

**Nonlinear canonical analysis of
grassland community data**

Peter Verboon

Jan de Leeuw

**Department of Data Theory
University of Leiden**

Introduction

In this paper we will compare a linear canonical analysis with a non-linear canonical analysis of an ecological data set. Linear canonical analysis is a relatively well known technique, also in ecology (Gittins, 1985). Nonlinear canonical analysis is far less well known, and we give a brief introduction to the particular version we have in mind. For further information we refer to Van der Burg and De Leeuw (1983) or De Leeuw (1986).

Canonical analysis

Canonical analysis can be described briefly as follows. We have two sets of vectors in \mathbb{R}^n , defined by two sets of variables measured on n individuals. These two sets of vectors define two linear subspaces, the linear combinations of the variables. The idea of canonical analysis is to compute the invariants of these two subspaces, which are the angles between certain canonical vectors. The first pair of canonical vectors are those linear combinations which make the smallest possible angle, the second pair makes the smallest possible angle with the restriction that it is orthogonal to the first pair, and so on. If X is the $n \times p$ matrix of observations on p variables in the first set, and Y is the $n \times q$ matrix of observations on q variables in the second set, then the canonical variables XA and YB satisfy $A'X'XA = B'Y'YB = I$ and $A'X'YB = \Lambda$, the diagonal matrix of canonical correlations. In order to study the relationship between canonical variables and the original variables we can compute the correlations between X and Y and the average canonical variable $(XA + YB)/2$. These correlations are the canonical loadings.

In nonlinear canonical correlation analysis the columns of X and Y are not considered as fixed, but they can be transformed. Optimal transformations are defined in terms of the canonical correlations. Thus we can find transformations of the variables which maximize the sum of the largest canonical correlation, or of the sum of the two largest canonical correlations.

Thus nonlinear canonical analysis gives us loadings, weights, and canonical correlations as in the linear technique, but in addition we compute the transformations of the variables. It is clear that this will generally lead to higher canonical correlations (smaller angles). In the application in this paper the transformation of the variables is restricted to be a monotonis step function, i.e. a monotonic function of a discretized variable. This is the simplest, but not necessarily the best, choice of the space of transformations. De Leeuw (1986) applies spline transformations in a nonlinear canonical analysis of an ecological example.

The data

The data used are described by Gittins (1985). From the four experiments he describes we will select the data from table A-2, of which the results are discussed in his chapter 7. The experiment concerns relationships between the abundance of several plant species and associated soil characteristics in a limestone grassland community in Anglesey, North Wales. The first set contains eight species: *H. pubescens*, *P. bertolonii*, *T. pratense*, *P. sanguisorba*, *R. squarrosus*, *H. pilosella*, *B. media*, *T. drucei*. The species were not randomly chosen, but in order of their expected response to variation in soil depth. In the second set we have three soil properties: depth (d), extractable phosphate (P) and exchangeable potassium (K), and furthermore the interactions between these soil properties, resulting in six variables for the second set. The sample size $N = 45$.

Linear analysis

The results of Gittins linear canonical analysis are summarized in table 1. The first two canonical correlations are .923 and .696.

Table 1a. Correlations between the species variables and the average canonical variates. (Gittins, 1985).

<i>H. pubescens</i>	.949	-.017
<i>P. bertolonii</i>	.489	-.613

T. pratense	.482	-.046
P. sanguisorba	.374	.413
R. squarrosus	.015	-.291
H. pilosella	-.290	.399
B. media	-.744	.184
T. drucei	-.878	-.012

Table 1b. Correlations between soil property variables and the average canonical variates.

depth(d)	.818	.268
phosphate(P)	.627	-.429
potassium(K)	.157	.029
d x P	.892	-.253
d x K	.624	.171
P x K	.406	-.319

We see that the first canonical variate is characterized in particular by depth and the interaction between depth and phosphate. According to Gittins the second variate might be an expression of phosphate. Considering the species variables we see a strong positive correlation with the first variate for *H. pubescens* and a strong negative correlation for *T. drucei*. Except for *R. squarrosus* all species have a sizeable correlation with the first variate. The second variate shows a contrast between *P. bertolonii* and *P. sanguisorba*. Gittins concludes that there is indeed a connection between soil depth in combination with the interaction between soil depth and phosphate and the abundance of the species, specifically in the order mentioned above. The second conclusion is that *P. bertolonii* and *P. sanguisorba* respond to variation in phosphate in opposite sense. Finally Gittins concludes that the explanatory power of the first and second variate in the species' domain is consistent with prior knowledge of the relative importance of depth, phosphate and the interaction between them on the behavior of the species.

Non-linear Analysis

By means of the CANALS program (Van der Burg, 1983) we did a non-linear canonical analysis on the data of table A-2. To do so the original

variables were recoded in such a way to obtain at most five categories with approximately equal frequencies. See table 2. The distribution of some variables was not appropriate to recode into five categories, for example the second and third variable, which have been recoded into two and three categories respectively. The measurement level of all variables was considered to be ordinal.

Table 2. Recoding of the variables

	1	2	3	4	5
H. pubescens	0-12	15-40	44-76	80-88	92-99
P. bertolonii	0	4-84			
T. pratense	0	4-12	20-60		
R. sanguisorba	8-24	28-36	40-48	52-60	68-88
R. squarrosus	0-8	12-16	20-24	28-40	44-68
H. pilosella	0-4	8	12	16-24	28-40
B. media	24-92	96-99			
T. drucei	0-4	8-32	36-76		
depth	32-39	40-46	49-62	64-83	
phosphate	3-6	7	8-9	10-16	
potassium	6-12	13	14-30		
d x P	147-238	252-360	364-450	462-576	636-1024
d x K	306-459	468-555	559-650	660-828	832-1550
P x K	27-63	65-84	89-108	117-144	150-405

The results in two dimensions, summarized in table 3 are fairly similar to the ones described by Gittins. The canonical correlations are .963 and .943. The first canonical variate correlates mainly with the interaction between depth and phosphate, and the second variate correlates with depth and both interactions of depth. In the first set we see H. pubescens have a strong positive correlation with the (dxP) interaction, while T. drucei and H. pilosella have rather strong negative correlations with (dxP). It is quite remarkable that B. media hardly correlates with the first variate. The second variate shows similar patterns: a strong positive correlation for H. pubescens and negative correlations for T. drucei and B. media, while now H. pilosella hardly correlates with the depth variables. Finally we mention that as in the linear analysis, R. squarrosus neither correlates with the first variate nor with the second.

Table 3a. Correlations between the species variables and the average canonical variates.

H. pubescens	-.580	-.677
P. bertolonii	-.501	-.212
T. pratense	-.529	-.450
P. sanguisorba	-.502	-.198
R. squarrosus	-.106	-.230
H. pilosella	.670	-.260
B. media	.179	.580
T. drucei	.672	.576

Table 3b. Correlations between the soil properties variables and the average canonical variates.

depth (d)	-.411	-.684
phosphate (P)	-.461	-.426
potassium (K)	-.112	-.092
d x P	-.707	-.665
d x K	-.225	-.769
P x K	-.326	-.129

So we have seen that the results of both technics are quite comparable, but that the two dimensions of CANALS split up the first dimension of Gittins' analysis. It is therefore that we discovered the difference between B. media and H. pilosella, the first one specifically reacting to depth and the second only to the interaction between depth and phosphate. The two-dimensional solution doesn't reveal the influence of potassium. The canonical correlations of the three dimensional CANALS solution are .948, .905 and .715. The first and second dimension are the same as before and the third dimension shows the small negative effect of phosphate and the (P x K) interaction on T. pratense. The one dimensional CANALS solution is identical to the first dimension of Gittins.

Figure 1 shows a plot of the correlations between the species variables and the average of the first canonical variate for the linear analysis (the underlined numbers) and the non-linear analysis. Figure 2 shows the same correlations for the second set. In both plots the signs of the first CANALS dimension have been changed for a better comparison.

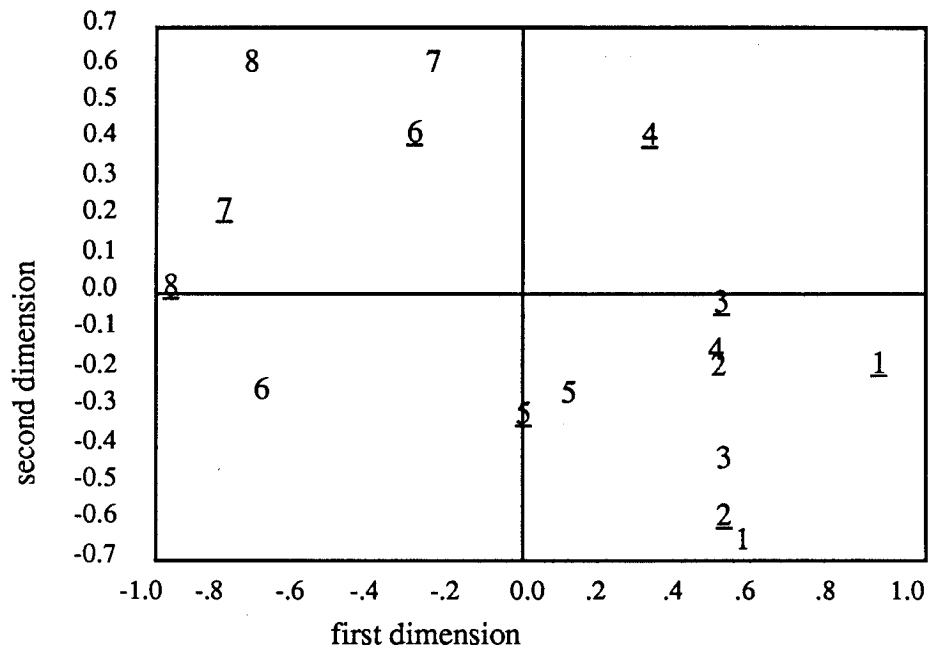


Figure 1. Correlations between the variables of the first set and the corresponding canonical variate in two dimensions. (underlined: linear analysis)

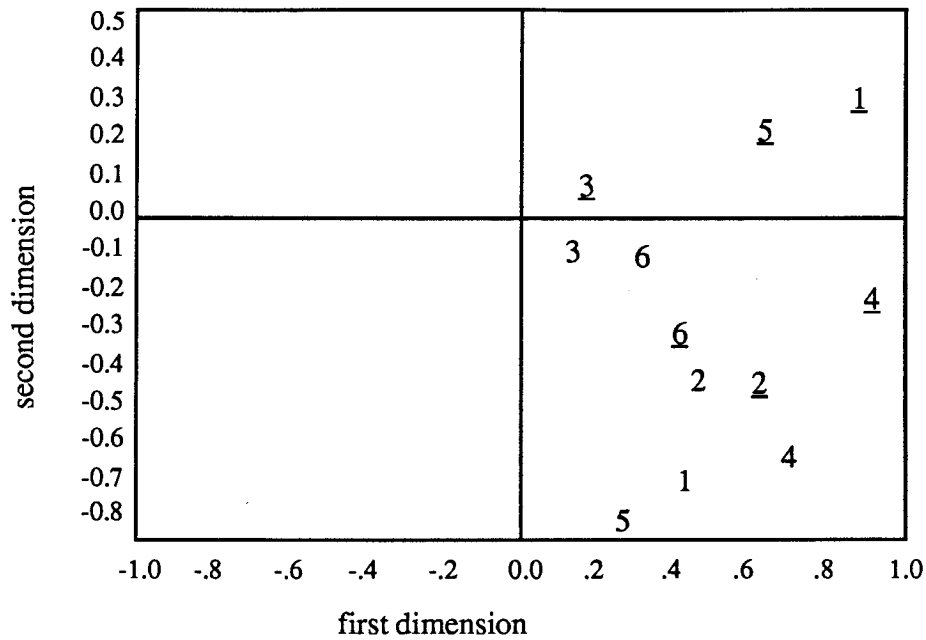
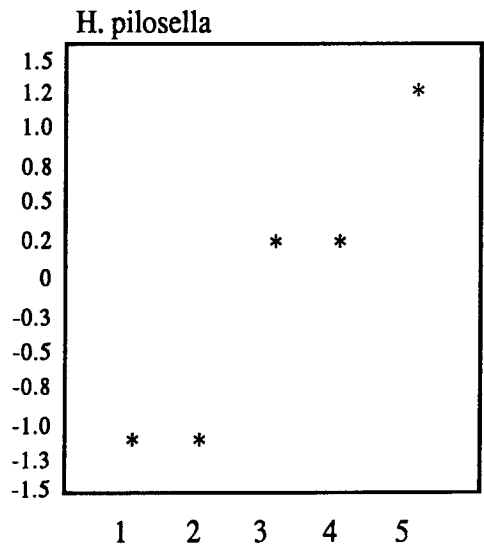
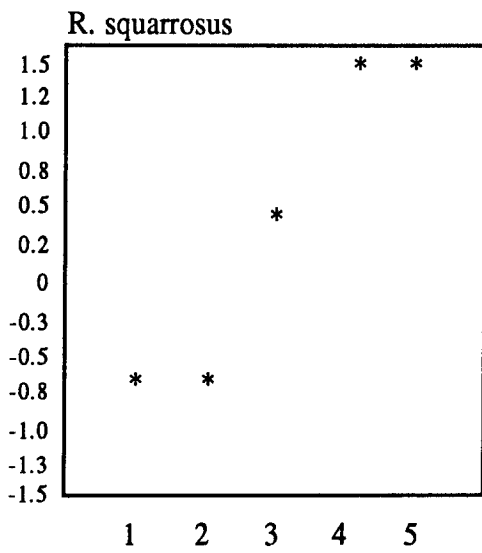
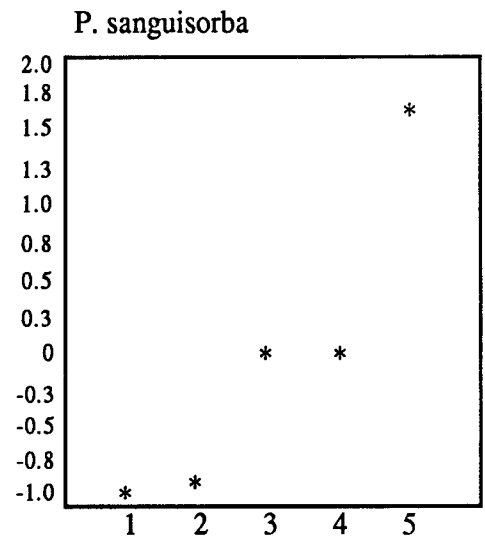
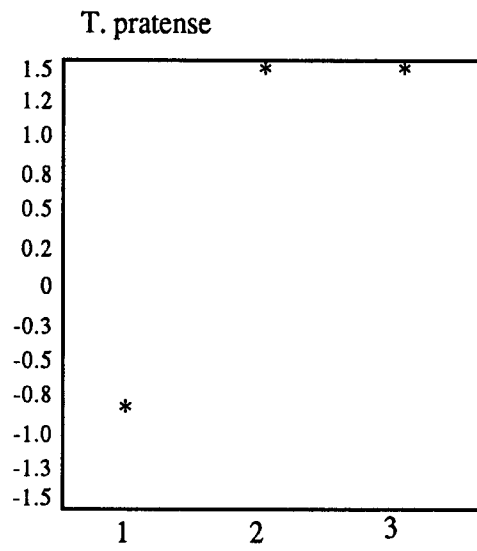
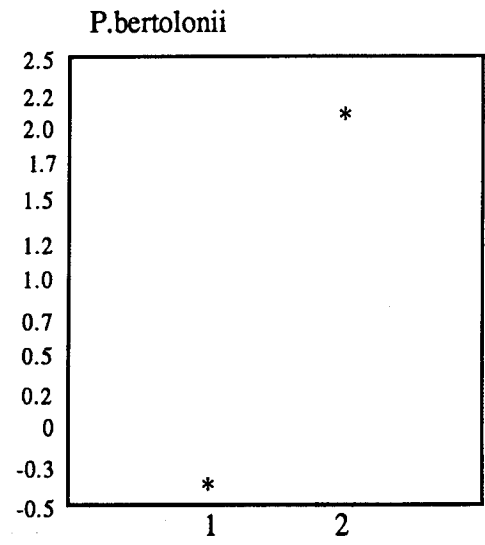
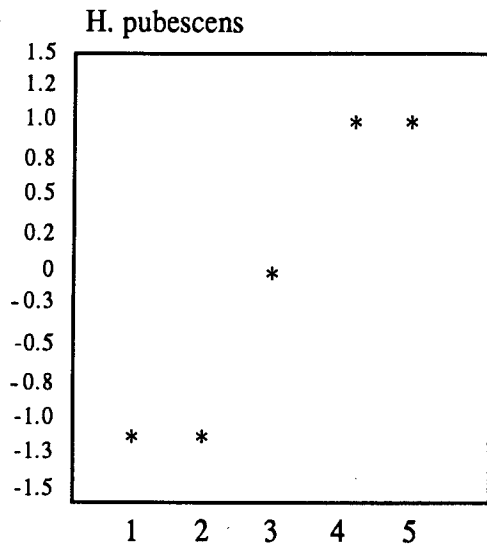
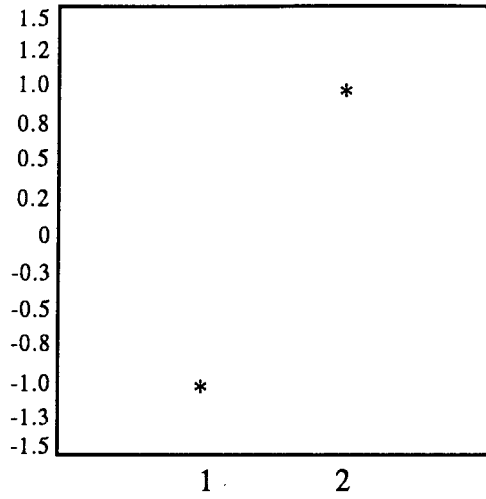


Figure 2. Correlations between the variables of the second set and the corresponding canonical variate in two dimensions. (underlined: linear analysis)

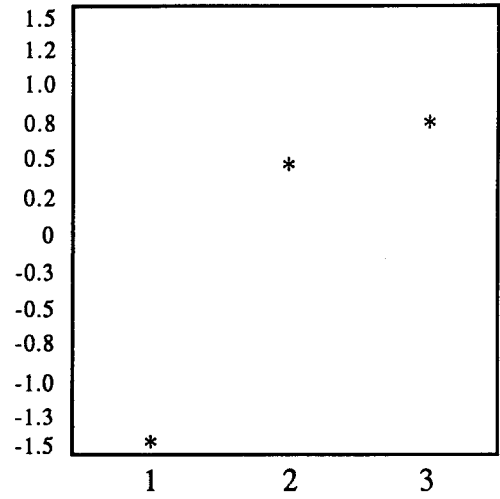
In the following figures the category quantifications of all variables are plotted. The horizontal axes represent the recoded categories and the vertical axes the quantifications, which are the non-linear transformations. We see that some categories of the variables obtain the same quantification, indicating a small difference between them. Considering for example the depth variable which has the same quantification for the first and second category, it might be concluded that there isn't must differentiation near the surface. On the other hand there are variables with a rather smooth transformation, for example the interaction between depth and phosphate.



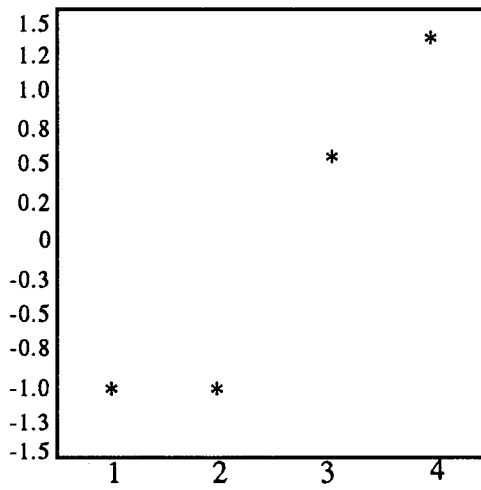
B. media



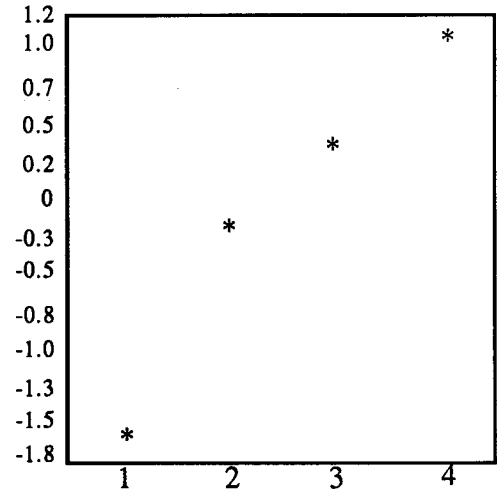
T. drucei



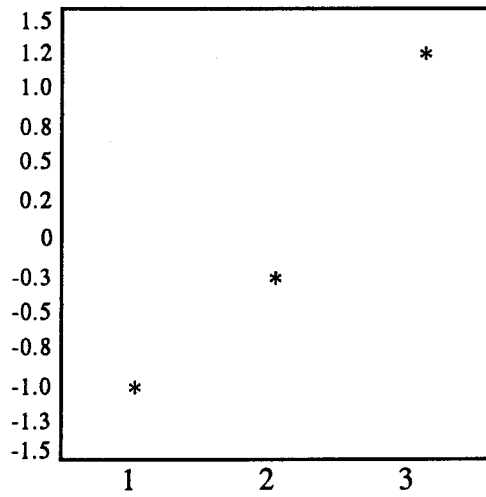
depth



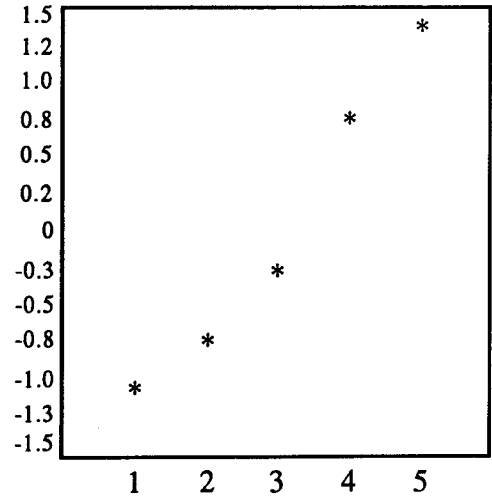
phosphate

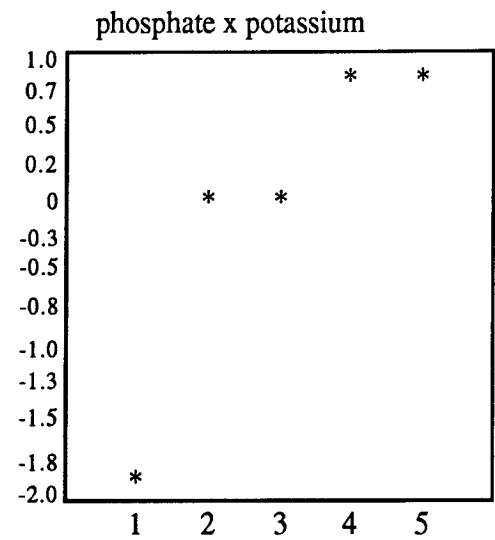
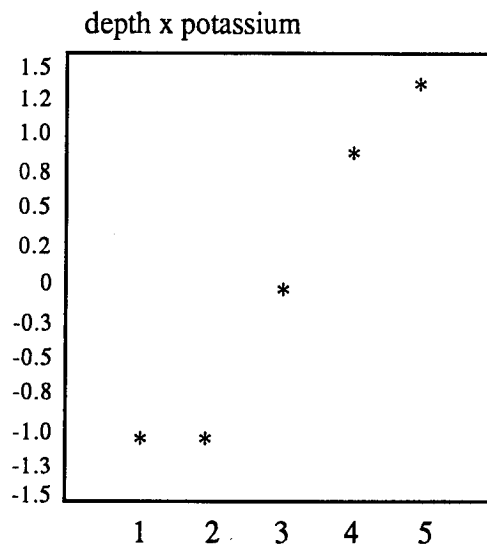


potassium



depth x phosphate





References

- De Leeuw, J. (1986). Nonlinear multivariate analysis with optimal scaling. In L. Legendre & P. Legendre, **Numerical Ecology** (in preparation).
- Van der Burg, E. (1985). **CANALS**. Report UG-85-05. Department of Data Theory, University of Leiden.
- Van der Burg, E., & De Leeuw, J. (1983). Nonlinear canonical analysis. **British J. Math. Statist. Psychol.**, 36, 54-80.
- Gittins, R. (1985). **Canonical analysis. A review with applications in ecology**. Berlin:Springer.