

Article

Predictive Mapping of Dwarf Shrub Vegetation in an Arid High Mountain Ecosystem Using Remote Sensing and Random Forests

Kim André Vanselow ^{1,*} and Cyrus Samimi ^{2,3}

¹ Institute of Geography, Friedrich-Alexander University Erlangen-Nuremberg, Wetterkreuz 15, D-91058 Erlangen, Germany

² Department of Geography, University of Bayreuth, Universitätsstr. 30, D-95440 Bayreuth, Germany; E-Mail: cyrus.samimi@uni-bayreuth.de

³ BayCEER, University of Bayreuth, D-95440 Bayreuth, Germany

* Author to whom correspondence should be addressed; E-Mail: kim.vanselow@fau.de; Tel.: +49-9131-85-23304; Fax: +49-09131-85-22013.

Received: 28 April 2014; in revised form: 4 July 2014 / Accepted: 9 July 2014 /

Published: 22 July 2014

Abstract: In many arid mountains, dwarf shrubs represent the most important fodder and firewood resources; therefore, they are intensely used. For the Eastern Pamirs (Tajikistan), they are assumed to be overused. However, empirical evidence on this issue is lacking. We aim to provide a method capable of mapping vegetation in this mountain desert. We used random forest models based on remote sensing data (RapidEye, ASTER GDEM) and 359 plots to predictively map total vegetative cover and the distribution of the most important firewood plants, *K. ceratoides* and *A. leucotricha*. These species were mapped as present in 33.8% of the study area (accuracy 90.6%). The total cover of the dwarf shrub communities ranged from 0.5% to 51% (per pixel). Areas with very low cover were limited to the vicinity of roads and settlements. The model could explain 80.2% of the total variance. The most important predictor across the models was *MSAVI2* (a spectral vegetation index particularly invented for low-cover areas). We conclude that the combination of statistical models and remote sensing data worked well to map vegetation in an arid mountainous environment. With this approach, we were able to provide tangible data on dwarf shrub resources in the Eastern Pamirs and to relativize previous reports about their extensive depletion.

Keywords: Central Asia; common pool resources; degradation; desertification; firewood; modeling; Pamir; post-Soviet transformation; rangeland ecology; Tajikistan

1. Introduction

Common pool resources play an important role in the sustainable development of many high mountain regions, as they provide crucial ecosystem goods and services. Despite the acquired knowledge about their importance, most “mountain commons are ecologically under-managed and suffer from the classic ‘commons syndrome’” [1] (p. 68). One important reason for this paradox is the lack of planning instruments that could support joint care of these natural resources, such as resource maps [1]. However, field-based generation of maps in high mountain regions is very time-consuming and costly for the usually large and often difficult-to-access areas. Therefore, in this study, we generate maps of common pool resources using remote sensing data that could provide a practical and economical alternative [2,3].

Despite the wide application of remote sensing data, arid regions and high mountains are underrepresented in studies, especially those focusing on more detailed aspects, such as species composition or species distribution [4–8]. However, even most of these studies do not work on single species, but mainly on plant communities. If single species are derived from remote sensing data, often hyperspectral data are used [9].

In mountainous environments, predominantly topographic gradients regulate species distributions and, thus, should be included in the mapping process [10]. The topographic conditions regulate the temperature, but also the water availability. The latter can hardly be quantified and differentiated for the whole region and is therefore neglected in the study. The altitude governs temperature, but also aspect and slope, all parameters easily derived from remote sensing-based DEMs. Furthermore, in arid environments, such as the Eastern Pamirs, vegetation is very open, and a large proportion of bare ground occurs. This can have a considerable influence on the spectral vegetation canopy reflectance; therefore, vegetation indices able to cope with soil noise should be included [11–13].

In this study, we use a machine learning method [14], derived from the RapidEye satellite system [15] and the ASTER Global Digital Elevation Model (GDEM) [16]. As such, we use the random forest algorithm [17] based on spectral, texture and topographic variables to predictively map dwarf shrub vegetation in the Eastern Pamirs as an example of an arid high mountain area.

We exemplify our approach with a case study from the arid mountain plateau of the Eastern Pamirs. There, dwarf shrubs represent an important resource, used for energy purposes (*i.e.*, heating and cooking) and livestock feeding (see Section 2.2).

The aim is to evaluate the suitability of different combinations of variables to create maps of the spatial distribution of the two most important firewood shrubs in this region, *Krascheninnikovia ceratoides* and *Artemisia leucotricha*, and to quantify the vegetative ground cover as an index of the degree of degradation.

2. Material and Methods

2.1. Study Area and Species

The study area is located in the Murghab District of the Gorno Badakhshan Autonomous Province (GBO), Tajikistan. It is part of the Eastern Pamirs that form a high plateau with elevations between 3500 and 4000 m a.s.l. located in the outermost east of Tajikistan, and partly in China and Afghanistan. Wide valleys, filled with alluvial sediments, are separated by relatively low ranges with peaks mostly around 5000 m a.s.l. The high altitude causes low temperatures all year round, with an annual average of -1 °C in the valleys. Westerlies, which represent the prevailing winds, are substantially intercepted by the Western Pamirs. The Hindukush, the Wakhan Range and the Karakorum block the possible influence of the Indian Monsoon. Therefore, the area is characterized by extreme aridity with a mean annual precipitation of less than 100 mm in some places, resulting in a high mountain desert with dwarf shrubs as the predominant growth form [18–21]. The most frequent dwarf shrub species are *Krascheninnikovia ceratoides* (*teresken*, Amaranthaceae) and *Artemisia leucotricha* (*shyvak*, Asteraceae). Since the dissolution of the Soviet Union and the associated ceasing of coal supply, these dwarf shrubs are increasingly used as an energy resource [22]. However, they are also important forage plants for livestock, as well as wild ungulates, particularly in winter [23,24]. Several authors assume that the dwarf shrub resources are becoming overused, resulting in a decrease of vegetation cover [23,25]; however, recent studies show heterogeneous degradation patterns [26].

Available data on the dwarf shrub vegetation in the Eastern Pamirs almost entirely dates back to the Soviet time. According to several authors, the total vegetative ground cover in *K. ceratoides*-dominated formations usually ranges between 3% and 15% [27,28], sometimes up to 25% [29,30]. Exceptions are associations where *A. leucotricha* (ground cover up to 60%) or the cushion-plants, *Acantholimon diapensioides* and *Sibbaldia tetandra* (ground cover 30%–40%), are the dominant species [18].

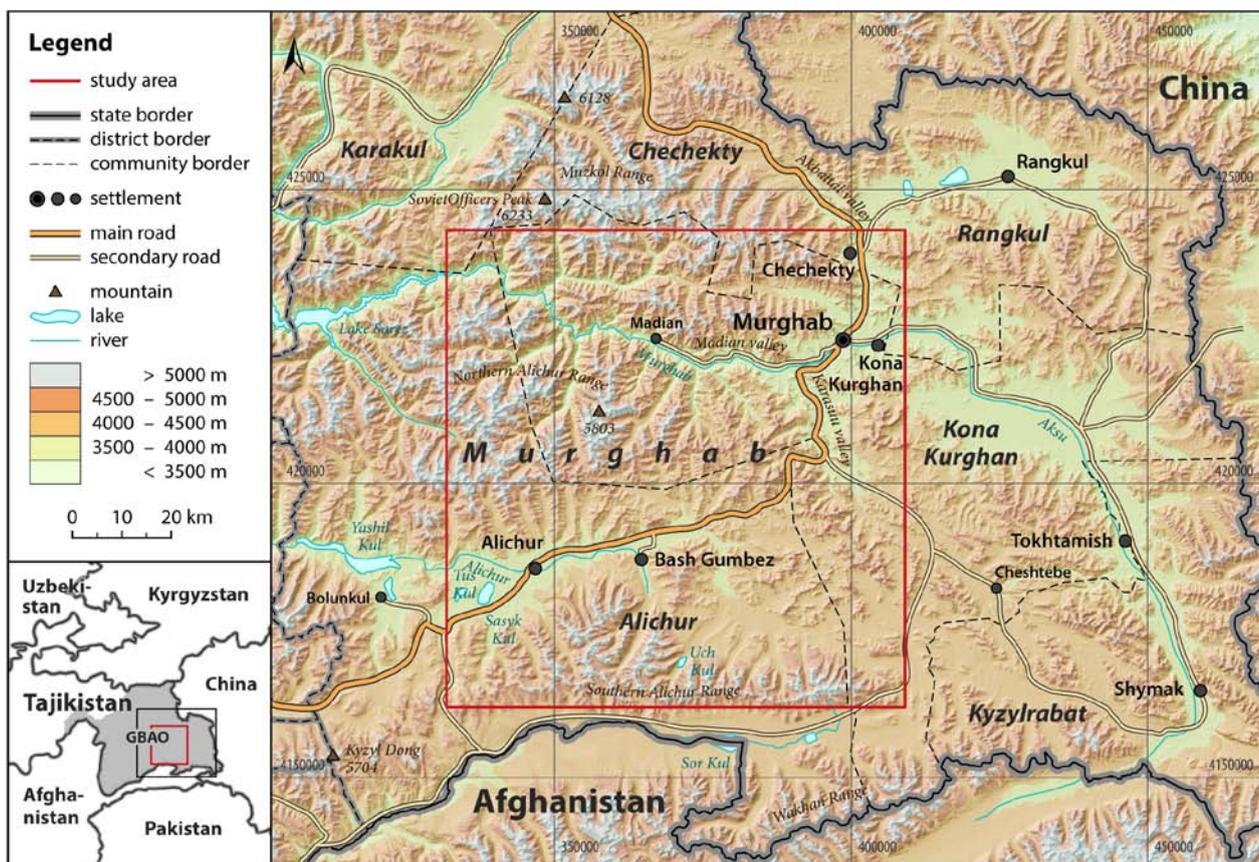
Due to the availability of satellite images, a rectangle of 9027 km² was selected as the focus area, which contains the region's major land cover types, as well as most of the settlements (Figure 1).

2.2. Explanatory Variable Data

In this work, exclusively remotely sensed data were used as explanatory variables. Altogether, 19 variables were produced and extracted from the ASTER Global Digital Elevation Model (GDEM) [16] and from RapidEye images [15] (Table 1). We used the georeferenced and orthorectified Level 3A product that is resampled to tiles of 25 km × 25 km with 5-m resolution [15]. Altogether, 16 tiles could be acquired that date between 2 and 21 August 2009, covering a total of 9027 km² of the Eastern Pamirs. First, the reflectances were calculated from the digital numbers (namely the pixel gray-level values) of the different tiles and bands. Second, because of the time differences between the tiles, these values were calibrated to one another for each band to exclude atmospheric differences, using linear regressions based on the values from intersections of multiple tiles. According to Song *et al.* [31], no further atmospheric correction was performed. Third, the calibrated tiles were mosaicked and used to produce a raster of Qi *et al.*'s [32] *MSAVI2* index, which is a spectral vegetation index that was invented especially to cope with soil noise. Furthermore, 6 rasters of texture parameters (Table 1) that were calculated by applying a range filter of 3 × 3 pixels to the rasters of the spectral bands and the

MSAVI2 were calculated. The associated standard deviation was used as a parameter for texture. Similar pixel values lead to a low standard deviation and indicated a homogeneous texture, whereas differing pixel values lead to high standard deviations that represented a heterogeneous texture [33].

Figure 1. Overview of the study area.



Topographic variables were calculated on the basis of the GDEM. First, the GDEM was preprocessed with the SAGA GIS modules to close gaps and fill sinks [34]. Then, elevation was directly extracted and slope and aspect were calculated using the *terrain* function of the R-package *raster* [35]. Aspect was transformed into north exposedness (\sin aspect) and east exposedness (\cos aspect) [36]. Furthermore, slope, aspect and the latitude were used to calculate the heat load of a plot [37].

Then, the median values of the field plots were extracted for the 17 rasters. To avoid redundant information, collinear variables were eliminated using an iterative VIF analysis (variance inflation factor; see [38]). The variable with the highest inflation factor was dropped, and the analysis was run again. This procedure was repeated, until all remaining variables had a VIF <3 [38]. Hence, 10 explanatory variables went into the analysis. This analysis caused the exclusion of common indices, like the NDVI.

Table 1. Description of the utilized explanatory variables (NIR = near-infrared, RE = red edge, R = red).

| No. | Explanatory Variable | Derived From | |
|-----|----------------------|---|------------------------------|
| 1 | Band 1 | Reflectance of RapidEye Band 1 (blue; 440–510 nm) | |
| 2 | Band 2 | Reflectance of RapidEye Band 2 (green; 520–590 nm) | |
| 3 | Band 3 | Reflectance of RapidEye Band 3 (red; 630–690 nm) | |
| 4 | Band 4 | Reflectance of RapidEye Band 4 (red edge; 690–730 nm) | |
| 5 | Band 5 | Reflectance of RapidEye Band 5 (near-infrared; 760–880 nm) | |
| 6 | NDVI | $\frac{\text{Band5} - \text{Band3}}{\text{Band5} + \text{Band3}}$ | |
| 7 | MSAVI2 | $= \frac{(2 \times \text{Band5} + 1 - \sqrt{(2 \times \text{Band5} + 1)^2 - 8 \times (\text{Band5} - \text{Band3})})}{2}$ | (Qi <i>et al.</i> [32]) |
| 8 | Band 1 texture | Standard deviation of a 3 × 3 pixel neighborhood calculated from Band 1 | |
| 9 | Band 2 texture | Standard deviation of a 3 × 3 pixel neighborhood calculated from Band 2 | |
| 10 | Band 3 texture | Standard deviation of a 3 × 3 pixel neighborhood calculated from Band 3 | |
| 11 | Band 4 texture | Standard deviation of a 3 × 3 pixel neighborhood calculated from Band 4 | |
| 12 | Band 5 texture | Standard deviation of a 3 × 3 pixel neighborhood calculated from Band 5 | |
| 13 | NDVI texture | Standard deviation of a 3 × 3 pixel neighborhood calculated from NDVI | |
| 14 | MSAVI2 texture | Standard deviation of a 3 × 3 pixel neighborhood calculated from MSAVI2 | |
| 15 | Elevation | derived from ASTER GDEM | |
| 16 | Slope | derived from ASTER GDEM using the terrain function | (Hijmans and van Etten [35]) |
| 17 | North exposedness | cos(aspect); aspect derived from ASTER GDEM using the terrain function | (Leyer and Wesche [36]) |
| 18 | East exposedness | sin(aspect); aspect derived from ASTER GDEM using the terrain function | (Leyer and Wesche [36]) |
| 19 | Heat load | $= -1.467 + 1.582 \times \cos(\text{latitude}) \times \cos(\text{slope}) - 1 \times \cos(\text{foldedaspect}) \times \sin(\text{slope}) \times \sin(\text{latitude}) - 0.262 \times \sin(\text{latitude}) \times \sin(\text{slope}) + 0.607 \times \sin(\text{foldedaspect}) \times \sin(\text{slope})$ | (McCune and Keon [37]) |

2.3. Sample Data and Statistical Analysis

For this study, data on the total vegetative ground cover, the dwarf shrub cover and the presence/absence of *Krascheninnikovia ceratoides* and *Artemisia leucotricha* were recorded. Altogether, 359 test plots were established, of which, 326 were compiled during four field campaigns (2007–2011) by GPS. Plots in water bodies (16) and snow-covered areas (17) were digitized on screen from RapidEye satellite images [15]. Because 31 plots were covered by clouds or located outside of the coverage of the available satellite images, information on total vegetative ground cover and dwarf shrub cover were present for 311 plots, while presence-absence data of dwarf shrub species were available for 281 plots. The plots were set up in selected valleys, distributed over the entire study area, with at least one plot on the valley bottom and one on each of the two valley slopes, wherever possible. The size of the plots

was 60×60 m. The size was calculated according to Justice and Townshend [39] and was related to the resolution of the ASTER GDEM. The presence/absence of the two relevant species, as well as the total vegetative ground cover and dwarf shrub cover were recorded.

The cover was estimated by percentage on four randomly selected 4×4 -m plots within each large plot. The data from the four plots were aggregated and their median considered as representative for the related larger 60×60 -m plot. Boxplots were used to visualize the sampled data and make comparisons to previous (Soviet) studies, as summarized under Section 2.1.

The mapping procedure was divided into four steps: First, spatial information on the vegetation attribute of interest (presence/absence of *K. ceratoides*/*A. leucotricha*; total and dwarf shrub vegetative ground cover) was compiled from the recorded field plots and used as a response variable. Second, digital maps derived from the satellite images served to determine the values of the explanatory variables at the locations of these plots. Third, these values were used to fit a model predicting the likelihood of the response variable. Finally, the model results were used to predict the likelihood of the response variable at all locations of the study area [40].

As vegetation-environment interaction behaves often non-linearly, traditional maximum likelihood methods, which assume the normal distribution of the training data, are ineligible [41]. In contrast, a feasible model approach that can deal with any type of distribution is the random forest algorithm introduced by Breiman [17], which is an advancement of single classification and regression trees. It is based on random subsets from the training data set, generated by bootstrap resampling [42,43]. This means that it generates a multitude of datasets by drawing a user-defined number of data points from the original data pool. After each iteration of the bootstrapping procedure, all values are returned to the data pool. Thus, it is possible for each value to be drawn more than one time during the generation of the entire bootstrap dataset. The data subsets are then used to grow a high number of single classification or regression trees (hence forest). The trees are successively split into smaller binary classes, utilizing the best fitting predictor within a randomly chosen subset of the explanatory variables. In this regard, a threshold value is chosen that maximizes the homogeneity of the two resulting classes with respect to the response variable [44]. Subsequently, the resulting single trees are combined to a stable final classification or regression [45]. If the response variable is a factor, random forests perform a classification. When processing continuous data, they execute regression analysis [42]. Another advantage of random forests is the so-called out-of-bag-validation. This is a bootstrap-based resampling technique intended to produce “honest” error estimates [42,45]. At each bootstrap iteration, the data that is not part of the bootstrap sample (*i.e.*, out-of-bag or OOB) is utilized as the test data. That is, it is predicted on the basis of the tree grown with the bootstrap sample. The predictions for all single trees grown with the bootstrap samples are aggregated, and the total error rate is calculated. As such, OOB accuracy estimates are efficiently cross-validated and unbiased, given that sufficient trees were fitted [17]. If regression analysis is performed, the “mean of squared residuals” (MSE) is computed with the OOB data (Equation (1)), where $y'_i{}^{OOB}$ is the average of the OOB predictions for the *i*-th observation. Based on this measure, the “percent variance explained” can be computed (Equation (2)), where σ_y^2 is computed with *n* as divisor (rather than n^{-1}).

$$MSE_{OOB} = n^{-1} \sum_1^n \{y_i - y'_i{}^{OOB}\}^2 \quad (1)$$

$$\% VarExplained = 1 - \frac{MSE_{OOB}}{\sigma_y^2} \quad (2)$$

In addition to the model fit, the function produces an importance value of the explanatory variables. It is based on the increment of the prediction error that occurs when the data for the examined variable is permuted, while the other variables remain unaltered. This value was used to reduce the set of explanatory variables and, thus, create simpler models. The less important variable was removed, and the model was calculated again. This step was repeated until the explained variance started to decrease [42].

All methodological steps were carried out with the R environment (<http://www.r-project.org>). Random forest models were calculated with the package, *randomForest* [46]. The package, *raster* [35], was used to apply the model results to the spatial data.

3. Results and Discussion

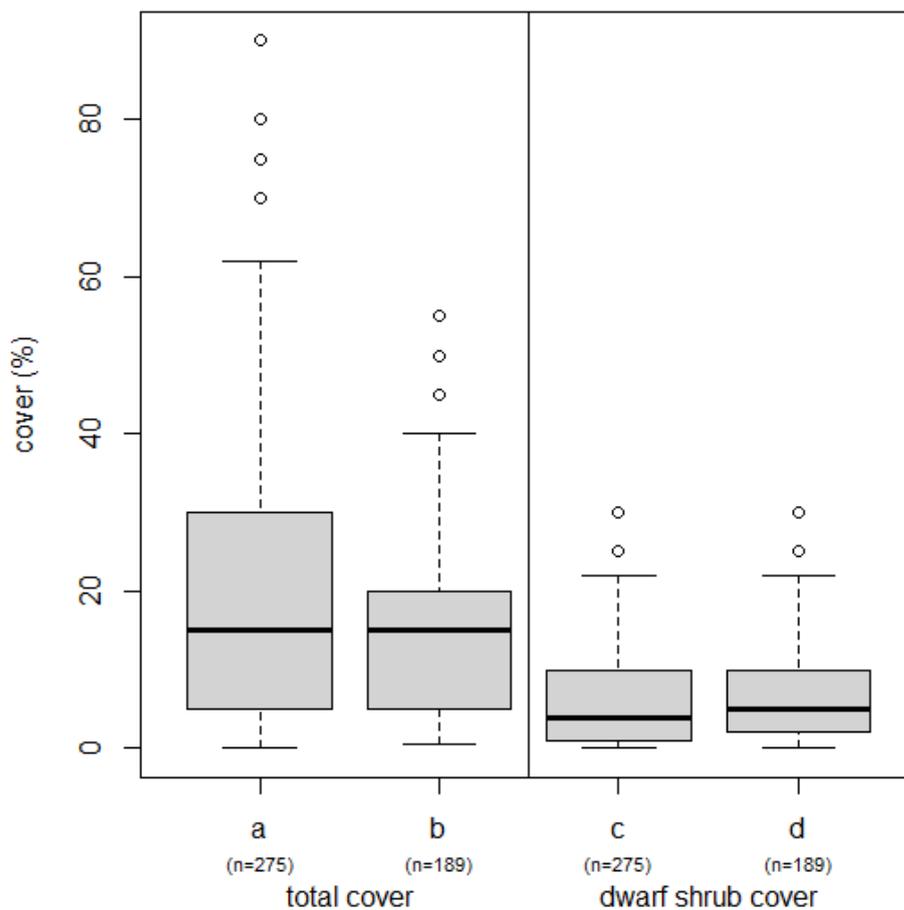
3.1. Descriptive Statistics

The total vegetative ground cover in all vegetation samples (Figure 2a, $n = 275$) ranged from 0% to 90%, while the majority of values (1.5 interquartile ranges) stayed below 62%. The median was 15%. Only 18 plots that all belong to the riparian meadows exceeded 70%. Focusing only on those samples with *K. ceratoides* and/or *A. leucotricha* (Figure 2b) the median was also 15%, and the majority of values ranged between 0.5% and 40%. Only five plots that are formed by dense *A. leucotricha* steppes show higher values of up to 55%. Dwarf shrub cover, regardless of whether for all or only the *K. ceratoides*/*A. leucotricha* plots, was estimated to values between near zero (only remnants of dwarf shrubs) in the vicinity of the town of Murghab and 30% in *A. leucotricha* dominated steppes on slopes in the southwest of the study area. Only six plots showed high coverage of more than 22%. The median was at 4% (all plots) and 5% (dwarf shrub plots), respectively. These results reflect very well those from the literature review (see Section 2.1).

3.2. Species Distribution Model

Altogether, three different classification models were calculated using the presence/absence of *K. ceratoides* and/or *A. leucotricha* as the response variable. The differentiation between the distribution areas of the two plants, based on optical remote sensing, turned out to be nearly impossible, due to a great overlap of their habitat and strong similarities regarding plant morphological and spectral properties. Particularly, the *A. leucotricha* model achieved a low accuracy (Table 2a). Although the *K. ceratoides* model performed much better (Table 2b), a joint model, which used the presence/absence of either of these plants as the response variable, was favored for further analysis. According to the variable importance, a model with four explanatory variables was selected, with *MSAVI2* as the most important predictor, followed by elevation, *MSAVI2* texture, and Band 5 reflectance (Figure 3a). The result was a common vegetation map, representing the presence and absence as discrete classes (Figure 4a).

Figure 2. Boxplots of total vegetative ground cover for all vegetation plots (a) and vegetation plots with the presence of *A. leucotricha* and/or *K. ceratoides* (b). Dwarf shrub ground cover for all vegetation plots (c) and vegetation plots with the presence of *A. leucotricha* and/or *K. ceratoides* (d).

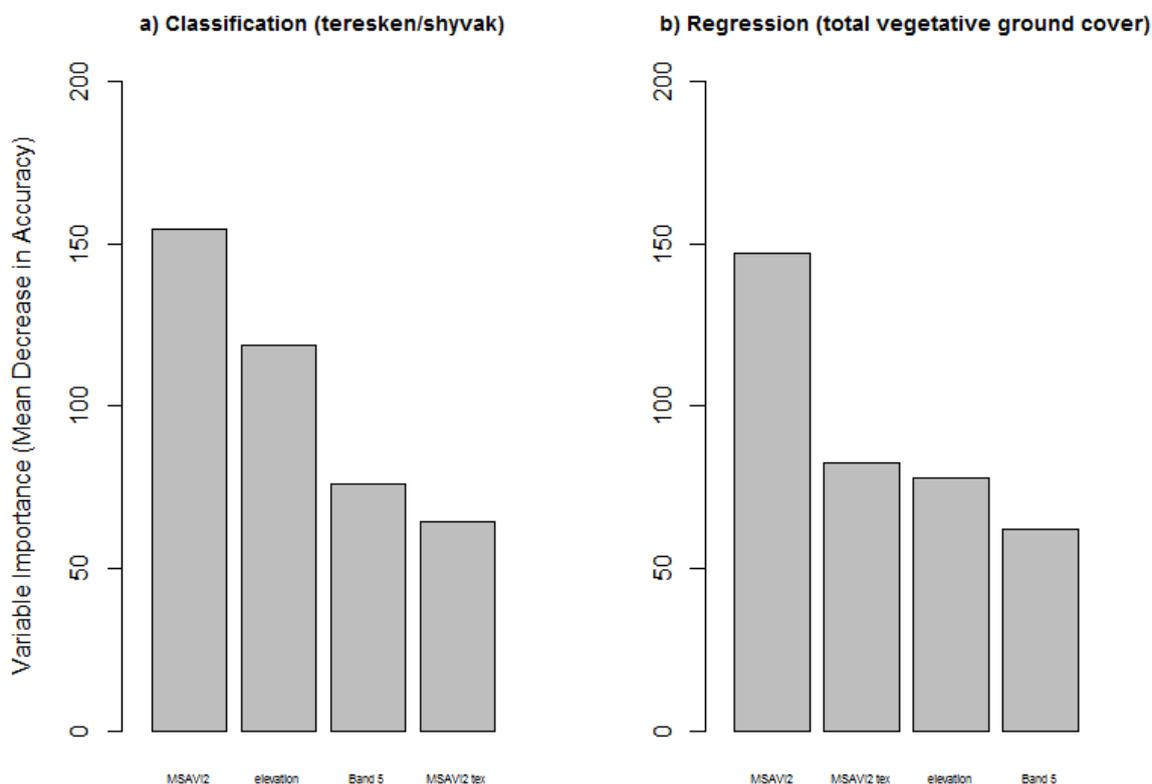


Compared with other studies about species composition, our results give a considerably accurate picture about the distribution of single species. The other studies on arid and similar mountain ecosystems do not map single species, but species composition or vegetation classes [4,5,7,8], plant richness [47]; however, using similar indices. Studies on species separation mainly use hyperspectral data often under experimental conditions [9]. As to expected, elevation was an important predictor apart from spectral information gained by remote sensing in our study region and in all studies in mountain areas [4,5,8].

Table 2. Description and evaluation of the calculated models (mtry: number of variables randomly sampled as candidates at each split; ntrees: number of trees grown for the model; OOB: out-of-bag).

| | a | b | c | d | e | f | g |
|---|-------------------------------------|--|---|---------------------|------------------------|---|---|
| Type | Classification | Class. | Class. | Class. | Class. | Regression | reg. |
| Response variable | pres./abs. <i>A. leucotricha</i> | pres./abs. <i>K. ceratoides</i> | pres./abs. <i>K. ceratoides</i> <i>A. leucotricha</i> | pres./abs. water | pres./abs. snow/ice | Total vegetative ground cover | dwarf shrub cover |
| <i>n</i> explanatory variables | 2 | 4 | 4 | 1 | 2 | 4 | 5 |
| Explanatory variables (in order of importance) | Heat load MSAVI2 | MSAVI2 Elevation MSAVI2 tex. B5 refl. | MSAVI2 Elevation MSAVI2 tex. B5 refl. | NDWI | Elevation B5 refl. | MSAVI2 MSAVI2 tex. Elevation B5 refl. | MSAVI2 Elevation Slope MSAVI2 tex. heat load |
| mtry | 2 | 2 | 2 | 1 | 1 | 2 | 3 |
| ntrees | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 | 5000 |
| <i>n</i> observations | 278 | 278 | 278 | 307 | 307 | 307 | 307 |
| <i>n</i> presence | 52 | 179 | 189 | 15 | 17 | | |
| <i>n</i> absence | 226 | 99 | 89 | 292 | 290 | | |
| correctly vs. wrongly classified (presence) | 30:22 | 170:9 | 181:8 | 15:0 | 17:0 | | |
| correctly vs. wrongly classified (absence) | 207:19 | 71:28 | 71:18 | 292:0 | 290:0 | | |
| OOB class. accuracy (%) (presence) | 57.7 | 95.0 | 95.8 | 100 | 100 | | |
| OOB class. accuracy (%) (absence) | 91.6 | 71.8 | 80.0 | 100 | 100 | | |
| OOB class. accuracy (%) (total) | 85.2 | 86.7 | 90.6 | 100 | 100 | | |
| OOB accuracy (%) var. expl. | | | | | | 80.2 | 47.5 |

Figure 3. Importance values of the explanatory variables included in the classification (a) and regression (b) model.

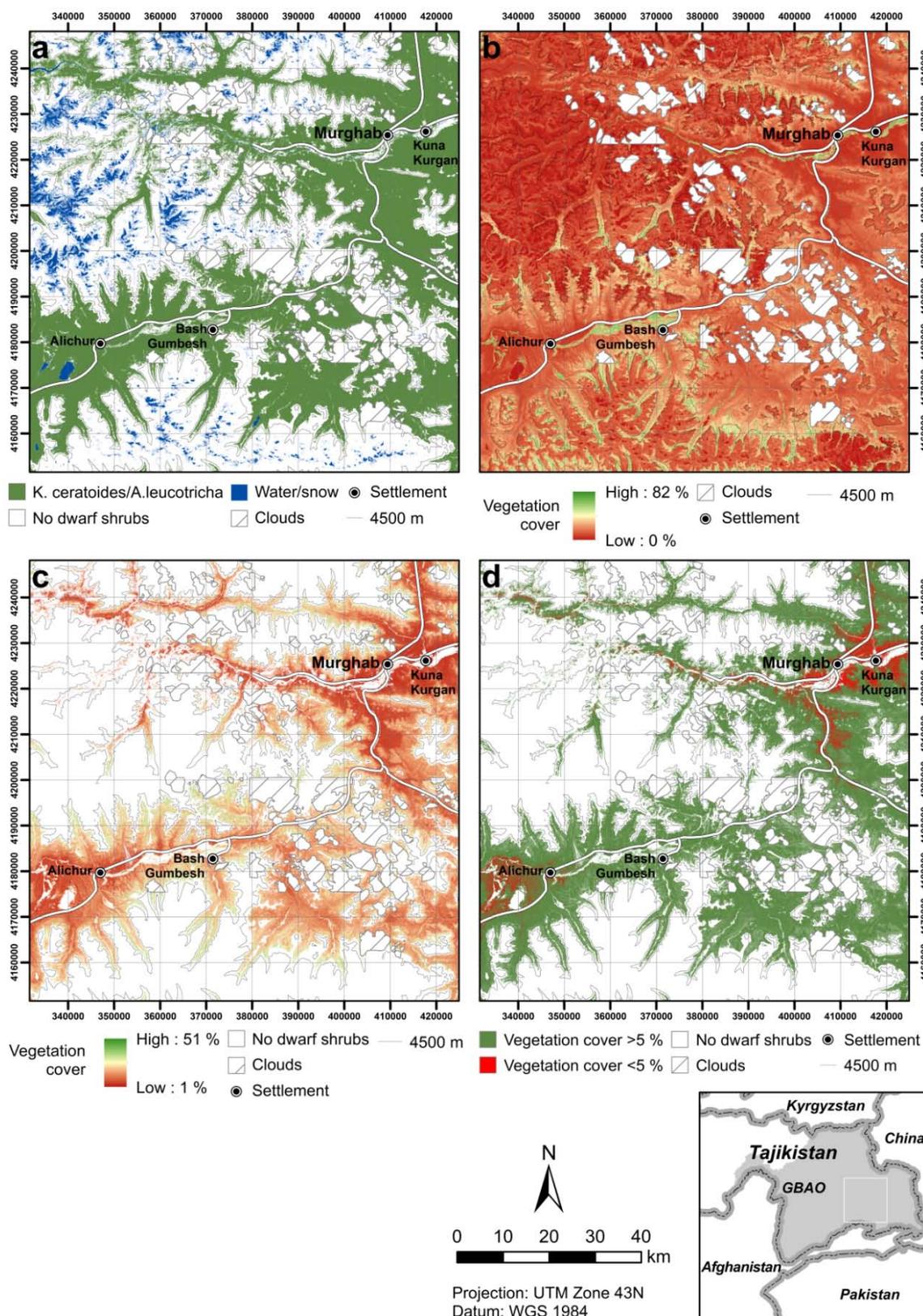


The map shows that firewood dwarf shrubs are still distributed over large parts of the Eastern Pamirs. They were mapped for an area of 3125 km², which is 33.8% of the entire study area. Their main occurrence is in the southwest (Alichur Pamir and adjacent valleys in the Northern and Southern Alichur Range below 4500 m a.s.l.), the southeast (high plateau of Uch Kul and Cheshtebe) and the northeast (Sarez Pamir, Kara Suu Valley, Madian Valley and Akbaital Valley). In the northwest, they are limited to narrow areas in valleys (Pshart and Madian Valley, adjacent valleys). Furthermore, they rarely grow in valley bottoms. In contrast, 5157 km² (57.1%) belong to other land cover classes. Within this area, 79 km² (0.9%) are covered by water bodies, of which, the lakes Tus Kul and Sasyk Kul (west of Alichur), and the eastern end of Lake Sarez (in the north-west), as well as Uch Kul (in the south) make the largest portion. Furthermore, the area covered by snow and ice accounts for 343 km² (3.8%) and is mainly located in the north-west of the study area. Eight percent was covered by clouds and, thus, could not be analyzed.

3.3. Vegetation Cover Model

In a second step, random forest regression models were performed with total vegetative ground cover and dwarf shrub ground cover as response variables. Altogether, information from 307 test plots was considered for the models. According to the variable importance, the final model for total vegetative ground cover was run with the same four explanatory variables, with *MSAVI2* being the most important, followed by *MSAVI2* texture, elevation and Band 5 reflectance (Figure 3b). The model did not perform as well in the prediction of dwarf shrub cover (see Section 3.3.).

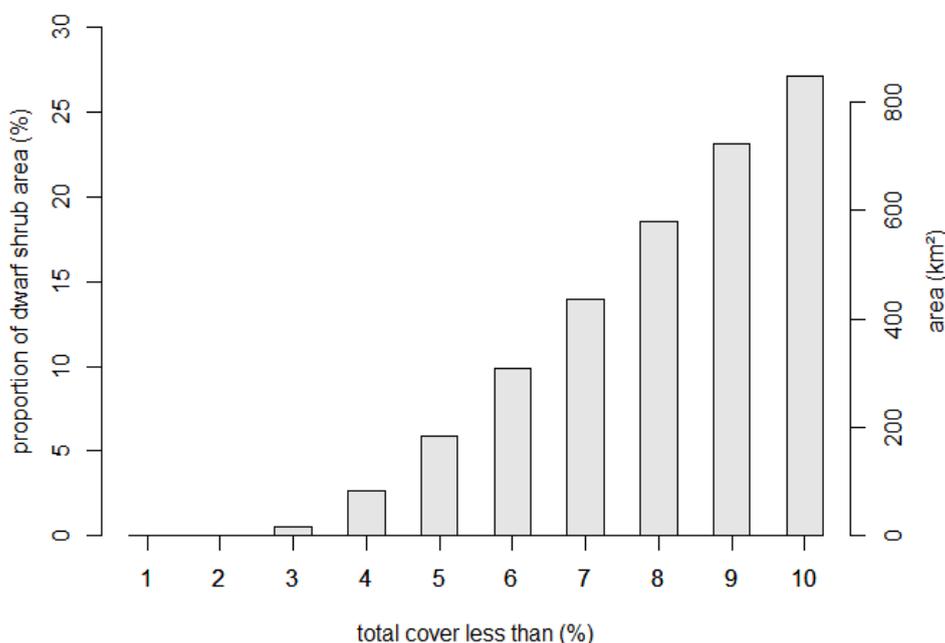
Figure 4. Vegetation map representing (a) the presence/absence of *A. leucotricha* and/or *K. ceratoides*. (b) The total vegetative ground cover. (c) The total vegetative ground cover for the area where *A. leucotricha* and/or *K. ceratoides* are present. (d) The area with less than 5% total vegetative ground cover for the area where *A. leucotricha* and/or *K. ceratoides* are present.



Therefore, in this paper, we omitted the dwarf shrub coverage and developed a continuous field map showing only total vegetative ground cover in percentage values (Figure 4b). The highest cover values were mapped in the valley bottoms and the large Pamirs. There, riparian *Carex* and *Kobresia* meadows with ground cover values sometimes exceeding 80% grow along rivers and streams (e.g., Alichur River between Alichur and Bash Gumbez; Murghab River in the Madian Valley west of the town of Murghab; see Figure 4b).

In a further step, non-dwarf shrub areas were masked using the output of the distribution model, so that Figure 4c shows the vegetation cover for the predictively mapped area of *K. ceratoides* and *A. leucotricha*. The values range from just above 0% to 51% (whereas values above 36% are rare). The lowest values (represented by bright colors in Figure 4c) predominantly occur around the larger settlements (Alichur, Kuna Kurgan/Murghab) and along the main roads, indicating the importance of facilitated access for degradation of dwarf shrub resources. It is very probable that these areas are the result of overexploitation due to dwarf shrub extraction [26]. However, they are also often dominated by unpalatable plants, such as *Zygophyllum obliquum* and *Neotorularia korolkowii*, which show that heavy grazing may also contribute to the situation. This would concur with the fact that the highest livestock numbers were recorded in the same area [22,24]. Social disparities among different groups of pastoralists and unsettled questions about formal user rights and rangeland stewardship are the main reasons for the concentration of livestock in this area [48,49]. On the whole, however, these areas are a minor share of the total area of dwarf shrub vegetation (Figure 5). For example, only 183 km², which is 5.9% of the dwarf shrub area or 2.0% of the entire study area, shows a total vegetative ground cover of less than 5% (Figure 4d). In contrast, 2278 km² (72.9%) of the dwarf shrub area shows cover values of more than 10%. Such high values were also mapped in the vicinity (less than 5 km) of all four settlements within the study area, at least on slopes.

Figure 5. Spatial extents of the areas with less than 1% to less than 10% total vegetative ground cover.



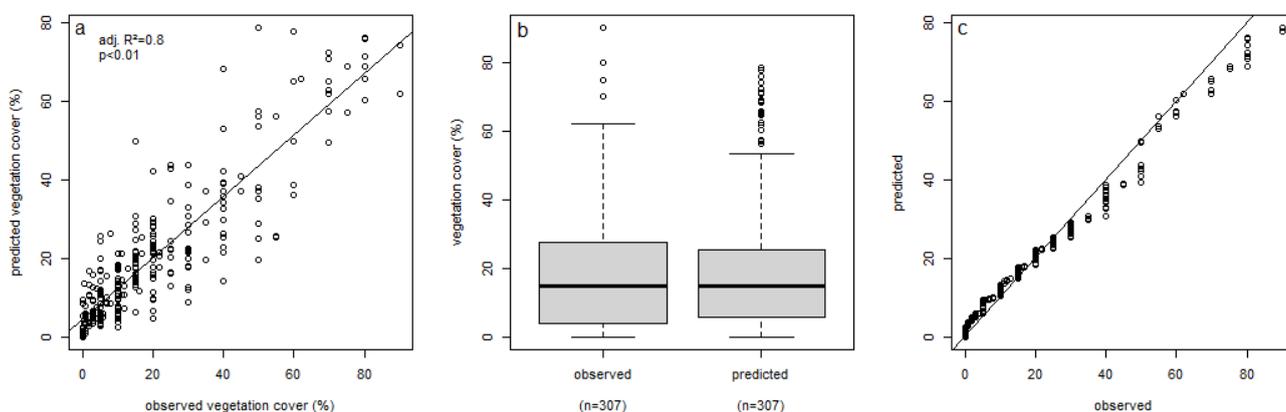
3.4. Model Evaluation

Validation of the combined *K. ceratoides/A. leucotricha* classification model, based on OOB estimation, indicated an overall accuracy of 90.6% (presence 95.8%, absence 80%, Table 2c). Water and snow/ice were mapped with individual models that were 100% accurate (Table 2d,e).

The variance of dwarf shrub cover could be explained to 47.5% (Table 2g). A better result was achieved for the total vegetative ground cover model, where 80.2% of the variance could be explained (Table 2f), a result similar to other studies in arid, respectively mountainous, regions, e.g., [5,47].

For the latter model, comparisons between observed and predicted values showed good agreement (adjusted $R^2 = 0.8$; $p < 0.01$; Figure 6a). However, it has to be mentioned that tree-based models, such as the random forest, tend to overestimate small and underestimate high values [50,51]. This was also the case in this study. Although the average and the median between observed and predicted values indicated a good agreement, significant differences were detected for low and high values. The 0.25 quantile of the predicted values exceeded its counterpart for the observed values by nearly 2%. In contrast, the 0.75 quantile and the maximum were underestimated by 1.3% and 11.5%, respectively (Figure 6b). The differences between observed and predicted values were visualized by quantile-quantile plots (Figure 6c) [52]. It can be clearly seen that values below 20% observed cover were slightly overestimated by the model, whereas values over 25% tend to be underestimated. Taking this finding into account, it is likely that the area displayed in Figure 4d represents stands with less than 3% vegetation cover, rather than 5%.

Figure 6. Comparison of observed and predicted values with linear regression (a), boxplot (b), and quantile-quantile plot (c).



4. Conclusions

In summary, it can be stated that the application of the presented mapping approach led to suitable and reproducible vegetation maps in an arid, mountainous environment, such as the Eastern Pamirs. The presence of firewood dwarf shrub areas is still widespread within the study area (33.8%). However, extremely low vegetative ground cover could be detected around the permanent population centers and along roads, which is an indication of the partial overuse of dwarf shrubs.

Therefore, the situation of the region's dwarf shrub resources has to be rated as vulnerable [24,26]. It stands to reason that degradation could expand with increasing pasture exploitation and firewood extraction. The maps produced in this study provide a first step towards an instrument for the sustainable management of the mountain commons in this region.

From the technical point of view, *MSAVI2* was identified as the most important explanatory variable for mapping dwarf shrub vegetation in this arid and mountainous environment. This is in contrast to other studies [47] and demonstrates the usefulness of this variable for remote sensing vegetation studies in arid lands. Among the single-band reflectances, only Band 5 reflectance (IR) significantly contributed to the model accuracies. Furthermore, *MSAVI2* texture increased the predictive power of most of the performed models. This reflects the spatial heterogeneity of vegetation cover in arid regions. Elevation, which is a proxy for temperature, was the only relevant topographic variable for the accuracy of the models. Additionally, slope and aspect were used to calculate heat load. Only the *Artemisia leucotricha* model strongly depended on heat load.

However, none of the implemented models were able to distinguish between the two major dwarf shrub species. Furthermore, the model accuracy for the estimation of dwarf shrub cover was relatively low (explained variance 47.5%). Hence, the study underlines the difficulty of predicting vegetation parameters under arid conditions with sparse vegetation cover, even with a spatial resolution of 6.5 m × 6.5 m and the red edge band (RapidEye). Future studies should therefore consider technical innovations, such as hyperspectral data, to achieve better performance.

Acknowledgments

Empirical research in the Pamirs was made possible by two projects funded by the Volkswagen Foundation (“Transformation Processes in the Eastern Pamirs of Tajikistan. Changing Land Use Practices, Possible Ecological Degradation and Sustainable Development”, AZ I/81 976) and Deutsche Forschungsgemeinschaft (“Distribution modeling of the most important forage and fuel plants in the Eastern Pamirs of Tajikistan”, VA 749/1-1). The authors acknowledge support by Deutsche Forschungsgemeinschaft and Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) within the funding programme Open Access Publishing. Furthermore, the authors would like to thank Tobias Kraudzun for his support during the field work. Finally, the authors are grateful to three anonymous reviewers for their constructive and helpful comments, which substantially improved earlier drafts of this paper.

Author Contributions

Kim André Vanselow and Cyrus Samimi recorded the field data. Kim André Vanselow analyzed the data and wrote the paper. Results and text were discussed with Cyrus Samimi who contributed important ideas and considerations.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. UNEP. *Global Environment Outlook 3—Past, Present and Future Perspectives*; Earthscan Publications: London, UK; Sterling, VA, USA, 2002.
2. Xie, Y.; Sha, Z.; Yu, M. Remote sensing imagery in vegetation mapping: A review. *J. Plant Ecol.* **2008**, *1*, 9–23.
3. Kerr, J.T.; Ostrovsky, M. From space to species: Ecological applications for remote sensing. *Trends Ecol. Evolut.* **2003**, *18*, 299–305.
4. Brinkmann, K.; Patzelt, A.; Schlecht, E.; Buerkert, A. Use of environmental predictors for vegetation mapping in semi-arid mountain rangelands and the determination of conservation hotspots. *Appl. Veg. Sci.* **2011**, *14*, 17–30.
5. Dobrowski, S.Z.; Safford, H.D.; Cheng, Y.B.; Ustin, S.L. Mapping mountain vegetation using species distribution modeling, image-based texture analysis, and object-based classification. *Appl. Veg. Sci.* **2008**, *11*, 499–508.
6. Münchow, J.; Bräuning, A.; Rodríguez, E.F.; von Wehrden, H. Predictive mapping of species richness and plant species' distributions of a peruvian fog oasis along an altitudinal gradient. *Biotropica* **2013**, *45*, 557–566.
7. Van Etten, E.J.B. Mapping vegetation in an arid, mountainous region of western Australia. *Appl. Veg. Sci.* **1998**, *1*, 189–200.
8. Von Wehrden, H.; Zimmermann, H.; Hanspach, J.; Ronnenberg, K.; Wesche, K. Predictive mapping of plant species and communities using GIS and Landsat data in a southern Mongolian mountain range. *Folia Geobot.* **2009**, *44*, 211–225.
9. Irisarri, J.G.N.; Oesterheld, M.; Verón, S.R.; Paruelo, J.M. Grass species differentiation through canopy hyperspectral reflectance. *Int. J. Remote Sens.* **2009**, *30*, 5959–5975.
10. Dirnböck, T.; Dullinger, S.; Gottfried, M.; Ginzier, C.; Grabherr, G. Mapping alpine vegetation based on image analysis, topographic variables and canonical correspondence analysis. *Appl. Veg. Sci.* **2003**, *6*, 85–96.
11. Mansour, K.; Mutanga, O.; Everson, T. Remote sensing based indicators of vegetation species for assessing rangeland degradation: Opportunities and challenges. *Afr. J. Agric. Res.* **2012**, *7*, 3261–3270.
12. Leprieux, C.; Kerr, Y.H.; Mastorchio, S.; Meunier, J.C. Monitoring vegetation cover across semi-arid regions: Comparison of remote observations from various scales. *Int. J. Remote Sens.* **2000**, *21*, 281–300.
13. Escadafal, R.; Huete, A.R. Soil Optical Properties and Environmental Applications of Remote Sensing. Available online: http://www.isprs.org/proceedings/xxix/congress/part7/709_XXIX-part7.pdf (accessed on 22 July 2014).
14. Franklin, J.; Miller, J.A. *Mapping Species Distributions. Spatial Inference and Prediction*; Cambridge University Press: Cambridge, UK, 2009.
15. Rapideye Homepage. Available online: <http://www.rapideye.com> (accessed on 19 March 2013).
16. Fujisada, H.; Bailey, G.B.; Kelly, G.G.; Hara, S.; Abrams, M.J. ASTER DEM performance. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2707–2713.
17. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

18. Walter, H.; Breckle, S. *Ökologie der Erde Bd3. Spezielle Ökologie der gemäßigten und arktischen Zonen Euro-Nordasiens. Zonobiom VI–IX*; Fischer: Stuttgart, Germany, 1991; Volume 3.
19. Agakhanjan, O.E. Besonderheiten in der natur der ariden gebirge der udssr. *Petermanns Geogr. Mitt.* **1979**, *123*, 73–77.
20. Miede, G.; Winiger, M.; Böhner, J.; Yili, Z. The climatic diagram map of high Asia. Purpose and concepts. *Erdkunde* **2001**, *55*, 94–97.
21. Vanselow, K.A. The High-Mountain Pastures of the Eastern Pamirs (Tajikistan)—An Evaluation of the Ecological Basis and the Pasture Potential. Ph.D. Thesis, University of Erlangen-Nuremberg, Erlangen, Germany, 2011.
22. Kraudzun, T. Livelihoods of the “new livestock breeders” in the eastern pamirs of tajikistan. In *Pastoral Practices in High Asia—Agency of “Development” Effected by Modernisation, Resettlement and Transformation*; Kreutzmann, H., Ed.; Springer: Dordrecht, The Netherlands, 2012; pp. 89–107.
23. Breckle, S.; Wucherer, W. Vegetation of the Pamir (Tajikistan): Land use and desertification problems. In *Land-Use Change and Mountain Biodiversity*; Spehn, E., Liberman, M., Körner, C., Eds.; CRC/Taylor & Francis: Boca Raton, FL, USA, 2006; pp. 225–237.
24. Vanselow, K.A.; Kraudzun, T.; Samimi, C. Grazing practices and pasture tenure in the eastern Pamirs: The nexus of pasture use, pasture potential and property rights. *Mt. Res. Dev.* **2012**, *32*, 324–336.
25. Akhmadov, K.; Breckle, S.; Breckle, U. Effects of grazing on biodiversity, productivity, and soil erosion of alpine pastures in Tajik Mountains. In *Land Use Change and Mountain Biodiversity*; Spehn, E., Liberman, M., Körner, C., Eds.; CRC/Taylor & Francis: Boca Raton, FL, USA, 2006; pp. 239–247.
26. Kraudzun, T.; Vanselow, K.A.; Samimi, C. Realities and myths of the teresken syndrome—An evaluation of the exploitation of dwarf Shrub resources in the eastern Pamirs of Tajikistan. *J. Environ. Manag.* **2014**, *132*, 49–59.
27. Jusufbekov, C.J.; Kasach, A.E. *Teresken na Pamire*; Izd. Donish: Dushanbe, Tajikistan, 1972.
28. Agachanjan, O.E. *Osnovnie Problemy Fizicheskoy Geografii Pamira*; Izd. Akademija Nauk Tadshikskoj SSR: Dushanbe, Tajikistan, 1966.
29. Ladygina, G.M.; Litvinova, N.P. Produktivnost Nadsemnoj i Podsemnoj Fitomassy Nekotorych Ractitel’nych Soobshchestv Vycokogorij Pamira. In *Biological Productivity and Mineral Cycling*; Rodin, L.E., Smirnof, N.N., Eds.; Akademija Nauk CCCP: Leningrad, Russian, 1971; pp. 150–156.
30. Litvinova, N.P. *Biologicheskaja Produktivnost’ Pustynych Soobshchestv Vostochnogo Pamira*; Akademija Nauk CCCP: Leningrad, Russian, 1969.
31. Song, C.; Woodcock, C.E.; Seto, K.C.; Lenney, M.P.; Macomber, S.A. Classification and change detection using Landsat TM data: When and how to correct atmospheric effects? *Remote Sens. Environ.* **2001**, *75*, 230–244.
32. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, *48*, 119–126.
33. Smith, R.B. *Filtering Images with Tntmips*; MicroImages: Lincoln, NE, USA, 2006.

34. Böhner, J.; Conrad, O.; Köthe, R.; Ringeler, A. *Saga—System for Automated Geoscientific Analyses*, 2.0.4; Göttingen: Hamburg, Germany, 2009.
35. Hijmans, R.J.; van Etten, J. *Raster: Geographic Analysis and Modeling*, 2.1–16; Available online: <http://cran.at.r-project.org/web/packages/raster/raster.pdf> (accessed on 22 July 2014).
36. Leyer, I.; Wesche, K. *Multivariate Statistik in der Ökologie. Eine Einführung*; Springer: Berlin/Heidelberg, Germany, 2007.
37. McCune, B.; Keon, D. Equations for potential annual direct incident radiation and heat load. *J. Veg. Sci.* **2002**, *13*, 603–606.
38. Zuur, A.F.; Ieno, E.N.; Elphick, C.S. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evolut.* **2010**, *1*, 3–14.
39. Justice, C.O.; Townshend, J.G. Integrating ground data with remote sensing. In *Terrain Analysis and Remote Sensing*; Townshend, J.G., Ed.; Allen and Unwin: London, UK, 1981; pp. 38–58.
40. Ferrier, S.; Watson, G.; Pearce, J.; Drielsma, M. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodivers. Conserv.* **2002**, *11*, 2275–2307.
41. Horning, N. Random forests: An algorithm for image classification and generation of continuous fields data sets. In Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Osaka, Japan, 9–11 December 2010.
42. Liaw, A.; Wiener, M. Classification and regression by randomforest. *R News* **2002**, *2*, 18–22.
43. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman & Hall: New York, NY, USA, 1993.
44. Prasad, A.; Iverson, L.; Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **2006**, *9*, 181–199.
45. Brenning, A. Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection. *Remote Sens. Environ.* **2009**, *113*, 239–247.
46. Liaw, A.; Wiener, M. *Randomforest: Breiman and Cutler's Random Forests for Classification and Regression*, 4.6–7; 2012. Available online: [online: http://cran.r-project.org/web/packages/randomForest/randomForest.pdf](http://cran.r-project.org/web/packages/randomForest/randomForest.pdf) (accessed on 22 July 2014).
47. Gaitán, J.J.; Bran, D.; Oliva, G.; Ciari, G.; Nakamatsu, V.; Salomone, J.; Ferrante, D.; Buono, G.; Massara, V.; Humano, G.; *et al.* Evaluating the performance of multiple remote sensing indices to predict the spatial variability of ecosystem structure and functioning in patagonian steppes. *Ecol. Indic.* **2013**, *34*, 181–191.
48. Vanselow, K.A.; Kraudzun, T.; Samimi, C. Land stewardship in practice: An example from the eastern Pamirs of Tajikistan. In *Rangeland Stewardship in Central Asia*; Squires, V., Ed.; Springer: Berlin, Germany, 2012; pp. 71–90.
49. Kreuzmann, H. The tragedy of responsibility in high Asia: Modernizing traditional pastoral practices and preserving modernist worldviews. *Pastor.: Res. Policy Pract.* **2013**, *3*, doi:10.1186/2041-7136-3-7.
50. Baccini, A.; Friedl, M.A.; Woodcock, C.E.; Warbington, R. Forest biomass estimation over regional scales using multisource data. *Geophys. Res. Lett.* **2004**, *31*, L10501.
51. Horning, N. Training Guide for Creating Percent Cover Images. Available online: <http://biodiversityinformatics.amnh.org> (accessed on 8 April 2013).

52. Caires, S.; Sterl, A. A new nonparametric method to correct model data: Application to significant wave height from the ERA-40 re-analysis. *J. Atmos. Ocean. Technol.* **2005**, *22*, 443–459.

© 2014 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 license (<http://creativecommons.org/licenses/by/3.0/>).