BioOne COMPLETE

Improving the Accuracy of Vegetation Classifications in Mountainous Areas

Authors: Gartzia, Maite, Alados, Concepción L., Pérez-Cabello, Fernando, and Bueno, C. Guillermo

Source: Mountain Research and Development, 33(1) : 63-74

Published By: International Mountain Society

URL: https://doi.org/10.1659/MRD-JOURNAL-D-12-00011.1

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Mountain Research and Development (MRD)
An international, peer-reviewed open access journal
published by the International Mountain Society (IMS)
www.mrd-journal.org

MountainResearch
Systems knowledge

Improving the Accuracy of Vegetation Classifications in Mountainous Areas

A Case Study in the Spanish Central Pyrenees

Maite Gartzia[1], Concepción L. Alados[2], Fernando Pérez-Cabello[3], and C. Guillermo Bueno[4]*

** Corresponding author: maitegartzia@yahoo.es*

[1] Pyrenean Institute of Ecology (CSIC), Avda De Nuestra Señora de la Victoria s/n, PO Box 64, Jaca E-22700, Spain

[2] Pyrenean Institute of Ecology (CSIC), Avda Montañana 1005, PO Box 13034, E-50192 Zaragoza, Spain

[3] University of Zaragoza, Department of Geography and Spatial Management, Pedro Cerbuna 12, E-5009 Zaragoza, Spain

[4] University of Alberta, Department of Biological Sciences, CW 405, Biological Sciences Building, Edmonton, Alberta T6G 2E9, Canada

Open access article: please credit the authors and the full source.

In recent decades, mountainous areas that contain some of the best-preserved habitats worldwide are experiencing significant, rapid changes. Efficient monitoring of these areas is crucial for impact assessments, understanding the key processes underlying the changes, and development of measures that mitigate degradation. Remote sensing is an efficient, cost-effective means of monitoring landscapes. One of the main challenges in the development of remote sensing techniques is improving classification accuracy, which is complicated in mountainous areas because of the rugged topography. This study evaluated the 3 main steps in the supervised vegetation classification of a mountainous area in the Spanish Pyrenees using Landsat-5 Thematic Mapper imagery. The steps were (1) choosing the training data sampling type (expert supervised or random selection), (2) deciding whether to include ancillary data, and (3) selecting a classification algorithm. The combination (in order of importance) of randomly selected training data, ancillary data (topographic and vegetation index), and a random forest classifier improved classification accuracy significantly (4–11%) in the study area in the Spanish Pyrenees. The classification procedure includes important steps that improve classification accuracies; these are often ignored in standard vegetation classification protocols. Improved accuracy is vital to the study of landscape changes in highly sensitive mountain ecosystems.

Keywords: *Supervised classification; remote sensing; accuracy improvement; training data; random forest; ancillary data; Spain.*

Peer-reviewed: *November 2012* ***Accepted:*** *December 2012*

Introduction

Mountainous areas contain some of the best-preserved natural habitats, but they are threatened by changes in land use and climate (Thuiller et al 2005). The trend in European mountains has been a reduction in grazing by extensive livestock (Luick 1998; Kozak 2003; Torta 2004; Lasanta-Martínez et al 2005), and its consequences in terms of the structural and compositional changes in ecosystems, eg woody encroachment and reduced biodiversity, are worrisome (Dirnböck et al 2003; Komac et al 2011a). Efficient monitoring of these ecosystems is crucial for assessing the impact of changes, detecting thresholds in ecosystem responses and degradation, and ultimately, for developing early measures to mitigate these changes. However, monitoring mountain areas is expensive and time consuming because of the inaccessibility and complexity of the terrain (highly variable aspect, slope, and elevation), which at the same time lead to different plant communities. In this respect, remote sensing and geographic information systems play an important and cost-effective role in monitoring landscape changes (Poyatos et al 2003; Pôças et al 2011b). Although several types of satellite imagery are used for vegetation mapping at the local or regional level, such as SPOT, IKONOS, ASTER, and Landsat (Lu and Weng 2007), we chose Landsat-5 Thematic Mapper (TM) because the platform has been providing images since 1984, which makes it the longest series available. The long-term data allow comparisons between periods, which increases the value of these comparisons (Pôças et al 2011a; Rodriguez-Galiano and Chica-Olmo 2012).

Since the 1970s, remote sensing data have been used for supervised vegetation classifications (Shalaby and Tateishi 2007; Aragón and Oesterheld 2008), and there have been efforts to improve classification accuracy (Chen and Stow 2002; Pal and Mather 2006; Tso and

 © 2013 by the authors

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Mather 2009). Although new classification methods have been developed, no single procedure has universal applicability (Lu and Weng 2007). The type of environment, scale, type and amount of training data, and availability of ancillary data influence the accuracy of classifications. In addition, classifications are more complicated and challenging in mountains because of the variable topography and potential for shadows (Shi et al 2009).

In mountainous areas, even small improvements in classification accuracy can be vital in detecting changes in vegetation, which tend to be prolonged and can thus go undetected in the short term; eg shrub encroachment on subalpine grassland can advance 2 m/y (Komac et al 2011b). At each step of the classification process, changes can be made to improvement classification accuracy. The type of training data used is important in the classification process, because it dictates the information that the algorithm will use to classify the vegetation. The training data must be mutually exclusive and exhaustive (Chuvieco and Huete 2009); therefore, selection should maximize the separability among categories and ensure that enough variability is retained so that the data are representative of each category. There is a tradeoff, because variability and separability are negatively correlated. Furthermore, the selection and addition of ancillary data can be critical steps for improving classification accuracy, especially in complex environments such as mountainous areas. In these areas, topography can influence the reflectance of the vegetation because of surface roughness, shadows, presence of rocks or bare soil, and steep slopes, which add inaccuracy to the values measured passively by satellite (Chuvieco and Huete 2009; Tso and Mather 2009; Balthazar et al 2012).

Methods of classifying vegetation have to be chosen, and numerous algorithms are available (Lu and Weng 2007; Tso and Mather 2009). In this study, we compared 2 classifiers, maximum likelihood (ML) and random forest (RF). The ML algorithm takes into account the variance and covariance of each category to be classified. When assigning each pixel to one of the categories, the algorithm assumes normality in the distribution of the values in each category. Therefore, given the mean and the covariance matrix of each category, the statistical probability is computed, and its ultimate category is the one that has the highest probability (ML) (Lillesand et al 2008). The ML classifier is commonly used in remote sensing (Lu and Weng 2007; Liu et al 2011), but it is a parametric classifier that is noise sensitive (Adams and Gillespie 2006).

The RF classifier is based on a combination of tree decision classifiers, in which each tree is generated using a random vector that is sampled independently of the training set of input variables (Breiman 2001). Although tree decision classifiers are independent of the distribution of the training data, the combination of the tree decision classifiers—which uses a majority voting scheme for each variable—from all trees provides a stable and stronger result (Tso and Mather 2009). The RF algorithm is used in ecology (Prasad et al 2006; Cutler et al 2007), remote sensing (Pal 2005; Rodriguez-Galiano and Chica-Olmo 2012), and vegetation classification (Ham et al 2005; Gislason et al 2006; Chan and Paelinckx 2008; Wessels et al 2011; Rodriguez-Galiano et al 2012). Because it is not dependent on parametric restrictions, it is easy to fine tune and executes quickly (Breiman 2001; Pal 2005; Prasad et al 2006; Cutler et al 2007). In addition, RF is robust to collinearity and overfitting of the variables (Breiman 2001), which can be particularly useful in mountain areas, where topographic variables tend to be highly collinear. RF can thus improve vegetation classifications, because topography and vegetation are highly correlated (Franklin et al 1994; De la Riva 1997; Shi et al 2009).

The objective of this study was to identify the best combination of classification steps to improve the accuracy of land-cover classifications, using Landsat-5 TM images, from among the selection of training data, ancillary data, and classifier. The choice of a classifier (Rodriguez-Galiano et al 2012) and the selection of ancillary data (Rodriguez-Galiano and Chica-Olmo 2012) can significantly improve the accuracy of the classification, but the effects of the selection of training data and the combinations of the options selected in each step need to be improved.

Methods

Study area

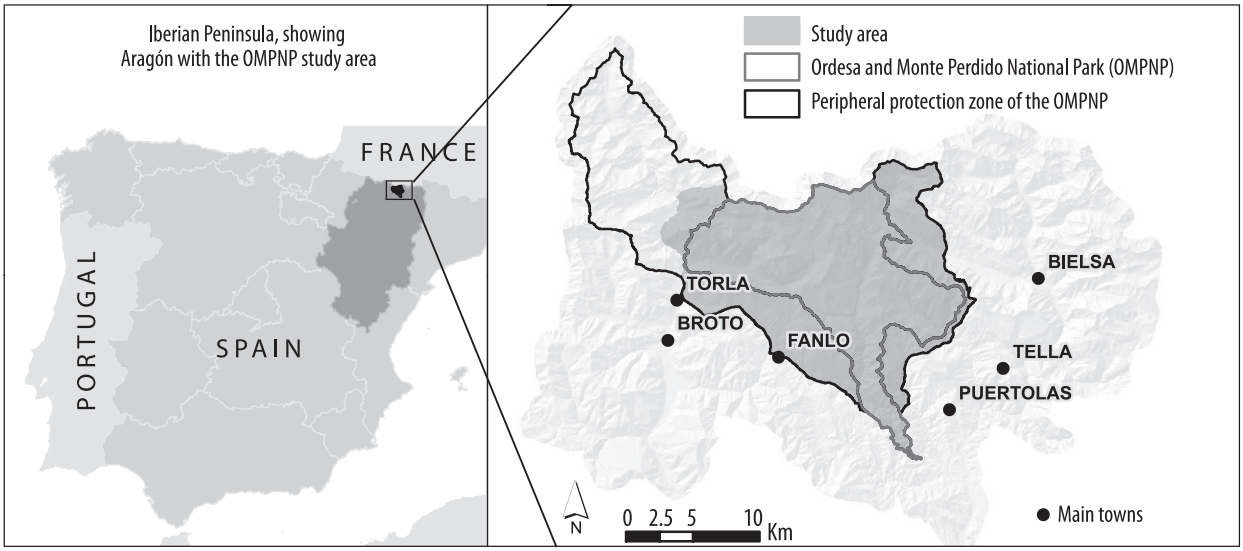
The 25,970-ha study area included Ordesa and Monte Perdido National Park and the surrounding protected area but excluded the northwest portion of the Ara River and Lapazosa Cliff in the Central Pyrenees, Aragón, Spain (42°36′N, 0°00′E) (Figure 1). Elevation was between 650 and 3355 masl. The area has a mountain climate characterized by an equinoctial precipitation regime and a continental influence. Average annual precipitation is 1657 mm, and the average daily maximum and minimum temperatures (at 2200 m) are 8.6 and 1.4°C, respectively.

Vegetation categories

Vegetation in the study area is of 4 main physiognomic types (Figure 2). These are strongly influenced by climate conditions and follow an elevation gradient: (1) high elevation (>2200 masl), sparse grassland (>60% rock cover); (2) dense grassland, which forms a continuum of vegetation cover that is interspersed with (3) shrubland at elevations below 2200 m; and (4) dense forests, which are below 1700 masl. Multiple plant communities are represented in each of the 4 categories. To identify the strata in the stratified sampling procedure, we used

http://dx.doi.org/10.1659/MRD-JOURNAL-D-12-00011.1
Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

FIGURE 1 Study area in Ordesa and Monte Perdido National Park (OMPNP), Spanish Central Pyrenees (42°36′N, 0°00′E). (Map by M. Gartzia)



existing vegetation maps of the study area (García-González et al 2007; Komac 2010; see also the section on random training).

Landsat-5 TM and multilayer composition

Since 1984, Landsat-5 has operated using the TM sensor, which was designed for thematic cartography and is useful for environmental monitoring (image spatial resolution = 30 m; 120 m in the thermal infrared). The analysis used cloud-free images taken on 4 August 2007 by the European Space Agency, which were orthorectified in Erdas 9.2 using a digital elevation model (DEM) with a 20-m resolution and orthophotographs from 2006. In the 2006 photographs, much of the study area up to 2000 m was covered with snow; therefore, the analysis included orthophotographs taken in 1997 and both had 0.5-m spatial resolution. The root mean square error of the orthorectified image was <0.5 pixels. An ATCOR 3 module was used to convert digital values to reflectance and remove atmospheric and topographic effects (based on the Lambertian assumption), and MODTRAN 4 was used to generate the lookup table. The correction was performed following the calibration parameters described by Chander et al 2009 for images taken after 31 December 1991, using 60-km scene visibility, a midland summer rural model for the solar region, a solar elevation angle of 57°, and a solar azimuth angle of 136°, as described in the image metadata. In ATCOR 3, a 10-m resolution DEM was used to correct errors caused by variations in elevation, slope, and aspect, because it was desirable to have a spatial resolution close to 0.25 times the pixel size (Geosystems 2009).

For the classification, we used 2 multilayers (Figure 3), which form a composite image that contains the bands that represent each of the variables included in the analysis. The layers differed in their spectral composition and use of ancillary data. Multilayer A included 6 of the 7 bands in the Landsat-5 TM images (all except the thermal band), and multilayer B included 14 layers: 7 bands from the Landsat-5 TM image, 2 vegetation indices (normalized difference vegetation index [NDVI] and soil adjust vegetation index [SAVI]), and 5 topographic features (elevation, slope, north–south aspect, east–west aspect, and insolation). NDVI = (ρ nir − ρ red)/(ρ nir + ρ red), where (ρ red) and (ρ nir) are the ground reflectance values of the red and near-infrared bands, respectively. SAVI = ([ρ nir − ρ red] * 1.5)/([ρ nir + ρ red] + 0.5) (Baret and Guyot 1991), which was derived from the ATCOR module. Topographic details were obtained from the 20-m resolution DEM; to obtain north–south and east–west orientations, aspects were transformed to cosine and sine, respectively (Felicísimo 1994). Potential insolation was calculated using ArcGis 9.3 software. Topographic information can complement remote sensing data because the rugged topography in mountainous areas has a strong influence on vegetation composition (Franklin et al 1994; Shi et al 2009).

Training data

The training data must be mutually exclusive and exhaustive (Chuvieco and Huete 2009). The 2 most common sampling strategies were used: expert supervised (see the section on selected training) and random (see the section on random training) (Figures 3, 4). Random sampling is the most exhaustive sampling procedure; it exhibits higher variability within vegetation categories and lower separability. Expert-supervised block selection is more exclusive and therefore exhibits lower variability and higher separability among vegetation categories.

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

FIGURE 2 Field photos of the vegetation studied in the study area. (A) Sparse grassland; (B) dense grassland; (C) shrub; (D) forest. (Photos by M. Gartzia)

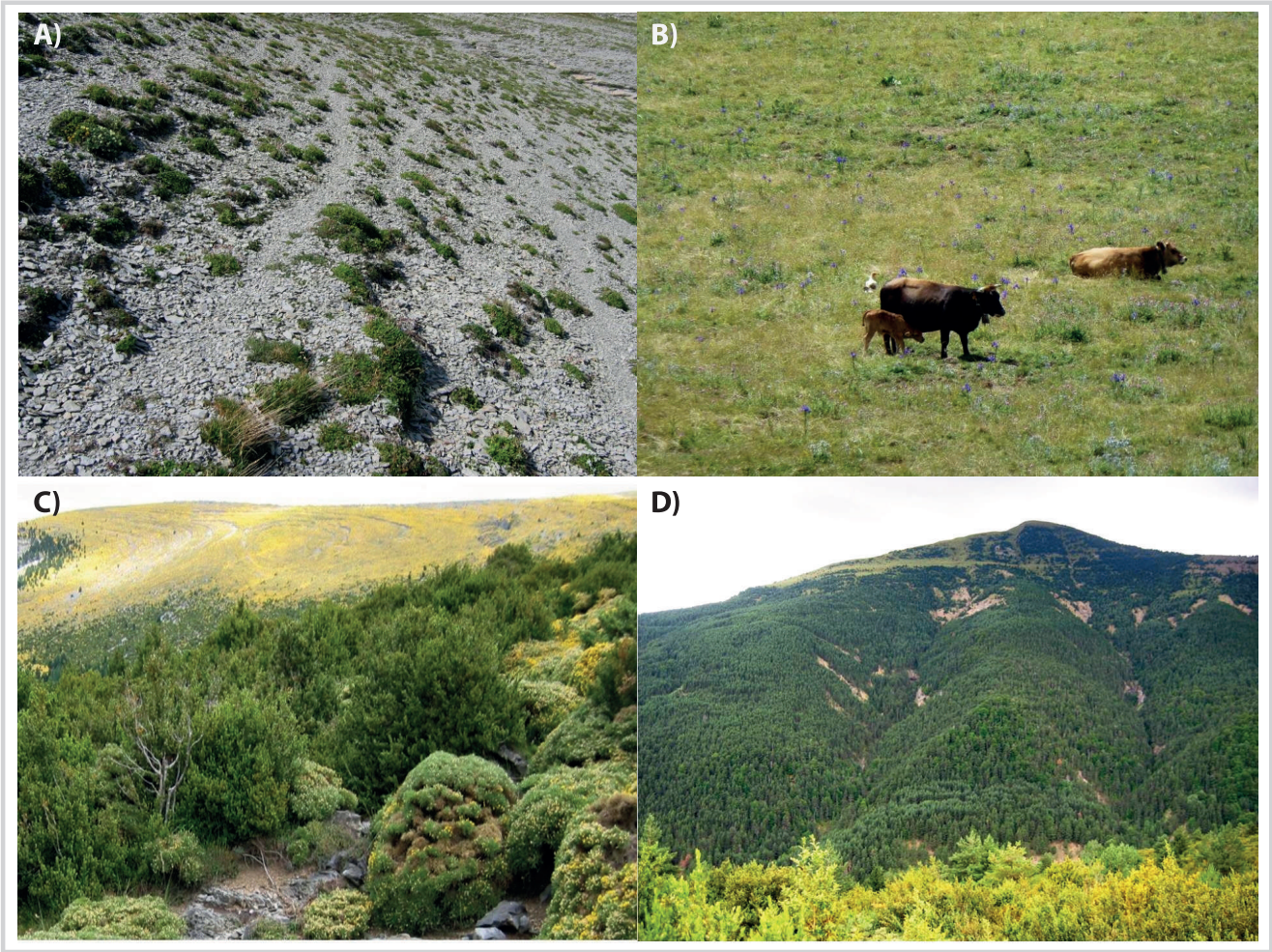
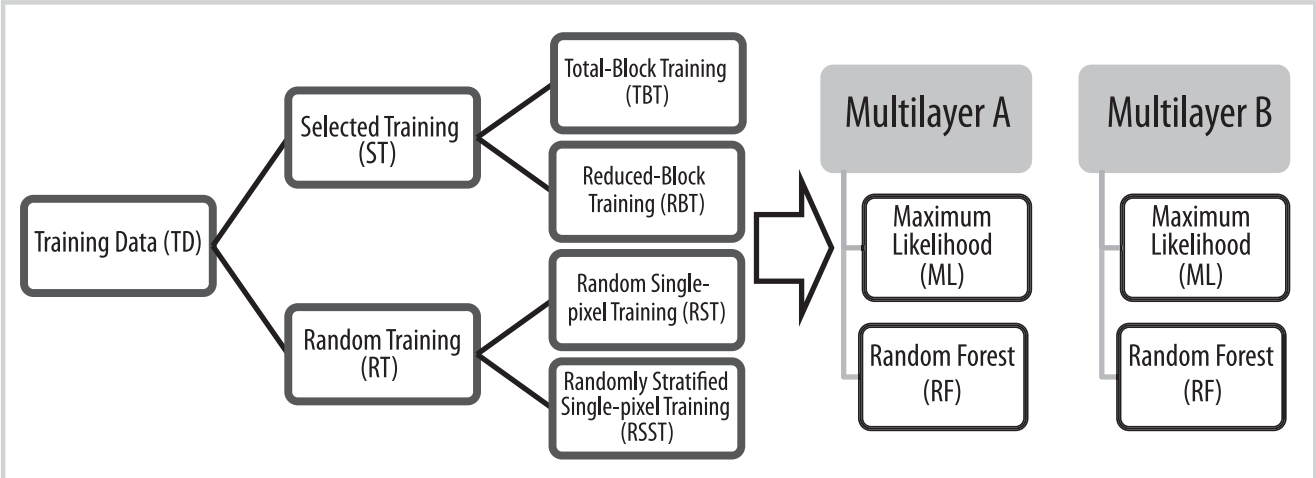
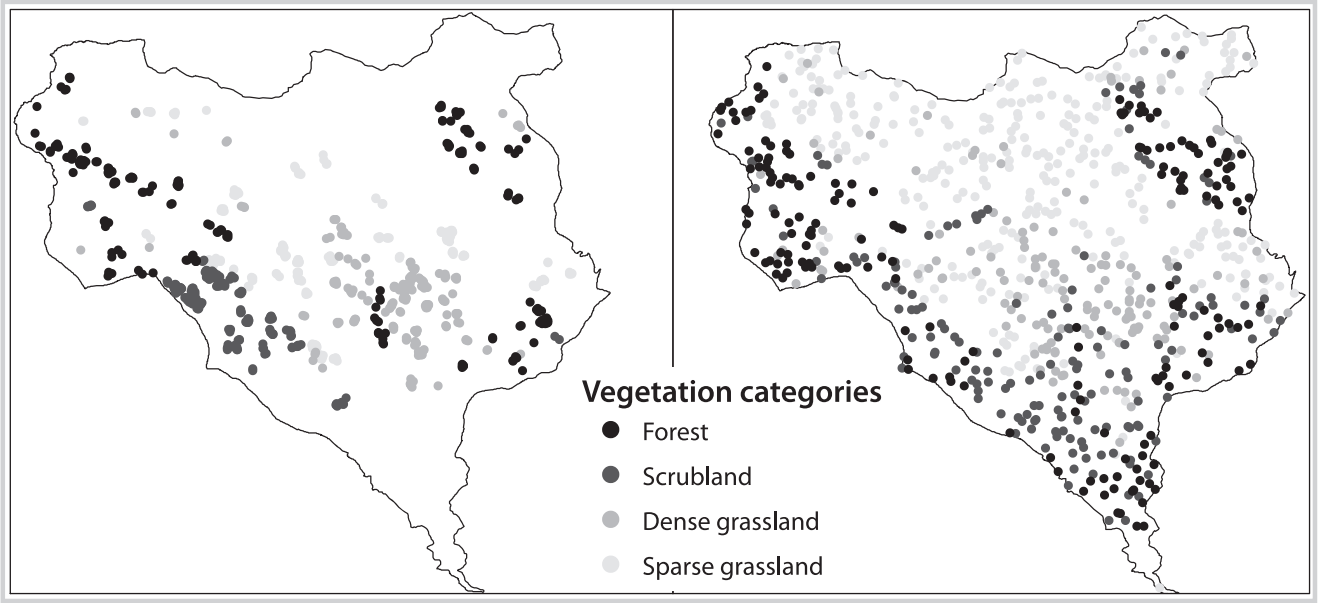


FIGURE 3 Scheme of the alternatives: training data, multilayer, and classification algorithm used for each methodological step analyzed. Four types of training data—2 RT and 2 ST—were compared. In addition, 2 multilayers (*A* and *B*), with different layer components, were used with ML or RF algorithms.



http://dx.doi.org/10.1659/MRD-JOURNAL-D-12-00011.1
Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

FIGURE 4 Distributions of the sources of the ST data (left) and the RT data (right) used for training the algorithms in supervised classifications. ST accounts for exclusivity, with lower data variance, and RT accounts for exhaustibility, with higher variance in the training data.



Transformed divergence values (TrDi) were used to quantify the separability between categories in the training data (Swain 1972). TrDi ≥ 1500 indicates acceptable separability (Haack et al 1987) and TrDi ≥ 1600 indicates good separability (Metternicht and Zinck 1998) between classes, with the maximum separability TrDi = 2000.

Selected training (ST): ST was based on a subjective selection (expert-supervised selection) of blocks of pixels

FIGURE 5 Pairwise comparisons of the accuracies (based on κ values) of the supervised classifications of vegetation in the Spanish Pyrenees, which were performed using combinations of procedures that included 2 multilayers, 4 sampling types (ST: TBT and RBT; RT: RST and RSST), and 2 classification methods (ML and RF).

				Maximum Likelihood (ML)								Random Forest (RF)							
				Multilayer A				Multilayer B				Multilayer A				Multilayer B			
				ST		RT		ST		RT		ST		RT		ST		RT	
				TBT	RBT	RST	RSST	TBT	RBT	RST	RSST	TBT	RBT	RST	RSST	TBT	RBT	RST	RSST
			Kappa	0.88	0.89	0.85	0.90	0.83	0.84	0.89	0.90	0.87	0.86	0.91	0.91	0.91	0.90	0.94	0.93
ML	Multilayer A	TBT	0.88																
		RBT	0.89	–															
		RST	0.85	–	–														
		RSST	0.90	–	–	s													
	Multilayer B	TBT	0.83	s	s	–	s												
		RBT	0.84	–	s	–	s	–											
		RST	0.89	–	–	–	–	s	s										
		RSST	0.90	–	–	s	–	s	s	–									
RF	Multilayer A	TBT	0.87	–	–	–	–	–	–	–	–								
		RBT	0.86	–	–	–	–	–	–	–	s	–							
		RST	0.91	–	–	s	–	s	s	–	–	s	s						
		RSST	0.91	–	–	s	–	s	s	–	–	s	s	–					
	Multilayer B	TBT	0.91	–	–	s	–	s	s	–	–	s	s	–	–				
		RBT	0.90	–	–	s	–	s	s	–	–	–	–	–	–	–			
		RST	0.94	s	s	s	s	s	s	s	s	s	s	s	–	s	s		
		RSST	0.93	s	s	s	s	s	s	s	–	s	s	–	–	–	s	–	

s: Z>1.96, significantly different at p=0.05 (Congalton 2008) – : not significantly different

http://dx.doi.org/10.1659/MRD-JOURNAL-D-12-00011.1
Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

from zones that were representative of each vegetation category in the study area and was influenced by the experience and knowledge of the observer (Figure 4). We used the following 2 versions of ST:

- Total-block training (TBT): For each vegetation category, different numbers of training data were selected using a nonstratified design that included 4000 training data. This is the most common way of selecting training data in remote sensing research.
- Reduced-block training (RBT): RBT data were obtained using a random selection of 800 training data from the TBT and based on a stratified design that included 200 training data for each vegetation category.

Random training (RT): RT was based on a random selection of a single pixel. The vegetation category to which each selected pixel corresponded was identified using a vegetation map of the area (given earlier), inspection in the field, and orthophotographs (Figure 4). The following 2 versions of RT were evaluated:

- Random single-pixel training (RST): In RST, 800 pixels within the study area were selected using a nonstratified design.
- Randomly stratified single-pixel training (RSST): In RSST, 200 pixels per vegetation category were selected in a stratified design. For the stratification, we used the vegetation map described in the section on vegetation categories.

Selection of classifiers

A supervised classification includes a multilayer image, training data, and a classifier (Figure 3). For each vegetation category, classifiers identify a known amount of training data and assign the remainder of the pixels in the multilayer to a vegetation category based on their similarity in the reflectance (Chuvieco and Huete 2009). In this study, we compared the ML method using Erdas 9.2 and the RF method using R from the randomForest package. ML is the most commonly used (standard) method (Liu et al 2011), and RF is one of the most promising methods for the classification of vegetation using remote sensing data (Gislason et al 2006). ML is a parametric method that is based on the probability that a pixel belongs to a given class, and the mean vector and the variance–covariance matrix are the 2 parameters that characterize each class (Adams and Gillespie 2006).

The nonparametric RF classifier is among the tree-based classifier methods that interactively use a hierarchical splitting mechanism to categorize the data. The RF can produce a large number of trees (ntree), each generated by selecting a subset of the sample (randomly drawn with replacement from the original dataset, bootstrapping), and uses a random set of predictors (mtry, the number of predictors used in each split). The trees that result are aggregated by averaging (Breiman 2001; Pal 2005). For the RF classifier, ntree and mtry are the tune parameters. The tuning parameters were adjusted by selecting an ntree in which the overall misclassification error stabilizes (ntree = 500), testing different numbers of predictors, and selecting the one that reduces significantly the misclassification error using the tuneRF function from the randomForest R package (Liaw and Wiener 2002).

Accuracy assessment

The accuracy of the classifications was tested using samples taken following the RST procedure (see the earlier section) (Congalton 1988, 1991). We selected 800 pixels, which provided more than the minimum number of samples per category recommended for accuracy assessment (50 samples per category) (Congalton 1991), specifically, 213 pixels for forest, 120 for shrub, 140 for dense grassland, and 327 for sparse grassland. To assess the accuracy of each vegetation classification, we used Cohen's kappa index (κ) of agreement (Cohen 1960), which is the standard method in remote sensing research (Congalton et al 1983; Rosenfield and Fitzpatricklins 1986; Chen and Stow 2002; Congalton and Green 2009).

Statistical analysis

To assess the effects of the different procedures—ie type of multilayer, method of obtaining training data, and classification methods—on classification accuracy, 2 complementary analyses were performed: a pairwise comparison of κ values and a multivariate comparison using generalized least squares (GLS) models (Zuur et al 2009). In the pairwise comparison, differences between each pair of independent kappa indices were tested statistically based on the standardized and normally distributed Z value (significant at $p = 0.05$ when $Z > 1.96$) (Congalton and Green 2009), which identified the classification pairs that differed significantly.

A multivariate analysis compared the following procedures: multilayer composition type, sampling procedures for selecting training data, classification methods, and the interactions among them. GLS models were used because variance heterogeneity was detected in model residuals and differences identified by a multivariate analysis of variance model were statistically significant (Zuur et al 2009). The 3 procedures (multilayer and sampling types and classification methods) and all 2-way interactions were included in the GLS. Sampling procedures were grouped into 2 categories (ST and RT), because optimal comparisons were obtained using categorical variables that had the same number of categories. The best model was identified based on forward and backward stepwise procedures using Akaike's information criterion (Zuur et al 2009). If a procedure or interaction term was significant, the factor was evaluated using analysis of variance (ANOVA) and a Tukey test.

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

TABLE 1 Importance of the procedures used to improve the accuracy of vegetation classification in the Spanish Pyrenees, as assessed by GLS analysis.

Procedure	*F* value	*P* value
Training data type (ST or RT)	55.39	<0.0001***
Multilayer type (A or B)	35.17	0.0001**
Classification method (ML or RF)	10.46	0.0080*
Classification method × multilayer type	6.04	0.0318

Significance: *** = $P < 0.0001$; ** = $P < 0.001$; * = $P < 0.01$; [no sign] = $P < 0.05$.

To identify the differences among the training data types in each layer used in the classification, spectral curves were created for each vegetation category using the mean percentage of reflectance and standard deviation of the different types of training data for 6 Landsat bands and the mean and standard deviation of the values of the ancillary data (NDVI and elevation).

Results

In Ordesa and Monte Perdido National Park and the surrounding protected area in the Spanish Central Pyrenees, the accuracy ($\kappa = 0.83$–0.94) of a supervised classification of the vegetation improved by 4–11%, depending on the choices made at each step (Figure 5). Sampling type (ST or RT) had the greatest effect on accuracy, and classification method (ML or RF) had the least effect (Table 1). The type of multilayer had an intermediate effect, and the interaction between multilayer type and classification method was marginally significant (Table 1).

RT produced significantly higher κ values than did ST (ANOVA $F = 2.6$, $P < 0.05$), even though the separability of RT (as reflected by the TrDi value) was lower than that of ST. The accuracy of the classification that used multilayer A (which included the Landsat-5 TM spectral bands only) was independent of the classification method (ML or RF); however, when topographic information and vegetation indices were included in the multilayer (multilayer B), the RF classifier was significantly better than was the ML classifier (Tukey post-hoc test $T = 2.9$, $P < 0.05$). The RF classification method produced significantly higher κ values than did ML (ANOVA $F = 5.09$, $P < 0.05$). The combination of RT data, multilayer B, and the RF method had the highest accuracy (Figure 5). In this case, the results based on stratified (RSST) and nonstratified (RST) RT data did not differ significantly (Figure 5).

Based on the κ values of each vegetation category, the RF algorithm produced stable, robust results (Table 2). Differences among sampling types (ST and RT) were never >10% (columns TBT-RST and RBT-RSST, Table 2). The use of RT sampling data with the 2 multilayers provided a small but consistent improvement (Table 2). The ML, however, produced results that were sensitive to changes in the type of sampling for either multilayer A or multilayer B (differences > 10%; columns TBT-RST and RBT-RSST, Table 2). Multilayer A produced the highest accuracies using ST data (TBT or RBT) (except for sparse grassland), while with multilayer B, the highest accuracies were obtained using RT data (RST or RSST) (Table 2).

The 2 types of ST (TBT, RBT) and the 2 types of RT (RST, RSST) had similar spectral curves (Figure 6); therefore, ST and RT were compared. In most cases, and especially in forest and sparse grassland, the RT data had higher variability or standard deviations than did the ST data (Figure 6). The spectral signatures of the forest, shrubland, and grassland were typical of these vegetation types (Figure 6A–D), and they all showed limited (<12%) reflectance in the visible bands (highest in sparse grassland). Increases in the reflectance of all vegetation types were detected in the near-infrared (NIR) region (band 4), and the increase was highest in dense grassland (up to 35%). Reflectance in the shortwave infrared (SWIR) region (bands 5 and 7) tended to decline because of an increase in radiance absorption by the water contained in plants. In sparse grassland, however, reflectance values in NIR and SWIR were 25–30% because of the high proportion of bare soil. In SWIR bands, the reflectance values from the RT data were higher than those from the ST data, and in NIR, except forests, the values from the RT data were lower than those from the ST data. In the visible region (bands 1–3), the performances of the 2 sampling types were similar.

An analysis of the training values of the ancillary data (Figure 6E, F) showed that the RT data had more contrasted values than did the ST data, eg for elevation, which increased the separability among categories, even though the RT data had higher variability than did the ST data. Except for shrubland, the RT data had the lowest NDVI. Like Landsat bands 2, 3, 5, and 7, elevation and

http://dx.doi.org/10.1659/MRD-JOURNAL-D-12-00011.1
Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

TABLE 2 Conditional κ values for each vegetation type and the overall κ (**in bold**), for the supervised classifications, using combinations of procedures that included 2 multilayer types (A and B), 2 methods of classification (ML and RF), and 4 sampling types, 2 of which not stratified, TBT (ST) and RST (RT), whereas the other 2 were stratified, RBT (ST) and RSST (RT). The third and sixth columns show a subtraction between the κ of the sampling types (columns 1, 2, 4, and 5).

	Not stratified			Stratified		
	TBT (ST)	*RST (RT)*	*TBT-RST*	*RBT (ST)*	*RSST (RT)*	*RBT-RSST*
Multilayer A						
ML	**0.88**	**0.85**	**0.03**	**0.89**	**0.90**	**−0.01**
F	0.96	0.80	0.15	0.99	0.89	0.10
S	0.89	0.83	0.07	0.89	0.89	0.00
DG	0.95	0.77	0.18	0.95	0.83	0.12
SG	0.80	0.98	−0.18	0.80	0.97	−0.16
RF	**0.87**	**0.91**	**−0.04**	**0.86**	**0.91**	**−0.05**
F	0.89	0.93	−0.04	0.99	0.97	0.02
S	0.83	0.82	0.01	0.71	0.79	−0.07
DG	0.88	0.89	−0.01	0.88	0.91	−0.03
SG	0.86	0.93	−0.08	0.86	0.94	−0.09
Multilayer B						
ML	**0.83**	**0.89**	**−0.06**	**0.84**	**0.90**	**−0.06**
F	0.80	0.98	−0.18	0.88	0.99	−0.10
S	0.92	0.86	0.06	0.96	0.85	0.11
DG	0.88	0.70	0.18	0.87	0.77	0.11
SG	0.80	0.99	−0.19	0.77	0.96	−0.20
RF	**0.91**	**0.94**	**−0.03**	**0.90**	**0.93**	**−0.04**
F	0.94	0.99	−0.05	0.99	0.98	0.01
S	0.92	0.88	0.04	0.79	0.86	−0.07
DG	0.90	0.95	−0.05	0.89	0.91	−0.02
SG	0.88	0.93	−0.05	0.89	0.95	−0.07

F, forest; S, shrubland; DG, dense grassland; SG, sparse grassland.

NDVI were the most important layers for the RF algorithm in the supervised classification when multilayer B was used.

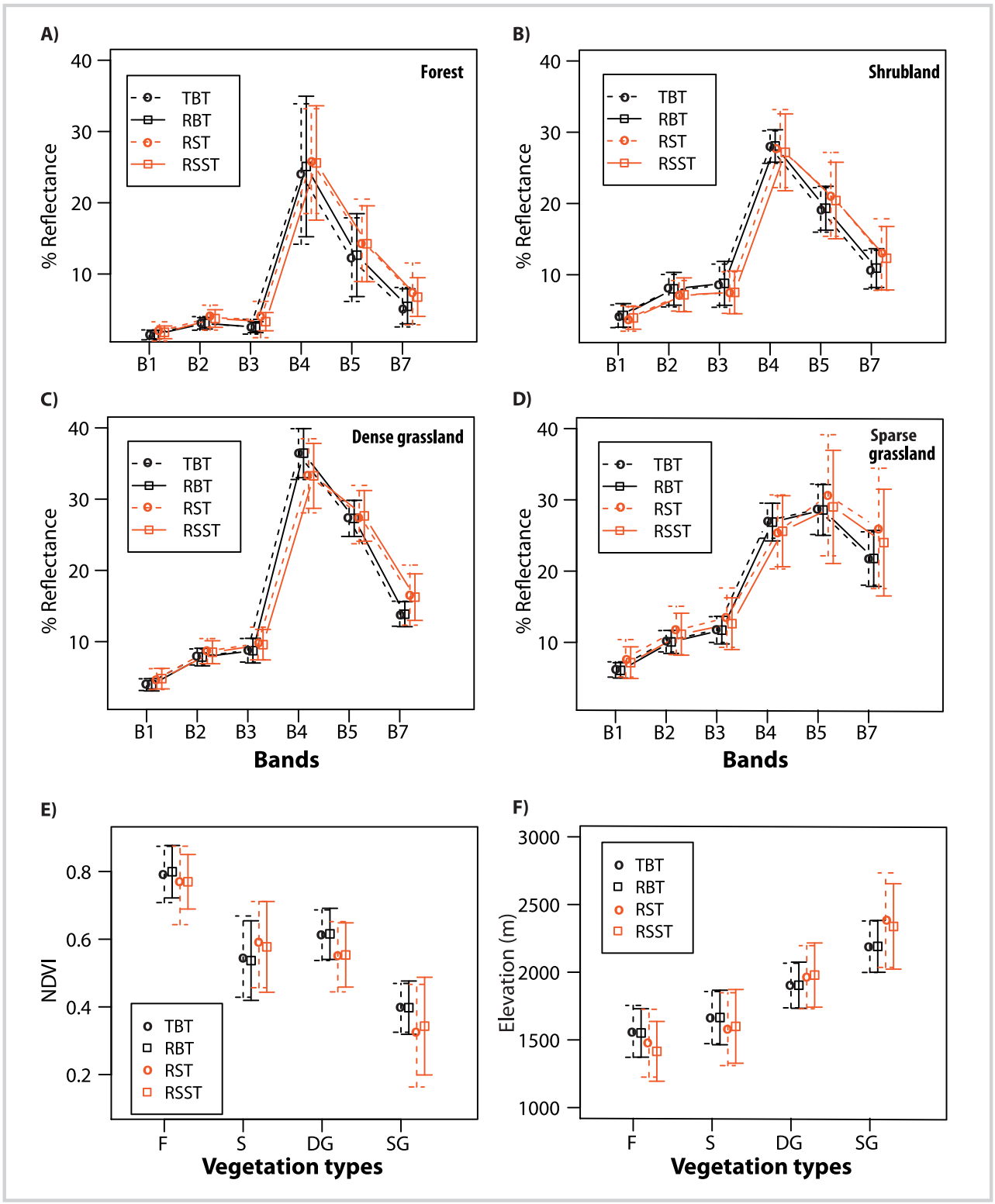
The combination of multilayer type and classification method, which was marginally significant (Table 1), was examined further because of the differences between the final maps obtained using multilayer A or B with the RF algorithm and RT data (RST) (Figure 7). In most cases, the pure pixels (which represented only 1 of the 4 vegetation categories) were classified correctly, and were similar with the 2 strategies (Figure 7). When the pixels included a combination of vegetation categories, the 2 classification procedures often produced different results. When multilayer A was used, at high elevations—where there are little more than sparse grassland—some shadows and steep slopes were classified as shrub (Figure 7). These errors were corrected by including topographic variables such as ancillary data in the multilayer (multilayer B).

Discussion

In Ordesa and Monte Perdido National Park and the surrounding protected area in the Spanish Central Pyrenees, the modest improvements (κ = 0.83–0.94) in the accuracies of the classification strategies might be essential for detecting changes in the vegetation in mountainous areas, particularly woody encroachment and loss of biodiversity (Dirnböck et al 2003; Lasanta-Martínez et al 2005; Komac 2010). The improvements came from considering options at all crucial steps in the classification procedure (Lu and Weng 2007), rather than the classification method only (Landgrebe 2003; Cutler et al 2007; Tso and Mather 2009).

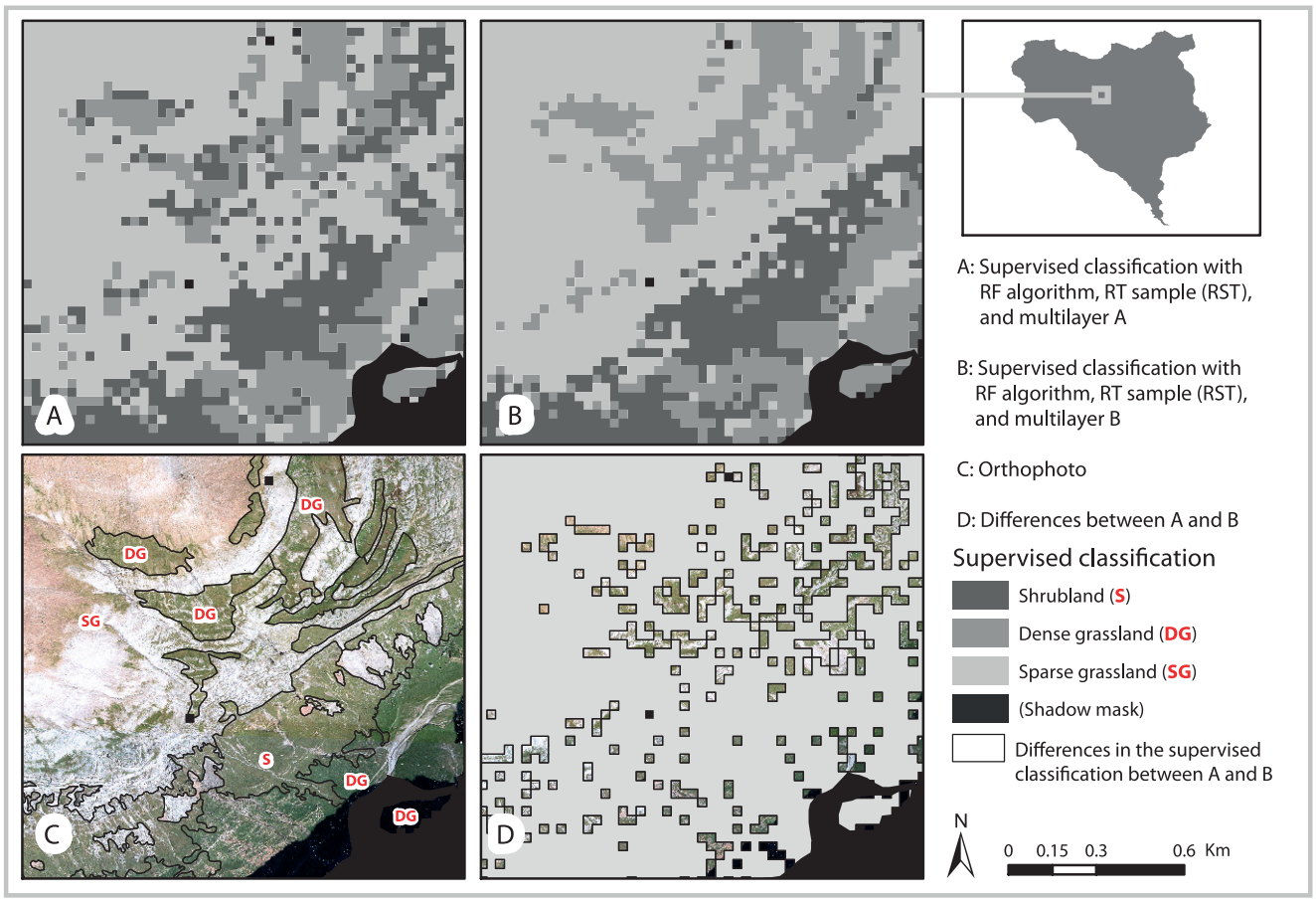
Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

FIGURE 6 Spectral curves showing the percentage of reflectance of the 6 Landsat bands for each vegetation category (A–D), and mean and standard deviation of the ancillary data values (E, NDVI, and F, altitude) for each vegetation category. Four sampling types: TBT, RBT, RST, and RSST. The black ones (circle and square) are ST data, and the orange ones are RT data. Vegetation categories: F, forest; S, shrubland; DG, dense grassland; SG, sparse grassland.



http://dx.doi.org/10.1659/MRD-JOURNAL-D-12-00011.1
Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

FIGURE 7 An example from the study area showing the differences between 2 of the methodologies used, differing in the multilayer used (A or B). In higher altitudes where there are no shrub communities, multilayer B lead to better classification of the mixed zones, taking into account the topographic data.



In our study, the selection of training data was the most important factor in increasing the accuracy of a supervised classification (Hixson et al 1980; Gog and Howarth 1990), and RT was the best sampling type (Gog and Howarth 1990; Gong et al 1996). The nonstratified RST provided slightly, but not significantly, better results than did the stratified RSST. Thus, stratification might not be recommendable except when a vegetation map exists and the vegetation categories are automatically identified, which is less time consuming.

The RT data provided high variation (as reflected in the standard deviation of the data), which can help to better define all vegetation communities, but it created overlap in the training data values of the 4 vegetation types, thus reducing the separability of those categories. In addition, training data that have high variability, such as the RT data, are better classified using nonparametric classifiers such as the RF classifier (Cortijo and De La Blanca 1997).

In our study, the second most important factor influencing the accuracy of the classifications was the multilayer compositions used. In mountainous areas, the inclusion of topographic information can improve classification accuracy (Franklin et al 1994; Hansen et al 1996; Gislason et al 2006; Shi et al 2009), because topography strongly influences vegetation composition. The benefits of the topographic layers were greatest in areas of mixed vegetation, in which case, with multilayer A (Landsat information only), information was insufficient to distinguish among the vegetation categories. In this case, the use of the topographic information (multilayer B) might improve the classification by adding more information that better defines the classification of each pixel. But too many layers can reduce classification accuracy (Price et al 2002) because not all classifiers work well with a large number of layers or correlated layers; in that case, nonparametric classifiers might be more appropriate (Lu and Weng 2007). That might be why, when we used the ancillary information in the multilayer, the RF classifier improved accuracy significantly; however, when the ML method was used, the improvement was not as pronounced as before.

The ML and other single-stage classifier methods are used extensively in remote sensing studies, but they can produce results that are less accurate than are those derived by multistage iterative classifiers

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

(San Miguel-Ayanz and Biging 1996; Friedl and Brodley 1997; Muchoney et al 2000; Rogan et al 2002; Shi et al 2009). In some cases, RF multistage classifiers had advantages and were more accurate (Na et al 2009; Waske and Braun 2009; Rodriguez-Galiano and Chica-Olmo 2012) than was the ML classifier. The RF classifier is a nonparametric classifier, robust to outliers and noise, in which the most effective feature is selected sequentially in each node-splitting stage, and it can be used with layers that have high collinearity (Breiman 2001); thus, it is suited to the classification of complex areas. The ML classifier is appropriate when there is enough information for the classification, few and noncorrelated layers, and the variability of the training data is low (eg ST data). In mountain areas, however, such conditions are infrequent and RF or other multistage classifiers are highly recommended.

Conclusions

This study aimed to optimize the outcome of the vegetation classifications derived from remote sensing images in mountain areas. We assessed the 3 main steps required for this procedure (selection of the training data, the ancillary data, and the classifier), evaluating the most widely used and promising options for each step. All analyses were applied to a case study area in a national park within the Spanish Pyrenees, representative of a mountain scenario. Each of the 3 steps played a significant role in improving the accuracy of the classification, and their relative importance was not influenced by the selection of the classifier. The results highlight the need to select more than 1 step to improve the accuracy of vegetation classifications in mountainous areas because of the complexity of the topography and its consequences, such as undesired shadows.

More research on vegetation classifications in other mountainous areas is required to identify the role of each classification step in improving accuracy. Ultimately, improvements in accuracy would improve the ability to detect change, which is directly related to our ability to manage and preserved these important and complex ecosystems.

The specific combinations of parameters and procedures used in this research improved the accuracies of vegetation classifications by 4–11%. RT data coupled with multilayers derived from Landsat-5 TM images and ancillary data, and a RF classifier method produced the highest accuracy for the supervised classification of vegetation. Some standard protocols for the classification of Landsat images include important steps that dictate the way in which the data are analyzed. The choice of sampling design, the use of ancillary data, and to a lesser extent, the choice of classifiers influenced the accuracies of vegetation classifications, especially in mountainous areas, where errors caused by a heterogeneous topography are intrinsically high.

ACKNOWLEDGMENTS

The Spanish government (Configuración Espacial de la Biodiversidad y Conservación del Ecosistemas project, I+D+I. CGL2008-00655/BOS, Ministry of Science and Innovation) and the European Community (Land and Ecosystem Degradation and Desertification: Assessing the Fit of Responses project, FW7 ENV.2009.2.1.3.2) funded this study. In addition, we thank the managers of Ordesa and Monte Perdido National Park for providing information and assistance. For assistance with the bibliographic search, we thank Cristina Pérez de Larraya. The suggestions of IC Barrio considerably improved earlier versions of the paper. We thank Bruce MacWhirter for his critical reading of the manuscript and for providing helpful suggestions. Finally, we thank the anonymous reviewers for their helpful comments.

REFERENCES

Adams JB, Gillespie AR. 2006. *Remote Sensing of Landscapes With Spectral Images: A Physical Modeling Approach*. New York, NY: Cambridge University Press.

Aragón R, Oesterheld M. 2008. Linking vegetation heterogeneity and functional attributes of temperate grasslands through remote sensing. *Applied Vegetation Science* 11:117–130.

Balthazar V, Vanacker V, Lambina EF. 2012. Evaluation and parameterization of ATCOR3 topographic correction method for forest cover mapping in mountain areas. *International Journal of Applied Earth Observation and Geoinformation* 18:436–450.

Baret F, Guyot G. 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment* 35:161–173.

Breiman L. 2001. Random forests. *Machine Learning* 45:5–32.

Chan JCW, Paelinckx D. 2008. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment* 112: 2999–3011.

Chander G, Markham BL, Helder DL. 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+ and EO-1 ALI sensors. *Remote Sensing of Environment* 113:893–903.

Chen D, Stow D. 2002. The effect of training strategies on supervised classification at different spatial resolutions. *Photogrammetric Engineering and Remote Sensing* 68:1155–1161.

Chuvieco E, Huete A. 2009. *Fundamentals of Satellite Remote Sensing*. Boca Raton, FL: CRC Press.

Cohen J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20:37–46.

Congalton RG. 1988. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing* 54:593–600.

Congalton RG. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37:35–46.

Congalton RG, Green K. 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices.* 2nd edition. Boca Raton, FL: CRC Press.

Congalton RG, Oderwald RG, Mead RA. 1983. Assessing Landsat classification accuracy using discrete multivariate statistical techniques. *Photogrammetric Engineering and Remote Sensing* 49:1671–1678.

Cortijo FJ, De La Blanca NP. 1997. A comparative study of some non-parametric spectral classifiers: Application to problems with high-overlapping training sets. *International Journal of Remote Sensing* 18:1259–1275.

Cutler DR*, *Edwards TC*, *Beard KH*, *Cutler A*, *Hess KT*, *Gibson J*, *Lawler JJ. 2007. Random forest for classification in ecology. *Ecology* 88:2783–2792.

De la Riva JR. 1997. *Los Montes de la Jacetania. Caracterización Física y Explotación Forestal*. Zaragoza, Spain: Consejo de Protección de la Naturaleza de Aragón.

Dirnböck T, Dullinger S, Grabherr G. 2003. A regional impact assessment of climate and land-use change on alpine vegetation. *Journal of Biogeography* 30: 401–417.

http://dx.doi.org/10.1659/MRD-JOURNAL-D-12-00011.1
Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use

Felicísimo AM. 1994. *Modelos Digitales del Terreno. Introducción y Aplicaciones en las Ciencias Ambientales*. Oviedo, Spain: Pentalfa Ediciones.
Franklin SE, Connery DR, Williams JA. 1994. Classification of alpine vegetation using Landsat Thematic Mapper, SPOT HRV and DEM data. *Canadian Journal of Remote Sensing/Journal Canadien de Télédétection* 20:49–58.
Friedl MA, Brodley CE. 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment* 61:399–409.
García-González R, ***Alados CL***, ***Bueno G***, ***Fillat F***, ***Gartzia M***, ***Gómez D***, ***Komac B***, ***Marinas A***, ***Saint-Jean N.*** 2007. Valoración ecológica y productiva de los pastos supraforestales en el Parque Nacional de Ordesa y Monte Perdido. *In:* Ramírez L, Asensio B, editors. *Proyectos de Investigación en Parques Nacionales: 2003–2006*. Madrid, Spain: Organismo Autónomo Parques Nacionales, Ministerio de Medio Ambiente, pp 105–128.
Geosystems. 2009. *ATCOR for ERDAS IMAGINE 2010. User Manual ATCOR 2 and ATCOR 3*. Germering, Germany: Geosystems.
Gislason PO, Benediktsson JA, Sveinsson JR. 2006. Random forests for land cover classification. *Pattern Recognition Letters* 27:294–300.
Gog P, Howarth PJ. 1990. An assessment of some factors influencing multispectral land-cover classification. *Photogrammetric Engineering and Remote Sensing* 56:597–303.
Gong P, Pu R, Chen J. 1996. Mapping ecological land systems and classification uncertainties from digital elevation and forest-cover data using neural networks. *Photogrammetric Engineering and Remote Sensing* 62:1249–1260.
Haack B, Bryant N, Adams S. 1987. An assessment of Landsat MSS and TM data for urban and near-urban land-cover digital classification. *Remote Sensing of Environment* 21:201–213.
Ham J, Chen Y, Crawford MM. 2005. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* 43:492–501.
Hansen M, Dubayah R, DeFries R. 1996. Classification trees: An alternative to traditional land cover classifiers. *International Journal of Remote Sensing* 17: 1075–1081.
Hixson M, Scholz D, Fuhs N, Akiyama T. 1980. Evaluation of several schemes for classification of remotely sensed data. *Photogrammetric Engineering and Remote Sensing* 46:1547–1553.
Komac B. 2010. *Effets des Modifications de l'Utilisation des Terres sur la Conservation des Pâturages Subalpins du Parc National de Ordesa Mont-Perdu*. Chambéry, France: Université de Savoie.
Komac B, Alados CL, Bueno CG. 2011a. Spatial patterns of species distributions in grazed subalpine grasslands. *Plant Ecology* 212:519–529.
Komac B, Alados CL, Camarero JJ. 2011b. Influence of topography on the colonization of subalpine grasslands by the thorny cushion dwarf *Echinospartum horridum. Arctic, Antarctic, and Alpine Research* 43:601–611.
Kozak J. 2003. Forest cover change in the western Carpathians in the past 180 years. *Mountain Research and Development* 23:369–375.
Landgrebe DA. 2003. *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ: John Wiley and Sons.
Lasanta-Martínez T, Vicente-Serrano SM, Cuadrat-Prats JM. 2005. Mountain Mediterranean landscape evolution caused by the abandonment of traditional primary activities: A study of the Spanish Central Pyrenees. *Applied Geography* 25:47–65.
Liaw A, Wiener M. 2002. Classification and regression by randomForest. *R News* 2:18–22.
Lillesand TM, Kiefer RW, Chipman JW. 2008. Remote sensing and image interpretation. 6th ed. New York, NY: Wiley.
Liu K, Shi W, Zhanga H. 2011. A fuzzy topology-based maximum likelihood classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 66:103–114.
Lu D, Weng Q. 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing* 28:823–870.
Luick R. 1998. Ecological and socio-economic implications of livestock-keeping systems on extensive grasslands in south-western Germany. *Journal of Applied Ecology* 35:979–982.
Metternicht GI, Zinck JA. 1998. Evaluating the information content of JERS-1 SAR and Landsat TM data for discrimination of soil erosion features. *ISPRS Journal of Photogrammetry & Remote Sensing* 53:143–153.
Muchoney D, ***Borak J***, ***Chi H***, ***Friedl M***, ***Gopal S***, ***Hodges J***, ***Morrow N***, ***Strahler A.*** 2000. Application of the MODIS global supervised classification model to vegetation and land cover mapping of Central America. *International Journal of Remote Sensing* 21:1115–1138.
Na X, Zang S, Wang J. 2009. Evaluation of random forest ensemble classification for land cover mapping using TM and ancillary geographical data. *In:* Chen Y, Deng H, Zhang D, Xiao Y, editors. *Sixth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2009, Tianjin*. Los Alamitos, CA: IEEE Computer Society, pp 89–93.
Pal M. 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing* 26:217–222.
Pal M, Mather PM. 2006. Some issues in the classification of DAIS hyperspectral data. *International Journal of Remote Sensing* 27:2895–2916.
Pôças I, Cunha M, Marcal ARS, Pereira LS. 2011a. An evaluation of changes in a mountainous rural landscape of northeast Portugal using remotely sensed data. *Landscape and Urban Planning* 101:253–261.
Pôças I, Cunha M, Pereira LS. 2011b. Remote sensing based indicators of changes in a mountain rural landscape of northeast Portugal. *Applied Geography* 31:871–880.
Poyatos R, Latron J, Llorens P. 2003. Land use and land cover change after farmland abandonment: The case of a Mediterranean mountain area (Catalan Pre-Pyrenees). *Mountain Research and Development* 23:362–368.
Prasad A, Iverson L, Liaw A. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* 9:181–199.
Price KP, Guo X, Stiles JM. 2002. Optimal Landsat TM band combinations and vegetation indices for discrimination of six grassland types in eastern Kansas. *International Journal of Remote Sensing* 23:5031–5042.
Rodriguez-Galiano V, Chica-Olmo M. 2012. Land cover change analysis of a Mediterranean area in Spain using different sources of data: Multi-seasonal Landsat images, land surface temperature, digital terrain models and texture. *Applied Geography* 35:208–218.
Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP. 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 67:93–104.
Rogan J, Franklin J, Roberts D. 2002. A comparison of methods for monitoring multitemporal vegetation change using Thematic Mapper imagery. *Remote Sensing of Environment* 80:143–156.
Rosenfield G, Fitzpatricklins K. 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing* 52:223–227.
San Miguel-Ayanz J, Biging GS. 1996. An iterative classification approach for mapping natural resources from satellite imagery. *International Journal of Remote Sensing* 17:957–981.
Shalaby A, Tateishi R. 2007. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the northwestern coastal zone of Egypt. *Applied Geography* 27:28–41.
Shi Q, Zhang X, Shi Q, Gao W. 2009. Investigations on methods of land cover classification of TM image in mountain area. *In:* Zhang X, Li J, Liu G, Yang X, editors. *SPIE [Society of Photographic Instrumentation Engineers] Proceedings. Second International Conference on Earth Observation for Global Changes*. Vol 7471. http://dx.doi.org/10.1117/12.836446.
Swain PH. 1972. Pattern recognition: A basis for remote sensing data analysis. *In: LARS Information Note 111572*. West Lafayette, IN: Laboratory for Applications of Remote Sensing, Purdue University, p 40.
Thuiller W, Lavorel S, Araújo MB, Sykes MT, Prentice C. 2005. Climate change threats to plant diversity in Europe. *Proceedings of the Natural Academy of Sciences* 102:8245–8250.
Torta G. 2004. Consequences of rural abandonment in a northern Apennines landscape (Tuscany, Italy). *In:* Mazzoleni S, Pasquale GD, Mulligan M, Martinio PD, Rego F, editors. *Recent Dynamics of the Mediterranean Vegetation and Landscape*. Chichester, United Kingdom: Wiley, pp 157–167.
Tso B, Mather PM. 2009. *Classification Methods for Remotely Sensed Data*. Boca Raton, FL: CRC Press.
Waske B, Braun M. 2009. Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 64:450–457.
Wessels K, Steenkamp K, von Maltitz G, Archibald S. 2011. Remotely sensed vegetation phenology for describing and predicting the biomes of South Africa. *Applied Vegetation Science* 14:49–66.
Zuur AF, Leno EN, Walker NJ, Saveliev AA, Smith GM. 2009. *Mixed Effects Models and Extensions in Ecology with R*. Berlin, Germany: Springer.

Downloaded From: https://complete.bioone.org/journals/Mountain-Research-and-Development on 29 Mar 2024
Terms of Use: https://complete.bioone.org/terms-of-use