# Spectral Identification of Plant Communities for Mapping of Semi-Natural Grasslands

by A. Jacobsen • A.A. Nielsen • R. Ejrnæs • G.B. Groom

## RÉSUMÉ

Cette étude a été réalisée sur des prairies caractérisées par un sol sableux et bien drainé au Danemark. Les données images comprenaient des images du capteur CASI (Compact Airborne Spectrographic Imager), géoréférencées et étalonnées par rapport à la réflectance apparente de surface. Les données écologiques incluaient une carte d'aménagement basée sur le travail de terrain, le repérage des espèces de plantes (vasculaires) et trente sites tests de 30m par 30m regroupés en fonction de sept classes d'aménagement identifiées sur le terrain et de sept classes floristiques modélisées à partir de l'analyse de correspondance normalisée. L'analyse spectrale a été réalisée à partir de la réflectance dérivée des images de 18 sites tests localisés à l'intérieur du couloir de visée du CASI. L'identification spectrale des communautés de plantes était basée sur une approche hiérarchisée établissant le rapport entre les sites test et i) l'aménagement (Ma), et ii) la flore (Fl) utilisant la cohérence spectrale et la séparabilité comme critère principal. L'évaluation de la cohérence spectrale était basés sur la classification non dirigée par agrégation des sites test des classes Ma de 1 à 7 suivie d'une analyse discriminante canonique. L'évaluation de la séparabillité spectrale était basée sur des mesures de la distance Jeffries-Matusita. la croissance des graines a généré des classes d'entraînement en fonction de l'aménagement et de la flore (classes MaFl). Une classification par maximum de vraisemblance a montré que les classes étaient bien définies statistiquement et spatialement cohérentes. La superposition de la classification des classes MaFl sur la carte d'aménagement a permis d'ajouter une information détaillée sur la variation de la végétation à l'intérieur des zones d'aménagement. L'inverse de la précision de classification utilisant la carte d'aménagement comme réalité de terrain a été interprétée comme une mesure de l'hétérogénéité de la communauté de plantes à l'intérieur des classes d'aménagement. L'analyse spectrale de même que la classification par maximum de vraisemblance ont démontré que la source de la variation spectrale à l'intérieur des classes d'aménagement pourrait être reliée à la composition de la végétation.

## SUMMARY

This study was performed on Danish grasslands on well-drained sandy soils. Image data included georeferenced Compact Airborne Spectrographic Imager (casi) data calibrated to apparent surface reflectance. Ecological data included a field-based management map, registration of (vascular) plant species and thirty 30 m by 30 m test sites with affinities corresponding to seven management classes identified in the field and seven floristic classes modelled from detrended correspondence analysis. Spectral analysis was performed on the derived image reflectance of 18 test sites positioned within the casi scanline. Spectral identification of plant communities was based on a hierarchical approach relating the test sites to i) management (Ma) and ii) flora (Fl) using spectral consistency and separability as the main criteria. Evaluation of spectral consistency was based on unsupervised clustering of test sites of Ma classes 1 to 7 followed by canonical discriminant analysis. Evaluation of spectral separability was based on measures of the Jeffries-Matusita distance. Seed growing generated training classes relating to management and flora (MaFl classes). Maximum likelihood classification showed that the classes were well-defined in statistical terms and also spatially coherent. Superimposition of the classification of MaFl classes on the management map added detailed information of vegetation variation within the management areas. The inverse of the classification accuracy, using the management map as 'ground truth', was interpreted as a measure of plant community heterogeneity within management classes. The spectral analysis as well as the maximum likelihood classification indicated that the source of spectral variation within management classes might be related to vegetation composition.

# INTRODUCTION

Only a minor part of Denmark comprises semi-natural terrestrial habitats such as dry grasslands (Ejrnæs and Bruun, 1995). Dry grasslands support a large diversity of native plant and invertebrate species and have experienced a dramatic decline in the Danish landscape in recent years. Furthermore the open grassland areas act as corridors in the landscape for dispersion of plants and animals. A major threat to semi-natural grasslands today is encroachment with woody species following the cessation of grazing. Natural processes such as wildfires and wildlife grazing that open up scrub areas and woodlands rarely occur and are actively prevented by forestry management practices. Furthermore, since most of Denmark comprises highly productive agricultural areas, succession of fallow land into low-productive, species rich semi-natural grasslands is unlikely to occur. Accordingly, management of the remaining semi-natural grasslands in terms of livestock grazing or cutting is crucial for long-term conservation of biodiversity in the Danish landscape.

Mapping at a detailed level is an important issue in monitoring landscapes for biodiversity. Automated classification of remote sensing data is a well-established technique used for classification of broad land cover classes (Fuller et al., 1994). However, separation within broad land cover classes of management and floristics is also likely to be possible, associated with the opportunities provided by airborne scanners of high spatial and spectral resolution. Satellite remote sensing data have proven useful in monitoring of grassland management (Mino et al., 1998) and identification of natural grassland and rare species habitats (Lauver and Whistler, 1993). Recent developments in spectral classification have resulted in the interpretation of discriminant functions relating to, for example, changes in vegetation due to ecological disturbances (Goodin and Henebry, 1997) and numeric vegetation classification of plant communities (Lewis, 1998).

The aim of this study was to investigate if the source of spectral heterogeneity within grasslands may be related to vegetation composition. The main approach was to study spectral consistency and spectral separability between groupings of test sites before they were accepted as classes. Spectral consistency ensured that the multivariate data in each assumed class comprised, in a statistical sense, just one class. Spectral separability ensured that the multivariate data of each assumed class were spectrally different. The final classes were further refined and seed growing was used to generate training classes for maximum likelihood classification. The study was performed within the DANish Multisensor Airborne Campaign (DANMAC) project (Groom et al., 1997).

## STUDY AREA

The Mols Bjerge study area is located on the Djursland peninsula in the eastern part of Jutland **Figure 1**. Mols Bjerge was formed during the last Ice Age approximately 16,000 years ago. The area was located between two tongues of the Young-Baltic Glacier (Ebeltoft Vig and Kalø Vig, **Figure 1**). Terminal
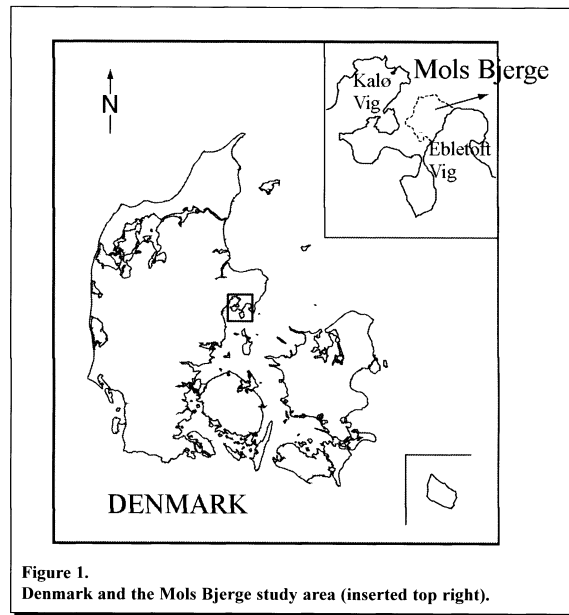
**Figure 1.**
**Denmark and the Mols Bjerge study area (inserted top right).**

moraines that were pushed from both sides formed long, parallel ridges; during the retreat of the ice tongues a belt of dead ice formed erosion valleys, dead ice kettles, and kames.
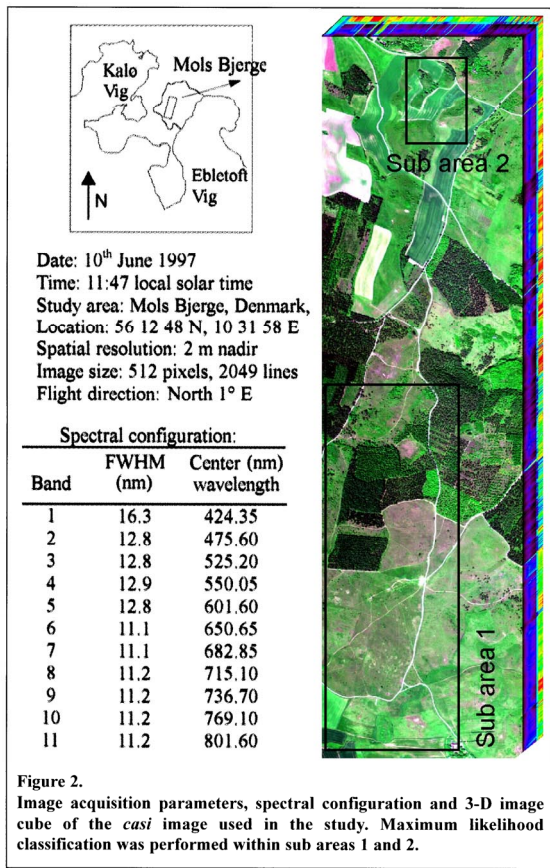
Cairns indicate that farming existed in Mols Bjerge already in the later Stone Age 6,000-5,000 years ago and many grave mounds indicate that the number of settlements increased in the early Bronze Age 3,000-2,500 years ago. From the Iron Age 2,500-900 years ago settlement was more permanent and relics of these farming systems are visible in the present landscape. Today, the earlier cultivated areas are replaced by grasslands with occasional domestic livestock grazing, and, apart from a few fields, the area is dominated by open, dry grasslands, with spontaneous shrub, thickets, deciduous forests and coniferous plantations. On account of the glacial history of the area, which is so clearly visible in the open landscape, its prehistoric remains and its rich flora and fauna, the northern part of Mols Bjerge was declared a protected area in 1977. The southern part was added to the protected area in 1992.

## DATA

The study was performed on image reflectance spectra derived from a casi (Compact Airborne Spectrographic Imager) image covering a minor part of Mols Bjerge and ecological data of management and plant species composition from the whole region of Mols Bjerge (**Figure 2**).

### Image Data

The casi is a linear array scanner recording 512 pixels across track in the spectral region from 400 to 900 nm with, in the present case, a field of view (FOV) of 42°. The spectral configuration of the scanner and the spatial resolution may be

**Figure 2.**
**Image acquisition parameters, spectral configuration and 3-D image cube of the *casi* image used in the study. Maximum likelihood classification was performed within sub areas 1 and 2.**

| Band | FWHM (nm) | Center (nm) wavelength |
|------|-----------|------------------------|
| 1 | 16.3 | 424.35 |
| 2 | 12.8 | 475.60 |
| 3 | 12.8 | 525.20 |
| 4 | 12.9 | 550.05 |
| 5 | 12.8 | 601.60 |
| 6 | 11.1 | 650.65 |
| 7 | 11.1 | 682.85 |
| 8 | 11.2 | 715.10 |
| 9 | 11.2 | 736.70 |
| 10 | 11.2 | 769.10 |
| 11 | 11.2 | 801.60 |

Date: 10th June 1997
Time: 11:47 local solar time
Study area: Mols Bjerge, Denmark,
Location: 56 12 48 N, 10 31 58 E
Spatial resolution: 2 m nadir
Image size: 512 pixels, 2049 lines
Flight direction: North 1° E

Spectral configuration:

user-defined – the one being the tradeoff of the other as a function of the integration time of the scanner. The scanner was flown with a 2 m spatial resolution along the nadir line and an 11-band spectral configuration. The spectral configuration (**Figure 2**) was defined for use in the classification of encroachment of forest onto grasslands (Jacobsen *et al.*, 1998) and grassland monitoring (this paper); consideration was taken of a pilot study of the spectral separability of grasslands (Jacobsen *et al.*, 1995) and a study of forest inventory and mapping (Baulies and Pons, 1995).

The image data (**Figure 2**) were acquired on 10th June 1997 as part of the DANMAC project. The image was acquired at 11:47 local solar time flying north 1° east, *i.e.* as close to solar noon as possible in order to diminish the effects of bi-directional reflectance. The image was acquired during good weather conditions – high pressure and no cloud cover. Spectral signatures over one calibration target were measured during image acquisition and used for atmospheric calibration of the image data to apparent surface reflectance (Jacobsen *et al.*, 2000).

The quality of the spectral and radiometric calibration of the *casi* data was assessed (Jacobsen *et al.*, 2000). It was found that the radiometric calibration was poor in the first two bands and that spectral calibration accuracy exceeded ± 0.25 nm in the left-most 135 columns across-track due to spectral alignment

problems. Retrieved surface reflectance factors were negative for a number of pixels in the first two bands and other studies showed that the spectral inaccuracy affected classification results (Jacobsen *et al.*, 1999a). In summary, due to the image quality problems, the image data were reduced to 377 columns and just nine bands of data from 525 to 802 nm.

The image data were georeferenced for integration of ecological and spectral data using a digital elevation model (DEM) and a triangular irregular network (TIN) resampling (Jacobsen *et al.*, 1999b). To avoid data resampling effects the spectral analysis and maximum likelihood classification were performed on non geo-referenced data.

## Ecological Data

In 1996, botanical field work including management mapping and registration of vascular plant species was performed on all grassland areas in Mols Bjerge that had a cover of grasses, forbs and/or dwarf shrubs of more than 25%; forest clearings were excluded.
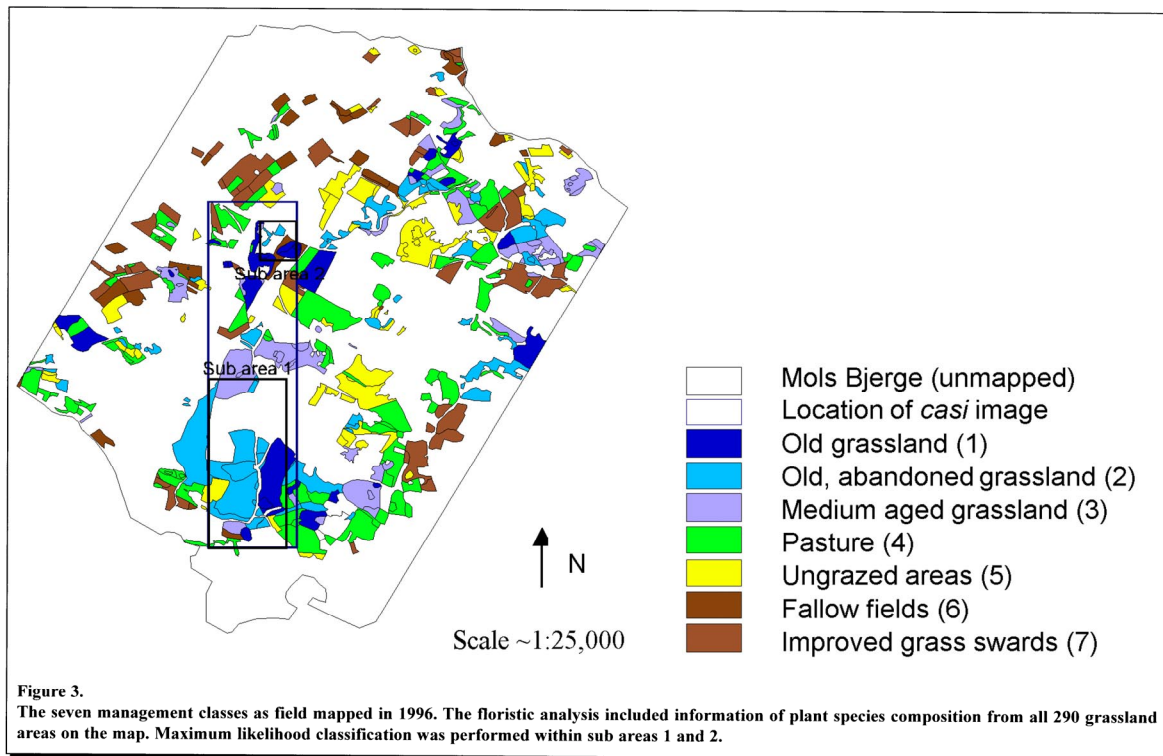
*Management*
The grassland areas were assigned to one of seven management (Ma) classes:

• old unimproved grasslands with continuous grazing (Ma 1),

• old, unimproved, but abandoned grasslands (Ma 2),

• medium aged grassland, previously cultivated, but now with spontaneous dry grassland vegetation (Ma 3),

• young, formerly cultivated areas with spontaneous grazed vegetation (Ma 4),

• young, formerly cultivated areas with spontaneous ungrazed vegetation (Ma 5),

• 1-5 years old 'set-aside' vegetation dominated by weed species (Ma 6), and

• improved, sown grass swards (Ma 7).

Assignment to management class was assessed on the basis of topography, vegetation, air photos (1945-1990) and landowner interviews. Ma classes 1, 2 and 3 were distinguished on the basis of their grazing pattern and an index of indicator species for old grasslands of conservation interest (Ejrnæs and Bruun, 1995). Ma classes 4, 5 and 6 were distinguished on the basis of grazing patterns and field assessment of short-term land management. A management map (**Figure 3**) was produced from the field surveys with the assistance of infrared airphotos acquired in 1995; for overview purposes this map was digitized via feature duplication onto a digital image acquired with a GER (Geophysical Environmental Research) 3715 scanner in 1995.

*Plant Species*
Vascular plant species were recorded using a robust abundance scale that was adapted to the species inventory of

**Figure 3.**
The seven management classes as field mapped in 1996. The floristic analysis included information of plant species composition from all 290 grassland areas on the map. Maximum likelihood classification was performed within sub areas 1 and 2.

these large and often heterogeneous areas. An abundance scale with four steps was used: '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'. Dominant species and indicative species of the seven management classes are listed in **Table 1**. The distinction between dominant and indicative species was based on the principle that dominant plant species are often widespread with wide ecological amplitude, which makes them less suited as indicators at a detailed level. Dominant species may nevertheless be informative for vegetation description. Dominance was estimated as the product of frequency and average abundance score; indicator value was estimated by a method proposed by Dufrêne & Legendre (1997).

### Test Sites

In 1997 (within ± 10 days of the *casi* over-flight) thirty 30m by 30m test sites, at the time regarded to be representative of the seven different management classes, were surveyed. For these test sites, areas with homogeneous vegetation and constant slope angle and aspect were selected. Vascular plant species were recorded using the same abundance scale as in 1996. Eighteen test sites were located in the *casi* scanline and used for the spectral analysis.

Integration of the image and ecological data was based on the georeferenced image data and geopositioning of the test sites using differential global positioning system (DGPS). The test sites were re-located by visual inspection in the non-georeferenced image (**Figure 4**) before spectral analysis of the reflectance data.

## VEGETATION ANALYSIS

Floristic analysis of the management classes in Mols Bjerge was performed using detrended correspondance analysis (DCA). Subsequent floristic classification, supervised by the management characteristic, was based on clustering of the DCA scores. The analysis was performed on floristic data from the management inventory in 1996 (290 plots) and the test site inventory in 1997 (30 plots). The vegetation analysis was performed independently from the other analyses as a means for interpretation of the subsequent spectral analysis and is presented here only in brief.

### Floristic Variation

The untransformed floristic data from the 1996 and 1997 inventories were combined and subjected to ordination using a standard DCA (Hill, 1979; Økland 1990) with down-weighting of rare species. PC-Ord (McCune and Mefford, 1997) was used for ordination. Three ordination axes were extracted, representing the major floristic gradients present in the data (Økland, 1990). The ordination scores on DCA axes 1 and 3 are

**Table 1.**
**Dominant species and indicative species of each of seven recognized management classes.**

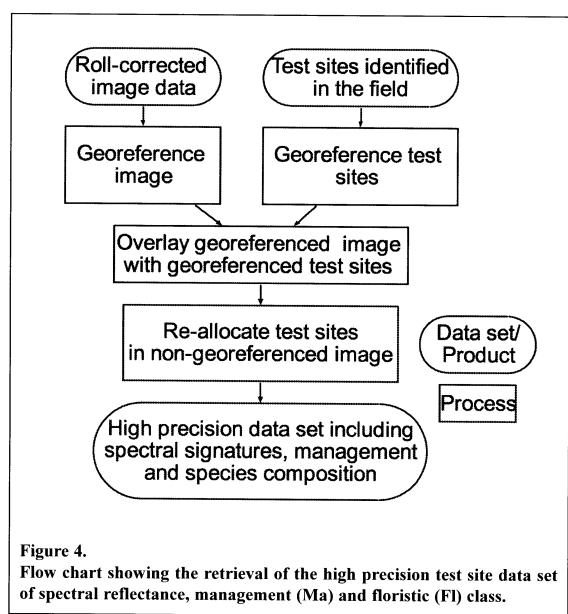| Management class | Most dominant species | Most indicative species |
|---|---|---|
| Old, unimproved grassland (grazed) (Ma 1) | Deschampsia flexuosa<br>Agrostis capillaris<br>Festuca ovina | Lathyrus montanus<br>Festuca ovina<br>Genista anglica |
| Old, derelict, unimproved grassland (ungrazed) (Ma 2) | Deschampsia flexuosa<br>Calluna vulgaris<br>Holcus mollis | Calluna vulgaris<br>Deschampsia flexuosa<br>Avenula pratensis |
| Medium-aged grassland (grazed/ungrazed) (Ma 3) | Hieracium pilosella<br>Deschampsia flexuosa<br>Agrostis capillaris | Luzula campestris<br>Hieracium pilosella<br>Lotus corniculatus |
| Young, formerly cultivated grassland (grazed) (Ma 4) | Festuca rubra<br>Agrostis capillaris<br>Bromus hordeaceus | Achillea millefolium<br>Agrostis capillaris<br>Festuca rubra |
| Young, formerly cultivated grassland (ungrazed) (Ma 5) | Deschampsia flexuosa<br>Elymus repens<br>Holcus mollis | Arrhenatherum elatius<br>Anthriscus sylvestris<br>Dactylis glomerata |
| Fallow fields (Ma 6) | Elymus repens<br>Senecio vernalis<br>Rumex acetosa | Viola arvensis<br>Senecio vernalis<br>Artemisia vulgaris |
| Improved grass swards (Ma 7) | Lolium perenne<br>Festuca rubra<br>Poa pratensis | Lolium perenne<br>Taraxacum sect. Vulgare<br>Trifolium repens |

shown in **Figure 5**. The numbers 1-7 in the figure indicate the Ma classes of the plots. **Figure 5** shows that Ma classes 1 and 2 (and to some extent Ma 3) and Ma classes 6 and 7 occupy opposite ends of DCA-1, the major floristic gradient of the ecological data in the study. This result is in agreement with **Table 1** that shows *Deschampsia flexuosa* to be a shared dominant of Ma classes 1, 2 and 3, whereas no other dominants co-occur in Ma classes 6-7 and Ma classes 1 to 3. Ma classes 4 and 5 appear to share dominants with both ends of the gradient; Ma 3 also shows this characteristic, but to a lesser extent.

### Floristic Classification

Floristic classification of the ecological data was performed using a multi-log linear classification model (Venables and Ripley, 1997). The Ma class, as a function of the ordination scores on DCA 1 to 3, was used to weight each observation in the model with the relative abundance of its management class. Modelling was performed with S-Plus 4.5 (MathSoft, 1998). The model was refined to optimize floristic similarity by fitting the model a second time using probability of membership to the

Ma class as the weight factor. This last procedure assigned more weight to typical floristic classes of a given Ma class, and allowed for untypical examples to be assigned to another class. The final model was used to predict the floristic class (Fl class) for each of the grassland areas and the test sites, with Fl classes 1 to 7 relating to Ma classes 1 to 7 respectively.

The cross table (**Table 2**) shows the agreement between the Ma classes and the Fl classes, and shows the floristic consistency of the field classification in terms of its Ma classes. It is seen from the table that the vegetation model gives 49 possible combinations of Ma and Fl classes (MaFl classes) of which the grassland areas and test sites together populate 35. Ma classes 1 to 3 (the target classes for conservation) are well separated from Ma classes 6 and 7 (areas of intensive human use). It may also be seen that Ma classes 4 and 5, and to some extent 3 (potential future conservation interest) are less well separated in floristical terms and there are several possible explanations for this. First, these areas are successional and many factors (*e.g.*, soil fertility and colonization opportunities) may influence the course and speed of succession. Second, the

**Figure 4.**
Flow chart showing the retrieval of the high precision test site data set of spectral reflectance, management (Ma) and floristic (Fl) class.

accuracy of the assignment to the management classes for these areas is considered less certain due to insufficiencies in the historical management data. In summary, 58 % of the samples could be classified to the floristic class characteristic of their management class, whereas 42 % were misclassified. Most misclassifications were however between management classes in the same conservation category, *viz.* "target" (Ma 1, 2 and 3), "potential" (Ma 3, 4 and 5) or "inferior" (Ma 6 and 7).

# SPECTRAL ANALYSIS

The subsequent spectral analysis was the core of this work and is presented in two main sections ('Methods' and 'Results and Discussion'). The treatment given here is more thorough than that of the vegetation analysis. Spectral analysis was performed on reflectance spectra derived from the 18 test sites that were located in the 377 columns of the image with acceptable data quality.

**Table 2.**
**Combinations of management (Ma) and modelled vegetation affinity (Fl).**

| Class | Fl1 | Fl2 | Fl3 | Fl4 | Fl5 | Fl6 | Fl7 |
|-------|-----|-----|-----|-----|-----|-----|-----|
| Ma1   | 6   | 8   | 7   | 1   | 0   | 0   | 0   |
| Ma2   | 4   | 33  | 2   | 2   | 6   | 1   | 1   |
| Ma3   | 4   | 7   | 16  | 6   | 3   | 0   | 0   |
| Ma4   | 1   | 2   | 7   | 51  | 17  | 5   | 8   |
| Ma5   | 0   | 7   | 3   | 18  | 36  | 6   | 1   |
| Ma6   | 0   | 0   | 0   | 0   | 4   | 19  | 2   |
| Ma7   | 0   | 0   | 0   | 3   | 0   | 2   | 34  |

# SPECTRAL ANALYSIS: METHODS

The spectral analysis included evaluation of spectral separability and consistency and subsequent seed growing to further refine the training classes to be used in the classification. Spectral class separability analysis was performed on Ma classes, Fl classes and MaFl classes. Spectral class separability analysis was performed again on MaFl classes after seed growing. Spectral consistency analysis and seed growing was performed on MaFl classes.

## Spectral Separability

Class spectral separability was analyzed by applying the pairwise Jeffries-Matusita (J-M) distance measure between classes and the average pairwise J-M distance between all classes (Matusita, 1966; Ersbøll, 1989). The J-M distance between perfectly separable classes is $\sqrt{2}$ (~ 1.41). Spectral class separability analysis was performed on Ma classes, Fl classes and MaFl classes. Spectral class separability analysis was performed again on MaFl classes after seed growing.
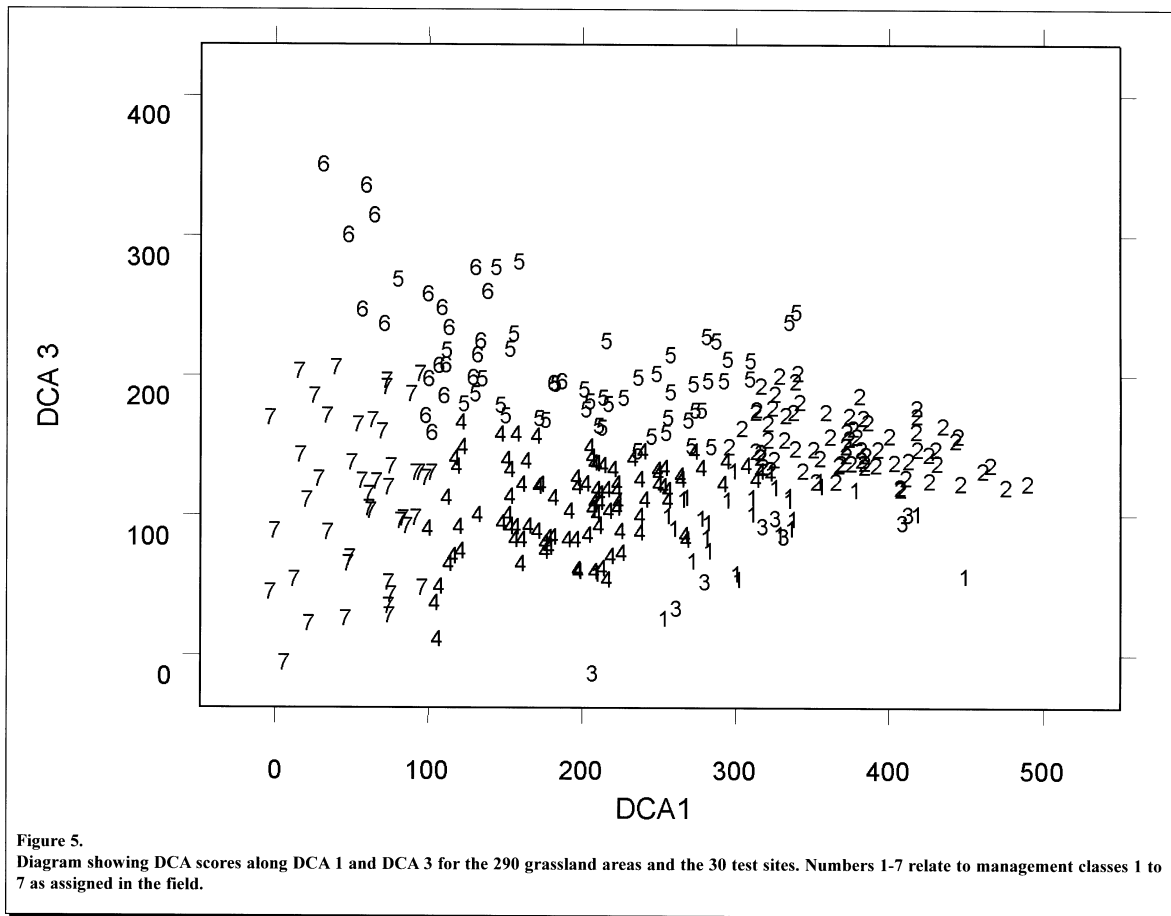
## Spectral Consistency

Spectral consistency was performed by artificially subdividing each of the Ma classes into a number of spectral clusters. Canonical discriminant analysis (CDA) was used to evaluate if the clusters were separable; CDA (Fisher, 1936) is a powerful tool to transform a set of features that optimize class separability (Richards, 1993). If a plot of canonical scores of an Ma class along CDA-axes 1 and 2 showed no sign of outliers or grouping it was accepted as a class.

An unsupervised clustering algorithm was used to sub-divide the Ma classes into the artificial clusters. Observations within each class, called 'cluster seeds' (one for each assumed cluster), were randomly selected as a first guess of potential cluster means. Clusters were formed by assigning all observations to the nearest seed as measured by Euclidean spectral distance in the feature space spanned by the original nine spectral bands. Once all observations were assigned, new cluster means were calculated. New clusters were formed by assigning all observations to the new cluster means. The last two steps were repeated until changes in cluster means became zero (or small).

## Generation of Training Classes

Since successful supervised classification relies on good training data, the statistically sound MaFl classes (see next section) that were identified from the evaluation of spectral consistency were further refined prior to classification. This involved use of a semi-automatic algorithm (Nielsen *et al.*, 1998, Larsen *et al.*, 1999 and Flesche *et al.*, 2000) that was applied for generation of training classes from the MaFl classes. The test site spectra of each MaFl class were used as seed points, and training classes were grown in a manner that ensured spatial and spectral closeness. Spatial closeness was obtained by requesting connectivity. Spectral closeness was obtained by restricting, during growing of the training set, the

**Figure 5.**
Diagram showing DCA scores along DCA 1 and DCA 3 for the 290 grassland areas and the 30 test sites. Numbers 1-7 relate to management classes 1 to 7 as assigned in the field.

spectral distance to the current mean value of the class. There were sufficient seed points per class to define an initial training set and to estimate the dispersion matrix. The dispersion matrix was used to i) exclude any outliers in the current training, and ii) increase the size of the training set further using the Mahalanobis distance method. The final training classes were used for classification using a maximum likelihood classifier.

## SPECTRAL ANALYSIS:
## RESULTS AND DISCUSSION

The spectral analysis was a long and rather complicated process. The process, including the results and discussion of the spectral analysis, interpreted with reference to the vegetation analysis, is described in detail in this section. A summary of the process is provided in table form by the end of the section.

### Spectral Separability of Ma and Fl Classes

As a first approach, the image reflectance spectra derived from the test sites were placed in one of seven groups based on their management class. The results of the spectral separability analysis are shown in **Table 3**.

Comparison of **Table 3** and the DCA diagram (**Figure 5**) indicates that the spectral separability of the Ma classes across the 18 test sites could be explained by the general grassland

**Table 3.**
J-M distance between classes related to management (Ma classes).

| Class | Ma1 | Ma2 | Ma3 | Ma4 | Ma5 | Ma6 | Ma7 |
|-------|------|------|------|------|------|------|------|
| Ma1 | 0.00 | | | | | | |
| Ma2 | 1.32 | 0.00 | | | | | |
| Ma3 | 1.29 | 1.21 | 0.00 | | | | |
| Ma4 | 1.40 | 1.36 | 1.41 | 0.00 | | | |
| Ma5 | 1.28 | 1.13 | 1.12 | 1.40 | 0.00 | | |
| Ma6 | 1.40 | 1.40 | 1.37 | 1.41 | 1.38 | 0.00 | |
| Ma7 | 1.41 | 1.39 | 1.32 | 1.41 | 1.39 | 1.40 | 0.00 |

113

floristic variation in the Mols Bjerge study area. **Figure 5** shows that Ma classes 6 and 7 and Ma classes 1 and 2 have extreme opposite DCA scores and **Table 3** shows that the spectral J-M distances between Ma classes 6 and 7 and Ma classes 1 and 2 are at or close to the value of 1.41. Similarly, Ma classes 6 and 7 are fairly well separated from each other along DCA 3 and they are also well separated from each other in spectral space. It is seen from **Figure 5** that Ma classes 3, 4 and 5 occupy the same floristic space to a large extent and that they are close to the other classes as well, in particular class 1. This pattern is reflected in spectral space by the J-M distances (**Table 3**), except for the case of Ma 4, a deviation that might be because this class is represented by only one test site.

The indication from the comparison between the spectral separability of the Ma classes of the 18 test sites and the floristic DCA analysis that vegetation composition is the source of spectral variation, was further investigated. For this, the 18 test sites were regrouped based on floristics (Fl classes) and the spectral separability analysis was performed again. J-M distances measured between the Fl classes showed that the Fl classes (**Table 4**) were more poorly separated than the Ma classes (already shown in **Table 5**). This result might be explained as the spectral signatures being markedly influenced by the recent, short-term management characteristics of the area e.g., the grazed/non-grazed status; conversely, the effect of such management on the floristics takes far longer (years) to be established. The comparatively species poor, acidic grasslands of Mols Bjerge are known to respond very slowly to changing land use (Bülow-Olsen, 1980).

**Table 4.**
**J-M distance between classes related to plant species composition (Fl classes).**

| Class | Fl1 | Fl2 | Fl3 | Fl4 | Fl5 | Fl6 | Fl7 |
|-------|------|------|------|------|------|------|------|
| Fl1 | 0.00 | | | | | | |
| Fl2 | 1.20 | 0.00 | | | | | |
| Fl3 | 1.23 | 1.11 | 0.00 | | | | |
| Fl4 | 1.40 | 1.34 | 1.27 | 0.00 | | | |
| Fl5 | 1.08 | 1.22 | 1.28 | 1.20 | 0.00 | | |
| Fl6 | 1.39 | 1.38 | 1.39 | 1.17 | 1.20 | 0.00 | |
| Fl7 | 1.40 | 1.41 | 1.34 | 1.40 | 1.38 | 1.41 | 0.00 |

### Identification of MaFl Classes

For further investigation of the vegetation composition as the source of spectral variation, spectral consistency analysis was performed on the data of the 18 test sites grouped into Ma classes. Each Ma class was clustered and subjected to CDA as described in the Spectral Analysis:Methods section.

Inspection of the CDA diagrams revealed that clusters within Ma classes could largely be explained by the variation in vegetation composition i.e., the modelled Fl classes from the vegetation analysis. Three test sites did not fit this pattern:

i) One test site in Ma 2, predicted by the vegetation model as Fl 5 was floristically atypical and placed at an extreme position in the DCA plot compared to other areas of the same MaFl class;

ii) Another test site in Ma 2, predicted by the vegetation model as Fl 2 was floristically as well as spectrally in between two other test sites belonging to MaFl 2.1 and MaFl 2.2;

iii) All four test sites in Ma 3 were predicted to belong to Fl 3. One of these test sites was spectrally different from the other three (**Figure 6**): from inspection of the plant species registered in the field it was clear that this site was floristically atypical for the whole area and accordingly less reliable in terms of floristic class prediction. Furthermore, two test sites belonging to MaFl 3.2 (labelled in **Figure 6** with respect to their slope aspects, NNW and SSE) were spectrally inconsistent, belonging to two spectral clusters. This was most likely due to the slope angle and aspect effects influencing the spectral reflectance through differences in the vegetation phenology and vigour as well as bi-directional effects. For some unexplained reason, this test site and MaFl 3.3 mentioned above belonged to the same spectral cluster.

For the above reasons the three test sites (MaFl 2.5, MaFl 2.2, and MaFl 3.3) were excluded from further analysis. The two test sites (MaFl 3.2 SSE and MaFl 3.2 NNW) of the same MaFl class that belonged to two different spectral classes were merged in order to investigate further the source of spectral variation as related to species composition. By doing this we deliberately overruled the principle of the model, i.e., the primacy of the spectral consistency (or unimodality) of spectral classes.

The results of the spectral analysis indicate that it is plausible to consider vegetation composition as a source of spectral variation within land use categories of grasslands. It is, however, open for discussion whether it is reasonable to expect unimodal spectral classes when it comes to semi-natural vegetation, which will always be heterogeneous, dependent on the scale we are looking at (as exemplified by the case of MaFl 3.2). We do find indications that the approach was reasonable, since **Figure 6** shows that the single test sites are spectrally homogeneous. We did not find any test sites with more than one statistically valid cluster; different clusters always indicated different test sites belonging to the same group of either Ma or MaFl. The ground pixel spacing of the casi image was 2 m by 2 m and an area of 4m² is considered optimal with regard to meeting the sampling criteria of representativeness and homogeneity in grassland studies (Økland, 1990). The dimensions of the test sites was 30 m by 30 m, which was regarded as optimal for the larger scale homogeneity of this study area. Either smaller pixel size or larger test sites are likely to have introduced spectral clusters that could not have been identified floristically.
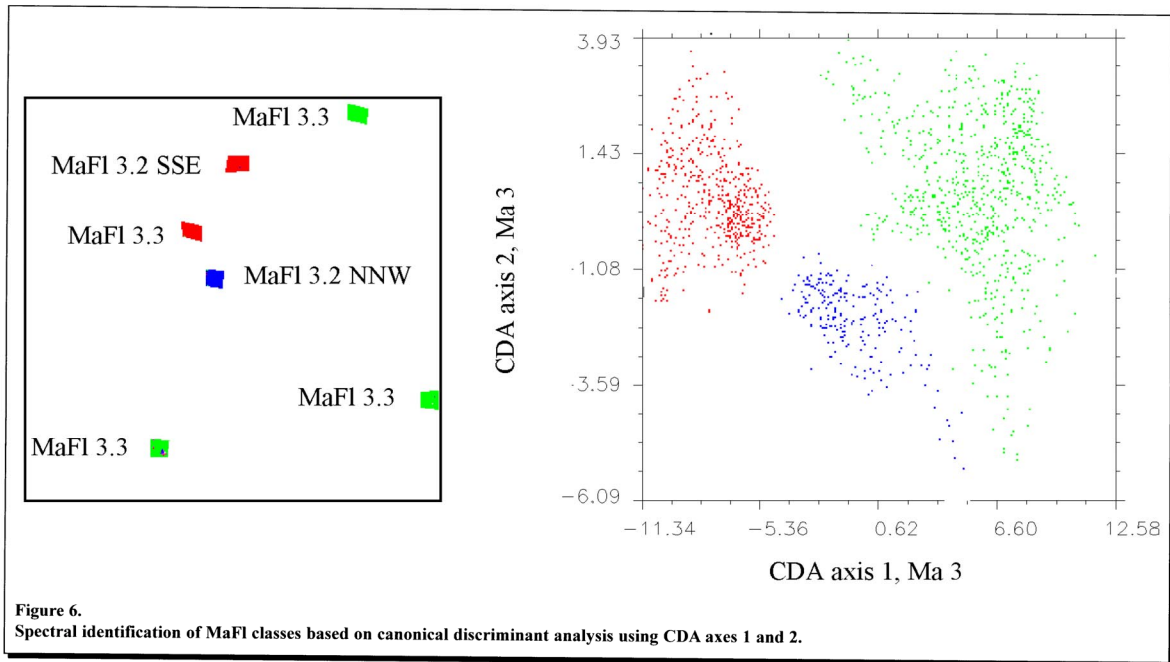
**Figure 6.**
**Spectral identification of MaFl classes based on canonical discriminant analysis using CDA axes 1 and 2.**

## Spectral Separability of MaFl Classes

The J-M distances calculated for 11 spectrally identified MaFl classes showed that the mean J-M distance including zeros on the diagonal was 1.28. For 34 of the 55 class combinations, the separability was 1.41 (**Table 5**).

## Training Class Generation

The MaFl classes were refined to training classes for maximum likelihood classification by seed growing. Using all observations in the test sites as seeds and restricting the Mahalanobis distance to $\chi^2 < 0.50$ improved the overall separability, with the resulting J-M distance being 1.41 for all except seven of class combinations (**Table 6**).

**Table 5.**
**J-M distance between 11 classes related to management and plant species composition (MaFl) before seed growing.**

| MaFl class | 1.1 | 1.2 | 2.1 | 2.2 | 3.2 | 3.3 | 4.4 | 5.3 | 5.5 | 6.5 | 7.7 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.1 | 0.00 | | | | | | | | | | |
| 1.2 | 1.41 | 0.00 | | | | | | | | | |
| 2.1 | 1.40 | 1.41 | 0.00 | | | | | | | | |
| 2.2 | 1.41 | 1.41 | 1.41 | 0.00 | | | | | | | |
| 3.2 | 1.41 | 1.41 | 1.40 | 1.20 | 0.00 | | | | | | |
| 3.3 | 1.40 | 1.23 | 1.41 | 1.41 | 1.39 | 0.00 | | | | | |
| 4.4 | 1.38 | 1.41 | 1.39 | 1.41 | 1.41 | 1.41 | 0.00 | | | | |
| 5.3 | 1.41 | 1.40 | 1.41 | 1.36 | 1.12 | 1.38 | 1.41 | 0.00 | | | |
| 5.5 | 1.37 | 1.37 | 1.38 | 1.41 | 1.41 | 1.30 | 1.39 | 1.41 | 0.00 | | |
| 6.5 | 1.41 | 1.41 | 1.41 | 1.41 | 1.41 | 1.32 | 1.41 | 1.41 | 1.34 | 0.00 | |
| 7.7 | 1.41 | 1.41 | 1.41 | 1.41 | 1.41 | 1.39 | 1.41 | 1.41 | 1.41 | 1.40 | 0.00 |

**Table 6.**
**J-M distance between 11 classes related to management and plant species composition (MaFl) after seed growing.**

| MaFl Class | 1.1 | 1.2 | 2.1 | 2.2 | 3.2 | 3.3 | 4.4 | 5.3 | 5.5 | 6.5 | 7.7 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.1 | 0.00 | | | | | | | | | | |
| 1.2 | 1.41 | 0.00 | | | | | | | | | |
| 2.1 | 1.41 | 1.41 | 0.00 | | | | | | | | |
| 2.2 | 1.41 | 1.41 | 1.41 | 0.00 | | | | | | | |
| 3.2 | 1.41 | 1.41 | 1.41 | 1.34 | 0.00 | | | | | | |
| 3.3 | 1.41 | 1.35 | 1.41 | 1.41 | 1.41 | 0.00 | | | | | |
| 4.4 | 1.41 | 1.41 | 1.41 | 1.41 | 1.41 | 1.41 | 0.00 | | | | |
| 5.3 | 1.41 | 1.41 | 1.41 | 1.40 | 1.30 | 1.41 | 1.41 | 0.00 | | | |
| 5.5 | 1.41 | 1.41 | 1.41 | 1.41 | 1.41 | 1.40 | 1.41 | 1.41 | 0.00 | | |
| 6.5 | 1.41 | 1.41 | 1.41 | 1.41 | 1.41 | 1.39 | 1.41 | 1.41 | 1.40 | 0.00 | |
| 7.7 | 1.41 | 1.41 | 1.41 | 1.41 | 1.41 | 1.41 | 1.41 | 1.41 | 1.41 | 1.41 | 0.00 |

### Schematic Summary

The list of steps in the spectral analysis is given in **Table 7**.

## IMAGE CLASSIFICATION: METHOD

The training class set was edited to remove any MaFl class that had a test site with an MaFl class different from that of the grassland area in which they were. This allowed superimposition of classification results on the management map, but reduced the number of training classes from eleven to six. Classification was performed on sub areas 1 and 2 masked to include only those grassland areas with the same MaFl class as the six remaining training classes. The maximum likelihood classification used training classes derived from the test sites of non-georeferenced data and was performed on non-georeferenced image data. The classification results were subsequently georeferenced for the analysis with the management map.
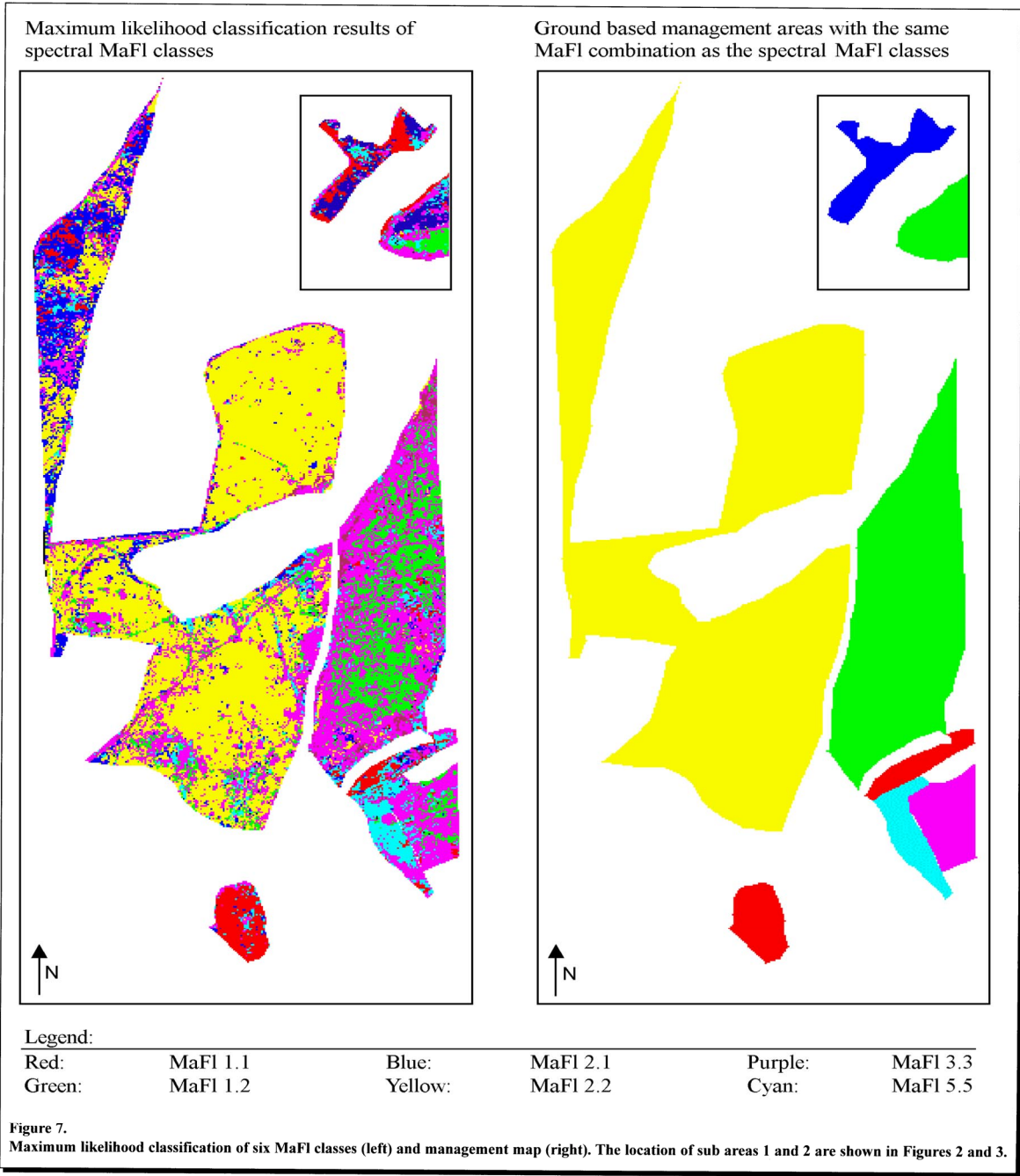
**Table 7.**
**Summary of the spectral analysis.**

| Step | Process and results of spectral analysis |
|---|---|
| 1 | The test sites were grouped in ground-based Ma classes 1-7. |
| 2 | The spectral separability of Ma-classes was indicated by plant species composition. |
| 3 | The test sites were grouped in modelled Fl classes 1-7. |
| 4 | The Fl-classes were distributed along a gradient and were less well spectrally separable than were Ma-classes. |
| 5 | The test sites were grouped again in ground-based Ma-classes 1-7. |
| 6 | Ma-classes were evaluated for spectral consistency. |
| 7 | The spectral analysis showed that unimodal sub-classes within Ma classes were related to plant species composition (Fl class). |
| 8 | Each test site was assigned a new class based on its combination of ground-based Ma class and modelled Fl class (MaFl classes). |
| 9 | The MaFl classes were spectrally consistent and spectrally well separated. |
| 10 | The MaFl classes were 'pruned'. Only some pixels were retained within the classes, retention based on spectral homogeneity. |
| 11 | The MaFl classes were further constrained to be spatially contiguous by the process of growing training classes based spectrally on the pruned MaFl classes |
| 12 | The MaFl training classes were demonstrated to be very spectrally separable. |

## IMAGE CLASSIFICATION: RESULTS AND DISCUSSION

The result of the maximum likelihood classification is shown for the masked sub areas 1 and 2 in **Figure 7**.

**Figure 7** shows that the generated training classes are not only consistent in terms of ecological and spectral variables, they are also spatially coherent, indicating that their selection criteria were successful. Comparison of the spectral classification (**Figure 7**, left) to the management map (**Figure 7**, right) gives detailed



Maximum likelihood classification results of spectral MaFl classes

Ground based management areas with the same MaFl combination as the spectral MaFl classes

Legend:

| | | | | | |
|---|---|---|---|---|---|
| Red: | MaFl 1.1 | Blue: | MaFl 2.1 | Purple: | MaFl 3.3 |
| Green: | MaFl 1.2 | Yellow: | MaFl 2.2 | Cyan: | MaFl 5.5 |

**Figure 7.**
Maximum likelihood classification of six MaFl classes (left) and management map (right). The location of sub areas 1 and 2 are shown in Figures 2 and 3.

117

information on the heterogeneity of the vegetation within large areas that are homogeneous with respect to their management. A measure of vegetation heterogeneity within management classes was obtained by inverting the classification accuracy. In practical terms, the vegetation heterogeneity (**Table 8**) was expressed as 100% minus the maximum likelihood classification accuracy, using the management map as 'ground truth' and the maximum likelihood classification as the classification result.

**Table 8** shows that the heterogeneity, in terms of the maximum likelihood classification, is largest for classes MaFl 1.2, MaFl 2.1 and MaFl 5.5. It is seen from the confusion matrix (**Table 9**) that the high measure of heterogeneity for class MaFl 1.2 largely relates to MaFl 2.2. Ma classes 1 and 2 are closely related floristically (**Figure 5**), and MaFl 2.1 might be seen as classifying as MaFl 1.2 and MaFl 2.2 for the same reason (**Table 9**). MaFl 5.5 is positioned in the centre area of the DCA plot (**Figure 5**) and the heterogeneity of this class could be due to its closeness to several other flora classes, as indicated in **Table 9**.

In the vegetation analysis, it was stated that only 58% of the grassland areas had their flora class predicted by their management class and hence the results in **Table 8** indicating considerable heterogeneity are not surprising. An interesting observation from this is that the source of heterogeneity may be interpreted from the position of the

Ma classes in the floristic diagram (**Figure 5**). This suggests that the source of spectral variation within the management classes may be related to vegetation composition, as was indicated by the spectral analysis of the test sites.

Topographic reflectance effects and variations in vegetation conditions not covered by the flora model are seen in the classification result for sub area 2 for MaFl 1.2. This area represents a small hill with a north-south ridge. The Ma class is the same on both sides of the ridge but due to differences in topography or micro-climate, the area is spectrally classified on either side mainly as MaFl 1.2 and MaFl 2.1. Studies incorporating digital elevation data could show if the pattern here of the image classification is more strongly related to the vegetation characteristics or to topographic effects.

Of the six MaFl classes, MaFl class 1.1 is spectrally classified with the least heterogeneity, which makes sense ecologically. MaFl 2.2 is, from field investigation, known to be fairly homogenous with respect to vegetation and is classified as such. The more textured classification of the northwestern part of MaFl 2.2 is field validated by its variability in terms of soil humidity and successional stage. The confusion between MaFl 1.2 and MaFl 3.3 is not readily recognized in the field, but may be explained floristically since Ma 1 and Ma 3 occupy the same space in the DCA diagram (**Figure 5**).

The clustering of the floristic and the reflectance spectra data is to some degree artificial since the phenomena that the data represent are not naturally clustered but continuous. The floristic gradient analysis (**Figure 5**) showed the continuous (interpreted as successional) development from Ma classes 6 and 7 to Ma classes 1 and 2. If, as shown in the study, the vegetation reflectance is governed by the species composition this feature could be exploited using spectral unmixing. As a first approach, unmixing could target classes distributed at the extreme ends of DCA 1, since these classes are well separated floristically as well as historically and with respect to their management. Spectral unmixing of the grassland areas,

**Table 8.**
**Measure of heterogeneity within management areas described as the inverse of the classification accuracy.**

| MaFl classes | Heterogeneity (%) |
|---|---|
| MaFl 1.1 | 37.54 |
| MaFl 1.2 | 74.48 |
| MaFl 2.1 | 66.52 |
| MaFl 2.2 | 40.52 |
| MaFl 3.3 | 32.87 |
| MaFl 5.5 | 66.55 |

**Table 9.**
**Confusion matrix of the classified image map and the ground reference map, showing the number of classified pixels per class.**

| Class | Ground truth management map | | | | | |
|---|---|---|---|---|---|---|
| | MaFl 1.1 | MaFl 1.2 | MaFl 2.1 | MaFl 2.2 | MaFl 3.3 | MaFl 5.5 |
| Unclassified | 154 | 1780 | 72 | 1627 | 148 | 238 |
| MaFl 1.1 | 1406 | 301 | 649 | 524 | 31 | 986 |
| MaFl 1.2 | 2 | 3474 | 4 | 1021 | 250 | 12 |
| MaFl 2.1 | 99 | 463 | 535 | 3394 | 4 | 68 |
| MaFl 2.2 | 1 | 175 | 6 | 19000 | 0 | 1 |
| MaFl 3.3 | 248 | 6270 | 148 | 5142 | 1195 | 879 |
| MaFl 5.5 | 290 | 605 | 164 | 936 | 144 | 1122 |

with Ma classes 6 and 7 and Ma classes 1 and 2 as training classes (endmembers) could potentially produce abundance maps that could be interpreted as the succession stages from sown grassland areas to old well-established grasslands.

# CONCLUSION

The study was a joint venture combining diverse scientific disciplines. The combination of ecological analysis based on management and floristic data using detrended correspondence analysis and subsequent floristic clustering and spectral analysis based on image derived reflectance spectra using canonical discriminant analysis performed well. A hierarchical approach identified management related plant community classes (MaFl classes) spectrally by dividing a set of test sites into management classes (Ma classes), based on field information, and then into floristic classes (Fl classes), based on modelling. The driving force in the analysis was the requirement for consistency of spectral classes in terms of their being unimodal and evaluation of spectral separability between classes using the J-M distance as a measure. Unimodality of spectral classes with respect to management and plant composition was achieved and apparently described the complex reality of the grasslands. The spatial resolution of the *casi* image data (2 m by 2 m) is recommended for further grassland studies whilst the size of the test sites (30 m by 30 m) should be field evaluated for every study.

Maximum likelihood classification of the image and its superimposition on the management map added detailed information on vegetation heterogeneity. The inverse of the classification accuracy described the vegetation heterogeneity within the field-mapped areas. Evaluation of the heterogeneity with respect to the confusion matrix was compared to the DCA results and indicated, as did the spectral analysis, that the source of the heterogeneity of the ground-mapped management areas might be related to plant species composition. The results indicate that airborne remote sensing with high-resolution spectral data could be an important tool in the conservation and monitoring of grasslands by adding information of plant community variation; this is even in spite of it being more labour intensive than satellite remote sensing. Transfer of the approach to include satellite data is not likely to be as powerful in relation to the spatial and spectral considerations. Airborne remote sensing data may be further exploited using spectral unmixing of the extreme ends of the MaFl class variation. Abundance maps of, in the present case, Ma classes 6 and 7 and Ma classes 1 and 2 could potentially produce vegetation composition abundance maps that could be interpreted in terms of successional stages from sown grassland areas to old well-established grasslands.

As a concluding remark, the authors would like to encourage further joint efforts between image processing specialists, data analysts and plant ecologists using advanced image and data processing techniques to extract and identify spectral information available from scanners at a highly detailed level.

# REFERENCES

Baulies, X. and X. Pons. (1995). "Approach to Forestry Inventory and Mapping by Means of Multi-Spectral Airborne Data", *International Journal of Remote Sensing*, Vol. 16, No. 1, pp. 61-80.

Bülow-Olsen, A. (1980). "Changes in Species Composition in an Area Dominated by Deschampsia Flexuosa (L.) Trin. as a Result of Cattle Grazing", *Biological Conservation*, Vol. 18, pp. 257-270.

Dufrêne, M. and P. Legendre. (1997). "Species Assemblages and Indicator Species: The Need for a Flexible Asymmetrical Approach", *Ecological Monographs*, Vol. 67, pp. 345-366.

Ejrnæs, R. and H.H. Bruun. (1995). "Prediction of Grassland Quality for Environmental Management", *Journal of Environmental Management*, Vol. 43, pp. 171-183.

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

Fisher, R.A. (1936). "The Utilisation of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, Vol. 7, pp. 179-188.

Flesche, H., A.A. Nielsen, and R. Larsen. (2000). "Supervised Mineral Classification with Semi-Automatic Training and Validation Set Generation in Scanning Electron Microscope Energy Dispersive Spectroscopy Images of Thin Sections," *Mathematical Geology*, Vol. 32, No. 3, pp. 337-366.

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

Goodin, D.G. and G.M. Henebrey. (1997). "A Technique for Monitoring Ecological Variables in Tallgrass Prairie Using Seasonal NDVI Trajectories and a Discriminant Function Mixture Model". *Remote Sensing of Environment*, Vol. 61, pp. 270-278.

Groom, G., H. Mikkelsen, H. Skriver, B. Hansen, A. Thompson, and A.A. Nielsen. (1997). "Activities and Early Results from DANMAC: The DANish Multisensor Airborne Campaign", Proceedings *Third International Airborne Remote Sensing Conference and Exhibition*, 7-10 July 1997, Copenhagen, Denmark, Vol. 2, pp. 617-24.

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

Jacobsen, A., N.H. Broge, and B.U. Hansen. (1995)."Monitoring Wheat Fields and Grasslands Using Spectral Reflectance Data". Proceedings *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.

Jacobsen, A., N. Drewes, M. Stjernholm, and T. Balstrøm. (1999b). "Generation of a Digital Elevation Model of Mols Bjerge, Denmark, and Georeferencing of Airborne Scanner Data", *Danish Journal of Geography*, Vol. 99, pp. 35-46.

Jacobsen, A., K.B. Heidebrecht, and A.F. Goetz. (2000). "Assessing the Quality of the Radiometric and Spectral Calibration of *casi* data and Retrieval of Surface Reflectance Factors". *Photogrammetric Engineering and Remote Sensing*, Vol. 66, no. 9, pp. 1083-0191.

Jacobsen, A., K.B. Heidebrecht, and A.A. Nielsen. (1998). "Monitoring Grasslands Using Convex Geometry and Partial Unmixing – A Case Study", In M. Shaepman, D. Schläpfer, and K. Itten (eds.) *Proceedings of 1st EARSel Workshop on Imaging Spectroscopy*, Remote Sensing Laboratories, University of Zürich, Switzerland, 6-8 October 1998, pp. 309-316.

Jacobsen, A., A.A. Nielsen, R. Ejrnæs, and G. Groom. (1999a). "Spectral Identification of Danish Grassland Classes Related to Management and Plant Species Composition", *Proceedings of 4th International Airborne Remote Sensing Conference and Exhibition*, Ottawa, Canada, 21-24 June, Vol. 1, pp. 74-81.

Larsen, R., A.A. Nielsen, and H. Flesche. (1999). "Sensitivity Study of a Semi-Automatic Supervised Classifier Applied to Minerals from X-ray Mapping Images. In B.K. Ersbøll and P. Johansen (eds.) *Proceedings of the Scandinavian Image Analysis Conference* (SCIA'99), Kangerlussuaq, Greenland, 7-11 June 1999, Vol. 2, pp. 785-792.

Lauver, C.L. and J.L. Whistler. (1993). "A Hierarchical Classification of Landsat TM Imagery to Identify Natural Grassland Areas and Rare Species Habitat". *Photogrammetric Engineering and Remote Sensing*, Vol. 59, pp. 627-634.

Lewis, M.M. (1994). "Species Composition Related to Spectral Classification in an Australian Spinifex Hummock Grassland", *International Journal of Remote Sensing*, Vol. 15, pp. 3223-39.

Matusita, K. (1966). "A Distance and Related Statistics in Multivariate Analysis." In P. R. Krishnaiah (ed.), *Multivariate Analysis*. Academic Press, New York, pp. 187-200.

McCune, B. and M.J. Mefford. (1997). *PC-ORD for Windows. Multivariate Analysis of Ecological Data*. Version 3.04. MjM Software, Gleneden Beach, Oregon, U.S.A.

Mino, N., Saito, G., and Ogawa, S. (1998). "Satellite Monitoring of Changes in Improved Grassland Management", *International Journal of Remote Sensing*, Vol. 19, No. 3, pp. 439-452.

Nielsen, A.A., H. Flesche, R. Larsen, J.M. Rykkje, and M. Ramm. (1998). "Semi-Automatic Supervised Classification of Minerals from X-ray Mapping Images", In A. Buccianti, G. Nardi, and R. Potenza (eds.) *Proceedings of the Fourth Annual Conference of the International Association for Mathematical Geology* (IAMG), Ischia, Naples, Italy, 6-8 October 1998, pp. 473-478.

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

Richards, J.A. (1993). *Remote Sensing Digital Image Analysis*. 2nd edition. Springer Verlag Berlin Heidelberg.

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

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

**383**