Article

Classification of Herbaceous Vegetation Using Airborne Hyperspectral Imagery

Péter Burai [1],*, Balázs Deák [2], Orsolya Valkó [2] and Tamás Tomor [1]

[1] Research Institute of Remote Sensing and Rural Development, Karoly Robert College,
   H-3200 Gyöngyös, Mátrai út 36, Hungary; E-Mail: tomor@karolyrobert.hu

[2] MTA-DE Biodiversity and Ecosystem Services Research Group, P.O. Box 71,
   H-4010 Debrecen, Hungary; E-Mails: debalazs@gmail.com (B.D.); valkoorsi@gmail.com (O.V.)

* Author to whom correspondence should be addressed; E-Mail: pburai@gmail.com;
   Tel.: +36-37-518-494; Fax: +36-37-518-297.

Abstract: Alkali landscapes hold an extremely fine-scale mosaic of several vegetation types, thus it seems challenging to separate these classes by remote sensing. Our aim was to test the applicability of different image classification methods of hyperspectral data in this complex situation. To reach the highest classification accuracy, we tested traditional image classifiers (maximum likelihood classifier—MLC), machine learning algorithms (support vector machine—SVM, random forest—RF) and feature extraction (minimum noise fraction (MNF)-transformation) on training datasets of different sizes. Digital images were acquired from an AISA EAGLE II hyperspectral sensor of 128 contiguous bands (400–1000 nm), a spectral sampling of 5 nm bandwidth and a ground pixel size of 1 m. For the classification, we established twenty vegetation classes based on the dominant species, canopy height, and total vegetation cover. Image classification was applied to the original and MNF (minimum noise fraction) transformed dataset with various training sample sizes between 10 and 30 pixels. In order to select the optimal number of the transformed features, we applied SVM, RF and MLC classification to 2–15 MNF transformed bands. In the case of the original bands, SVM and RF classifiers provided high accuracy irrespective of the number of the training pixels. We found that SVM and RF produced the best accuracy when using the first nine MNF transformed bands; involving further features did not increase classification accuracy. SVM and RF provided high accuracies with the transformed bands, especially in the case of the

aggregated groups. Even MLC provided high accuracy with 30 training pixels (80.78%), but the use of a smaller training dataset (10 training pixels) significantly reduced the accuracy of classification (52.56%). Our results suggest that in alkali landscapes, the application of SVM is a feasible solution, as it provided the highest accuracies compared to RF and MLC. SVM was not sensitive in the training sample size, which makes it an adequate tool when only a limited number of training pixels are available for some classes.

1. Introduction

Ecosystem functions and services are among the key factors of sustainable development on Earth, since ecosystems support both the survival and the quality of human life. Scientists, policy makers and stakeholders are all increasingly interested in the assessment of ecosystems, which is the basis for predicting future scenarios of landscape-scale changes; thus, it contributes to the planning of resource management and sustainable land use [1]. Habitat mapping is an essential tool of the nature conservation inventory, which provides data on the locality and distribution of habitats on a landscape scale. Habitat maps support the landscape-level planning of nature conservation actions, biodiversity monitoring, and scientific purposes [2]. Given these multiple functions, the need for high-precision large-scale habitat maps has been rapidly increasing all over Europe.

Remote sensing techniques offer a viable solution for mapping extended, complex and hardly accessible areas [3]. This type of habitat mapping is based on remotely sensed data such as multispectral images [4], airborne hyperspectral imagery [5], light detection and ranging (LiDAR) [6,7], radar [8] and in some cases even a combination of these [9,10]. Nowadays, hyperspectral sensors have a high potential for monitoring of the environment [11–14]. Airborne hyperspectral imagery can produce multiple narrow and contiguous spectral bands of less than 10 nm with a high geometric resolution (0.5–1 m). Hyperspectral imagery can be a suitable method for a detailed vegetation classification based on the dominant or subdominant genera or species [15–18]. In general, for processing hyperspectral data it is essential to reduce the high dimensionality and inherent multi-collinearity of a dataset. Several advanced feature extraction techniques (e.g., MNF, PCA, ICA and DBFE) have been developed for this purpose [13,19,20].

For testing the potential of remote sensing in mapping complex and extended areas, open alkali landscapes provide an excellent possibility. Alkali landscapes of the Pannonian Basin are one of the most extended semi-natural open landscapes of the European Union. They are characterized by a fine-scale mosaic of different vegetation types, including open alkali vegetation, dry steppic grasslands, tall-grass meadows, and sedge vegetation together with alkali and non-alkali marshes [21,22]. Alkali landscapes hold an extremely fine-scale mosaic of several vegetation types with similar characteristics regarding their biomass, vegetation structure and environmental conditions, thus it seems challenging to separate these classes by remote sensing. Our aim was to test the applicability of different image classification

methods of hyperspectral data in these complex conditions. To reach the highest classification accuracy, we tested traditional image classifiers *versus* the machine learning algorithm, feature extraction (MNF-transformation) and different sizes of training dataset.

We asked the following questions: (i) Does the use of feature extraction (MNF-transformation) increase classification accuracies? (ii) Does the application of traditional image classifier (maximum likelihood classifier) or machine learning algorithm (support vector machine, random forest) result in higher classification accuracies? (iii) How does the size of training dataset affect classification accuracies?

2. Methodology

2.1. Description of the Study Site

Our study site is situated in Pentezug-puszta (N47°34′, E21°6′) which is part of Hortobágy National Park (Eastern Hungary). The area is characterized by continental climate, the mean annual temperature is 9.5 °C, while the mean annual precipitation is 550 mm [23]. The total area of the study site is 23.49 km^2 (Figure 1). Pentezug-puszta holds a diverse complex of alkali vegetation: steppes, meadows, and marshes as well as the narrow floodplain and the neighboring riparian forests of the Hortobágy River. Roads, buildings, and woody vegetation were excluded from image classification using the digital database of the Hortobágy National Park Directorate.

2.2. Airborne Data Collection

We applied an AISA EAGLE II type hyperspectral sensor, which produced images with 128 contiguous bands (400–1000 nm), a spectral sampling of 5 nm bandwidth, and a ground pixel size of 1 m. The sensor was mounted to a Piper Aztec aircraft. Data acquisition took place in good weather conditions from 09:11 to 09:53 GMT, on 7 July 2013. OxTS RT 3003 GPS/INS system was used to record the navigation data.

2.3. Field Data Collection

We collected field reference data from all representative vegetation types of the study site in a one-week interval after the flight. Before field data collection, we made a preliminary survey and listed the typical vegetation types and estimated their share in the study area. In alkali landscapes the total area and distribution of different vegetation types and the average size of their stands generally show a high variability. While dry steppic grasslands, sedges, meadows, and marshes are usually well-represented with extended patches, open alkali grasslands are generally rare and are represented with smaller patches [24]. For representing patches of rare vegetation types as well, we decided to make a representative sampling based on preliminary field surveys.

During the field survey we marked—by measuring their corner points—altogether 98 homogenous vegetation patches with a Leica Viva GS15 GNSS (Figure 1). In each patch we recorded the list and percentage cover scores of vascular plant species with a cover above 10% and the mean canopy height and total cover of the vegetation. For the calculations we classified the species as dominant (>50%) and subdominant species (10%–50%) based on their relative cover.
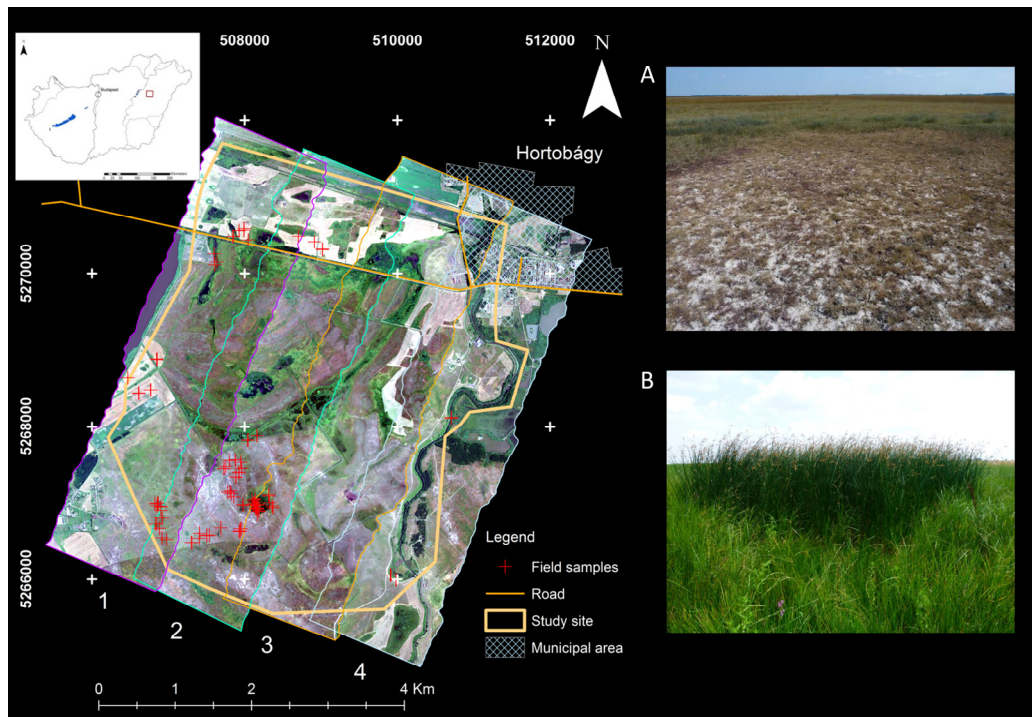
Figure 1. Overview map of the study site (Pentezug-puszta). Projection of the image is WGS 84 UTM 34 North. Red cross marks indicate the positions of field measurement plots in the RGB mosaic of the hyperspectral image. Typical alkali vegetation of study site: (**A**) An open alkali grassland with a dry steppic grassland in the background and (**B**) patch of a *Schoenoplectus* marsh surrounded by *Bolboschoenus* marsh.

2.4. Vegetation Classes

In the present study our aim was to classify the herbaceous vegetation (grasslands and wetlands) as these vegetation types were best represented (woody vegetation covered only ~0.5% of the total area) and have the highest importance for nature conservation and site managers. For the classification we built up twenty vegetation classes (Table 1) based on the dominant species of the vegetation, their canopy height and total vegetation cover. To fine-tune the classification, in some cases we used the subdominant species as well. The vegetation classes were assigned to the following vegetation groups: dry steppic grasslands (CYN, FAC and FAR), open alkali grasslands (CAM, PHO and ART), meadows and sedge vegetation (ELY, ALO, BEC, ACI and CAR) and marshes and reeds (GLY, TYP, SAL, BOL, SCH and PHR). The vegetation classes are in line with the phytosociological categories [23] but are easier to interpret and more applicable to site managers.

The CYN, FAC and FAR classes are dry steppic grasslands characterized by short grass species (*Cynodon dactylon* and *Festuca pseudovina*) and they also harbor some forb species, such as *Achillea collina* or *Artemisia santonica*. They have a high vegetation cover but low biomass [25,26].

The CAM, PHO and ART comprise the open alkali grasslands mainly composed of halophyte and stress-tolerant species (*Camphorosma annua, Pholiurus pannonicus* and *Artemisia santonica*). These classes are characterized by low vegetation cover and a high cover of bare ground, which is often covered by salt [26,27].

Table 1. Attributes of the studied vegetation classes. Dominant, subdominant species, mean canopy height, and mean total coverage of vegetation (*vs.* open soil surfaces) are based on field measurements. * classified by land use type; ** classified by the high cover of bare muddy soil. One square meter corresponds to one pixel.

Abbreviation	Dominant Species	Subdominant Species	Canopy Height (cm)	Total Coverage of Vegetation (%)	Measured Area (m²)
CYN	*Cynodon dactylon*	*Achillea collina*	21.2	96.2	211
FAC	*Festuca pseudovina*	*Achillea collina*	3.0	80.0	141
FAR	*Festuca pseudovina*	*Artemisia santonica*	28.3	80.8	96
CAM	*Camphorosma annua*	-	4.4	28.0	118
PHO	*Pholiurus pannonicus*	-	18.6	47.0	142
ART	*Artemisia santonica*	*Pholiurus pannonicus*	13.7	43.7	64
ELY	*Elymus repens*	-	96.0	64.0	402
ALO	*Alopecurus pratensis*	*Agrostis stolonifera*	48.3	93.3	531
BEC	*Beckmannia eruciformis*	*Agrostis stolonifera, Cirsium brachycephalum*	87.5	91.2	552
ACI	*Alopecurus pratensis*	*Cirsium arvense Elymus repens*	140.0	85.0	82
CAR	*Carex* spp.	-	100.0	90.0	253
GLY	*Glyceria maxima*	-	40.0	90.0	229
TYP	*Typha angustifolia*	*Salvinia natans*	200.0	70.0	63
SAL	*Salvinia natans*	*Typha angustifolia, Utricularia vulgaris*	133.0	70.0	65
BOL	*Bolboschoenus maritimus*	-	76.2	78.8	179
SCH	*Schoenoplectus lacustris* ssp. *tabernaemontani*	-	166.0	87.0	121
PHR	*Phragmites communis*	-	250.0	100.0	297
FMM *	*Alopecurus pratensis*	-	10.0	80.0	351
ARA *	*Gypsophyla muralis, Polygonum aviculare*	-	8.0	80.0	123
MUD **	not relevant	-	10.0	8.0	158

Meadows (ELY, ALO and BEC) are typical in sites with moist and moderately salty soil. They are tall grasslands with a medium amount of biomass, characterized by the high cover of a few grass species, such as *Agrostis stolonifera, Alopecurus pratensis, Beckmannia eruciformis and Elymus repens* [26,28]. In some parts of the study area, we detected meadow vegetation with weedy species, such as *Cirisum arvense*; we assigned them to the ACI class. The CAR class represents the sedge vegetation in our study sites. Its typical species are *Carex* species, which compose a very dense vegetation cover. This class was present under moist soil conditions [23].

Marshes and reed beds (GLY, TYP, SAL, BOL, SCH and PHR) were situated in the wet depressions of the study site, their canopy height can reach 4 m, and they are characterized by high biomass [29]. We divided the *Typha angustifolia* dominated vegetation into two classes based on the

cover of the subordinate species. While the TYP class was characterized by the monodominance of *Typha*, in the SAL class *Salvinia natans* had a high cover as well.

Patches of open muddy surface (class MUD) were present in the edge zone of the wetlands. They are characterized by a high cover of bare soil surface and a very low vegetation cover. The FMM class comprised mown meadows. They are characterized by a low vegetation height and low biomass. The ARA class consisted of fallow arable lands, characterized by weedy species.

2.5. Image Processing

ENVI/IDL 5.0 (Exelis, Inc., Boulder, CO, USA) and CaliGeoPro (Spectral Imaging Ltd., Oulu, Finland) softwares were used to calculate radiometric and geometric corrections of hyperspectral images. ENVI FLAASH and the empirical line method were applied for radiometric and atmospheric corrections to reflectance. During the flight there was a 30 min time lag between Stripe 1 and Stripe 2, which was the likely reason for the remaining radiometric line shift. In order to explore the information content of the hyperspectral datasets considered, Minimum Noise Fraction (MNF) were calculated. MNF transformation [19] was applied to achieve noise reduction, and the new artificial channels having the largest explained variance were used in the classifications. This method projects data into a subspace where the signal-to-noise ratio is maximized.

2.6. Separating the Classes Using Narrow Band NDVI

Normalized Difference Vegetation Index (NDVI) is the most widely used vegetation index in recent decades. NDVI can provide relevant information on the distribution of plant species, plant growth patterns, and plant physiological status [30]. In our study, we calculated narrow-band NDVI values using the selected red (679 nm) and near-infrared (NIR) bands (800 nm) [31] in order to test the separability of classes based on only NDVI scores. Fifty pixels were selected from each class from the whole training dataset by randomization.

2.7. Image Classification

2.7.1. Applied Classification Methods

ENVI/IDL 5.0 (Exelis, Inc., Boulder, CO, USA) and EnMap Box [32,33] softwares were used to classify hyperspectral images. Three supervised classification methods (MLC, RF and SVM) were selected as classifiers, since their efficiency was proven in vegetation mapping in recent studies [34,35].

Maximum likelihood classification (MLC) is one of the most commonly used image classification methods. In MLC the pixel is labeled to the classes with the highest probability. The algorithm relies on the statistic of Gaussian probability density function model [36,37]. The MLC method cannot be used directly when the number of training samples is smaller than the number of features. Applicability of traditional classification methods like the MLC or nearest neighbor can be limited when using a small number of training samples and high-dimensional feature space of hyperspectral dataset [20]. Despite this limitation, MLC generally produces similar or better classification accuracy than other classifiers [38]. In our study, for MLC classification, thresholds for the probability were not specified.

Recently, the random forest (RF) algorithm which was developed by Breiman [39] has been successfully used as a variable selection and classification algorithm for hyperspectral data [40]. Random Forest algorithm creates individual decision trees whose diversity is ensured by the use of random samples derived from the training dataset [39]. In our study, for RF classification 100 trees were computed; the minimum number of samples in a node was 1. In all classifications with RF, we used Gini coefficient for the node impurity function.

The Support Vector Machine (SVM) is a high-complexity classifier that has been widely used for classification of hyperspectral images [41,42]. In the current literature of hyperspectral remote sensing, SVM generally outperforms other classification methods [10,43]. In our study, SVM classification was performed with Gaussian Radial Basis Function kernel. SVM parameters ($C = 100$ and $\gamma = 0.11$) were selected by fivefold cross validation.

2.7.2. Image Classification Using Original Spectral Bands

SVM and RF methods were used for supervised classification, using the original dataset. We could not use MLC on original bands, as it needs at least $n + 1$ training data per class. To test the effects of pixel numbers on overall accuracy, image classifications were repeated on reduced training datasets, which consisted 10, 15, 20, 25 and 30 pixels from each vegetation class, respectively. Image classification was repeated five times using random sampling methods. The maximum size of randomly selected training pixels was 30, because some classes were poorly represented in the field, thus in these cases the number of pixels were limited.

After testing the effect of pixel numbers on overall accuracies, we applied SVM and RF classification methods using 10 and 30 randomly selected training pixels for each vegetation class. To test the classification accuracies (Producer's, User's and Overall Accuracy) of the applied algorithms, a confusion matrix was computed [44].

We used four categories—corresponding to the measured number of field samples (pixels)—for computing validation datasets (Table 2). Our concept was to have at least a 1:1 ratio of field samples and validation dataset. Pixels were selected using random sampling method from field samples. The validation dataset was not used for training. We used the same validation dataset for all analyses.

Table 2. Validation dataset (number of pixels).

Field Samples (Pixel)	Random Samples (Pixel)
60–80	30
81–100	40
101–200	50
201–600	100

2.7.3. Image Classification Using MNF-Transformed Bands

In order to select the optimal number of features, SVM, RF and MLC classifications were applied on MNF-transformed bands, the number of bands varied from 2 up to 15. To test the effect of number of pixels on overall accuracy using MNF-transformed bands, image classifications were repeated on five reduced training datasets. We applied SVM, RF and MLC classification, using 9 MNF-transformed

bands with 10 to 30 (10, 15, 20, 25 and 30) random training pixels from each vegetation class respectively. Image classification was repeated five times using random sampling methods. To test the classification accuracies (Producer's, User's and Overall Accuracy) of the applied algorithms, a confusion matrix was computed.

3. Results

3.1. Separating the Classes Using Narrow Band NDVI

In the scatter plot (Figure 2) it is apparent that even though some classes (MUD and SAL) are well-separated by the Red and Near Infrared bands, most of the classes overlap, which does not allow for an accurate classification based on the NDVI scores solely. On the other hand, it is obvious that the classes of the same vegetation groups are clustered together. In some cases even the groups with similar attributes (mostly living biomass and soil moisture) were positioned next to each other, for example, open alkali grasslands were situated next to sparsely vegetated areas and meadows next to marshes.
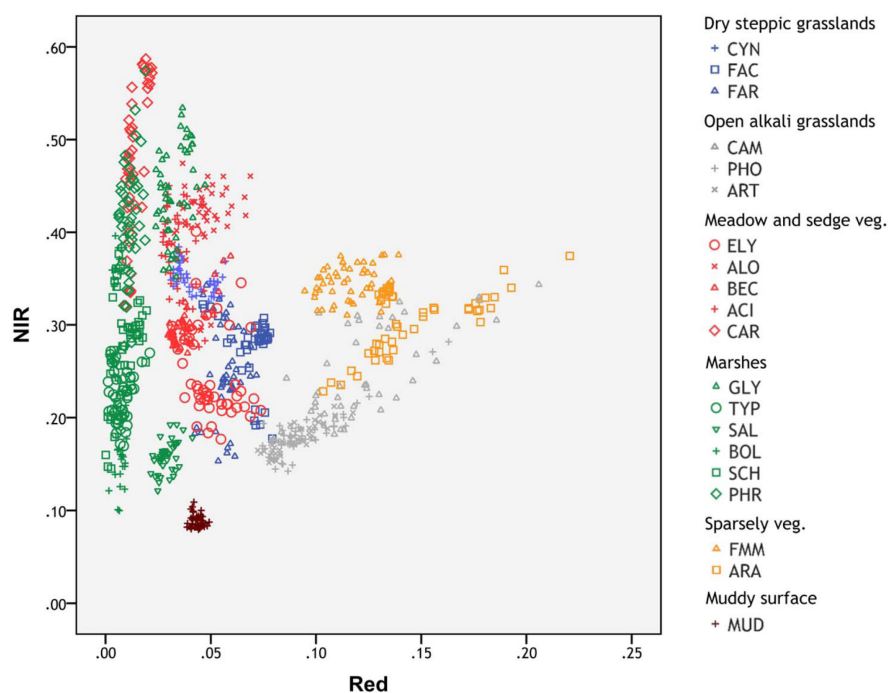

Figure 2. Scatter plot of the randomly selected data (50 pixels from each vegetation class) in two bands: Red (679 nm) and NIR (800 nm). Identical vegetation groups are represented by the same color.

NDVI scores of the vegetation classes showed a gradient from the vegetation classes characterized by low biomass and low vegetation cover to the classes with high biomass and closed vegetation (Figure 3). Accordingly, mown meadows (FMM), harvested arable lands (ARA), muddy surfaces (MUD) and the open alkali grasslands (CAM, PHO and ART) had low NDVI scores. Dry steppic grassland vegetation (FAC, FAR and CYN) with moderate biomass and closed vegetation was situated

in the middle. Meadows (ELY, BEC, ALO and ACI), sedge vegetation (CAR) and marshes (SAL, GLY, TYP, BOL, SCH and PHR)—characterized by high biomass and closed vegetation—had the highest NDVI scores.

Accurate separation of classes using only the NDVI scores was only possible for the FAC, SAL, ACI and GLY classes (Figure 3). Classes within the vegetation group of open alkali grasslands (CAM, PHO and ART) had similar NDVI scores, thus their separation was not possible. We detected significant differences between the CAM and ART classes, but it was not possible to separate the CAM and PHO classes from the ARA class. The MUD and FMM classes, both characterized by low biomass, could not be separated. The FAR and ELY classes could not be separated, either. CYN class (assigned to the dry steppic grasslands) could not be separated from two meadow classes (ALO and BEC). We detected an overlap between three marsh classes (BOL, SCH and TYP) and between two wetland classes (PHR and SCH).
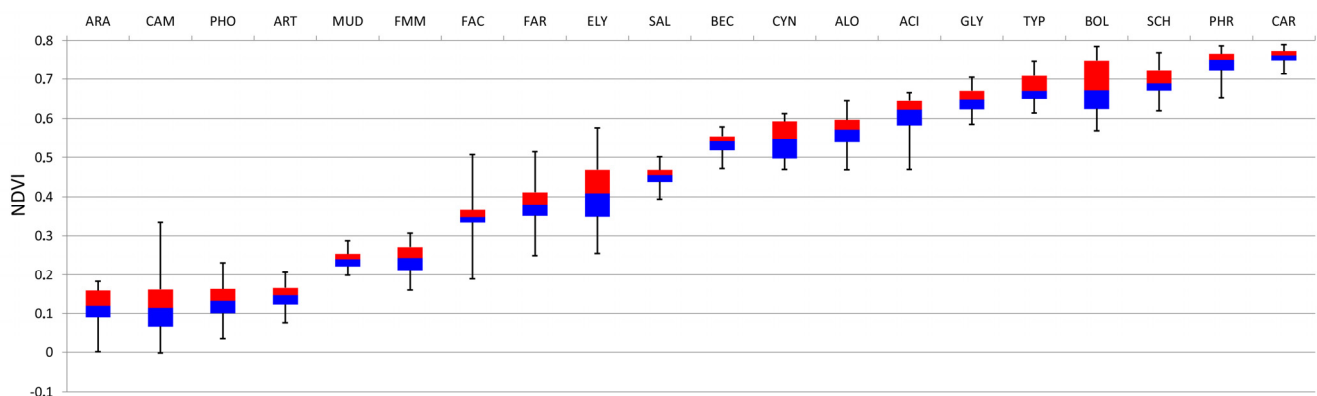
Figure 3. Boxplot of NDVI scores of random samples (50 pixels from each vegetation class).

3.2. Image Classification Using Original Spectral Bands

The overall accuracies provided by both SVM and RF classifiers increased slightly when increasing the number of training pixels (Figures 4 and 5; Table 4). We found that in case of both classifiers, the standard errors of overall accuracies were similarly low independently of the number of training pixels.

Both classifiers had the highest overall accuracies using original bands with 30 training pixels (SVM: 72.84%; RF: 72.89%). Classification using 10 training pixels decreased the overall accuracy for both the RF (OA = 70.95%) and the SVM (OA = 70.94%) classifiers (Figure 5). In case of SVM classification using 30 training pixels we got high classification accuracies for several classes: CYN, FAR, ELY, ALO, BEC, CAR, GLY, SAL, FMM, ARA and MUD. In some cases, the accuracy was low, like in the classes of PHO, TYP, and BOL (Figure 5).

The confusion matrix showed that misclassified pixels were usually distributed in the correct vegetation group (e.g., within dry steppic grasslands, open alkali grasslands, meadows and sedge or marshes and reeds; see Table 3). We found that 76 pixels (6.55%) were misclassified to another vegetation group. For example, only 13 pixels of the PHO class were classified correctly; other pixels assigned to the PHO class in the field were classified as CAM and ART classes (all within the vegetation group of open alkali grasslands), which are also characterized by low vegetation cover and accordingly very low NDVI scores (Figure 3). Several classes were misclassified within the vegetation

group of marshes. For example, TYP and BOL classes are characterized by different dominant species, but similar biomass scores and wet soil, which is the likely reason for their misclassification.
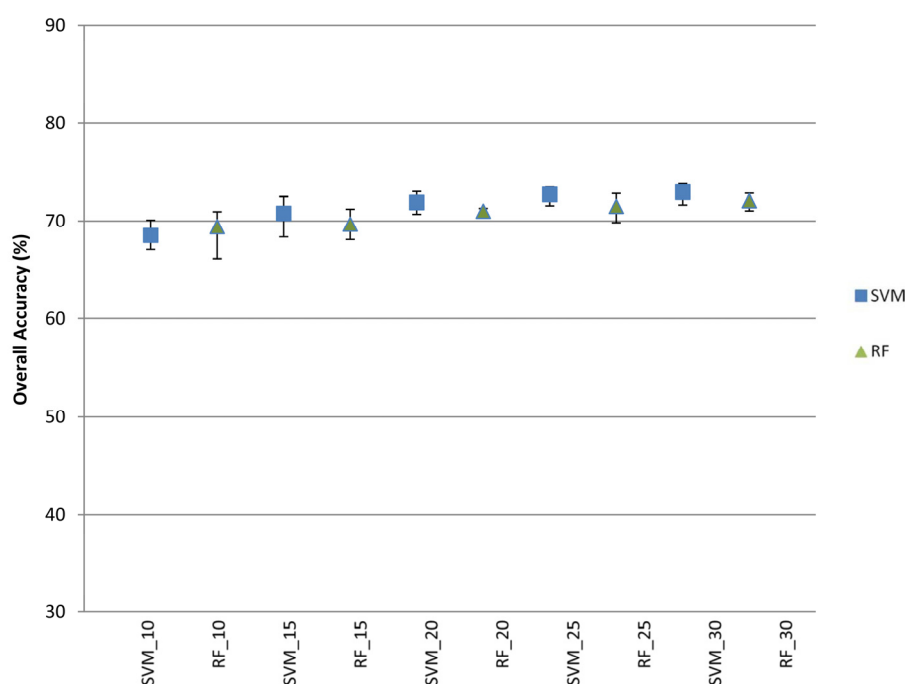
Figure 4. Overall accuracies of random forest (RF) and the Support Vector Machine (SVM) classifiers using original bands and different number of random training pixels (N = 10; 15; 20; 25 or 30) from each vegetation class (mean ± SD).
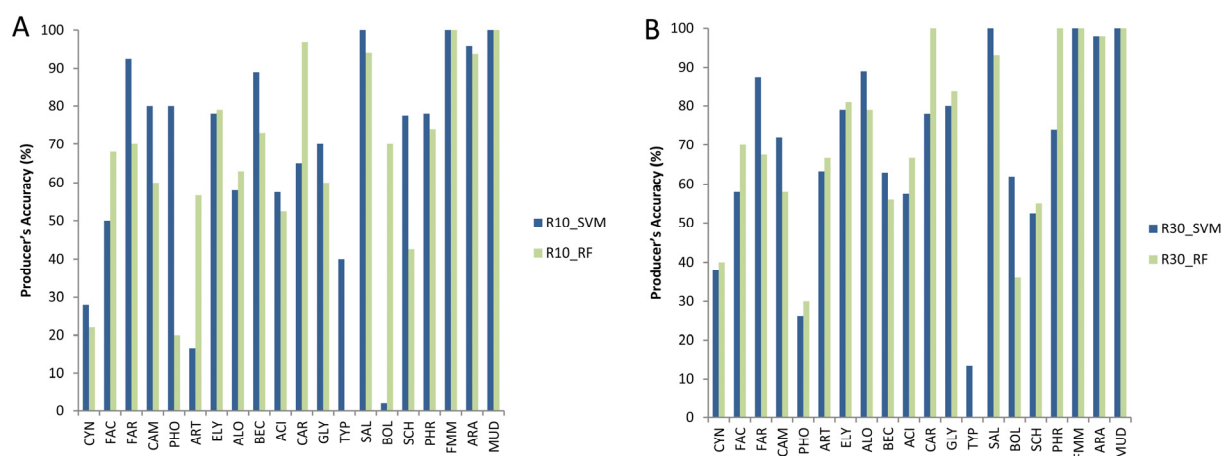
Figure 5. Producer's accuracy (%) of the classes of SVM and RF classifiers using original bands and 10 (**A**) and 30 (**B**) random training pixels.
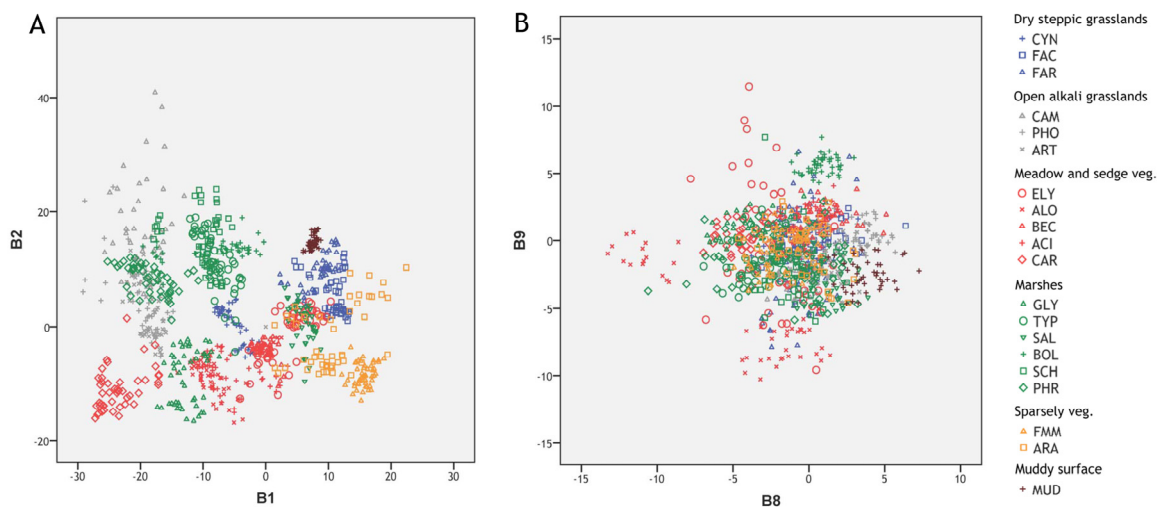
Table 3. Confusion matrix of SVM classification using original bands with 30 training pixels. Identical groups are represented by the same color (cyan—dry steppic grasslands; gray—open alkali grasslands; red—meadow and sedge vegetation; green—marshes; orange—sparsely vegetated areas; brown—muddy surface). Notations: UA: User's Accuracy; PA: Producer's Accuracy.

| Class	CYN	FAC	FAR	CAM	PHO	ART	ELY	ALO	BEC	ACI	CAR	GLY	TYP	SAL	BOL	SCH	PHR	FMM	ARA	MUD	Total
CYN	19	0	0	0	0	0	8	0	7	0	0	0	0	0	0	0	0	0	0	0	34
FAC	0	29	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	34
FAR	0	21	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	56
CAM	0	0	0	36	8	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	46
PHO	0	0	0	9	13	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32
ART	0	0	0	5	29	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	53
ELY	0	0	0	0	0	0	79	0	30	7	0	0	0	0	0	0	0	0	0	0	116
ALO	0	0	0	0	0	0	4	89	0	0	0	2	0	0	0	0	0	0	0	0	95
BEC	3	0	0	0	0	0	7	0	63	9	3	0	0	0	0	0	0	0	0	0	85
ACI	28	0	0	0	0	0	1	11	0	23	5	5	0	0	0	0	1	0	0	0	74
CAR	0	0	0	0	0	0	0	0	0	0	78	3	0	0	0	0	2	0	0	0	83
GLY	0	0	0	0	0	0	1	0	0	1	12	40	0	0	0	0	0	0	0	0	54
TYP	0	0	0	0	0	0	0	0	0	0	1	0	4	0	7	3	4	0	0	0	19
SAL	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	30
BOL	0	0	0	0	0	0	0	0	0	0	0	0	26	0	31	16	3	0	0	0	76
SCH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	21	3	0	0	0	33
PHR	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3	0	37	0	0	0	41
FMM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	100
ARA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	0	49
MUD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	50
Total	50	50	40	50	50	30	100	100	100	40	100	50	30	30	50	40	50	100	50	50	1160
PA (%)	38.0	58.0	87.5	72.0	26.0	63.3	79.0	89.0	63.0	57.5	78.0	80.0	13.3	100.0	62.0	52.5	74.0	100.0	98.0	100.0	
UA (%)	55.9	85.3	62.5	78.3	40.6	35.8	68.1	93.7	74.1	31.1	94.0	74.1	9.3	100.0	40.8	63.6	90.2	100.0	100.0	100.0	

3.3. Image Classification Using MNF-Transformed Bands

Distribution of pixels using the first two MNF-transformed bands showed a good feasibility for separation of classes (Figure 6). However vegetation classes could not be separated using the eighth and ninth MNF-transformed bands.

We found the highest classification accuracies for SVM and RF classifier using the MNF 1–9 transformed bands (overall accuracies: 82.06%; for SVM and 79.14% for RF; Figure 7). The classification results indicated that additional features over the first 9 MNF bands did not significantly improve the accuracy. In case of MLC, we found that additional features over the first 5 MNF bands did not improve the accuracy.


Figure 6. Scatterplot of the selected data in two MNF-transformed bands (**A**) B1 and B2; (**B**) B8 and B9. Identical vegetation groups are represented by the same color (cyan—dry steppic grasslands; gray—open alkali grasslands; red—meadow and sedge vegetation; green—marshes; orange—sparsely vegetated areas; brown—muddy surface).
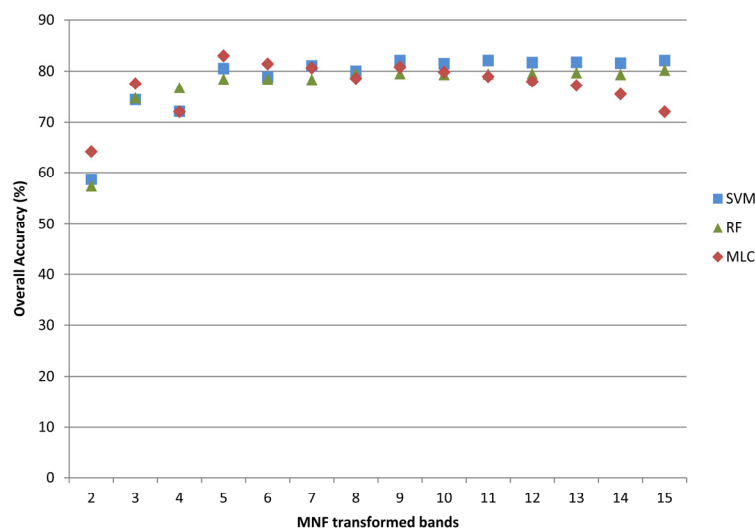
Figure 7. Overall accuracy (%) of classified image using SVM, RF and MLC with various MNF-transformed bands using 30 randomly selected training pixels.

Our results showed that SVM and RF classifiers provided high accuracies even with a low number of training pixels. MLC classifier provided low classification accuracies (with high standard error) when using less than 20 training pixels. MLC classifier provided accuracies comparable to SVM and RF only when at least 20 training pixels were used (Figure 8).
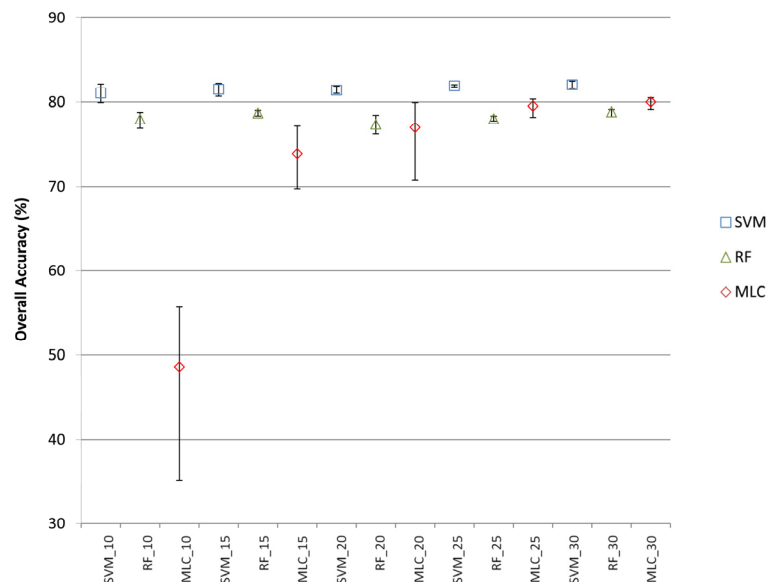
Figure 8. Overall accuracies of SVM, RF and MLC classifier using nine MNF-transformed bands and different number of random training pixels (*N* = 10; 15; 20; 25 or 30) from each vegetation class.

Each classifier provided similar high overall accuracies when using 30 random training pixels (SVM: 82.06%; RF: 79.14%; MLC: 80.78%) (Figure 9). When using 10 random training pixels SVM and RF classifiers provided considerably high overall accuracies (79.57% and 76.55%, respectively), but the accuracy of the MLC classifier decreased considerably (52.56%).

All classifiers provided high accuracies for the classes ELY, ALO, BEC, CAR, GLY, SAL, FMM, ARA and MUD, when using 30 random training pixels. We found that in case of classes CYN, FAR, ACI and TYP the three classifiers provided considerably different accuracies (Figure 9).

Similarly to the results of SVM classifier using original bands, we found that the majority of misclassified pixels were distributed within the correct vegetation group in the confusion matrix (Table 4). We found that only 15 pixels (1.29%) were misclassified to another vegetation group.

Table 4. Confusion matrix of SVM classification using 9 MNF-transformed bands with 30 training pixels. Identical groups are represented by the same color (cyan—dry steppic grasslands; gray—open alkali grasslands; red—meadow and sedge vegetation; green—marshes; orange—sparsely vegetated areas; brown—muddy surface). Notations: UA: User's Accuracy; PA: Producer's Accuracy.

Class	CYN	FAC	FAR	CAM	PHO	ART	ELY	ALO	BEC	ACI	CAR	GLY	TYP	SAL	BOL	SCH	PHR	FMM	ARA	MUD	Total
CYN	42	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	43
FAC	0	25	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27
FAR	0	25	38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	63
CAM	0	0	0	36	8	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45
PHO	0	0	0	9	13	10	0	0	0	0	0	0	0	0	0	0	0	0	1	0	33
ART	0	0	0	5	29	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	53
ELY	0	0	0	0	0	0	95	0	1	7	0	0	0	0	0	0	0	0	0	0	103
ALO	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	99
BEC	1	0	0	0	0	0	0	0	99	9	0	0	0	0	0	0	0	0	0	0	109
ACI	7	0	0	0	0	0	5	0	0	23	0	0	0	0	0	0	0	1	0	0	36
CAR	0	0	0	0	0	0	0	0	0	0	97	0	0	0	0	0	0	0	0	0	97
GLY	0	0	0	0	0	0	0	0	0	1	3	50	0	0	0	0	6	0	0	0	60
TYP	0	0	0	0	0	0	0	0	0	0	0	0	3	0	23	2	5	0	0	0	33
SAL	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	30
BOL	0	0	0	0	0	0	0	0	0	0	0	0	27	0	21	9	4	0	0	0	61
SCH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	29	0	0	0	0	35
PHR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	34	0	0	0	34
FMM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	100
ARA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	0	49
MUD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	50
Total	50	50	40	50	50	30	100	100	100	40	100	50	30	30	50	40	50	100	50	50	1160
PA (%)	84.0	50.0	95.0	72.0	26.0	63.3	95.0	99.0	99.0	57.5	97.0	100.0	10.0	100.0	42.0	72.5	68.0	100.0	98.0	100.0	
UA (%)	97.0	92.6	60.3	80.0	39.4	35.8	92.2	100.0	90.8	63.9	100.0	83.3	9.1	100.0	34.4	82.9	100.0	100.0	100.0	100.0	
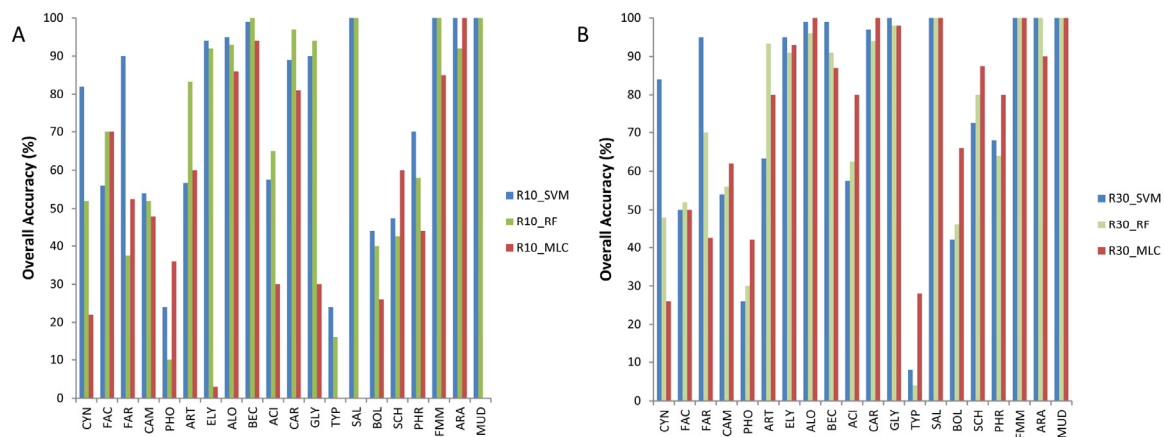
Figure 9. Producer's accuracy (%) of the classes using SVM, MLC and RF classifiers using 9 MNF-transformed bands and 10 (**A**) and 30 (**B**) random training pixels.

Classification accuracy of both SVM and RF classifiers increased with MNF-transformed bands compared to original bands, both for classes and for groups (Table 5).

Table 5. Classification accuracy of the vegetation classes and vegetation groups with respect to three classifiers using both original and nine MNF-transformed bands for 30 random training pixels.

	SVM		RF		MLC	
	Original Bands	MNF Bands	Original Bands	MNF Bands	Original Bands	MNF Bands
Overall accuracy of vegetation classes (%)	72.85	82.06	72.89	79.14	-	80.78
Overall accuracy of vegetation groups (%)	93.30	98.70	90.70	95.77	-	95.77

We produced a vegetation map of the study area using SVM classification with 30 random training pixels per class (Figure 10).

The total computational time is the sum of the classification time and the MNF transformation time. The MNF transformation for the AISA images (128 spectral bands) took 170 min. If the SVM or RF classification is applied on the original images, no MNF transformation is needed and the total time will only be the classification time (total computation times were 55 min for SVM and 7 min for RS). The total computational time essentially depends on the number of MNF bands used in the classification. The classification time is decreased when fewer bands are used in the classification. Calculation times for the three classifiers on the same dataset were approximately 16 min, 3 min and 8 min for SVM, RF and MLC, respectively (Processor: Intel Core i7—3740QM CPU @2.70 GHz; RAM: 16 GB, 64 bit).
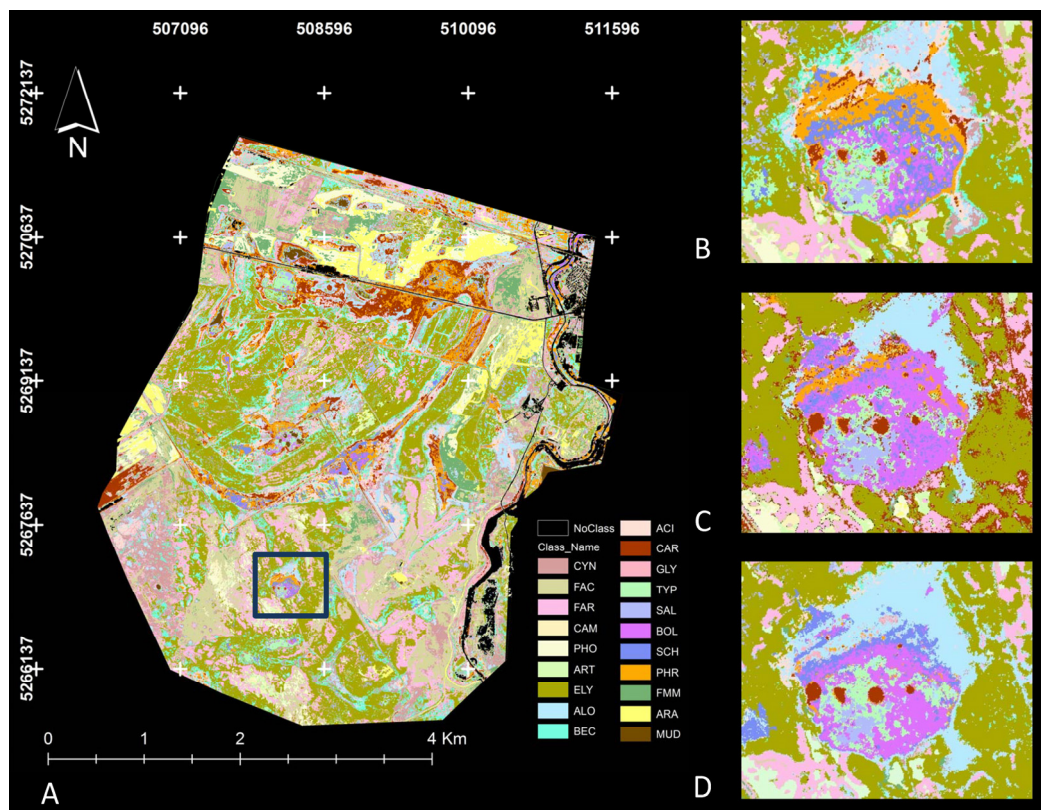
Figure 10. Vegetation map (**A**) of the study site produced using SVM classification with 30 random training pixels per class. Detailed maps produced by SVM (**B**), RF (**C**) and MLC (**D**) classification are provided in the subset image (right side).

4. Discussion

We detected a marked gradient in the NDVI scores of the studied classes (see Figure 3). Despite the fact that certain classes (FAC, SAL, ACI and GLY) could be well-separated, it was not possible to distinguish most of the classes based on the NDVI scores alone. The reason for that might be that classes within the same vegetation group are generally characterized by a similar structure (species composition, plant life form, biomass, environmental moisture and cover of bare soil surface, see [26]), which likely resulted in considerable overlaps between their NDVI scores.

For further testing, we used SVM and RF classifiers with original bands on reduced training samples between 10 and 30 pixels. Efficiency of both classifiers decreased when applying low numbers of training pixels (Figure 4). Other studies found that SVM is not sensitive to the limited number of training data (see also [13]). We found that both SVM and RF classifiers provided a high accuracy with 30 training pixels.

SVM classification using original bands resulted in a high Producer's accuracy for most of the vegetation classes, but we detected low accuracy in some cases (e.g. PHO, TYP and BOL) when most of the pixels were assigned to a class with similar biotic and environmental attributes (Figure 5). We found that most of the misclassified pixels were assigned to the correct vegetation group, and only 6.55% of the misclassified pixels were assigned to a different vegetation group (Table 3). Overall accuracies for the vegetation groups were above 90 % with all classifiers (see Table 5). Other studies also found that merging vegetation groups increases classification accuracies. [45], who classified the

vegetation of Belgian heathlands using CHRIS/Proba image with a spatial resolution of 36 m had an overall accuracy of 45.3% for the 22 vegetation classes; and merging them into 8 habitat classes resulted in a higher overall accuracy of 62.4%. Usually, vegetation classification comprises a number of classes that differ more subtly than the broader categories of vegetation groups (see also [46]). Thus, classes within vegetation groups (i) generally have dominant (and often also subdominant) species with similar height, biomass and metabolism and (ii) are characterized by similar environmental conditions (soil moisture, cover of open soil surface, amount of salt accumulation on the soil surface), which are the likely reasons for the resulted high classification accuracies.

In order to select the optimal number of transformed features, we applied the three classifiers on 2–15 MNF-transformed bands. We found that the SVM and RF algorithm produced the best accuracy performance using the first nine MNF-transformed bands. Involving further features did not increase the classification accuracy significantly, due to the Hughes phenomenon [20]. In case of MLC we found the best accuracy performance using the first five MNF-transformed bands which indicates that MLC is more sensitive to unstructured datasets compared to the other classifiers. However, the accuracy of MLC can be high in case of well-structured dataset and proper feature extraction methods.

Application of MNF-transformed bands increased the accuracies of the classification for the SVM and RF in general compared to classification using original bands (Table 5; see also [19]). While both SVM and RF classification using MNF-transformed bands resulted in constantly high accuracies irrespective of the number of the training pixels, results of the MLC were not so obvious. Even though MLC provided similarly high overall accuracy to those of SVM and RF in the case of 30 training pixels (80.78%), the use of a smaller training dataset (10 training pixels) significantly reduced the accuracy (52.76%) of the classification and increased its instability (higher standard errors when using smaller training dataset), thus the estimated covariance matrix of MLC became unstable.

Our results supported that MNF transformation increased classification stability both in the level of vegetation classes and vegetation groups. Although certain classes were misclassified with the SVM classifier, similarly when applying original bands, most of the misclassified pixels were assigned to another vegetation class within the same vegetation group. We found that the application of MNF-transformed bands further decreased the pixels assigned to a different vegetation group (1.29% *vs.* 6.55% when using original bands).

We found that SVM provided the highest accuracies both for the classes and vegetation groups compared to the RF and MLC when using MNF transformed bands. Besides the detected high classification accuracies with SVM classifier using MNF-transformed bands, in case of some classes (CAM and PHO) it provided the lowest accuracy among the three classifiers. These classes both belong to the vegetation group of open alkali grasslands, which are characterized by a high ratio of open soil surface which is likely responsible for the detected low classification accuracies of SVM.

5. Conclusions

Our aim was to evaluate the classification accuracy of three classifiers in a highly mosaic open landscape with a special emphasis on the applicability of the methods. We classified open vegetation characterized by grasslands and wetlands with three classifiers and we could separate 20 vegetation classes with a maximum overall accuracy of 82.06% (using SVM with MNF-transformed bands). An

important practical point can be that the computing process was the fastest in case of RF classifier which gives it an advantage over other classifiers (see also [13]). For the optimal performance of classifiers, in most of the cases it would be optimal to collect a huge amount of field data, but during the practical application of remote sensing techniques it is not always possible [13]. Collecting huge amounts of field data is labor- and time-consuming, especially in areas with difficult accessibility. In many cases, certain vegetation classes are sparsely distributed and are present in small fragments, thus it is impossible to collect high number of field samples from every class. Given the limitations of field data collection, robust classification methods can support vegetation classification in such landscapes. We found that SVM was not sensitive for the training sample size, which makes it an adequate tool when only a limited number of training pixels are available from some classes. Our results suggest that extended areas can be mapped using even a limited number of training pixels with robust image classification. These findings have a high potential for environmental monitoring and warrant further testing in other types of open landscapes or complex landscapes with mosaic herbaceous and woody vegetation.

Acknowledgments

Author Contributions

P.B, B.D. and T.T. designed the experiment. B.D. and V.O. performed the botanical sampling. P.B. and T.T. performed the analyses. All authors took part in manuscript preparation.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Mace, G.M.; Norris, K.; Fitter, A.H. Biodiversity and ecosystem services: A multilayered relationship. *Trends Ecol. Evol.* **2012**, *27*, 19–26.
2. Lengyel, S.; Déri, E.; Varga, Z.; Horváth, R.; Tóthmérész, B.; Henry, P.Y.; Kobler, A.; Kutnar, L.; Babij, V.; Seliskar, A.; *et al.* Habitat monitoring in Europe: A description of current practices. *Biodivers. Conserv.* **2008**, *17*, 3327–3339.
3. Burai, P.; Lövei, G.; Lénárt, Cs.; Nagy, I.; Enyedi, P. Mapping aquatic vegetation of the Rakamaz-Tiszanagyfalui Nagy-Morotva using hyperspectral imagery. *Acta Geogr. Debr. Landsc. Environ. Ser.* **2010**, *4*, 1–10.

4. Kobayashi, T.; Tsend-Ayush, J.; Tateishi, R.A. New tree cover percentage map in Eurasia at 500 m resolution using MODIS data. *Remote Sens*. **2014**, *6*, 209–232.

5. Cole, B.; McMorrow, J.; Evans, M. Empirical modelling of vegetation abundance from airborne hyperspectral data for upland Peatland restoration monitoring. *Remote Sens*. **2014**, *6*, 716–739.

6. Alexander, C.; Bøcher, P.K.; Arge, L.; Svenning, J.C. Regional-scale mapping of tree cover, height and main phenological tree types using airborne laser scanning data. *Remote Sens. Environ*. **2014**, *147*, 156–172.

7. Zlinszky, A.; Schroiff, A.; Kania, A; Deák, B.; Mücke, W., Vári, Á.; Székely, B.; Pfeifer, N. Categorizing grassland vegetation with full-waveform airborne laser scanning: A feasibility study for detecting Natura 2000 habitat types. *Remote Sens*. **2014**, *6*, 8056–8087.

8. Beamish, D. Peat mapping associations of airborne radiometric survey data. *Remote Sens*. **2014**, *6*, 521–539.

9. Li, C.; Wang, J.; Hu, L.; Yu, L.; Clinton, N.; Huang, H.; Yang, J.; Gong, P.A. Circa 2010 thirty meter resolution forest map for China. *Remote Sens*. **2014**, *6*, 5325–5343.

10. Dalponte, M.; Bruzzone, L.; Gianelle, D. Fusion of hyperspectral and LiDAR remote sensing data for classification of complex forest areas. *IEEE Trans. Geosci. Remote. Sens*. **2008**, *46*, 1416−1427.

11. Thenkabail, P.S. *Hyperspectral Remote Sensing of Vegetation*; Taylor and Francis: New York, NY, USA, 2011; p. 781.

12. Adam, E.; Mutanga, O.; Rugege, D. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review. *Wetlands Ecol. Manag*. **2010**, *18*, 281−296.

13. Plaza, A.; Benediktsson, J.A.; Boardman, J.W.; Brazile, J.; Bruzzone, L.; Camps-Valls, G.; Chanussot, J.; Fauvel, M.; Gamba, P.; Gualtieri, A.; *et al*. Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ*. **2009**, *113*, 110–122.

14. Vanden Borre, J.; Paelinckx, D.; Mücher, C.A.; Kooistra, L.; Haest, B.; de Blust, G.; Schmidt, A.M. Integrating remote sensing in Natura 2000 habitat monitoring: Prospects on the way forward. *J. Nat. Cons*. **2011**, *19*, 116−125.

15. Pu, R.; Bell, S. A protocol for improving mapping and assessing of seagrass abundance along the West Central Coast of Florida using Landsat TM and EO-1 ALI/Hyperion images. *ISPRS J. Photogramm. Remote Sens*. **2013**, *83*, 116–129.

16. Stratoulias, D.; Balzter, H.; Zlinszky, A.; Toth, V.R. Assessment of ecophysiology of lake shore reed vegetation based on chlorophyll fluorescence, field spectroscopy and hyperspectral airborne imagery. *Remote Sens. Environ*. **2015**, *157*, 72–84.

17. Huang, C.; Asner, G.P. Applications of remote sensing to alien invasive plant studies. *Sensors* **2009**, *9*, 4869–4889.

18. Mirik, M.; Ansley, R.J.; Steddom, K.; Jones, D.C.; Rush, C.M.; Michels, G.J., Jr.; Elliott, N.C. Remote distinction of a noxious weed (musk thistle: *Carduus nutans*) using airborne hyperspectral imagery and the Support Vector Machine Classifier. *Remote Sens*. **2013**, *5*, 612–630.

19. Green, A.A.; Berman, M.; Switzer, P.; Craig, M.D. A transformation for ordering multispectral data in terms of image quality with implications for noise removal, *IEEE Trans. Geosci. Remote Sens*. **1988**, *26*, 65–74.

20. Landgrebe, D.A. *Signal Theory Methods in Multispectral Remote Sensing*; John Wiley & Sons: New York, NY, USA, 2003; p. 503

21. Molnár, Z.; Bölöni, J.; Biró, M.; Horváth, F. Distribution of the Hungarian (semi-) natural habitats I. Marshes and grasslands. *Acta Bot. Hung.* **2008**, *50*, 59–105.

22. Eliáš, P.; Sopotlieva, D.; Dítě, D.; Hájková, P.; Apostolova, I.; Senko, D.; Melečková, Z.; Hájek, M. Vegetation diversity of salt-rich grasslands in Southeast Europe. *Appl. Veg. Sci.* **2013**, *16*, 521–537.

23. Török, P.; Kapocsi, I.; Deák, B. Conservation and management of alkali grass-land biodiversity in Central-Europe. In *Grasslands: Types, Biodiversity and Impacts*; Zhang, W.J., Ed.; Nova Science Publishers Inc.: New York, NY, USA, 2012; pp. 109–118.

24. Borhidi, A.; Kevey, B.; Lendvai, G. *Plant communities of Hungary*; Akadémiai Kiadó: Budapest, Hungary, 2012; p. 544.

25. Kelemen, A.; Török, P., Valkó, O.; Miglécz, T.; Tóthmérész, B. Mechanisms shaping plant biomass and species richness: Plant strategies and litter effect in alkali and loess grasslands. *J. Veg. Sci.* **2013**, *24*, 1195–1203.

26. Deák, B.; Valkó, O.; Alexander, C.; Mücke, W.; Kania, A.; Tamás, J.; Heilmeier, H. Fine-scale vertical position as an indicator of vegetation in alkali grasslands—Case study based on remotely sensed data. *Flora- Morphol. Distribut. Funct. Ecol. Plants* **2014**, *209*, 693–697

27. Valkó, O.; Tóthmérész, B.; Kelemen, A.; Simon, E.; Miglécz, T.; Lukács, B.; Török, P. Environmental factors driving vegetation and seed bank diversity in alkali grasslands. *Agric. Ecosyst. Environ.* **2014**, *182*, 80–87.

28. Deák, B.; Valkó, O.; Török, P.; Tóthmérész, B. Solonetz meadow vegetation (*Beckmannion eruciformis*) in East-Hungary—An alliance driven by moisture and salinity. *Tuexenia* **2014**, *34*, 187–203.

29. Deák, B.; Valkó, O.; Tóthmérész, B.; Török, P. Alkali marshes of Central-Europe—Ecology, management and nature conservation. In *Salt Marshes: Ecosystem, Vegetation and Restoration Strategies*; Shao, B., Ed.; Nova Science Publishers Inc.: New York, NY, USA, 2014; pp. 1–11.

30. Pettorelli, N. *The Normalised Difference Vegetation Index*; Oxford University Press: Oxford, UK; 2013; p. 194.

31. Hurcom, S.J.; Harrison, A.R. The NDVI and spectral decomposition for semi-arid vegetation abundance estimation. *Int. J. Remote Sens.* **1998**, *19*, 3109–3126.

32. Rabe, A.; Jakimow, B.; Held, M.; van der Linden, S.; Hostert, P. EnMAP-Box, Version 2.0, 2014. Available online: www.enmap.org (accessed on 28 January 2015).

33. Heldens, W.; Heiden, U.; Esch, T.; Stein, E.; Muller, A. Can the future EnMAP mission contribute to urban applications? A literature survey. *Remote Sens.* **2011**, *3*, 1817–1846.

34. Mansour, K.; Mutanga, O.; Everson, T.; Adam, E. Discriminating indicator grass species for rangeland degradation assessment using hyperspectral data resampled to AISA Eagle resolution. *ISPRS J. Photogramm. Remote Sens.* **2012**, *70*, 56–65.

35. Burai, P.; Laposi, R.; Enyedi, P.; Schmotzer, A.; Kozma, B.V. Mapping invasive vegetation using AISA Eagle airborne hyperspectral imagery in the Mid-Ipoly-Valley. In Proceedings of the 3rd IEEE GRSS Workshop on Hyperspectral Image and Signal Processing-WHISPERS'2011, Lisboa, Portugal, 6–9 June 2011.

36. Maselli, F.; Conese, C.; Petkov, L.; Resti, R. Inclusion of prior probabilities derived from a nonparametric process into the maximum likelihood classifier. *Photogramm. Eng. Remote Sens.* **1992**, *58*, 201–207.

37. Richards, J.A. *Remote Sensing Digital Image Analysis*; Springer-Verlag: Berlin, Germany, 1999; p. 240.

38. Yang, C.; Everitt, J.H.; Fletcher, R.S.; Jensen, R.R.; Mausel, P.W. Evaluating AISA+ hyperspectral imagery for mapping black mangrove along the South Texas Gulf, Coast. *Photogramm. Eng. Remote Sens.* **2009**, *75*, 425–435.

39. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

40. Lawrence, R.; Wood, S.; Sheley, R. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sens. Environ.* **2006**, *100*, 356–362.

41. Vapnick, V.N. *Statistical Learning Theory*; John Wiley and Sons Inc.: Hoboken, NJ, USA, 1988; p. 768.

42. Camps-Valls, G.; Bruzzone, L. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1351−1362.

43. Camps-Valls, G.; Gomez-Chova, L.; Calpe-Maravilla, J.; Martin-Guerrero, J.D.; Soria-Olivas, E.; Alonso-Chorda, L.; Moreno, J. Robust support vector method for hyperspectral data classification and knowledge discovery. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1530–1542.

44. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46.

45. Chan, J.C.-W.; Spanhove, T.; Ma, J.; Vanden Borre, J.; Paelinckx, D.; Canters, F. Natura 2000 habitat identification and conservation status assessment with superresolution enhanced hyperspectral (CHRIS/Proba) imagery. In Proceedings of GEOBIA 2010 geographic object-based image analysis, Ghent, Belgium, 29 June–2 July 2010.

46. Chopping, M.J.; Rango, A.; Ritchie, J.C. Improved semi-arid community type differentiation with the NOAA AVHRR via exploitation of the directional signal. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 1132–1149.