

Article

# Classification of High-Mountain Vegetation Communities within a Diverse Giant Mountains Ecosystem Using Airborne APEX Hyperspectral Imagery

Adriana Marcinkowska-Ochtyra \* , Bogdan Zagajewski , Edwin Raczko , Adrian Ochtyra   
and Anna Jarocińska 

Department of Geoinformatics, Cartography and Remote Sensing, Faculty of Geography and Regional Studies, University of Warsaw (UW), Krakowskie Przedmieście 30, 00-927 Warsaw, Poland; bogdan@uw.edu.pl (B.Z.); edwin.raczko@uw.edu.pl (E.R.); adrian.ochtyra@uw.edu.pl (A.O.); ajarocinska@uw.edu.pl (A.J.)

\* Correspondence: adriana.marcinkowska@uw.edu.pl (A.M.-O.); Tel.: +48-2255-21507

Received: 15 February 2018; Accepted: 3 April 2018; Published: 7 April 2018



**Abstract:** Mapping plant communities is a difficult and time consuming endeavor. Methods relying on field surveys deliver high quality data but are usually limited to relatively small areas. In this paper we apply airborne hyperspectral data to vegetation mapping in remote and hard to reach areas. We classified 22 vegetation communities in the Giant Mountains on 3.12-m Airborne Prism Experiment (APEX) hyperspectral images, registered in 288 spectral bands (10 September 2012). As the classification algorithm, Support Vector Machines (SVM) was used. APEX data were corrected geometrically and atmospherically, and three dimensionality reduction methods were performed to select the best dataset. As reference we used a non-forest vegetation map containing vegetation communities of Polish Karkonosze National Park from 2002, orthophotomap and field surveys data from 2013 to 2014. We obtained the post-classification maps of 22 vegetation communities, lakes and areas without any vegetation. Iterative accuracy assessment repeated 100 times was used to obtain the most objective results for individual communities. The median value of overall accuracy (OA) was 84%. Fourteen out of twenty-four classes were classified of more than 80% of producer accuracy (PA) and sixteen out of twenty-four of user accuracy (UA). APEX data and SVM with the use of iterative accuracy assessment are useful for the mountain communities classification. This can support both Polish and Czech national parks management by giving the information about diversity of communities in the whole transboundary area, helping with identification especially in changing environment caused by humans.

**Keywords:** hyperspectral; iterative accuracy assessment; vegetation communities; mountain ecosystem; support vector machines

---

## 1. Introduction

Mountain ecosystems are an important indicator of climate change [1] because, within a small altitude gradient, mountain vegetation goes from deciduous forests, through pine forests, dwarf pine thickets and grassland communities to tussock grassland communities including bryophytes and lichens [2]. The distribution of mountain plant communities is influenced by height above sea level, exposure, slope, substrate, the degree of development and moisture of soil, and the length of growing season and snow retention [3].

Traditional vegetation mapping is most often done in the field by methods according to the Braun–Blanquet floristic approach and by interpreting aerial photographs [4]. These methods are

time consuming and require an investment of resources and labour. Studies and monitoring of high-mountain vegetation are made difficult by poor accessibility and the short vegetation period, and so remote sensing is becoming increasingly popular as a base for ecological analyses. Spectral characteristics reflect species' physiological states, and the anatomical and morphological structure of species, such as the content of photosynthetic dyes, water, cellulose, and the structural state of cells [5]. Therefore, hyperspectral techniques are a great support for traditional vegetation mapping [6–9]. The first vegetation classification of hyperspectral data was conducted by American researchers [10–12]. In Europe, hyperspectral data were used for classification of land cover, wetlands and Natura 2000 sites [13–15], tree species [16–19] and, less commonly, of plant communities [6,20]. There is still a lack of typical mountain vegetation mapping works using hyperspectral data [6,9,10], some studies have covered upland areas [13,14], and the vast majority describe lowland [7,11,12,15,16].

The algorithms based on the idea of a single spectral signature representing a given spectral class for an analysed image pixel [21] are used to classify, inter alia, plant species [20]. In the case of plant communities consisting of different species, this would require the spectral characteristics of each species to be obtained. For high-resolution images, the object-oriented approach is useful, although proper image segmentation should be performed. This is complicated and often requires a more individual approach, limiting the automation desired in monitoring studies. With data of high spatial and spectral resolutions, detailed vegetation classification is successfully performed using non-parametric machine learning methods [22,23], which allow for high classification accuracies.

Comparative analyses of different hyperspectral data classification methods have evaluated Support Vector Machines (SVM) as one of the best [22–25]; also, in detailed vegetation analyses [8,26], moreover, it performs better in terms of speed and ability to correctly distinguish classes when trained on small training dataset [8] and a large number of bands [25].

Evaluation of classification reliability should exclude the randomness and subjectivity associated with the selection of training and validation samples and how they are divided. This is especially important for such diverse classes as a mosaic of plant communities comprising different species. This variety can be recorded through multiple repetitions of classification and validation process with a randomised change in the training dataset and the validation dataset each time [16,19] in order to objectivise the results. This approach in the literature is used for less numerous and more homogeneous classes as tree species, there is no such elaboration performed for numerous detailed vegetation communities as we studied here.

The Airborne Prism Experiment (APEX) sensor has been operational since 2010. Existing publications described the nature of the instrument [27], the processing of acquired images [28–30], or the environmental analysis, e.g., identification of: cover type in urban areas [31]; of tree species [18,19]; or of non-forest vegetation [9,32–34]. The Giant Mountains (Polish Karkonosze and Czech Krkonoše) are an area of high environmental value and are characterised by the greatest diversity of endemic species which are sensitive to environmental changes and should thus be permanently monitored. The vegetation is currently undergoing rapid changes and is in danger due to human induced environmental threats (pollution, increased human activity due to tourism, deforestation, etc.).

In the Czech part of the Giant Mountains, one-metre aerial images allowed seven vegetation classes to be distinguished with an overall accuracy (OA) of 81% [4]. It was noted that vegetation consisting of similar species compositions was confused in classification, such as vegetation along trails, and herbaceous plants and grasses. It was pointed out that the large diversity of vegetation and the images' low spectral resolution had caused the confusion of classes. Nine and twelve vegetation classes above the tree line were classified on 30-m Landsat-8 satellite data, 2-m WorldView-2 data and 12.5-cm aerial orthophotomaps [35]. The best OAs obtained were 83.6% for the nine classes, and 72.0% for the 12 detailed classes built by the most important grassland vegetation classes and other vegetation. The authors suggest that the usefulness of lower-spatial-resolution data are better for general identification of land cover, and for detailed analyses of vegetation the use of high-resolution data is suggested.

Classification of APEX data assumed that vegetation would be distinguishable at the level of communities, and these objects were also analyzed on HyMap data (Hyperspectral Mapper, 126 spectral bands in 400–2500 nm) [7]; additionally, 55 species were classified. The comparative analysis of results showed that classification of communities yields better results (about 90%) than a more detailed classification at the species level, although those accuracies were also high (about 80%). This is confirmed by the classification of 20 species of herbaceous vegetation on AISA EAGLE II (128 spectral bands in 400–1000 nm) [8], where OA was approximately 80%.

Previous studies of vegetation communities on Karkonosze [32,34] and vegetation types on Krkonoše [9] conducted separately for each country on APEX data presented smaller areas and the results were obtained only once, however for different datasets. In Polish Karkonosze, analyses were based on original 288 APEX bands [32] and original 252 and 30 Minimum Noise Fraction (MNF) bands [34]. In the Czech area research, APEX data were used in original form (288 bands) and were reduced spectrally with Principal Component Analyses (PCA) [9]. Here we used three dimensionality reduction methods to identify the best way to process APEX data: PCA, MNF and original bands selected with backward elimination approach. Additionally, we performed the sensitivity analysis of training dataset to classification accuracy. The SVM method resulting in high accuracies of classifications on diverse classes of non-forest vegetation [7–9,34] was selected as a proper tool to identify high-mountain plant communities in a repeatable way. In light of this, the aerial flight over the Giant Mountains using an APEX hyperspectral scanner to classify even 22 plant communities of the whole Polish–Czech transboundary area with SVM represents a novel approach to obtain valuable documentary and reference material for future studies. It is the first elaboration where numerous accuracies for such detailed and complex high-mountain vegetation classes were obtained. Noteworthy is the fact that this study also presents the first result of classification performed for both the Polish and Czech Giant Mountains vegetation with the same repeatable method.

The aim of the research is to present a methodology for mapping mountain plant communities in the Polish and Czech Giant Mountains and the study was designed to answer the following questions:

1. Can APEX hyperspectral image data and SVM method be used for classification of high-mountain vegetation communities in both the Polish and Czech parts of Giant Mountains?
2. Does the iterative assessment of accuracy allow to distinguish communities in terms of the difficulty of identifying them?
3. How does the preparation of the dataset and number of sample pixels affect the accuracy?
4. What is the consistency of the classification results with the reference data?
5. How can the evaluated algorithm and the results be useful for national parks administration?

## 2. Materials and Methods

### 2.1. Study Area

We conducted the research in the Giant Mountains, protected by the creation of cross-border Polish and Czech national parks belonging to the UNESCO Man and Biosphere program (Figure 1).

The study area covers the whole high-mountain ecosystem (above 1200 m a.s.l.) of both countries characterised by their diversity of species and communities typical of alpine and subalpine zones (Figure 2) and it covers an area 34.5 km long and 7.6 km wide. The species' occurrences are mainly determined by geological, climatic, soil and water conditions [36]. The Giant Mountains were formed late in the Alpine orogeny, which was followed by a period of erosion which resulted in rocky-walled postglacial corries (cirques), and characteristic corrie lakes, such as Mały Staw and Wielki Staw. In the ridge areas of the mountains the landscape resembles that of two regions, subarctic and high-mountain, and is referred to as "arctic–alpine tundra" [37]. A local mountain climate has been created here with long, harsh winters, large temperature fluctuations and violent winds [38].

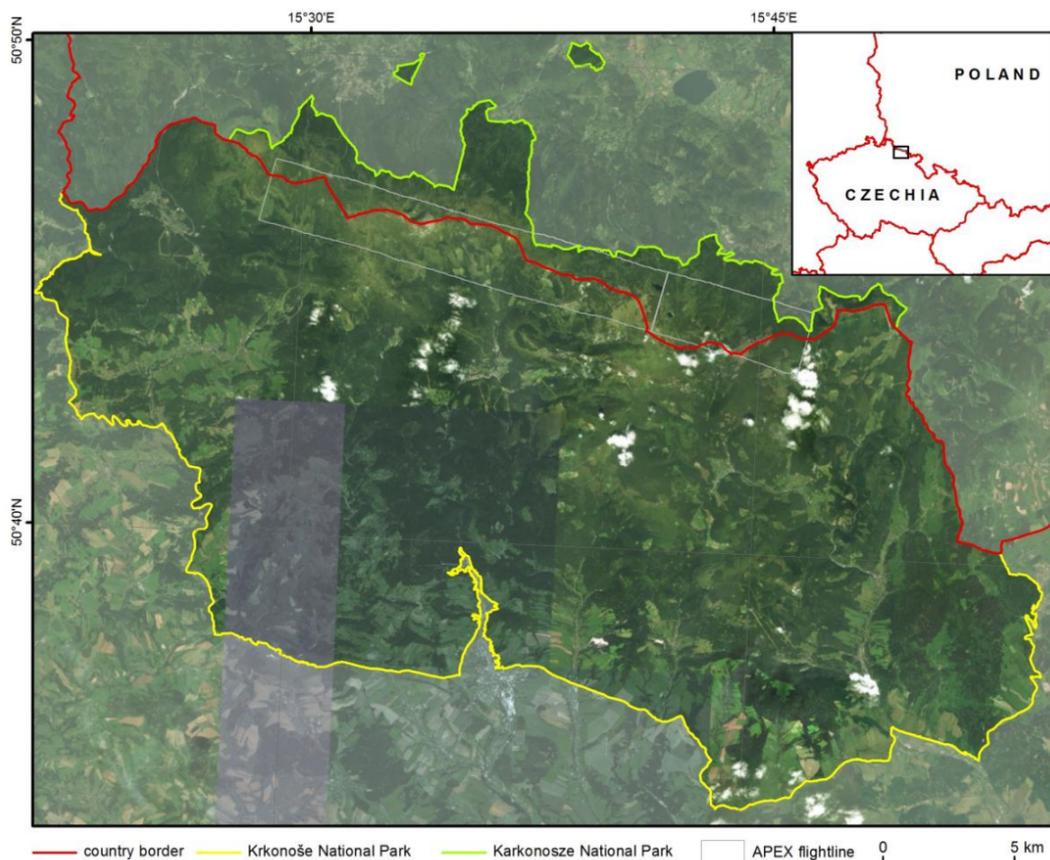


Figure 1. Giant Mountains area.



Figure 2. Mosaics of diverse vegetation communities in western (a) and eastern (b) Giant Mountains.

The most widespread community in the subalpine zone is *Pinetum mugo sudeticum* thickets, whose physiognomy is largely determined by the dominant species *Pinus mugo* [36]. *Pado–Sorbetum* thickets are dominated by the mountain sub-species of rowanberry, *Sorbus aucuparia* subsp. *glabrata*, which is accompanied by bird cherry, *Padus petraea*. Both endemic and a relict are *Salicetum lapponum* communities, which occur almost exclusively in the eastern Giant Mountains, which are dominated by a remnant of the glaciation, *Salix lapponum*, a heavily protected species. In stream channels, gullies and springs, there is *Athyrietum distentifolii*, in which *Athyrium distentifolium* dominates. Since 2010, they have been observed to be dying off in the Polish part of the Karkonosze Mountains, which has not yet been seen in any other mountain or northern region where this community occurs [39]. The high-mountain herbaceous communities also include *Adenostyletum alliariae*,

in which the characteristic species is *Adenostyles alliariae*, and which occur mainly within postglacial corries with a northern exposure [40]. On steep avalanche-prone slopes and leeward slopes of postglacial corries with long-lasting snow cover, *Calamagrostion* communities occur [41]. In the most common type of grass community, *Crepido–Calamagrostietum villosae*, which often occurs in complexes, the species *Calamagrostis villosa* dominates. One characteristic feature is Vaccinium communities: *Vaccinium myrtillus*, with European blueberry dominating, accompanied by *Calamagrostis villosa*. The community covers vast areas on the slopes and channels of postglacial corries. Patches of Vaccinium communities are often found in complexes with, for example, *Crepido–Calamagrostietum villosae* or *Carici (rigidae)–Nardetum*. A smaller share of the area is covered by *Empetro–Vaccinietum* crowberry communities, which include, among others, *Empetrum nigrum* and *Vaccinium myrtillus*. They occupy exposed locations, including the ridges of corries. *Carici (rigidae)–Nardetum* grass communities comprise the species *Nardus stricta*; they are mainly anthropogenic in nature, and also occur in complexes with *Vaccinium myrtillus*, among others. For over 50 years, the species *Molinia caerulea* has been observed to encroach on areas of *Carici (rigidae)–Nardetum* grasslands, especially on the Czech side of the Krkonoše Mountains [42]. On account of the diversity of species, *Carici (rigidae)–Festucetum airoidis* mountain grasslands are divided into two altitude categories: alpine and subalpine [43]. The alpine form is richer and is distinguished by the presence of the species *Juncus trifidus* and lichens, while the poorer subalpine form is characterised by *Festuca airoides* grasslands co-occurring with alpine heaths with the alliance *Loiseleurio procumbentis–Vaccinion* forming small patches dominated by *Calluna vulgaris*, *Vaccinium myrtillus* blueberry, grasses and lichens. Grasslands occur in the alpine zone above 1450 m a.s.l. and partly in the subalpine zone in areas exposed to strong winds. Large areas of rocks and screes are occupied by vegetation—mainly lichen communities, whose determining factor for diversity is snow cover [41]. Large areas of subalpine and alpine vegetation level in the Giant Mountains are covered by “snow-loving” communities of the alliance *Rhizocarpon alpicolae* and slightly smaller areas by “snow-avoiding” communities of the alliance *Umbilicarium cylindrica*. On the former there is snow cover during the winter months, which disappears to reveal a colourful mosaic on screes which is, for example, yellow–green thanks to lichens of the genus *Rhizocarpon*. Communities not covered by snow give the rock walls a dark grey colour. In small scattered areas there are spring-habitat plants in non-limestone communities of the alliance *Cardamino–Montion*, which include the species *Pedicularis sudetica*, which is endemic and a glacial relict, and occurs only in the Giant Mountains [43]. The subalpine zone is characterised by the presence of subalpine–subarctic mountain bogs. Raised and transitional bogs with a total area of 40 ha have been entered in the global list of areas covered by the Ramsar Convention. The non-forest vegetation of the Giant Mountains also includes meadow communities in the form of alpine tundra, which are relics of the pastoral economy which functioned from the 17th to the end of the 19th century [41]. The composition and structure of species have undergone changes, but they remain refuges of biodiversity. Tourist areas have synanthropic vegetation, including areas around mountain shelters, where nitrophilous perennials occur. The list of vegetation communities that are the object of the study is presented in Table 1.

The vegetation of the Giant Mountains is under threat due to increasing human activity in the region and environmental changes impairing its ability to grow. In the 1980s and 90s it died back due to an ecological disaster caused by, among other factors, strong winds, air pollution from industrial areas and the precipitation of nitrogen compounds [44]. A relatively new danger to the fern community on the Polish side of the Giant Mountains is a weevil, which is the carrier of the microbiological pathogen or nematodes, eating fern leaves resulting in reduced growth in the next year. Another high-mountain vegetation threat is mass tourism in summer and winter. In terms of human influence, the synanthropisation of vegetation is observed, allowing invasive species as *Rumex alpinus* to encroach and displace natural vegetation. Expansion of native grasses *Calamagrostis villosa* and *Molinia caerulea* encroachment is also noticed.

## 2.2. Reference Data

We extracted the samples for classification from APEX data based on field surveys conducted at the end of August 2013 and 2014 with GPS Trimble GeoXT receiver (Trimble Inc., Sunnyvale, CA, USA). The flight over the Giant Mountains with the use of the APEX scanner was carried out in September, so it was important that the field research was done at a similar time, due to the vegetation phase. The mapping area covered two parts of the eastern (16 km<sup>2</sup>) and western (12 km<sup>2</sup>) Giant Mountains, for which we collected a total of 812 identified patches of vegetation communities.

As the reference material, we used non-forest vegetation map provided by Wojtuń and Żołnierz from 2002 [45]. The map contained 48 vegetation communities and 13 vegetation types for the whole Polish Karkonosze National Park. Most communities are above the tree line, but the map also include meadow ecosystems and peat bogs on the foothills, as well as communities occurring in the area of forest floors, not being in the interest of our study. Previous studies [34] identified the possibilities for combining certain classes on account of, for example, spectral similarity. We aggregated pixels representing physiognomic classes *Artemisietea vulgaris* and *Urtica dioica* communities into one class named *Artemisietea vulgaris*. Similarly, we aggregated *Pinetum mugo sudeticum* and *Pino mugo–Sphagnetum* into one *Pinetum mugo sudeticum* class, because differences existed only in the properties of the substrate. Areas with no vegetation and those in the early stages of succession we also combined into a single class due to similar spectral reflectances, which resulted in the occurrence of erroneous pixels representing the early stages of succession around lakes or on trails. Based on previous work [34] we selected *Deschampsia caespitosa* class in clearing near Mały Staw lake because it was larger than the anthropogenic community class identified on non-forest vegetation map from 2002. We decided to exclude from the final classification classes occupied small areas (*Peucedanum ostruthium* included a total of fewer than 20 pixels, as was the case with *Deschampsia flexuosa*, which was significantly overestimated). We included in the final study only those classes considered representative on the basis of test classifications. Based on an assessment of the possibility to classify units and the APEX spatial resolution available, the spectral differentiation of classes, and occurring in the subalpine and alpine vegetation belt, we selected 22 classes of vegetation communities, lakes and non-forest vegetation covering the study area. We extracted additional information based on photointerpretation of 12 cm orthophotomaps collected in 2012 for the Polish side of the Giant Mountains.

We used reference data collected during field surveys, the existing plant communities map and image data interpretation for creation of training and validation datasets. In this step each reference point was either put into a training or validation dataset. Next, we extracted pixel samples representing each class based on reference data location in the APEX image. Before model training we decided to perform two steps that would benefit our training and validation process. To balance the training dataset depending on available sample size for each class, we decided to use up to 200 pixels per class in the training dataset (Table 1). This created a more balanced dataset which minimized the influence of more numerous classes on model training. Most classifiers tend to favor classes with more samples, treating less numerous classes as noise. This effect is also more visible in datasets where samples for several classes dominate the whole dataset (more than 50% of total number of samples).

Similar steps were performed when creating validation dataset. We selected up to 400 pixels per class (Table 1). Its balancing was aimed at reducing OA. This allowed us to achieve somewhat lower OA but at the same time we were able to provide more accurate measure of post-classification image accuracy. All of the classes were labelled according to the vegetation communities from the non-forest vegetation map developed by Wojtuń and Żołnierz [45].

**Table 1.** The number of samples used for classification and validation.

No.	Class	Training Samples		Validation Samples	
		Original	Randomly Selected	Original	Randomly Selected
1	<i>Rhizocarpon alpicolae</i> all. <sup>1</sup>	582	200	1116	400
2	<i>Umbilicaria cylindrica</i> all.	170	150	341	300
3	<i>Carici (rigidae)</i> – <i>Nardetum</i>	546	200	1079	400
4	<i>Carici (rigidae)</i> – <i>Festucetum airoidis</i> (subalpine)	267	200	389	300
5	<i>Carici (rigidae)</i> – <i>Festucetum airoidis</i> (alpine)	116	100	235	200
6	<i>Cardamino</i> – <i>Montion</i> all.	30	30	62	60
7	<i>Scheuchzerio</i> – <i>Caricetea nigrae</i> cl. <sup>2</sup>	236	200	503	400
8	<i>Oxycocco</i> – <i>Sphagnetea nigrae</i> cl.	345	200	509	400
9	<i>Crepido</i> – <i>Calamagrostietum villosae</i>	456	200	668	400
10	<i>Deschampsia caespitosa</i>	496	200	1005	400
11	<i>Molinia caerulea</i>	347	200	525	400
12	<i>Calamagrostion</i> all.	331	200	593	400
13	<i>Pado</i> – <i>Sorbetum</i>	193	200	440	400
14	<i>Salicetum lapponum</i>	58	50	111	100
15	<i>Adenostyletum alliariae</i>	12	10	25	20
16	<i>Athyrietum distentifolii</i>	503	200	1108	400
17	<i>Empetro</i> – <i>Vaccinietum</i>	159	150	322	300
18	<i>Vaccinium myrtillus</i>	668	200	1351	400
19	<i>Calluna vulgaris</i>	67	50	128	100
20	<i>Artemisietea vulgaris</i> cl.	39	30	80	60
21	<i>Pinetum mugo sudeticum</i>	2463	200	5501	400
22	<i>Calamagrostio villosae</i> – <i>Piceetum</i>	1000	200	2002	400
23	lakes	524	200	1076	400
24	areas without vegetation	239	200	471	400

<sup>1</sup> alliance, <sup>2</sup> class.

### 2.3. Hyperspectral Data Processing

APEX is a pushbroom dispersive imaging spectrometer designed to simulate, calibrate and validate existing and planned satellite missions, and was developed under the ESA-PRODEX (European Space Agency-PROgramme de Développement d'Expériences scientifiques) program [46]. Data are collected in 288 spectral bands within the range 380.5–2501.5 nm. Images of 3.12-m spatial resolution were acquired for the Polish and Czech areas of the Giant Mountains on 10 September 2012 from Dornier Do 288 plane belonging to the German Aerospace Centre (DLR). It was a part of the project Hyperspectral Remote Sensing for Mountain Ecosystem (HyMountEcos), co-financed by the European Facility for Airborne Research Transnational Access (EUFAR TA).

Concurrently to the aerial pass, we performed calibration spectrometric measurements on surfaces of relatively uniform spectral reflectance, such as water reservoirs, asphalt surfaces and homogeneous meadows. They were obtained using an ASD FieldSpec 3 spectrometer (ASD Inc., Longmont, CO, USA) registering within the range 350–2500 nm. Image preprocessing was carried out at VITO (the Flemish Institute for Technological Research), the owner of the APEX scanner. This included radiometric calibration, geometric and atmospheric correction using spectrometric data and a Digital Terrain Model (DTM). We extracted the area above 1200 m a.s.l. from the APEX data using DTM. It includes some very steep terrain, such as postglacial cirques, and so topographic shadows were masked to avoid false-positive identifications of the objects spectrally similar to shaded areas [47]. We built the shadow mask on the basis of spectral reflection values from band 104 (946 nm), because in comparison to other bands, near infrared allows to clearly distinguish shadows, where they have very low reflectance opposite to other objects. Because we developed the mask by selecting the threshold of reflectance values to be assumed for pixels representing shadows, it slightly encompassed pixels representing spectrally similar objects, such as *Umbilicaria cylindrica* class.

#### 2.4. Dimensionality Reduction

The 288 hyperspectral bands of APEX allow for detailed analysis, but a high correlation of neighbouring bands is nonetheless observed, leading to redundancy of information and lengthening the algorithm run-time. An important aspect is the reduction of the number of bands and the selection of the most informative. Based on visual analysis of all APEX bands we removed 36 bands associated with the absorption of electromagnetic waves by water vapour (ranges: 1335–1422 and 1759–1954 nm).

We used three methods of dimensionality reduction: two based on transformations PCA [48] and MNF [49] available in ENVI software, and one band selection by backward elimination [50], executed using IDL language. PCA determines a new principal axis for the coordinate system along the largest possible variance of data [48]. Using covariance matrix, new, uncorrelated bands were calculated and first 40 bands were selected as the most informative based on the highest eigenvalues. MNF transformation required noise estimation based on image statistics; therefore, two separate statistics files were created. During the transformation, the average for each band of the original image was calculated and covariance matrices and noise correlations were created. Bands with eigenvalues higher than 1 contained information, while bands with values close to zero contained noise. Thirty MNF bands were selected for further analysis, which were read from the information curve. We analyzed the information on APEX spectral bands by backward elimination, carrying out 288 classifications using the SVM method on the area, where polygons with training and validation samples has been previously designated. During each classification one band was subtracted from the image to assess how the OA changes in the absence of a specific band. We selected the bands that enhanced the accuracy and finally, two sets of bands were selected, consisting of 70 and 18 bands, respectively.

#### 2.5. Support Vector Machine Classification

The classification used a Support Vector Machine (SVM) machine learning algorithm—a non-parametric method that enables the supervision of pattern recognition [51]. The algorithm for distinguishing classes in feature space  $\mathbb{R}^n$  tries to determine the best hyperplane, which separates classes while attaining the maximum possible margin between classes. It is assumed that the data set is composed of pairs  $(x_i, y_i)$ , where:  $i = 1, x_i \in \mathbb{R}^n, y_i \in \{-1; 1\}$ . Each point from a given set has the value  $x_i$  describing the elements of the training sample and  $y_i$  indicating the membership in one of the classes. The hyperplane  $H$ , by means of which individual classes are separated, is defined by the following formula [52]:

$$w \times x - b = 0, \quad (1)$$

where  $x$  is the point on the hyperplane,  $w$  is the  $n$ -dimensional normal vector for the hyperplane, and  $b$  is the distance from the nearest point on the hyperplane to the starting point. Figure 3 shows two sets of points separable linearly. Determination of two hyperplanes  $H_1$  and  $H_2$  in such a way that there are no points from the set of training data between them, and the distance between them (margin) was as large as possible, can be described by the following equations:

$$w \times x - b = -1, \quad (2)$$

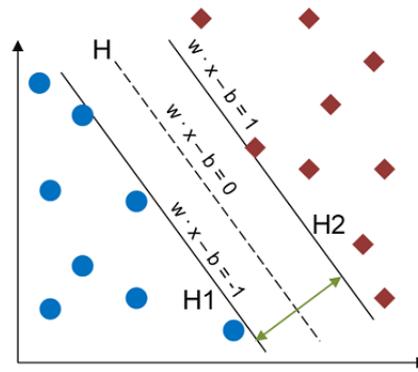
$$w \times x - b = 1, \quad (3)$$

assuming that:

$$w \times x_i + b \geq 1, \text{ if } y_i = 1, \quad (4)$$

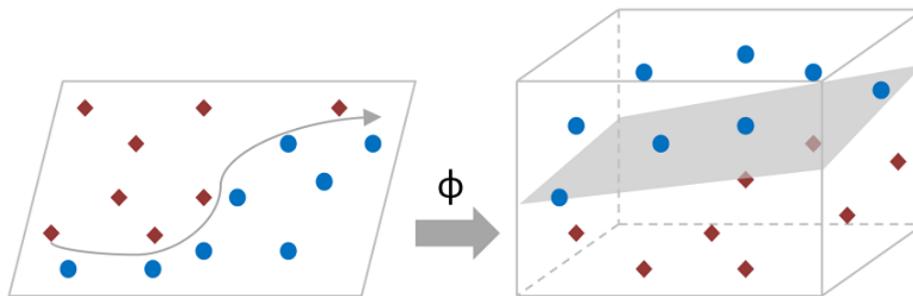
$$w \times x_i + b \leq -1, \text{ if } y_i = -1. \quad (5)$$

The sample closest to the hyperplane are the support vectors. The result is dependent on the width of the boundary between classes: the greater it is, the lower the classification error.



**Figure 3.** Hyperplane and parallel H1 and H2 hyperplanes that separate classes while maintaining the maximum margin [53], modified.

If the feature vectors  $x \in \mathbb{R}^n$  are not linearly separable they should be mapped into higher dimensional feature space with non-linear vector function  $\Phi: \mathbb{R}^n \mapsto F$ , where F is a higher dimensional Euclidean space [52]. Subsequently, the SVM algorithm is used in the new feature space, and the resulting linearly separating hyperplane plane corresponds to the non-linear surface of the original space (Figure 4).



**Figure 4.** Feature space transformation [54], modified.

The scalar product of the vectors  $x_i x_j$  in the transformed space is equivalent to the kernel function [52]:

$$\Phi(x_i) \times \Phi(x_j) = K(x_i x_j). \tag{6}$$

A central element is the selection of function, as follows:

linear:

$$K(x_i x_j) = x_i^T x_j, \tag{7}$$

radial:

$$K(x_i x_j) = \exp(-\gamma \|x_i - x_j\|^2), \tag{8}$$

polynomial:

$$K(x_i x_j) = (\gamma(x_i^T x_j) + 1)^p, \tag{9}$$

sigmoid:

$$K(x_i x_j) = \tanh(\gamma(x_i^T x_j) + 1), \tag{10}$$

where  $\gamma$  is the gamma parameter, p is the polynomial degree, and 1 is the bias parameter.

SVM can achieve good results, even with noisy bands; it is not time-consuming, despite using significant amounts of data [22].

We performed the classification in package e1071 [55] in the R software [56]. Our choice of kernel function was based on a series of experiments where all of the kernel functions were tested and then results were compared. The aim was to select a kernel function that yielded highest accuracy.

In machine learning it is always good to perform hyper parameter tuning before training of model. Tuned parameters were: penalty (C), which controls the trade-off between classification errors and forcing margins; and gamma, which expresses the width of the Gaussian function. The parameters were fine-tuned using a grid search method which performs a series of training and validation of a model (trained classifier) for each combination of parameters one desires to test and then selects those that give the highest accuracy. Values for gamma from 0.1 to 0.9 and for C from 1 to 100 were tested. Analysis concluded that highest accuracies were acquired for the following combination of gamma and C: 0.1 and 9 respectively.

It was an assumption of the classification algorithm for APEX hyperspectral images that the analyses would be highly objectivised, so the selection of training and validation samples was repeated 100 times with a random selection of pixels from polygons collected during field tests using a high-accuracy GPS device with support of other reference materials.

## 2.6. Accuracy Assessment

We provided the assessment of accuracies obtained for the results by creating error matrices from which the overall and class accuracies were calculated [57]. An error matrix express the number of pixels assigned to a particular class relative to the verified class. Overall accuracy (OA) is calculated by dividing total correct ( $x_{kk}$ ) by total number of pixels ( $N$ ) in the error matrix:

$$OA = \frac{\sum_{k=1}^r x_{kk}}{N}. \quad (11)$$

The accuracy for each class was determined by producer accuracy (PA), which is calculated by dividing the number of pixels correctly classified into a class ( $x_{kk}$ ) by the total number of pixels in the reference class ( $x_{k\_val}$ ):

$$PA = \frac{x_{kk}}{x_{k\_val}}, \quad (12)$$

and user accuracy (UA), which expresses the number of pixels correctly classified into a class ( $x_{kk}$ ), divided by the number of pixels in the class according to the classification result ( $x_{k\_class}$ ):

$$UA = \frac{x_{kk}}{x_{k\_class}}. \quad (13)$$

To measure sensitivity of the classifier to changing training dataset and to provide more detailed accuracy metrics of our post-classification images we decided to use an iterative accuracy assessment procedure. The whole procedure consists of 100 repetitions of following steps:

1. Randomly select training and validation pixels without replacement from all available samples. These will create training and validation datasets. During this process, a specific number of pixels was selected for each class, as described in Table 1.
2. Perform model training using the training dataset.
3. Perform model accuracy assessment using the validation dataset. Calculate PA and UA for each class and OA.
4. Remove all samples from training and validation dataset. Repeat step 1.

The iterative accuracy assessment procedure provided the information about median class accuracy and estimated distribution shape and width of calculated accuracy measures for each class. The variation within classes we assessed by analyzing the absolute values (the difference between the minimum and maximum accuracy), and focusing on the values between the first and the third quartile in each boxplot.

Additionally, we analyzed the accuracy of the classification depending on variable number of training pixels, from 50, through 100, 150, 200, 250 to 300 for all classes which have more than 300 pixels in original training samples set (for 13 classes, Table 1). The number of validation pixels for this approach was constant and amounted to 400 samples.

A minimal classification unit was defined as APEX pixel size (3.12 m) and accuracy assessment was based on this unit. To present the map cartographically we decided to generalize the result to 1:15,000 scale using minimal mapping area defined as  $5 \times 5$  pixels and the boundaries of plant communities were smoothed using the majority analysis function in ENVI 5.3 software (Harris Geospatial Solutions, Broomfield, CO, USA), which removes individual pixels from within larger classes. In addition, we used the aggregation function to delete groups of pixels of less than a set size (25 pixels). For the design of the legend, we consulted with phytosociologists and performed cartographic work using ArcGIS 10.3 software (Environmental Systems Research Institute, Redlands, US).

### 3. Results

Dimensionality reduction methods provided us with the best dataset selection. Table 2 presents the accuracies obtained for the two best kernel functions in term of overall accuracy where five datasets were used. Linear function was used to compare based on previous studies of Polish Karkonosze vegetation [34]. Additionally, we added the information on the resulting file size to present the significance of the reduction.

**Table 2.** Overall accuracies obtained for different datasets and two kernel functions. PCA: Principal Component Analysis; MNF: Minimum Noise Fraction.

Dataset	Overall accuracy (OA, %)		File Size (MB)
	Linear	Radial	
252 spectral bands	82.69	83.11	465
40 PCA bands	81.04	84.49	65
30 MNF bands	80.76	82.02	48
70 spectral bands *	76.68	77.16	113
18 spectral bands *	68.14	69.01	31

\* backward elimination approach.

The highest OA were achieved for 40 PCA bands with radial function, next for 252 original spectral bands and MNF transforms. Lower than transformed values were obtained for selected spectral bands; for 70 and 18 bands, respectively. In linear function, the highest OA was achieved for the original 252 bands, then for 40 PCA bands; the lowest accuracy value was also taken by the image composed of 18 selected spectral bands. It has been noticed that a significant reduction in the size of the image, like in the case of the PCA method (more than seven times smaller by volume), took place: the accuracy dropped by less than two percent, which allowed to determine the high suitability of the use of spectral space reduction methods, increasing the operability of the classification process.

The analysis of the impact of the training set size on the accuracy of the classification we executed on the best dataset: 40 PCA bands on radial function. We noticed an increase in OA as the number of training sets increased, obtained for 13 classes of communities consisting of 300 samples (Supplementary Materials, Figure S1). For such sets of data median accuracy of 86% was obtained and the distribution of values was small, as for sets consisting of 250 and 200 training pixels. With 50 samples, median OA value was less than 74% and for this dataset the widest distribution of values was obtained.

The maps of the vegetation of the Giant Mountains and the error matrices for each 100 iterations were obtained also for the best dataset, which were 40 PCA bands on radial function (examples of eastern and western parts resulting in the best iteration based on OA (85.5%) are presented in Figure 5 and Supplementary Materials, Table S1).

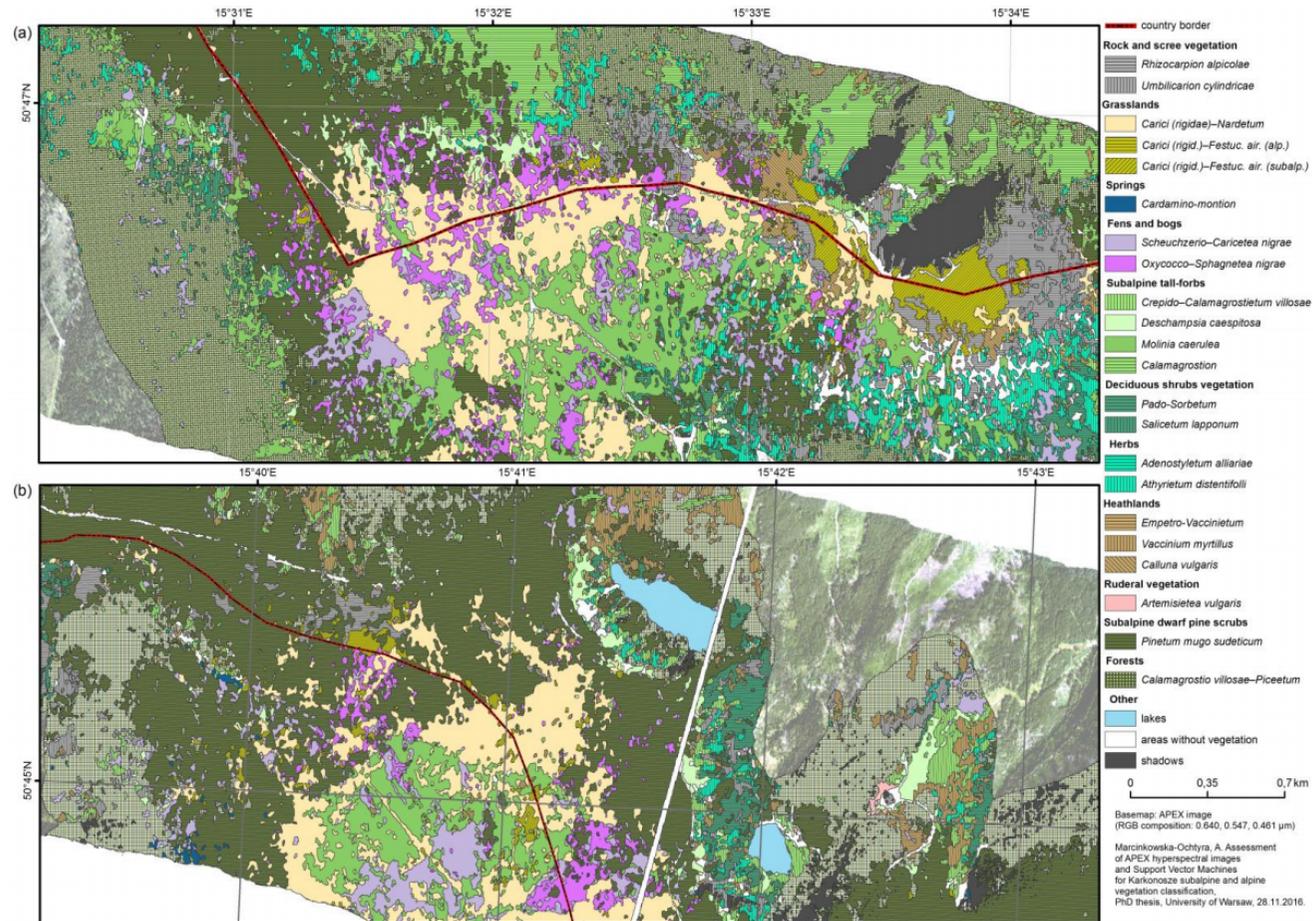
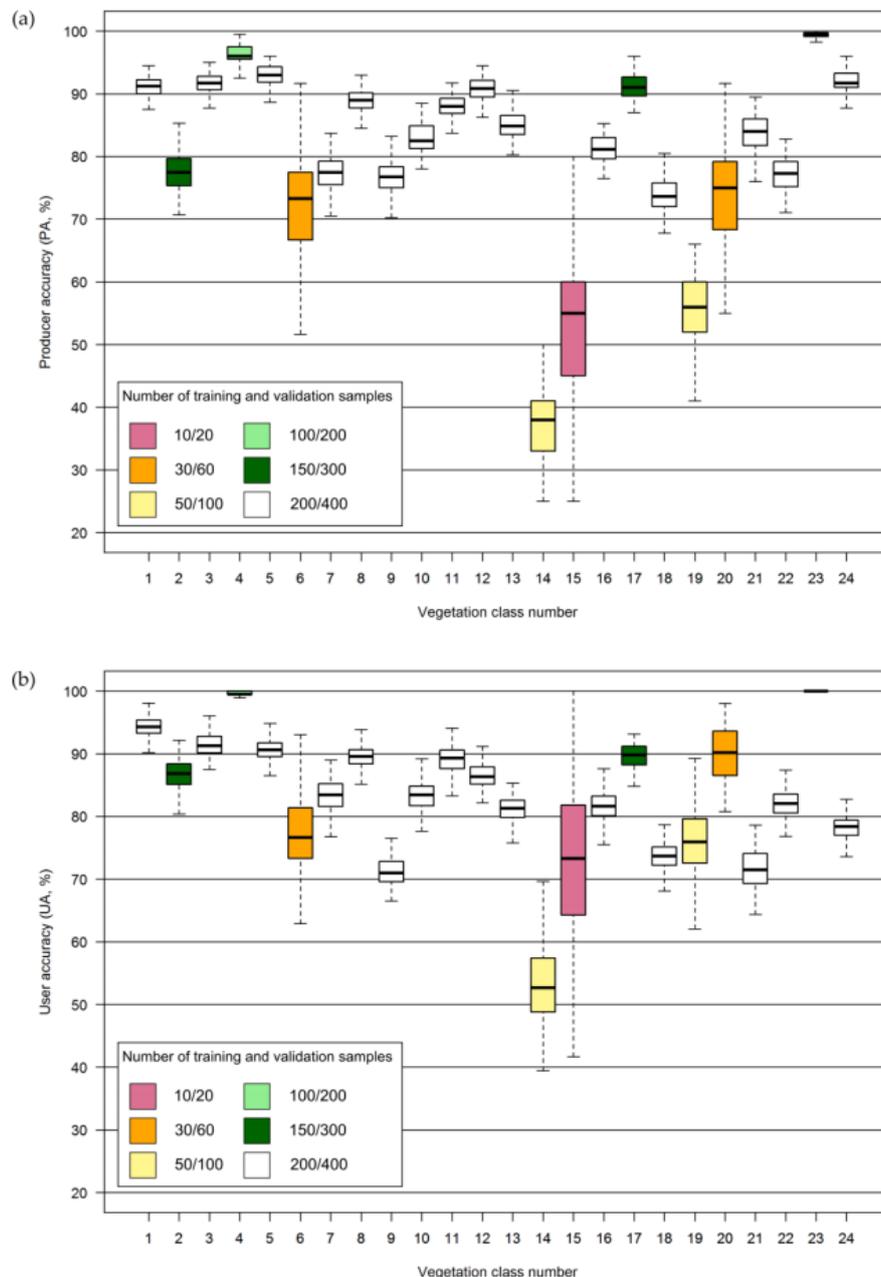


Figure 5. Map of vegetation communities of the Giant Mountains: (a) western part, (b) eastern part ([58], modified).

Median value of OA obtained during 100-fold accuracy assessment was equal to 84%. In total, median values of PA and UA were high for most classes: 14/24 classes were distinguished with a PA of above 80%, while 16/24 classes were also identified with a UA of over 80%.

To show the diversity within individual plant communities, we presented their accuracies as boxplots showing the ranges of values adopted for each class during the 100 iterations of the algorithm (Figure 6).



**Figure 6.** Accuracies for each for each class obtained during 100-fold iterative accuracy assessment: (a) Producer accuracies; (b) User accuracies. (1—*Rhizocarp. alp. all.*, 2—*Umbilic. cyl. all.*, 3—*Carici (rig.)–Nard.*, 4—*Carici (rig.)–Fest. air. (alp.)*, 5—*Carici (rig.)–Fest. air. (subalp.)*, 6—*Cardamino–Mont. all.*, 7—*Scheuch.–Caric. nigrae cl.*, 8—*Oxyc.–Sphagn. nigrae cl.*, 9—*Crepido–Calamagr. vill.*, 10—*Deschamp. caesp.*, 11—*Molinia caer.*, 12—*Calamagrostion all.*, 13—*Pado–Sorb.*, 14—*Salic. lapp.*, 15—*Adenos. alliar.*, 16—*Athyrietum distent.*, 17—*Empetro–Vacc.*, 18—*Vacc. myrt.*, 19—*Calluna vulg.*, 20—*Artemisietea vulg. cl.*, 21—*Pinetum mugo sud.*, 22—*Calamagr. vill.–Piceetum*, 23—lakes, 24—areas without vegetation).

The highest median accuracies were for lakes (100% PA and UA), alpine *Carici (rigidae)*–*Festucetum airoidis* in alpine (96% and 99%, respectively) and subalpine form (93% and 91%) and *Carici (rigidae)*–*Nardetum* (92% and 91%). High accuracies were noted for other large, homogeneous communities as *Rhizocarpion alpicolae* (91% PA, 94% UA) and *Calamagrostion* (91% PA, 86% UA). *Empetro*–*Vaccinietum*, diversified in terms of species achieved a high accuracy (91% PA, 90% UA). Also high accuracies were obtained for *Oxycocco*–*Sphagnetetea nigrae* (89% PA, 90% UA). *Scheuchzerio*–*Caricetea nigrae* were classified less accurately (78% and 83%, respectively). With about 89% of both PA and UA were classified expansive *Molinia caerulea* which covers vast areas of subalpine belts in the Czech part of the Giant Mountains. Slightly lower accuracies was obtained for *Athyrietum distentifolii* (81% and 82%, respectively). The values oscillating around 75% were obtained for *Crepido*–*Calamagrostietum villosae* (77% PA, 71% UA) and *Vaccinium myrtillus* (74% PA, 74% UA). *Calamagrostio villosae*–*Piceetum* (77% PA, 82% UA) in most cases were confused with *Pinetum mugo sudeticum* (84% PA and 72% UA). The *Artemisietea vulgaris* class achieved lower PA (75%) than UA (90%). A relatively low PA median value was also observed for *Calluna vulgaris* (56%). The least accurately classified were *Salicetum lapponum* (38% PA, 53% UA).

Within most classes (19/24) PAs show little variation (less than 15% difference in accuracy between the lowest and the highest value within a boxplot), which shows that training and validation patterns were well selected and that they can be properly classified. The biggest in-class variation was for *Adenostyletum alliariae*, which in the best iteration achieved a PA of 80%, while the lowest observed value was only 25% and the difference within boxplot was 41%. The class was most often confused with *Cardamino*–*Montion* with also broad distribution of accuracies (23%). In addition, higher variability of PA values (over 20%) was observed for the three classes whose median PA score was lowest: *Salicetum lapponum*, *Calluna vulgaris* and *Adenostyletum alliariae*. The results of UA correlate well for most classes, with the variation within individual plant communities being small (less than 15% difference between minimum and maximum values within boxplots). More variation was noticed for the same classes described by analyzing their PA distributions above. The widest distribution was noted for *Adenostyletum alliariae* (55%), then for *Salicetum lapponum* (26%) and *Cardamino*–*Montion* (23%), which were most often confused with each other and for *Calluna vulgaris* (23%).

In Figure 6 we also marked number of training and validation samples as colors of boxplots. Generally, high median accuracies (above 85%) and low level of variation (less than 15%) were noted for communities for which collecting a representative number of training and validation pixels (200/400) was possible, as for *Carici (rigidae)*–*Festucetum airoidis* in subalpine form, *Carici (rigidae)*–*Nardetum*, *Rhizocarpion alpicolae*, *Calamagrostion*, *Oxycocco*–*Sphagnetetea nigrae*, and *Molinia caerulea*. However, high accuracy did not always go hand in hand with a larger training set; e.g., for *Carici (rigidae)*–*Festucetum airoidis* in alpine form represented by 100 training pixels was 96% PA and 100% UA or for *Empetro*–*Vaccinietum*, represented by 150 pixels was about 91% PA and 90% UA, while accuracies for *Crepido*–*Calamagrostietum villosae* or *Vaccinium myrtillus* with 200 training pixels oscillated around 75%. Nevertheless, a relationship between the size of the training set and the width of the distribution was noticeable. For *Calluna vulgaris* and *Salicetum lapponum* the sets of training and validation pixels (50/100) was too little to achieve high accuracies and their PA values distributions were wide. The *Artemisietea vulgaris* and *Cardamino*–*Montion* classes had very few training (30) and validation pixels (60) resulting in lower PA and higher levels of this accuracy variation. For *Adenostyletum alliariae* the smallest number of training and validation samples was collected (10 and 20) which caused the widest distribution of both accuracy values.

#### 4. Discussion

Comparison results reported by the team of Marcinkowska-Ochtyra [26] present the OA as well as PA and UA for the same 22 vegetation communities of the whole Giant Mountains area above 1200 m a.s.l. classified using SVM and Random Forest (RF) classifier. It was observed that general accuracies were almost the same (84.62% and 84.87% OA, respectively). The difference was in the

median values obtained in iterative way; for the worst classified communities represented by smaller training sample size as *Salicetum lapponum*, *Adenostyletum alliariae* and *Calluna vulgaris* median PA for RF were much lower than for SVM (11%, 25% and 23%, respectively) and they were characterised by similar to SVM level of dispersion. It is confirmed by comparative studies of using SVM, RF and Maximum Likelihood to classify herbaceous vegetation [8], where the authors also find out that SVM is not sensitive in the training sample size, which makes it useful when only a limited number of training pixels are available. The accuracies for vegetation communities classes of the Polish Tatras obtained for DAIS7915 data (Digital Airborne Imaging Spectrometer, Geophysical and Environmental Research, New York, US, 79 spectral bands in 400–12,600 nm,) on August 2002 with Fuzzy ARTMap classifier [6], were comparable to Giant Mountains vegetation: PA above 90% for, among others, mountains grasslands and *Empetro–Vaccinietum*, and UA below 70% for willow thicket, *Chamaenerion angustifolium–Salix silesiaca* community and *Calamagrostietum villosae tatricum* in the Tatras, physiognomically comparable to *Salicetum lapponum* and *Crepido–Calamagrostietum villosae* in the Giant Mountains. For *Vaccinium myrtillus* in a complex with tall herb communities in the Tatra mountains, PA of 72% and UA of 78% were achieved; in the Giant Mountains, for *Vaccinium myrtillus*, also co-occurred with grasslands, very similar PA and UA of 74% and 74%, respectively, were achieved. Complementary studies with hyperspectral data were provided by Kupková et al. [9] where APEX and AISA Dual data (Airborne Imaging Spectrometer, 494 spectral bands in 400–2300 nm) were used to classify 11 vegetation classes in Eastern Tundra of the Krkonoše Mountains, which provided very similar OAs to our results (83% and 84%, based on SVM, respectively and 69% and 73%, based on ANN). For individual classes, PA above 90% were obtained for, among others, block fields and anthropogenic areas, *Pinus mugo* scrubs, and below 70% for *Calamagrostis villosa* stands, wetlands and peat bogs, and also *Molinia caerulea* stands, which is confirmed in this work, but a higher PA in analysis of the Czech part of the Giant Mountains was found for alpine heathlands (90% on APEX and 82% on AISA image).

There are classifications by other authors using APEX data relating to five species of trees in Forêt de Hardt, France (OA 74%) [18] and the Polish Karkonosze (OA 68% and 79%) [19,59]. Preliminary classifications of Western Karkonosze non-forest communities on raw APEX data allowed to achieve approximately 75% OA [32]. Classifications of subalpine and alpine vegetation around the Mały Staw lake in the Eastern Karkonosze Mountains have achieved an accuracy of around 75% [34]. Nine types of vegetation of the Karkonosze National Park were classified at 78% OA based on simulated EnMAP (Kayser-Threde GmbH, Munich, Germany) satellite data and there is a similarity to the results obtained in this work: large, homogeneous classes were classified better, e.g., dwarf pine thickets (over 90% PA and UA), and grasslands (approximately 90% PA and UA). Generally, the majority of classes were classified at PA and UA levels above 70%, while the lowest PA was for the herbaceous community class (44%), consisting of heather and *Vaccinium* communities. These analyses are particularly important in the event of EnMAP launch, which is to be made available free of charge from 2019. This will open up new possibilities for extracting information on the environment, in its broadest sense.

In the literature, the most common is using radial kernel function [23,60,61]. In our research it was noted that it gave the best results with PCA transformed bands and with 252 original spectral bands it took second place in accuracy. In some cases this function is not the best: when the number of variables is very large, there is no need to transform the space into higher dimension and it is appropriate to use the linear function. We presented this approach in the previous studies of Polish Karkonosze vegetation classification where the linear function with original spectral bands presented the best dataset to use [34].

Hyperspectral data provide greater possibilities in the analysis of a mosaic of species forming a diverse community. Comparative analyses of the classification results obtained from original and spectrally reduced hyperspectral data show higher accuracies for the transformed data [7,8,16,17]. Limiting the size of the resultant file optimises the method for further applications.

The sensitivity analysis of the classification to the training data set showing an increase in the accuracy of vegetation classification along with an increase in the training set was confirmed by

other authors [7,8,24,62]. The random selection of training and validation pixels made the results more objective. Thanks to the iterative accuracy assessment method, it was possible to obtain 100 accuracy metrics for each individual class, which indicated the accuracy of randomly selected pixels for communities of similar assumed values. In the literature, this approach appeared only recently, with authors classifying species of trees [16,17,19] more spectrally pure than other plant communities. It should be underlined that this method has not yet been carried out for many classes of complex plant communities.

#### 4.1. Difficulties in Classification of Mountain Vegetation Communities

The methodological problems of mapping mountain vegetation stem mainly from the specific nature of mountain vegetation cover, which is made up of small and spatially diverse communities [63]. We aggregated some classes due to their spectral similarity. From the aerial level it is not possible to detect the differences between *Pinetum mugo sudeticum* and *Pino mugo–Sphagnetum*. Despite their different structures, *Pinetum mugo sudeticum* and *Calamagrostio villosae–Piceetum* are also spectrally similar. Mapping communities near the upper forest border was difficult due to the smooth boundaries of classes mixing with particular trees. Because *Calamagrostio villosae–Piceetum* do not belong de facto to non-forest communities, they were found on the Wojtuń and Żolnierz (2002) map only at the borders of other communities. LiDAR data could be helpful in distinguishing trees, thanks to information on altitude.

We also noted a confusion of co-occurring classes, such as *Vaccinium myrtillus* with heathers or *Crepido–Calamagrostietum villosae* because of occurrence in small, dispersed patches, most often forming complexes. *Crepido–Calamagrostietum villosae* was mixed with *Deschampsia caespitosa* due to their spectral similarity. *Vaccinium myrtillus* was also confused with grasslands, and less with *Athyrietum distentifolii* encroaching on areas formerly occupied by ferns. *Calluna vulgaris* was most often confused with *Vaccinium myrtillus*, due to its co-occurrence in small patches with subalpine *Carici (rigidae)–Festucetum airoidis* dominated by heather and also with *Athyrietum distentifolii*. *Oxycocco–Sphagnetea nigrae* classified on high accuracy is unique to the Giant Mountains, and stand out spectrally from among homogeneous areas. *Scheuchzerio–Caricetea nigrae* were identified in the field as occurring alongside *Deschampsia caespitosa* and *Crepido–Calamagrostietum villosae* communities encroached on them, thus making class less distinguishable. In the eastern part of the Giant Mountains these fens were replaced by bogs from *Oxycocco–Sphagnetea nigrae* class.

Finding representative samples of *Salicetum lapponum* in the field was difficult, as they often occupy areas of no more than nine square meters, and some occur on steep, shady slopes. Therefore, they were represented by a smaller set of training (50) and validation pixels (100) resulting both median PA and UA were the smallest and the high variation within this class. *Cardamino–Montion* was also difficult to identify due to the linear courses of its occurrence on the steep slopes of postglacial carries near water seepages; therefore, the training to validation pixels ratio were only 30:60 and the variation of PA and UA was high.

In general, PA and UA for each class were similar, but for two classes PA was lower than UA (*Pinetum mugo sudeticum* and areas without vegetation) and they were underestimated. Higher values of the reference communities have been correctly identified as these classes than in the case of their actual representation. Also, four classes were classified on higher UA than PA level (*Artemisieta vulgaris*, *Calluna vulgaris*, *Adenostyletum alliariae* and *Salicetum lapponum*) and they were overestimated. Higher percent values of the areas identified as these classes were actually found. However, they were represented by few training and validation pixels (100/200 or less, for *Adenostyletum alliariae* only 10/20) and fitting to the reference data was more difficult than for more representative dataset. The *Artemisieta vulgaris* class achieved lower PA (75%) because it had undergone dynamic changes and its current coverage did not coincide with that of 2002, making it difficult to determine appropriate patterns. It was most often confused with *Crepido–Calamagrostietum villosae*. However, the pixel fit for all the pixels classified to this class was high (90% UA), due to its lower spectral reflectance in the red

radiation range and much higher reflectance in the near infrared, clearly distinguishing it from the surrounding *Deschampsia caespitosa*. *Salicetum lapponum* with lower PA were incorrectly identified in places where *Pado-sorbetum* was found.

Difficulties in classification of some communities were also caused by different dates of the data used affected the accuracy results. The non-forest vegetation reference map was ten years older than the APEX image. The ruderal vegetation became one of the most dynamic and related to human activity. *Peucedanum ostrythium* during the field inventory in 2013 occupied small patches, smaller than half the size of the APEX pixel. In 2014, it was not possible to detect even individual one. On the reference map *Athyrietum distentifolii* community occupies several major patches, mainly in the western part of Karkonosze, and to a lesser extent of eastern part. Since 2010, the fern's death has been observed in the Polish part of the Giant Mountains and encroachment of grassland or shrub communities in these areas. The discrepancies in acquisition date of reference and hyperspectral data highly affect the accuracy of mapping of fern communities [32] and only field validation in 2013 and 2014 allowed to observe the disappearance of the community in places of potential occurrence. Communities that have changed in the past decade needed to be inspected in the field in a phenological period close to the date of the imaging. In September, when the APEX data was acquired, vegetation begins to change color, which allows it to be differentiated well. For example, *Juncus trifidus*, which belongs to the alpine grasslands, has a reddish color in late summer, which distinguishes it from other species by its different spectral signature. *Carici (rigidae)–Nardetum* communities cover extensive homogeneous areas and are also dominated by one species—*Nardus stricta*, stands out from the surrounding thickets of dwarf pine or *Vaccinium myrtillus* for its much higher reflectance of radiation in the visible bands. Use of an image from spring or summer might not allow for the unambiguous distinction of the characteristics of communities spectrally similar during their optimal vegetation phase.

One problem associated with mapping mountain areas is topographical shadows on the images. A shadow mask excluded such areas from the analysis; otherwise, classes would have been erroneously identified. The application of masks likely excluded the part of sites representing *Umbilicarium cylindricae*, which typically grows in the shadows of rocks not covered by snow.

Some phenomena are similar in different parts of the Giant Mountains. In both the Polish and Czech parts, spreading of expansive *Calamagrostis villosa* is noticed. Some different phenomena are also observed: in the Polish part, dying of the *Athyrietum distentifolii* community, and the encroachment of invasive species *Rumex alpinus* and expansive *Molinia caerulea* in Czech. Our results might help in identifying them and support management of national parks areas.

## 5. Conclusions

In summary, it can be stated that:

1. The high spatial and spectral resolution of APEX data and the use of SVM allowed to accurately classify vegetation communities with high accuracies. Iterative random selection of training and validation samples made it possible to avoid the subjectivity of a single selection of data. The most varied values were obtained for classes represented by smaller training sets, heterogeneous classes difficult to identify due to small spatial extent (less than nine square metres) and location in shaded areas (*Adenostyletum alliariae*, *Salicetum lapponum*). Higher accuracies were achieved for classes consisting of more training and validation pixels, such as mountain grasslands, *Rhizocarpon alpicolae*, *Calamagrostion*, *Molinia caerulea*.
2. The PCA reduction allowed us to keep all the information and accelerate the classification process. Increasing the size of the training set resulted in higher OA because of a more representative dataset (the highest accuracy at 300 training pixels).
3. The majority of the plant communities' extent of the Giant Mountains was similar to the reference map from 2002. However, we underline the importance of the up-to-dateness of the data. Anthropogenic communities are subject to dynamic changes: *Athyrietum distentifolii* suffer damage from pest outbreaks, which can be observed within few years. Some classes were aggregated

into larger ones (*Artemisietea vulgaris*, *Pinetum mugo sudeticum*, areas without vegetation) due to spectral similarities.

The map of high-mountain plant communities developed using the same repeatable method for the Giant Mountains may support monitoring works in both national parks. With the results collected in previous studies based on APEX data and the Polish and Czech parts [32,34,35] this work is an objective material for the management of the whole transboundary area. In terms of vegetation synanthropisation, the results can be used for invasive or expansive species encroachment analysis. Differences between communities caused by phenology are highlighted due to the date of APEX data acquisition proper to map specific community for nature conservation purposes (such as *Juncus trifidus* in grassland communities). The iterative accuracy may be useful in detailed analyses of each community, allowing proper identification of the information about diversity and possibility thanks to their distribution widths.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2072-4292/10/4/570/s1>, Figure S1: Increased accuracy of classification caused by the increase in the number of training pixels, Table S1: Error matrix for the best iteration of 40 PCA bands, in %.

**Acknowledgments:** We extend our thanks to Prof. Bronisław Wojtuń and Lidia Przewoźnik for their help in identifying plant communities, the authorities of the Polish Karkonoski Park Narodowy and the Czech Krkonošský Národní Park for providing access to research and accommodation facilities, EUFAR, DLR and VITO for participating in the HyMountEcos project and obtaining APEX data and the Anna Pasek Foundation for the scholarship awarded to Adriana Marcinkowska-Ochtyra for the academic year 2013/2014. Article processing charges (APC) have been financed from the grant No. 501-D119-64-0180200-15 awarded by the Polish Ministry of Science and Higher Education.

**Author Contributions:** A.M.-O. and B.Z. designed the study. They with A.O. performed the field spectroscopy measurements to the APEX data corrections. A.M.-O. conducted the data classification and results interpretation. E.R. helped with classification and iterative accuracy assessment and A.J. helped with hyperspectral data processing methods. A.M.-O. wrote the manuscript and with A.O. and E.R. prepared the Figures. All authors read and revised the submitted manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Körner, C. Impact of atmospheric changes on high mountain vegetation. In *Mountain Environments in Changing Climates*; Routledge: London, UK, 1994; pp. 155–166. [CrossRef]
2. Kozłowska, A. Strefy przejścia między układami roślinnymi-analiza wieloskalowa (na przykładzie roślinności górskiej). In *Prace Geograficzne*; IGiPZ PAN: Warsaw, Poland, 2008; Volume 215, pp. 1–150.
3. Piękoś-Mirkowa, H.; Mirek, Z. Zbiorowiska roślinne. In *Przyroda Tatrzańskiego Parku Narodowego*; Mirek, Z., Ed.; Tatrzański Park Narodowy: Zakopane-Kraków, Poland, 1996; pp. 237–272.
4. Müllerová, J. Use of digital aerial photography for sub-alpine vegetation mapping: A case study from the Krkonoše Mts., Czech Republic. *Plant Ecol.* **2005**, *175*, 259–272. [CrossRef]
5. Jensen, J.R. Biophysical remote sensing. *Ann. Assoc. Am. Geogr.* **1983**, *73*, 111–132. [CrossRef]
6. Zagajewski, B. Assessment of neural networks and Imaging Spectroscopy for vegetation classification of the High Tatras. *Teledetekcja Środowiska* **2010**, *43*, 1–113.
7. Zhang, C.; Xie, Z. Object-based vegetation mapping in the Kissimmee River watershed using HyMap data and machine learning techniques. *Wetlands* **2013**, *33*, 233–244. [CrossRef]
8. Burai, P.; Deák, B.; Valkó, O.; Tomor, T. Classification of herbaceous vegetation using airborne hyperspectral imagery. *Remote Sens.* **2015**, *7*, 2046–2066. [CrossRef]
9. Kupková, L.; Červená, L.; Suchá, R.; Jakešová, L.; Zagajewski, B.; Březina, S.; Albrechtová, J. Classification of Tundra Vegetation in the Krkonoše Mts. National Park Using APEX, AISA Dual and Sentinel-2A Data. *Eur. J. Remote Sens.* **2017**, *50*, 29–46. [CrossRef]
10. Kokaly, R.F.; Despain, D.G.; Clark, R.N.; Livo, K.E. Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data. *Remote Sens. Environ.* **2003**, *84*, 437–456. [CrossRef]
11. Filippi, A.M.; Jensen, J.R. Fuzzy learning vector quantization for hyperspectral coastal vegetation classification. *Remote Sens. Environ.* **2006**, *100*, 512–530. [CrossRef]

12. Jollineau, M.Y.; Howarth, P.J. Mapping an inland wetland complex using hyperspectral imagery. *Int. J. Remote Sens.* **2008**, *29*, 3609–3631. [[CrossRef](#)]
13. Chan, J.; Paelinckx, D. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sens. Environ.* **2008**, *112*, 2999–3011. [[CrossRef](#)]
14. Olesiuk, D.; Zagajewski, B. Wykorzystanie obrazów hiperspektralnych do klasyfikacji pokrycia terenu zlewni Bystrzanki. *Teledetekcja Środowiska* **2008**, *40*, 125–148.
15. Delalieux, S.; Somers, B.; Haest, B.; Kooistra, L.; Múcher, C.A.; Vanden Borre, J. Monitoring heathland habitat status using hyperspectral image classification and unmixing. In Proceedings of the 2nd Whispers on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Iceland, 14–16 June 2010; pp. 50–54. [[CrossRef](#)]
16. Ghosh, A.; Fassnacht, F.E.; Joshi, P.K.; Koch, B. A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 49–63. [[CrossRef](#)]
17. Fassnacht, F.E.; Neumann, C.; Förster, M.; Buddenbaum, H.; Ghosh, A.; Clasen, A.; Joshi, P.K.; Koch, B. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2547–2561. [[CrossRef](#)]
18. Tagliabue, G.; Panigada, C.; Colombo, R.; Fava, F.; Cilia, C.; Baret, F.; Vreys, K.; Meuleman, K.; Rossini, M. Forest species mapping using airborne hyperspectral APEX data. *Misc. Geogr.* **2016**, *20*, 28–33. [[CrossRef](#)]
19. Raczko, E.; Zagajewski, B. Comparison of Support Vector Machine, Random Forest and Neural Network Classifiers for Tree Species Classification on Airborne Hyperspectral APEX images. *Eur. J. Remote Sens.* **2017**, *50*, 144–154. [[CrossRef](#)]
20. Wang, C.; Menenti, M.; Stoll, M.P.; Belluco, E.; Marani, M. Mapping mixed vegetation communities in salt marshes using airborne spectral data. *Remote Sens. Environ.* **2007**, *107*, 559–570. [[CrossRef](#)]
21. Dehaan, R.; Louis, J.; Wilson, A.; Hall, A.; Rumbachs, R. Discrimination of blackberry (*Rubus fruticosus* sp. agg.) using hyperspectral imagery in Kosciuszko National Park, NSW, Australia. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 13–24. [[CrossRef](#)]
22. 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. [[CrossRef](#)]
23. Chan, J.C.W.; Beckers, P.; Spanhove, T.; Borre, J.V. An evaluation of ensemble classifiers for mapping Natura 2000 heathland in Belgium using spaceborne angular hyperspectral (CHRIS/Proba) imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *18*, 13–22. [[CrossRef](#)]
24. Pal, M.; Mather, P.M. Some issues in the classification of DAIS hyperspectral data. *Int. J. Remote Sens.* **2006**, *27*, 2895–2916. [[CrossRef](#)]
25. Melgani, F.; Bruzzone, L. Classification of Hyperspectral Remote Sensing Images with Support Vector Machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790. [[CrossRef](#)]
26. Marcinkowska-Ochtyra, A.; Zagajewski, B.; Ochtyra, A.; Jarocińska, A.; Raczko, E.; Wojtuń, B. Mapping subalpine and alpine vegetation using APEX hyperspectral data. In Proceedings of the 36th EARSeL Symposium 2016, Bonn, Germany, 20–24 June 2016.
27. Itten, K.I.; Dell’Endice, F.; Hueni, A.; Kneubühler, M.; Schläpfer, D.; Odermatt, D.; Seidel, D.; Huber, S.; Schopfer, J.; Kellenberger, T.; et al. APEX—The hyperspectral ESA airborne prism experiment. *Sensors* **2008**, *8*, 6235–6259. [[CrossRef](#)] [[PubMed](#)]
28. Hueni, A.; Schlaepfer, D.; Jehle, M.; Schaepman, M.E. Impacts of dichroic prism coatings on radiometry of the airborne imaging spectrometer APEX. *Appl. Opt.* **2014**, *53*, 5344–5352. [[CrossRef](#)] [[PubMed](#)]
29. Sterckx, S.; Vreys, K.; Biesemans, J.; Iordache, M.D.; Bertels, L.; Meuleman, K. Atmospheric correction of APEX hyperspectral data. *Misc Geogr.* **2016**, *20*, 16–20. [[CrossRef](#)]
30. Vreys, K.; Iordache, M.D.; Biesemans, J.; Meuleman, K. Geometric correction of APEX hyperspectral data. *Misc Geogr.* **2016**, *20*, 11–15. [[CrossRef](#)]
31. Demarchi, L.; Canters, F.; Cariou, C.; Licciardi, G.; Chan, J.C.W. Assessing the performance of two unsupervised dimensionality reduction techniques on hyperspectral APEX data for high resolution urban land-cover mapping. *ISPRS J. Photogramm. Remote Sens.* **2014**, *87*, 166–179. [[CrossRef](#)]

32. Marcinkowska, A.; Zagajewski, B.; Ochtyra, A.; Jarocińska, A.; Raczko, E.; Kupková, L.; Stych, P.; Meuleman, K. Mapping vegetation communities of Karkonosze National Park using APEX hyperspectral data and Support Vector Machines. *Misc. Geogr.* **2014**, *18*, 23–29. [[CrossRef](#)]
33. Jarocińska, A.; Kacprzyk, M.; Marcinkowska-Ochtyra, A.; Ochtyra, A.; Zagajewski, B.; Meuleman, K. The application of APEX images in the assessment of the state of non-forest vegetation in the Karkonosze Mountains. *Misc. Geogr.* **2016**, *20*, 21–27. [[CrossRef](#)]
34. Marcinkowska-Ochtyra, A.; Zagajewski, B.; Ochtyra, A.; Jarocińska, A.; Wojtuń, B.; Rogass, C.; Mielke, C.; Lavender, S. Subalpine and Alpine Vegetation Classification based on Hyperspectral APEX and Simulated EnMAP images. *Int. J. Remote Sens.* **2017**, *38*, 1839–1864. [[CrossRef](#)]
35. Suchá, R.; Jakešová, L.; Kupková, L.; Červená, L. Classification of vegetation above the tree line in the Krkonoše Mts. National Park using remote sensing multispectral data. *Acta Univ. Carol. Geogr.* **2016**, *51*, 113–129. [[CrossRef](#)]
36. Żołniercz, L.; Wojtuń, B. Roślinność Subalpejska i Alpejska. In *Przyroda Karkonoskiego Parku Narodowego*; Knapik, R., Raj, A., Eds.; Karkonoski Park Narodowy: Jelenia Góra, Poland, 2013; pp. 241–278.
37. Soukupová, L.; Kociánová, M.; Jeník, J.; Sekyra, J. Arctic-alpine tundra in the Krkonoše, the Sudetes. *Opera Corcon.* **1995**, *32*, 5–88.
38. Sobik, M.; Błaś, M.; Migala, M.; Godek, M.; Nasiółkowski, T. Klimat. In *Przyroda Karkonoskiego Parku Narodowego*; Knapik, R., Raj, A., Eds.; Karkonoski Park Narodowy: Jelenia Góra, Poland, 2013; pp. 147–186.
39. Dunajski, A. Huge dieback of the tall fern community (*Athyrium distentifolii*). In Proceedings of the 9th Conference Geocological Problems of the Krkonoše/Karkonosze Mountains: The International Scientific Conference: Past, Present and Future of Transboundary Cooperation in Research and Management, Karkonoski Park Narodowy, Szklarska Poręba, Poland, 12–14 October 2016.
40. Pawłowski, B. *Flora Polska*; PWN: Warszawa-Kraków, Poland, 1967; Volume 11, p. 371.
41. Żołniercz, L.; Wojtuń, B.; Przewoźnik, L. *Ekosystemy Nieleśne Karkonoskiego Parku Narodowego*; Karkonoski Park Narodowy: Jelenia Góra, Poland, 2012.
42. Hejzman, M.; Češková, M.; Pavlí, V. Control of *Molinia caerulea* by cutting management on sub-alpine grassland. *Flora Morphol. Distrib. Funct. Ecol. Plants* **2010**, *205*, 577–582. [[CrossRef](#)]
43. Matuszkiewicz, W.; Matuszkiewicz, A. Mapa zbiorowisk roślinnych Karkonoskiego Parku Narodowego. *Ochr. Przyr.* **1974**, *40*, 45–109.
44. Danielewicz, W.; Raj, A.; Zientarski, J. Lasy. In *Przyroda Karkonoskiego Parku Narodowego*; Knapik, R., Raj, A., Eds.; Karkonoski Park Narodowy: Jelenia Góra, Poland, 2013; pp. 241–278.
45. Wojtuń, B.; Żołniercz, L. Plan Ochrony Ekosystemów Nieleśnych—Inwentaryzacja Zbiorowisk. In *Plan Ochrony Karkonoskiego Parku Narodowego*; Bureau for Forest Management and Geodesy: Brzeg, Poland, 2002.
46. Itten, K.I.; Schaepman, M.; De Vos, L.; Hermans, L.; Schlaepfer, H.; Space, O.C.; Droz, F. APEX—airborne prism experiment a new concept for an airborne imaging spectrometer. In Proceedings of the Third International Airborne Remote Sensing Conference and Exhibition, Copenhagen, Denmark, 7–10 July 1997; Volume 1, pp. 181–188.
47. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94. [[CrossRef](#)]
48. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educat. Psychol.* **1933**, *24*, 417–441. [[CrossRef](#)]
49. 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. [[CrossRef](#)]
50. John, G.H.; Kohavi, R.; Pfleger, K. Irrelevant features and the subset selection problem. In Proceedings of the Eleventh International Machine Learning Conference, Rutgers University, New Brunswick, NJ, USA, 10–13 July 1994; pp. 121–129. [[CrossRef](#)]
51. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995. [[CrossRef](#)]
52. Gualtieri, J.A.; Crompton, R.F. Support vector machines for hyperspectral remote sensing classification. In Proceedings of the 27th AIPR Workshop: Advances in Computer Assisted Recognition; SPIE: Bellingham, WA, USA, 1999; pp. 221–232. [[CrossRef](#)]
53. Hsu, C.W.; Chang, C.C.; Lin, C.J. *A Practical Guide to Support Vector Classification*; National Taiwan University, 2010; Available online: <https://ntu.csie.org/~cjlin/papers/guide/guide.pdf> (accessed on 13 February 2017).

54. Liao, R. Support Vector Machines, CSC 411 Tutorial. 2015. Available online: <https://.cs.toronto.edu/~urtasun/courses/CSC411/tutorial9.pdf> (accessed on 5 September 2016).
55. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R Package Version 1.6-8. 2017. Available online: <https://CRAN.R-project.org/package=e1071> (accessed on 15 December 2017).
56. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017; Available online: <https://www.R-project.org/> (accessed on 15 January 2018).
57. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [[CrossRef](#)]
58. Marcinkowska-Ochtyra, A. Assessment of APEX Hyperspectral Images and Support Vector Machines for Karkonosze Subalpine and Alpine Vegetation Classification. Ph.D. Thesis, University of Warsaw, Warsaw, Poland, 28 November 2016.
59. Raczko, E.; Zagajewski, B.; Ochtyra, A.; Jarocińska, A.; Marcinkowska-Ochtyra, A.; Dobrowolski, M. Forest species identification of Mount Chojnik (Karkonoski National Park) using airborne hyperspectral APEX data. *Sylvan* **2015**, *159*, 593–599.
60. Foody, G.M.; Mathur, A. A relative evaluation of multi-class image classification by support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1335–1343. [[CrossRef](#)]
61. Lin, S.W.; Lee, Z.J.; Chen, S.C.; Tseng, T.Y. Parameter determination of support vector machine and feature selection using simulated annealing approach. *Appl. Soft Comput.* **2008**, *8*, 1505–1512. [[CrossRef](#)]
62. Waske, B.; van der Linden, S.; Benediktsson, J.A.; Rabe, A.; Hostert, P. Sensitivity of Support Vector Machines to Random Feature Selection in Classification of Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2880–2889. [[CrossRef](#)]
63. Kozłowska, A. Problemy kartowania roślinności wysokogórskiej w skali szczegółowej (na przykładzie map roślinności Kotła Gąsienicowego i Goryczkowego Świńskiego). In *Badania Geoekologiczne w Otoczeniu Kasprowego Wierchu*; Kotarba, A., Kozłowska, A., Eds.; Prace Geograficzne, Volume 174; Wydawnictwo Continuo: Wrocław, Poland, 1999; pp. 37–43.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).