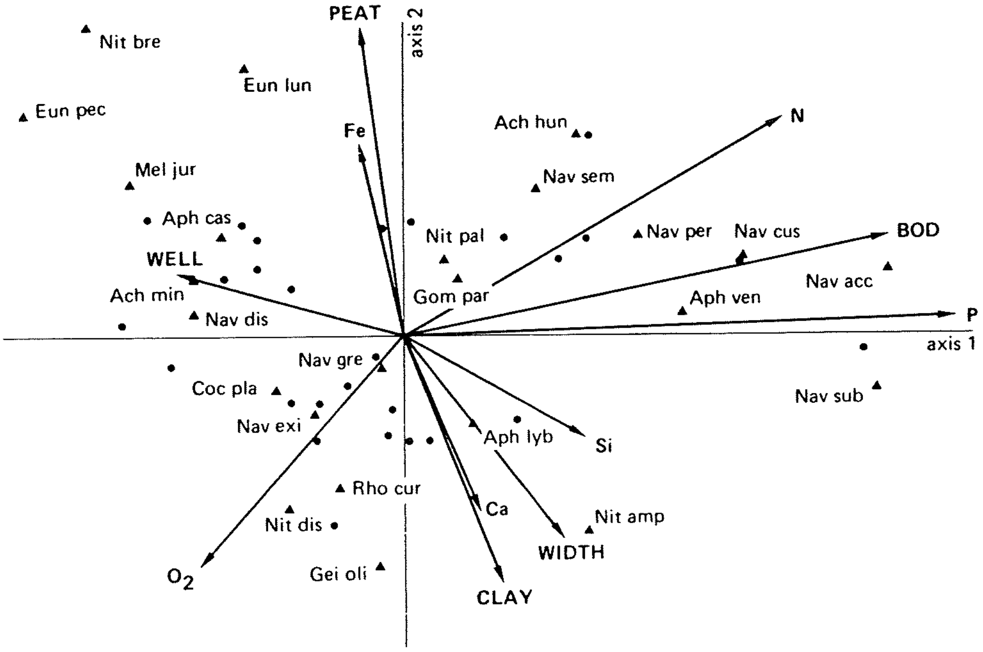


Fig. 2. Ordination diagram of a partial canonical correpondence analysis of diatom species (Δ) in dykes with as explanatory variables 24 variables-of-interest (arrows) and 2 covariables (chloride concentration and season). The diagram is symmetrically scaled [23] and shows selected species and standardized variables and, instead of individual dykes, centroids (o) of dyke-clusters. The variables-of-interest shown are: BOD = biological oxygen demand, Ca = calcium, Fe = ferrous compounds, N = Kjeldahl-nitrogen, O_2 = oxygen, P = ortho-phosphate, Si = silicium-compounds, WIDTH = dyke width, and soil types (CLAY, PEAT). All variables except BOD, WIDTH, CLAY and PEAT were transformed to logarithms because of their skew distribution. The diatoms shown are: Ach hun = Achnanthes hungarica, Ach min = A. minutissima, Aph cas = Amphora castellata Giffen, Aph lyb = A. lybica, Aph ven = A. veneta, Coc pla = Cocconeis placentula, Eun lun = Eunotia lunaris, Eun pec = E. pectinalis, Gei oli = Gomphoneis olivaceum, Gom par = Gomphonema parvulum, Mel jur = Melosira jürgensii, Nav acc = Navicula accomoda, Nav cus = N. cuspidata, Nav dis = N. diserta, Nav exi = N. exilis, Nav gre = N. gregaria, Nav per = N. permitis, Nav sem = N. seminulum, Nav sub = N. subminuscula, Nit amp = Nitzschia amphibia, Nit bre = N. bremensis v. (brunsvigensis, Nit dis = N. dissipata, Nit pal = N. palea, Rho cur = Rhoicosphenia curvata.
(Adapted from H. Smit, province of Zuid Holland, in prep.)

[1] Anderson, J.A., Regression and ordered categorical variables. J. R. Statist. Soc. B 46 (1984) 1-30.

[2] Coombs, C.H., A theory of data. Wiley, New York, 1964.

[3] Daudin, J.J., Partial association measures and an application to qualitative regression. Biometrika 67 (1980) 581-590.

[4] DeSarbo, W.S., and Rao, V.R., GENFOLD2: a set of models and algorithms for the general unfolding analysis of preference/dominance data. J. of Classification 1 (1984) 147-186.

[5] Gabriel, K.R., The biplot graphic display of matrices with application to principal component analysis. Biometrika 58 (1971) 453-467.

[6] Gauch, H.G., Multivariate analysis in community ecology. Cambridge Univ. Press, Cambridge, 1982.

[7] Gifi, A. Nonlinear multivariate analysis. Dept. of Data Theory, Univ. of Leiden, Leiden, 1981.

[8] Goodman, L.A., Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. Int. Statist. Rev. 54 (1986) 243-309.

[9] Greenacre, M.J., Theory and applications of correspondence analysis. Academic Press, London, 1984.

[10] Heiser, W.J., Unfolding analysis of proximity data. Thesis. University of Leiden, Leiden, 1981.

[11] Heiser, W.J., Undesired nonlinearities in nonlinear multivariate analysis. In: Data Analysis and Informatics 4 (E. Diday et al, eds.), North-Holland, Amsterdam, 1986, pp. 455-469.

[12] Heiser, W.J., Joint ordination of species and sites: the unfolding technique. In: New developments in numerical ecology, (P. Legendre and L. Legendre, eds.), Springer-Verlag, Berlin, 1987, in press.

[13] Ihm, P. and Van Groenewoud, H., Correspondence analysis and Gaussian ordination, COMPSTAT lectures 3 (1984) 5-60.

[14] Israëls, A.Z., Redundancy analysis for qualitative variables. Psychometrika 49 (1984) 331-346.

[15] Lauro, N. and D'Ambra, L., Multiple and partial non symmetrical correspondence analysis, this volume.

[16] McCullagh, P. and Nelder, J.A., Generalized linear models. Chapman and Hall, London, 1983.

[17] Meulman, J., and Heiser, W.J., Constrained multidimensional scaling: more directions than dimensions. COMPSTAT 1984, Physica-Verlag, Vienna, 1984, pp. 137-142.

[18] Rao, C.R., Linear statistical inference and its applications. 2nd ed. Wiley, New York, 1973.

[19] Ter Braak, C.J.F., Correspondence analysis of incidence and abundance data: properties in terms of a unimal response model. Biometrics 41 (1985) 859-873.

[20] Ter Braak, C.J.F., Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67 (1986) 1167-1179.

[21] Ter Braak, C.J.F., The analysis of vegetation-environment relationships by canonical correspondence analysis. Vegetatio 69 (1987), 69-77.

[22] Ter Braak, C.J.F., Ordination. In: Data analysis in community and landscape ecology (R.H.G. Jongman, C.J.F. Ter Braak and O.F.R. Van Tongeren, eds.). Pudoc, Wageningen, 1987, p.p. 91-173.

[23] Ter Braak, C.J.F., CANOCO - a FORTRAN program for canonical community ordination by [partial][detrended][canonical] correspondence analysis, principal components analysis and redundancy analysis (version 2.1). TNO Institute of Applied Computer Science, Wageningen, 1987.

[24] Torgerson, W.S., Theory and methods of scaling. Wiley, New York, 1958.

[25] Yanai, H., Some generalizations of correspondence analysis in terms of projection operators. In: Data Analysis and Informatics 4. (E. Diday et al, eds.), North-Holland, Amsterdam, 1986, pp. 193-207.