SPECTRAL IDENTIFICATION OF DANISH GRASSLAND CLASSES RELATED TO MANAGEMENT AND PLANT SPECIES COMPOSITION *

A. Jacobsen [1,2], A. A. Nielsen [3], R. Ejrnæs[1], G. B. Groom[1]

[1]National Environmental Research Institute, Department of Landscape Ecology,
Grenåvej 12-14, DK-8410 Rønde, Denmark

[2]Institute of Geography, University of Copenhagen,
Øster Voldgade 10, DK-1350 Copenhagen K, Denmark

[3]Department of Mathematical Modelling, Technical University of Denmark,
Building 321, DK-2800 Lyngby, Denmark

ABSTRACT

The study presents a hierarchical means of obtaining well separable training sites that describe natural vegetation of different management and species composition. The study area is Mols Bjerge, Denmark, with dry grasslands assigned to one of seven management classes. A June 1997 Compact Airborne Spectrographic Imager image with 11 bands from 400-900 nm was used for the image analysis. Spectral clustering and canonical discriminant functions (CDFs) showed that the management classes were not spectrally unique and that spectral subclasses existed. The recognition of a spectral differentiation within management classes led to a floristical clustering based on plot scores in ordination space. The result implied that 29% of the test sites were predicted (by means of their vegetation) to belong to a cluster defined by the floristic composition of a different management class. The spectral clustering and analysis of CDFs within management classes and the floristic modelling showed that management classes could be identified spectrally with respect to their plant species composition. Separability between management classes and subclasses respectively using the Jeffries-Matusita distance measure showed that separability was improved in both cases after seed growing using the Mahalanobis distance and that separability was related to plot scores in floristic ordination space.

1.0 INTRODUCTION

The mapping of major land cover classes by automated classification of remote sensing data is a well-established technique (Fuller et al., 1994). However, separation within broad land cover classes in terms of management and floristics is becoming increasingly important, associated with the opportunities offered from airborne scanners of high spatial and spectral resolution. Classification at such a detailed level involves challenging discussions of identification of and separation between classes since discrimination of classes along more or less continuous gradients is a well-known problem not only in the spectral domain but also with respect to the vegetation communities.

At the same time, vegetation ecologists are increasingly using ordination and multivariate discriminant analysis in stead of discrete classes when describing floristic variation. Attempts have been made to combine such methods with spectral analysis (Lewis, 1997; Wessman et al., 1993). These studies, and others (Jacobsen et al., 1995, Lauver and Whistler, 1993) indicate that both management and biodiversity affect the spectral signature. The aim of the study is to examine a hierarchical way of identifying grassland classes from their spectral signatures with consideration to the mutual inter-dependency between management and biodiversity.

---

* Presented at the Fourth International Airborne Remote Sensing Conference and Exhibition/21st Canadian Symposium on Remote Sensing, Ottawa, Canada, 21-24 June 1999.

In: *Proceedings of 4th International Airborne Remote Sensing Conference and Exhibition/21st Canadian Symposium on Remote Sensing*, Vol. 1, 21-24 June 1999, Ottawa, Ontario, Canada, pp. 74-81.

The study was performed within the DANish Multisensor Airborne Campaign (DANMAC) project, which includes interdisciplinary studies on various physical and biological conditions and processes using ground, airborne and spaceborne optical and radar sensors (Groom et al., 1997).

2.0 STUDY AREA

Mols Bjerge (Figure 1) (approx. 15 sq. km) is dominated by open grasslands with occasional scrubs, thickets and otherwise deciduous forests and coniferous plantations. The dry, uncultivated areas can be divided into seven management classes: old unimproved grasslands with continuous grazing (class 1), old, unimproved, but derelict grasslands (class 2), medium aged grassland, previously cultivated, but now with spontaneous dry grassland vegetation (class 3), young, ex-arable areas with spontaneous grazed vegetation (class 4), young, ex-arable areas with spontaneous ungrazed vegetation (class 5) 1-5 years old 'set-a-side' vegetation dominated by weed species (class 6), and improved, sown grass leys (class 7).

Seen in a landscape planning perspective, areas of management classes 1 or 2 can be considered as important for the conservation of biodiversity, class 3 can be considered as potentially important, whereas classes 4 to 7 are less important although potentially useful for wildlife.

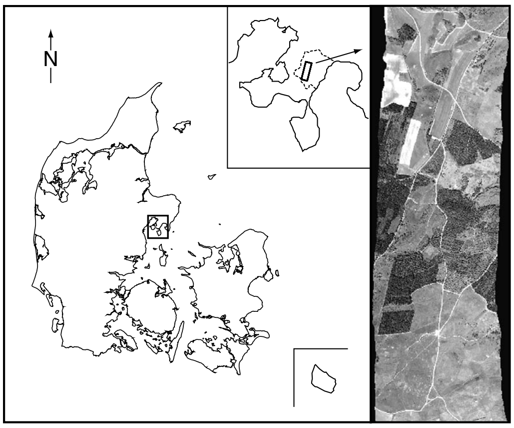

Figure 1: Study area Mols Bjerge (inserted top right) and the location of the *casi* flight line used in the analysis. Light grey areas are grasslands and agricultural fields, dark areas are forests. Gravel roads transect the flight line.

3.0 DATA

3.1 IMAGE DATA

Compact Airborne Spectrographic Imager (*casi*) data were acquired over Mols Bjerge, 10th June 1997, during a DANish Multisensor Airborne Campaign. The study is performed on scan line K3, which is a spatial mode (2m spatial resolution) 11-band image in the spectral range from 400 to 900 nm that was calibrated to surface reflectance.

The quality of the spectral and radiometric calibration of the *casi* data were assessed and it was found that the radiometric calibration was poor in bands 1 and 2 and that spectral calibration exceeded ± 0.25 nm in the left-most 135 columns across track due to spectral alignment problems. Bands 1 and 2 were excluded from the analysis since a large part of the image had negative reflectance in this spectral region. Since it was not known to what extent the spectral calibration accuracy would affect the spectral classes, initially the entire image was included in the analysis; subsequent analysis excluded the 135 columns with poor spectral calibration.

A digital elevation model (DEM) and triangular irregular network (TIN) resampling using Delaunay triangles was applied for georeferencing the scanner data (Jacobsen et al., 1999).

3.2 VEGETATION DATA

Extensive fieldwork was carried out in 1996. All areas, excluding forest clearings, with more than 25% cover of grasses, forbs and dwarf shrubs were identified and surveyed. 290 areas fell within one of the seven management classes. Assignment to management class was assessed in the field on the basis of topography and vegetation. Classes 1, 2 and 3 were distinguished with the help of indicator species for old grasslands of conservation interest (Ejrnæs and Bruun, 1995). Vascular plant species were recorded using a robust abundance scale adapted to an inventory of undulating, mainly extensive areas. Simple abundance scoring was made, in terms of: 'present with low cover', 'frequent with moderate cover', 'frequent with high cover over at least part of the area and 'high cover over the majority of the area'.

In 1997 thirty 30m by 30m test sites, representative of the seven different management classes, were placed among the previously surveyed areas, mainly in areas covered by the *casi* scan-line used in the study. Test sites were placed in areas with homogeneous vegetation and constant slope and aspect, and was geo-positioned using differential global positioning system (DGPS). The positions of these test sites were identified in the georeferenced image from their DGPS coordinates but the image analysis was performed on the non-georeferenced, roll-corrected image to avoid resampling distortion.

4.0 THEORY AND METHODS

4.1 TRAINING SET GENERATION

Successful supervised classification relies heavily on good training data. Statistically sound training data are not necessarily obtained when hand-drawn by a human operator. One reason for this is the human inability to overview multidimensional space. Also, training sets need to be extracted in a consistent way and irrespective of the operator and the image structure. In this study a semi-automatic algorithm was applied for generation of training sets from a series of seed points (Flesche et al., 1999). From these points, training classes were grown in a fashion that ensured spatial and spectral closeness. Spatial closeness was obtained by requesting connectivity. Spectral closeness was obtained by restricting the spectral distance to the current mean value of the class while growing the training set. There were enough seed points per class to define an initial training set and to estimate the dispersion matrix. The dispersion matrix was first used to exclude any outliers in the current training and second, to grow that training set further using the Mahalanobis distance method.

4.2 TRAINING CLASS CONSISTENCY AND SEPARABILITY

Training data were checked for consistency to make sure that the multivariate data in each assumed class comprised, in a statistical sense, just one class. For this, a method based on an unsupervised clustering algorithm was applied: (1) observations within each management class called cluster seeds were selected as a first guess of the sub-class means; (2) clusters were formed by assigning observations to the nearest seed as measured by Euclidean distance; (3) after all observations were assigned, new cluster means were calculated. This last step was repeated until changes in cluster means became zero (or small). The clustering was based on 9 cluster seeds and followed by a canonical discriminant analysis (Fisher, 1936), which combined the original variables into new orthogonal variables or 'canonical discriminant functions' (CDFs). The CDFs are the best possible linear discriminators between the sub-classes into which the training data have been clustered. If a scatter plot of the first two CDFs showed no outliers and no sign of grouping, the training data were considered consistent. Otherwise a number of spectral subclasses were identified from the CDFs scatterplots linked to an image view of the test sites.

Training data were checked for class separability by applying the Jeffries-Matusita (J-M) (pairwise) distance between classes i and j , J_{ij}, (or equivalently the Bhattacharyya distance, a_{ij}) and the average J-M distance between all classes, (Matusita, 1966); Ersbøll, 1989). The J-M distance between perfectly separable classes is $\sqrt{2}$ (1.41).

4.3 MULTIVARIATE TREATMENT OF FLORISTIC DATA

The untransformed floristic data combining the 1996 and 1997 inventories were subjected to ordination in order to reduce the dimensionality of the species*plot matrix. Detrended correspondence analysis (DCA) (Hill, 1979) was used with downweighting of rare species and otherwise default options (McCune and Mefford, 1997). Three ordination axes were extracted and chosen to represent the major floristical gradients present in the data (Økland, 1990). In order to produce a floristically based clustering of the plots, optimized to reflect the management classes, a multinomial log-linear model fit was applied (Venables and Ripley, 1997) predicting class membership as a function of plot scores on the three ordination axes. The modelling was performed in S-Plus 4.5 (S-Plus 4.5) using multinom and predict.multinom included in the NNET library add-in to S-Plus (Ripley, Unpubl.).

5.0 RESULTS AND DISCUSSION

5.1 MANAGEMENT CLASS SEPARABILITY

J-M distances between management classes 1-7, based on all pixels within all test sites, are shown in Table 1. It is seen from the table that no management class is spectrally perfectly separated from any other class. The best bands for separation are, in order of importance: 4, 11, 8, 10, 7, 9, 3, 5 and 6. Band 4 (550 nm) is the green peak, bands 11 (802 nm) and 10 (769 nm) are the NIR shoulder, 8 (715 nm) and 9 (737 nm) are on the red edge and 7 (683 nm) is a chlorophyll absorption band. The separability was only slightly increased beyond these first six bands.

J-M distances were recalculated excluding the areas with calibration in-accuracy exceeding ± 0.25 nm. The separability between classes was improved and increased until eight bands were included. These eight bands were: 3, 4, 5, 7, 8, 9, 10 and 11. The order of importance was different from above but again band 6, at 650 nm (the back slope of the green peak), was the poorest for separation between the classes (bands 1 and 2 were excluded from the analysis). The results indicate that including vegetation indices and/or red edge parameters as additional variables might improve separability between classes. Generally it is seen that classes 6 and 7 are the two classes that separate best from the other classes.

Table 1: J-M separability measures between management classes.

All test sites							
Class	1	2	3	4	5	6	7
1	0.00						
2	1.31	0.00					
3	1.34	1.24	0.00				
4	1.10	1.28	1.27	0.00			
5	1.26	1.14	1.18	1.27	0.00		
6	1.40	1.40	1.38	1.35	1.38	0.00	
7	1.30	1.36	1.27	1.16	1.33	1.37	0.00

Test sites within 0.25 nm spectral accuracy							
Class	1	2	3	4	5	6	7
1	0.00						
2	1.39	0.00					
3	1.35	1.24	0.00				
4	1.30	1.24	1.28	0.00			
5	1.37	1.13	1.17	1.20	0.00		
6	1.41	1.40	1.38	1.36	1.38	0.00	
7	1.41	1.39	1.30	1.40	1.39	1.40	0.00

The overall improvement may be indicative of heterogeneity between test sites within the management classes but it may also reflect that a calibration in-accuracy exceeding ± 0.25 nm has a negative effect on the classification results. This indication was supported by a match filtering test, in which it was seen that the columns of a calibration in-accuracy exceeding ± 0.25 nm were assigned to the same endmember irrespective of the management class. The affected scan lines were excluded leaving 16 test sites for the further analysis.

The separability of the management classes in spectral space is well explained from the ordination diagram in floristic space (Figure 2). Classes 6 and 7 and classes 1 and 2 have opposing extreme plot scores along DCA1, whereas the other classes are placed along a gradient. In the spectral space, classes 6 and 7 separate better from class 1 than class 2 due to the fact that the test site of management class 1 belongs to floristic class 2, whereas the test site of management class 2 belongs to floristic class 1 (Table 2). Classes 3, 4, and 5 are poorest separated since they occupy largely the same space in the ordination diagram but since the floristic centers of the classes are markedly different, a relatively high J-M distance is found between some classes.

5.2 SUBCLASS IDENTIFICATION

Each management class was clustered into nine spectral classes and CDFs were calculated. Two-dimensional scatter plots of the spectral clustering linked to image spatial space showed that the management classes were not spectrally unique and twelve subclasses were identified from the test site data.

A floristically based clustering of the 290 plots reflecting the management classes were approached applying a multinomial log-linear model fit by feed forward neural nets (Ripley & Venables 1997) to predict class membership as a function of plot scores on the three ordination axes extracted from the DCA. Figure 2 shows the distribution of sample plots and their cluster-membership along DCA1 and DCA3, the two axes discriminating best between the 7 clusters.

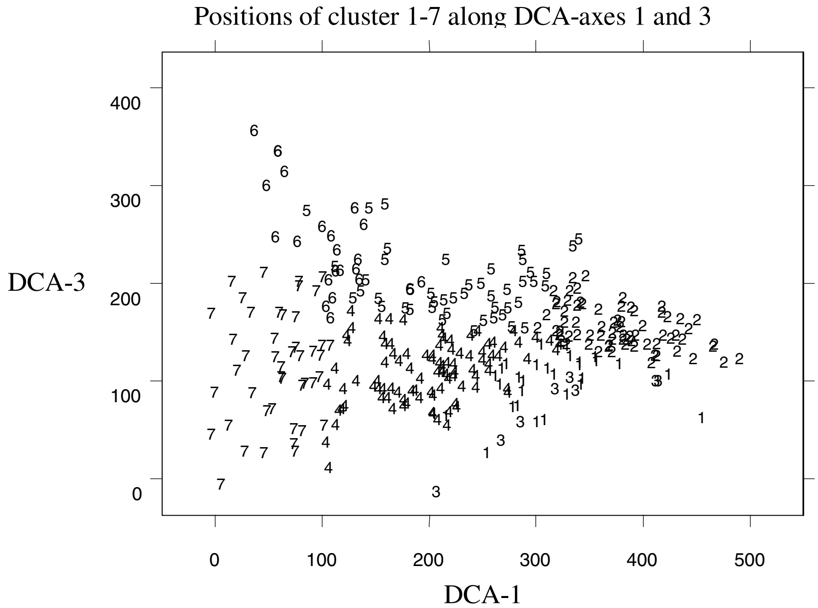

Figure 2: Cluster membership of the 290 field samples and the 30 test sites.

The initial model was refined to optimize floristical similarity of the resulting clusters by fitting the model a second time using only plots with a probability > 0.5 (predicted from the first model) of membership to one of the 7 management classes. An exception was made for the poorly represented and predicted management class 1 (only 18 plots) where plots with a probability > 0.2 were included in the second model. The second model was then used to predict the membership of all plots to one of the seven management related floristical clusters. The resulting floristical clusters were perfectly separable in terms of the three DCA-axes and had a 50% similarity with the original management classes, implying that 50% of the plots were predicted (by means of their vegetation) to belong to a cluster defined by the typical floristic composition of a different management class (Table 2).

Table 2: Cross tabulation showing the relation between field based management class assignment and modelled floristic class assignment.

Management classes	Floristic classes						
	1	2	3	4	5	6	7
1	7	8	0	2	1	0	0
2	9	31	0	1	5	0	1
3	7	8	3	12	1	0	0
4	4	13	3	52	8	3	10
5	2	13	2	23	25	4	1
6	0	0	0	1	4	14	6
7	0	0	0	6	1	2	27

The original 30 test sites had 29% similarity between management classes and floristical clusters, indicating that management class membership of test sites were less well defined floristically than the average. The 12 spectral subclasses defined from the clustering and CDFs explained the whole variation within the 16 test sites used in the analysis.

5.3 SUBCLASS SEPARABILITY

The J-M separabilities were calculated between subclasses including all pixels in the test site and after statistical outliers in the regions of interest were identified and removed and the Mahalanobis distance was used to grow a spectrally coherent training data set in the image data (Table 3). The first digit in the subclass number refers to management class, the second digit to the floristic class.

The overall separability between subclasses improved after seed growing; in particular, class 4.1 showed better results. For class 4.1 93 pixels were included in the training area after seed growing compared to more than 200 in the initial training set. The reduced number of pixels included after seed growing indicates that the test site belongs to more than one floristic class and that seed growing performs well. Subclass 4.3 and 4.5 have also fewer pixels included after seed growing but in this case it turns out to be less related to the J-M distance. This indicates that these two subclasses spectrally dominate their respective test sites, even though the test sites are heterogeneous.

Table 3: J-M distance between subclasses using Mahalonobis distance with $\chi^2 = 0.50$.

Subclasses seed grown ($\chi^2 = 0.50$)												
	1.2	2.1	3.2	3.3	3.4	4.1	4.3	4.5	5.1	5.5	6.5	7.7
1.2	0.00											
2.1	1.41	0.00										
3.2	1.41	1.37	0.00									
3.3	1.41	1.41	1.41	0.00								
3.4	1.41	1.41	1.41	1.41	0.00							
4.1	1.39	1.41	1.41	1.41	1.41	0.00						
4.3	1.41	1.41	1.41	1.41	1.41	1.41	0.00					
4.5	1.41	1.41	1.41	1.41	1.41	1.41	1.41	0.00				
5.1	1.41	1.38	1.14	1.41	1.41	1.41	1.41	1.41	0.00			
5.5	1.41	1.41	1.41	1.41	1.41	1.30	1.41	1.41	1.41	0.00		
6.5	1.41	1.41	1.41	1.41	1.39	1.41	1.41	1.41	1.41	1.39	0.00	
7.7	1.41	1.41	1.41	1.41	1.41	1.41	1.41	1.41	1.41	1.41	1.41	0.00

Table 3 shows that there is a very high separability between the subclasses. Class 1.2 has its smallest J-M distance to class 4.1, whilst class 2.1 has the smallest distances to 3.2 and 5.1. This may be explained by the fact that floristic classes 1 and 2 are difficult to separate floristically, as seen from their plot scores in ordination space (Figure 2). Generally speaking the two subclasses separate well from any other class. Class 3.2 mixes with 5.1 and

4.1 mixes with 5.5. It is seen from Figure 2 that classes 3, 4 and 5 largely occupy the same floristical space in the DCA diagram. Classes 6 and 7 have extreme plot scores in ordination space and separates well from every other class. J-M distance measured between floristic classes alone parallel to the J-M distance between management classes alone showed that floristic classes were poorer separated than management classes. The spectral signatures are immediately influenced by the management characteristics of e.g. grazing/non-grazing whereas the effects on the floristics need years to be established.

6.0 CONCLUSION

The hierarchical approach to define statistically sound spectral subclasses describing management and species composition is promising. Clustering of data, and definition of spectral subclasses within management classes using two-dimensional scatterplots of canonical discrimination functions 1 and 2 were well explained by the species composition of the management class. The floristic clustering governed by management class membership explained separability between subclasses in spectral space. Separability evaluated using the Jeffrey-Matusita distance was improved when the training sites were based on seed growing using Mahalanobis distance. The analysis was based on 16 test sites but despite of the good results a more robust set of data including field measured spectral signatures of higher spectral resolution and vegetation sampling from the same area is required in order to verify the indications of this study.

At a management level, classes 1 and 2 of importance for conservation of biodiversity were well separated from class 6 and 7, dominated by weed species and sown grass leys, in both ordination and spectral space. Management class 3, potentially important for conservation of biodiversity, and management classes 4 and 5, less important but still useful for wild life, were less well separated in floristical and spectral space. At a subclass level, all classes but a few were well separated and in good agreement with the plot scores in ordination space.

A next step in this analysis would be to map the grassland areas. The high quality set of training areas established from the hierarchical approach shows that a maximum likelihood classifier would map the 12 subclasses to a high degree of accuracy. Maximum likelihood assigns pixels to a discrete class but floristical and environmental gradients are seldom discrete leading to a a lack of information regarding the exact conditions of the areas. Matched filtering could be a better means to extract the floristical variation along environmental gradients across management classes using the training sets as endmembers. The match filtering would, however, only perform well if the endmembers are extreme pixels, hence covariance drivers, in spectral space. In any case, field work should be carried out to add training sets for the remaining 18 combinations fo management classes and floristic clusters to perform a full mapping of the grassland areas.

The interpretation based on spectral signatures and floristics indicated that the importance of management overrules the importance of floristics in a remote sensing perspective: the spectral signatures are influenced very soon after changes in significant management characteristics such as grazing/non-grazing whereas there is a timelag before the floristic effects of management changes is established. The results look promising in a monitoring perspective since operative methods for planning of management priorities demand information about not only the present management and management history but also the dominant plant species of an area. We encourage the development of methods that take advantage of identifying spectrally unique classes to be interpreted with continuous as opposed to discrete environmental and floristical data.

7.0 ACKNOWLEDGEMENTS

Dr. Bjarne K. Ersbøll, Department of Mathematical Modelling, Technical University of Denmark, wrote the J-M distance computer program. Dr. Rasmus Larsen, Novo Nordisk A/S, wrote the seed computer program in cooperation with A. A. Nielsen.

8.0 REFERENCES

R. Ejrnæs and H.H. Bruun, "Prediction of grassland quality for environmental management", *J. Environmental Management* 43: 171-183, 1995.

B.K. Ersbøll, "*Transformations and Classifications of Remotely Sensed Data:Theory and Geological Cases,*" Ph.D. Thesis, Department of Mathematical Modelling, Technical University of Denmark, pp. 297, 1989.

R.A. Fisher, "The utilisation of multiple measurements in taxonomic problems," *Annals of Eugenics*, 7: 179-188, 1936.

R.M. Fuller, Groom, G.B., Jones, A.R., "The Land Cover Map of Great Britain: an automated classification of Landsat Thematic Mapper Data", *Photogrammetric Engineering and Remote Sensing*, 60:553-562, 1994.

H. Flesche, A.A. Nielsen, and R. Larsen, "Supervised mineral classification with semi-automatic training and validation set generation in scanning electron microscope energy dispersive spectroscopy images of thin sections," accepted for *Mathematical Geology*, 1999.

G. Groom, H. Mikkelsen, H. Skriver, B. Hansen, A. Thompson, A.A. Nielsen., "Activities and Early Results from DANMAC: The DANish Multisensor Airborne Campaign", Proc. *Third International Airborne Remote Sensing Conference and Exhibition*, 7-10 July 1997, Copenhagen, Denmark, vol.-II:617-24

M.O. Hill, *DECORONA: A FORTRAN program for detrending correspondance analysis and reciprocal averaging.* Cornell University, Ithaca, N.Y., 1979.

A. Jacobsen, N.H. Broge and B.U. Hansen, "Monitoring Wheat Fields and Grasslands Using Spectral Reflectance Data". Proc. *International Symposium on Spectral Sensing Research (ISSSR)*, Melbourne, Victoria, Australia, 26 Nov – 1 Dec, 1995. Available on CD-rom from Commonwealth Information Services, Australian Government Publishing Service, GPO Box 84, Canberra ACT 2601. ISBN No. 0 644 39625.

A. Jacobsen, N. Drewes, M. Stjernholm and T. Balstrøm, "Generation of a Digital Elevation Model of Mols Bjerge, Denmark, and Georeferencing of Airborne Scanner Data", accepted for *Danish Journal of Geography*, 1999.

C.L. Lauver and J.L. Whistler, "A hierarchical Classfication of Landsat TM Imagery to Identify Natural Grassland Areas and Rare Species Habitat. *Photogrammetric Engineering and Remote Sensing,* 59: 627-634, 1993.

M.M. Lewis, "Species Composition Related to Spectral Classification in an Australian Spinifex Hummock Grassland", *Int. J. Remote Sensing*, 15:3223-39, 1994.

K. Matusita, "A distance and related statistics in multivariate analysis." In ed. P. R. Krishnaiah, *Multivariate Analysis*. Academic Press, New York, pp. 187-200, 1966.

B. McCune and M.J. Mefford, *PC-ORD for windows. Multivariate analysis of ecological data*. Version 3.04. MjM Software, Gleneden Beach, Oregon, U.S.A, 1997.

B.D. Ripley, *The NNET library*. Obtained from http: //www.stats.ox.ac.uk/~ripley/MASS2/Windows.html. Last edited on Monday 22 June 1998. Unpublished.

S-Plus 4.5 Professional. Release 2. 1988-1998. MathSoft, Inc.

W.N. Venables and B.D. Ripley , *Modern Applied Statistics with S-PLUS*, Springer, 1997.

C.A. Wessman, C.A. Bateson, T.L. Benning, "Detecting Fire and Grazing Patterns in a Tallgrass Prairie Using Spectral Mixture Analysis", *Ecological Applications*, 7:493-511, 1997.

R.H. Økland, "Vegetation Ecology: Theory, methods and applications with reference to Fennoscandia", *Sommerfeltia* suppl. vol 1, 1990.