International Journal of Applied Earth Observation and Geoinformation 49 (2016) 151–162

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Using a stochastic gradient boosting algorithm to analyse the effectiveness of Landsat 8 data for *montado* land cover mapping: Application in southern Portugal

Sérgio Godinho[a,*], Nuno Guiomar[a], Artur Gil[b]

[a] ICAAM—Instituto de Ciências Agrárias e Ambientais Mediterrânicas, LDSP—Landscape Dynamics and Social Process Research Group, Universidade de Évora, Ap. 94, 7002-554 Évora, Portugal

[b] Centre for Ecology, Evolution and Environmental Changes (CE3C), Azorean Biodiversity Group, Department of Biology, University of the Azores, 9501-801 Ponta Delgada, Portugal

ARTICLE INFO

Article history:
Received 2 October 2015
Received in revised form 19 February 2016
Accepted 22 February 2016
Available online 27 February 2016

Keywords:
Vegetation indices
Montado
Mediterranean
Multi-seasonal data
LULC mapping

ABSTRACT

This study aims to develop and propose a methodological approach for *montado* ecosystem mapping using Landsat 8 multi-spectral data, vegetation indices, and the Stochastic Gradient Boosting (SGB) algorithm. Two Landsat 8 scenes (images from spring and summer 2014) of the same area in southern Portugal were acquired. Six vegetation indices were calculated for each scene: the Enhanced Vegetation Index (EVI), the Short-Wave Infrared Ratio (SWIR32), the Carotenoid Reflectance Index 1 (CRI1), the Green Chlorophyll Index (CI_{green}), the Normalised Multi-band Drought Index (NMDI), and the Soil-Adjusted Total Vegetation Index (SATVI). Based on this information, two datasets were prepared: (i) Dataset I only included multi-temporal Landsat 8 spectral bands (LS8), and (ii) Dataset II included the same information as Dataset I plus vegetation indices (LS8 + VIs). The integration of the vegetation indices into the classification scheme resulted in a significant improvement in the accuracy of Dataset II's classifications when compared to Dataset I (McNemar test: Z-value = 4.50), leading to a difference of 4.90% in overall accuracy and 0.06 in the Kappa value. For the *montado* ecosystem, adding vegetation indices in the classification process showed a relevant increment in producer and user accuracies of 3.64% and 6.26%, respectively. By using the variable importance function from the SGB algorithm, it was found that the six most prominent variables (from a total of 24 tested variables) were the following: EVI_summer; CRI1_spring; SWIR32_spring; B6_summer; B5_summer; and CI_{green}_summer.

1. Introduction

The so-called "*montado*" constitutes an agro-silvo-pastoral system dominated by cork oak trees (*Quercus suber*) and/or holm oaks (*Q.* [*ilex*] *rotundifolia*) presenting high levels of spatial variability in tree densities, usually with an understory mosaic of annual crops, grasslands, and shrublands (Joffre et al., 1999; Doorn et al., 2007). This ecosystem covers an area of about 3.5×104 to 4.0×104 km^2 in the south-western part of the Iberian Peninsula, and is therefore of great relevance to the Mediterranean biogeographical region (Olea and San Miguel-Ayanz, 2006). *Montado* is described as a multifunctional system, as it supports a variety of goods and services that are valued by society today (Surová et al., 2011). Aside from cork and firewood, this system also provides acorns and pasture for livestock feeding, and other ecosystem services such as soil conservation, carbon sequestration and biodiversity conservation (Bugalho et al., 2009). Changes in *montado* landscapes are mainly related to environmental constraints (*e.g.*, soil type and hydrological conditions, drought, and wildfires), ineffective land management, the vulnerability of the agricultural economy, and also modifications in the organisation of farming labour (*e.g.*, Godinho et al., 2014; Pinto-Correia, 2000). Monitoring these changes is therefore a pressing concern for society and governmental institutions, as well as for the scientific community.

The availability of accurate and up-to-date spatial information on the *montado* is crucial to understanding the patterns and trends of this ecosystem. Consistent and regular *montado* land cover information with high spatial resolution is required to support the decision-making process regarding ecosystem management and conservation. Established methods, such as field inventories and aerial photographic interpretation, can be used for land cover

* Corresponding author.
E-mail addresses: sgodinho@uevora.pt (S. Godinho), nunogui@uevora.pt (N. Guiomar), artur.jf.gil@uac.pt (A. Gil).

http://dx.doi.org/10.1016/j.jag.2016.02.008
0303-2434/

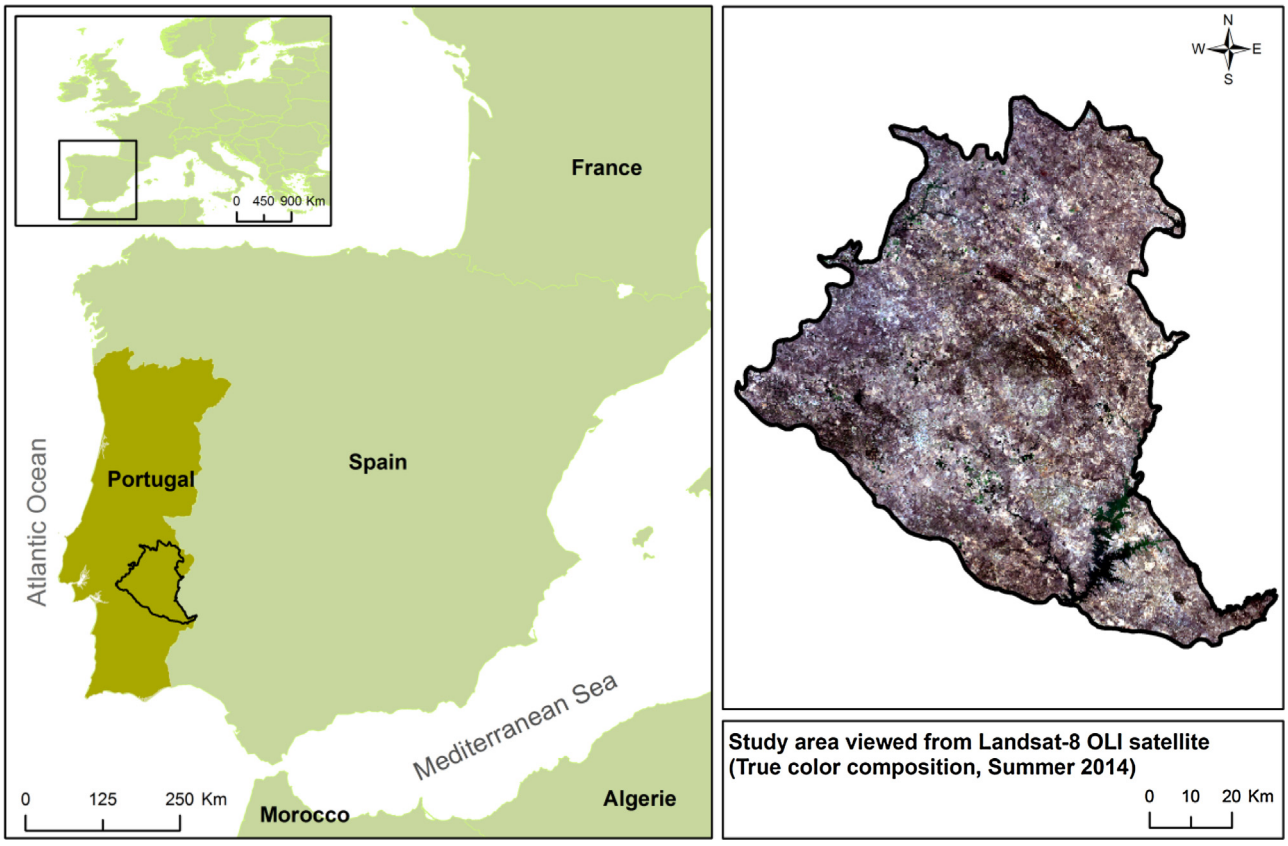

Fig. 1. Study area.

Table 1
Spectral vegetation indices calculated from Landsat 8 to be used in this study.

Vegetation index	Band formula	Reference
Green chlorophyll index	$CIgreen = \frac{\rho NIR}{\rho Green} - 1$	Gitelson et al. (2003)
SWIR32*	$SWIR32 = \frac{\rho SWIR2}{\rho SWIR1}$	Guerschman et al. (2009)
Carotenoid reflectance index 1	$CRI1 = \left(\frac{1}{\rho Blue}\right) - \left(\frac{1}{\rho Green}\right)$	Gitelson et al. (2002)
Enhanced vegetation index	$EVI = 2.5 \times \left(\frac{\rho NIR - \rho Red}{1 + \rho NIR + 6 \times \rho Red - 7 \times \rho Blue}\right)$	Huete et al. (1997)
Normalized multi-band drought index	$NMDI = \frac{\rho NIR - (\rho SWIR1 - \rho SWIR2)}{\rho NIR + (\rho SWIR1 - \rho SWIR2)}$	Wang and Qu (2007)
Soil-adjusted total vegetation index	$SATVI = \left(\frac{\rho SWIR1 - \rho Red}{\rho SWIR1 + \rho Red + L}\right) \times (1 + L) - \left(\frac{\rho SWIR2}{2}\right)$	Marsett et al. (2006)

Note: $L = 0.5$ was applied in SATVI index. *SWIR1 and SWIR2 bands in the case of Landsat 8. Original configuration corresponds to SWIR2 and SWIR3 bands of MODIS sensor (Guerschman et al., 2009).

Table 2
List of land cover classification categories.

Class code	Class name	Number of sample points
MO	*Montado*	420
EF	Eucalyptus forest	117
SL	Shrubland	221
PF	Pine forest	80
WT	Water	89
OG	Olive grove	266
IA	Irrigation agriculture	101
C/P	Dry crops/pastures	213
BS	Bare soil	81
UB	Urban	80
VI	Vineyards	235

mapping, but these tasks are often time-consuming, prohibitively expensive, and limited in their ability to provide spatially continuous information over large territories (Xie et al., 2008). Using remote sensing technology, land cover mapping can be gathered utilising a reduced amount of field data, making it more cost-effective (Rogan and Chen, 2004).

Landsat Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and Landsat 8 Operational Land Imager (OLI) sensors have been collecting imagery data in the visible, near infrared

Table 3
Confusion matrix (in percentage) obtained with the SGB algorithm applied to selected multi-seasonal Landsat 8 OLI multi-spectral bands (Dataset I).

		Reference data												User's accuracy
		MO	EF	SL	PF	WT	OG	IA	C/P	BS	UB	VI	Total	
Classified data	MO	**18.2**	0.7	0.9	0.4	0.0	2.4	0.2	0.6	0.0	0.0	1.0	24.4	74.59
	EF	0.1	**4.5**	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	90.00
	SL	0.4	0.5	**10.3**	0.1	0.0	0.4	0.0	0.1	0.1	0.0	0.0	11.9	86.55
	PF	0.1	0.3	0.1	**2.9**	0.0	0.1	0.0	0.0	0.0	0.0	0.1	3.6	80.56
	WT	0.0	0.0	0.0	0.0	**4.7**	0.0	0.0	0.0	0.0	0.0	0.0	4.7	100.0
	OG	1.8	0.2	0.2	0.2	0.0	**9.6**	0.1	0.8	0.1	0.2	0.8	14.0	68.57
	IA	0.1	0.0	0.0	0.0	0.0	0.0	**4.9**	0.1	0.0	0.0	0.0	5.1	96.08
	C/P	0.8	0.0	0.1	0.1	0.0	0.6	0.1	**9.2**	0.9	0.1	0.3	12.2	75.41
	BS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	**2.8**	0.4	0.0	3.4	82.35
	UB	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.1	0.3	**3.3**	0.0	3.9	84.62
	VI	0.5	0.0	0.0	0.1	0.0	0.6	0.0	0.2	0.0	0.1	**10.3**	11.8	87.29
	Total	22.0	6.2	11.6	4.2	4.7	13.9	5.3	11.3	4.2	4.1	12.5	**100**	
Producer's accuracy		82.73	72.58	88.79	69.05	100.00	69.06	92.45	81.42	66.67	80.49	82.40		
Overall accuracy		**80.70**		Kappa	**0.78**									

Table 4
Comparisons of per-class kappa values obtained with KNN and SGB algorithms applied to Dataset I (LS8). Values in brackets represent the overall accuracy.

Class name	KNN (74.90%)	SGB (80.70%)	Kappa improvement
Montado	0.59	0.67	0.09
Eucalyptus forest	0.78	0.89	0.12
Shrubland	0.79	0.85	0.06
Pine forest	0.78	0.80	0.02
Water	1.00	1.00	0.00
Olive grove	0.53	0.63	0.11
Irrigation agriculture	0.96	0.96	0.00
Dry crops/pastures	0.66	0.72	0.06
Bare soil	0.79	0.82	0.02
Urban	0.76	0.84	0.08
Vineyards	0.79	0.85	0.07
Mean	0.77	0.82	0.06
Std. deviation (SD)	0.13	0.11	0.04

Table 5
Confusion matrix (in percentage) obtained with the SGB algorithm applied to selected multi-seasonal Landsat 8 OLI multi-spectral bands and Vegetation Indices (Dataset II).

		Reference data												User's accuracy
		MO	EF	SL	PF	WT	OG	IA	C/P	BS	UB	VI	Total	
Classified data	MO	**19.0**	0.4	1.2	0.3	0.0	1.9	0.0	0.3	0.0	0.0	0.4	23.5	80.85
	EF	0.1	**5.0**	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.4	92.59
	SL	0.4	0.4	**10.3**	0.2	0.0	0.3	0.0	0.1	0.1	0.0	0.0	11.8	87.29
	PF	0.0	0.3	0.0	**3.1**	0.0	0.1	0.0	0.0	0.0	0.0	0.1	3.6	86.11
	WT	0.0	0.0	0.0	0.0	**4.7**	0.0	0.0	0.0	0.0	0.0	0.0	4.7	100.0
	OG	1.6	0.1	0.1	0.2	0.0	**10.9**	0.0	0.4	0.2	0.0	0.8	14.3	76.22
	IA	0.0	0.0	0.0	0.0	0.0	0.0	**5.1**	0.1	0.0	0.0	0.0	5.2	98.08
	C/P	0.4	0.0	0.0	0.1	0.0	0.4	0.1	**9.9**	0.6	0.1	0.2	11.8	83.90
	BS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	**3.1**	0.3	0.0	3.6	86.11
	UB	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.2	**3.6**	0.1	4.1	87.80
	VI	0.5	0.0	0.0	0.0	0.0	0.2	0.1	0.2	0.0	0.1	**10.9**	12.0	90.83
	Total	22.0	6.2	11.6	4.2	4.7	13.9	5.3	11.3	4.2	4.1	12.5	**100**	
Producer's accuracy		86.36	80.65	88.79	73.81	100.00	78.42	96.23	87.61	73.81	87.80	87.20		
Overall accuracy		**85.60**		Kappa	**0.84**									

Table 6
Comparison of SGB classification accuracies (per-class kappa values) obtained using Dataset I (LS8) and Dataset II (LS8 + VIs). Values in brackets represent the overall accuracy.

Class Name	LS8 (80.70%)	LS8 + VIs (85.60%)	Kappa improvement
Montado	0.67	0.75	0.08
Eucalyptus forest	0.89	0.92	0.03
Shrubland	0.85	0.86	0.01
Pine forest	0.80	0.86	0.06
Water	1.00	1.00	0.00
Olive grove	0.63	0.72	0.09
Irrigation agriculture	0.96	0.98	0.02
Dry crops/pastures	0.72	0.82	0.10
Bare soil	0.82	0.86	0.04
Urban	0.84	0.87	0.03
Vineyards	0.85	0.90	0.05
Mean	0.82	0.87	0.04
Std. deviation (SD)	0.11	0.08	0.03

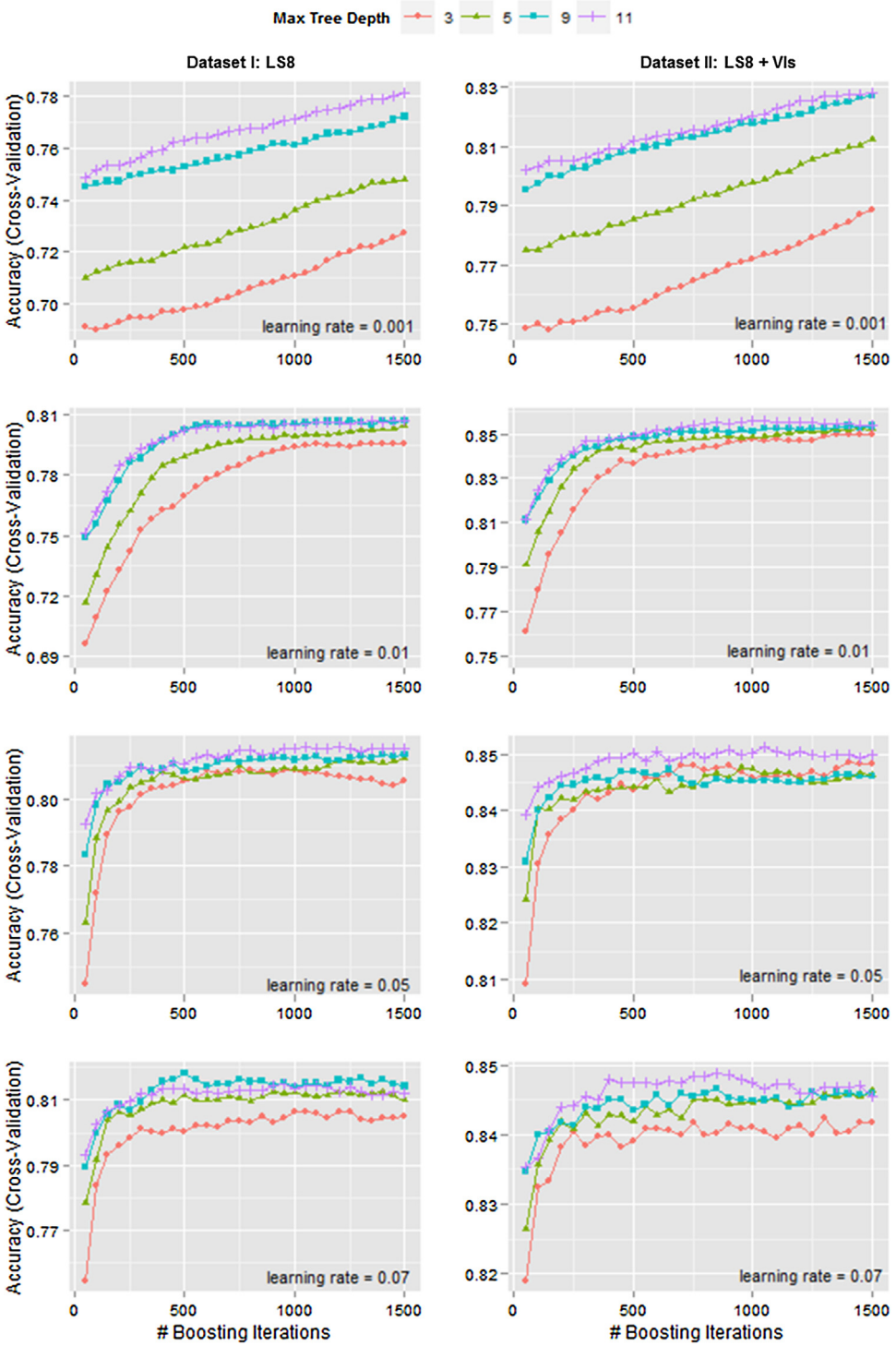

Fig. 2. Overall accuracy levels for tested tree complexity (tc) and learning rate (lr) values using bag fraction 0.50.

(NIR), and shortwave infrared (SWIR) portions of the electromagnetic spectrum, making them appropriate for vegetation studies across a wide range of environments (Cohen and Goward, 2004; Jia et al., 2014; Li et al., 2014). Compared with its predecessors (TM and ETM+), Landsat 8 OLI has some new features, of which the evident narrowing of the NIR band (0.845–0.885 μm) to avoid the effects of water vapour absorption at 0.825 μm that occurs in ETM+ band 4 (0.775–0.900 μm), and the radiometric quantisation of 16 bits are two of the most relevant improvements. These technical advancements mean the Landsat 8 OLI performance levels are superior to those of the previous Landsat sensors (Irons et al., 2012; Roy et al., 2014).

Given the land cover characteristics and the predominance of a dry climate and soil conditions, the Mediterranean landscape presents low levels of inter-class separability (Berberoglu et al., 2000). In these types of landscapes, bare soils have a significant level of spatial occurrence, presenting high reflectance that can mask reflected components from sparse vegetation (Berberoglu et al., 2007; Rodriguez-Galiano et al., 2012). As a Mediterranean ecosystem, obtaining *montado* cover maps constitutes a difficult

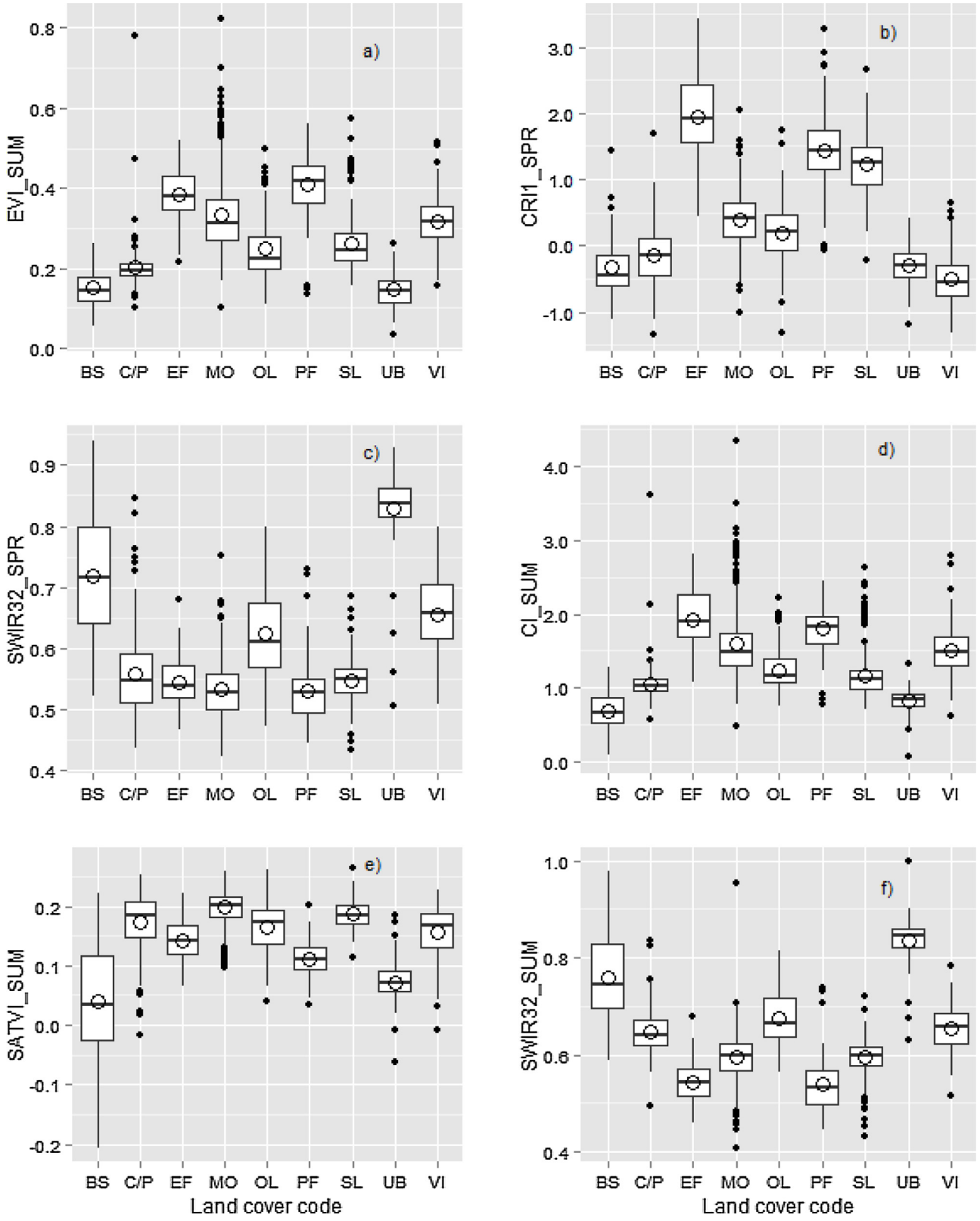

Fig. 3. Distribution of spectral vegetation indices values between land cover categories. Land cover code: BS—bare soil; C/P—Dry crops/pastures; EF—eucalyptus forest; MO—*montado*; OG—olive grove; PF—pine forest; Sl—shrubland, UB—urban; VI—vineyards.

and complex task due to the spatial fuzziness caused by its tree density variability (Doorn and Pinto-Correia, 2007).

Several approaches have been tested and applied to deal with these constraints in order to increase the separability between land covers that present similar spectral behaviour, such as the integration of multi-seasonal images and vegetation indices into the classification process (*e.g.*, Carrão et al., 2008; Rodriguez-Galiano et al., 2012; Senf et al., 2015). Using multi-seasonal images improves

the differentiation between vegetation types thanks to their ability to capture variations in their phenological state (Julien et al., 2011; Rodriguez-Galiano et al., 2012). The integration of vegetation indices into classification models is useful to attain reliable spatial and temporal comparisons of variations in canopy structural, phenological, and biophysical parameters, which can facilitate vegetation inter-class separability (Huete et al., 2002). However, the accuracy of these indices in quantifying vegetation parameters is affected by their sensitivity to atmospheric water vapour content, particularly those that use the NIR band in their computation (Kerekes, 1994). Recently, Li et al. (2014) reported that, depending on the land cover type, there are subtle differences in vegetation indices derived from ETM+ and OLI sensors. Therefore, further research is needed to evaluate the reliability of the vegetation indices resulting from the new Landsat sensor in specific remote sensing applications, such as evaluating their performance in improving land cover classifications in Mediterranean environments.

In the past decade, machine learning algorithms such as Support Vector Machines (SVM) (Cortes and Vapnik, 1995), Random Forest (RF) (Breiman, 2001), and Stochastic Gradient Boosting (SGB) (Friedman, 2002) have gained prominence in improving the performance of classification and regression processes in the remote sensing field (Huang et al., 2002; Lawrence et al., 2004; Rodriguez-Galiano et al., 2012). Nevertheless, to the best of our knowledge, only a few studies have reported applications of SGB for remote sensing classification purposes (Chirici et al., 2013; Lawrence et al., 2004). Therefore, more research is needed to assess the potential and effectiveness of the SGB algorithm for land cover mapping, especially using Landsat 8 OLI data.

Despite the environmental and socio-economic importance of *montado* and its spatial distribution in the Mediterranean Basin, the development of remote sensing-based approaches for mapping this ecosystem with the use of medium spatial and spectral resolution imagery is rarely addressed in scientific literature (Carreiras et al., 2006; Godinho et al., 2015; Joffre and Lacaze, 1993). It is as such that this study aims to develop and propose a methodological approach for *montado* ecosystem mapping using a classification scheme which integrates Landsat 8 OLI multi-spectral data, vegetation indices, and an advanced machine learning algorithm. This study addresses three research goals directly:

1 Assessing the overall suitability and effectiveness of Landsat 8 OLI imagery for *montado* ecosystem mapping;
2 Assessing the improvement of *montado* classification accuracy by adding vegetation indices as proxy variables of plants physiological processes;
3 Evaluating the performance of the SGB classifier for *montado* land cover classification.

2. Material and methods

2.1. Study area and data

This study was conducted in southern Portugal (Fig. 1), a region with a markedly Mediterranean climate characterised by hot and dry summers (August: 31–32 °C T_{max}) and wet and cold winters (January: 6–7 °C T_{min}). Mean annual precipitation varies from 550 mm to 650 mm. The elevation range varies from 40 m to 645 m, and the mean slope is 3.52°, corresponding to a low roughness zone. The study area, covering roughly 8567 km^2, is mainly located in the biogeographic Luso-Extremadurense Province. *Montado* covers about 44.8% of the study area, being the predominant land use system in the region, followed by arable land (27.9%).

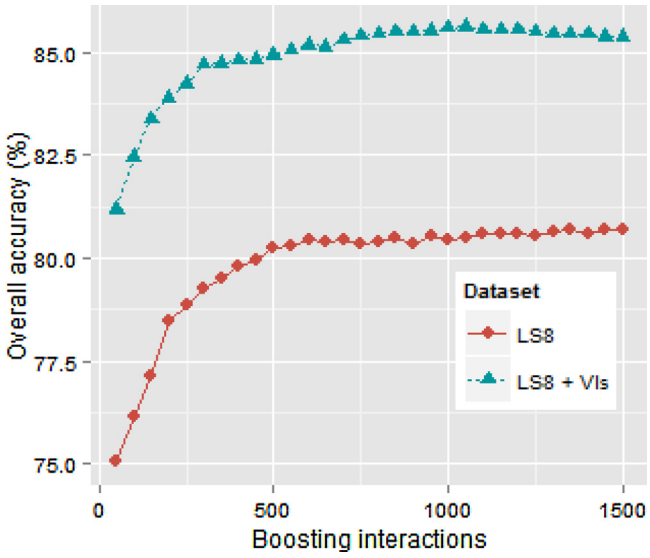

Fig. 4. Results of overall accuracy.

Two Landsat 8 images of the same area in southern Portugal were acquired in order to develop a multi-seasonal-based classification scheme. One image was taken in spring (15th May 2014) and the other in summer (19th August 2014), to ensure inter-class separability benefits from the phenological variation of the vegetation cover (Rodriguez-Galiano et al., 2012). May and August represent peaks in productivity and are also prominent periods in the phenological development of the major vegetation types in the study area (*e.g.*, Vaz et al., 2010; Senf et al., 2015). Landsat 8 imagery (path 203, row 33) entirely covers the study area. The processing level of the images corresponds to the "Standard Level 1 Terrain Corrected" (L1T). Both are cloud-free images. For the image classification procedure, only OLI bands 2 (Blue), 3 (Green), 4 (Red), 5 (NIR), 6 (SWIR1) and 7 (SWIR2) were used in this study.

2.2. Satellite data pre-processing

An atmospheric correction was applied to both images on selected bands, using the FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) method. FLAASH is an ENVI atmospheric correction algorithm based on MODTRAN4 radiative transfer code (Berk et al., 2002). Before running the FLAASH algorithm the images were radio-metrically calibrated using the "Radiometric Calibration" ENVI tool which corrects Digital Numbers (DN) to radiance values. The radiance images were then atmospherically corrected using MODTRAN4 code, resulting in surface reflectance images. In the FLAASH toolbox the "Mid-Latitude Summer" and "Rural" input parameters were selected, respectively, as atmospheric and aerosol models. These two parameters were selected because they are the most appropriate atmospheric and aerosol models for this study region.

2.3. Vegetation indices

Six vegetation indices were calculated from the selected Landsat 8 multi-spectral bands of both the spring and summer images: the Enhanced Vegetation Index (EVI), the Short-Wave Infrared Ratio (SWIR32), the Carotenoid Reflectance Index 1 (CRI1), the Green Chlorophyll Index (CI_{green}), the Normalised Multi-band Drought Index (NMDI), and the Soil-Adjusted Total Vegetation Index (SATVI) (Table 1). These vegetation indices were selected based on their respective sensitivity and effectiveness for monitoring vegetation cover and for retrieving vegetation parameters in semi-arid

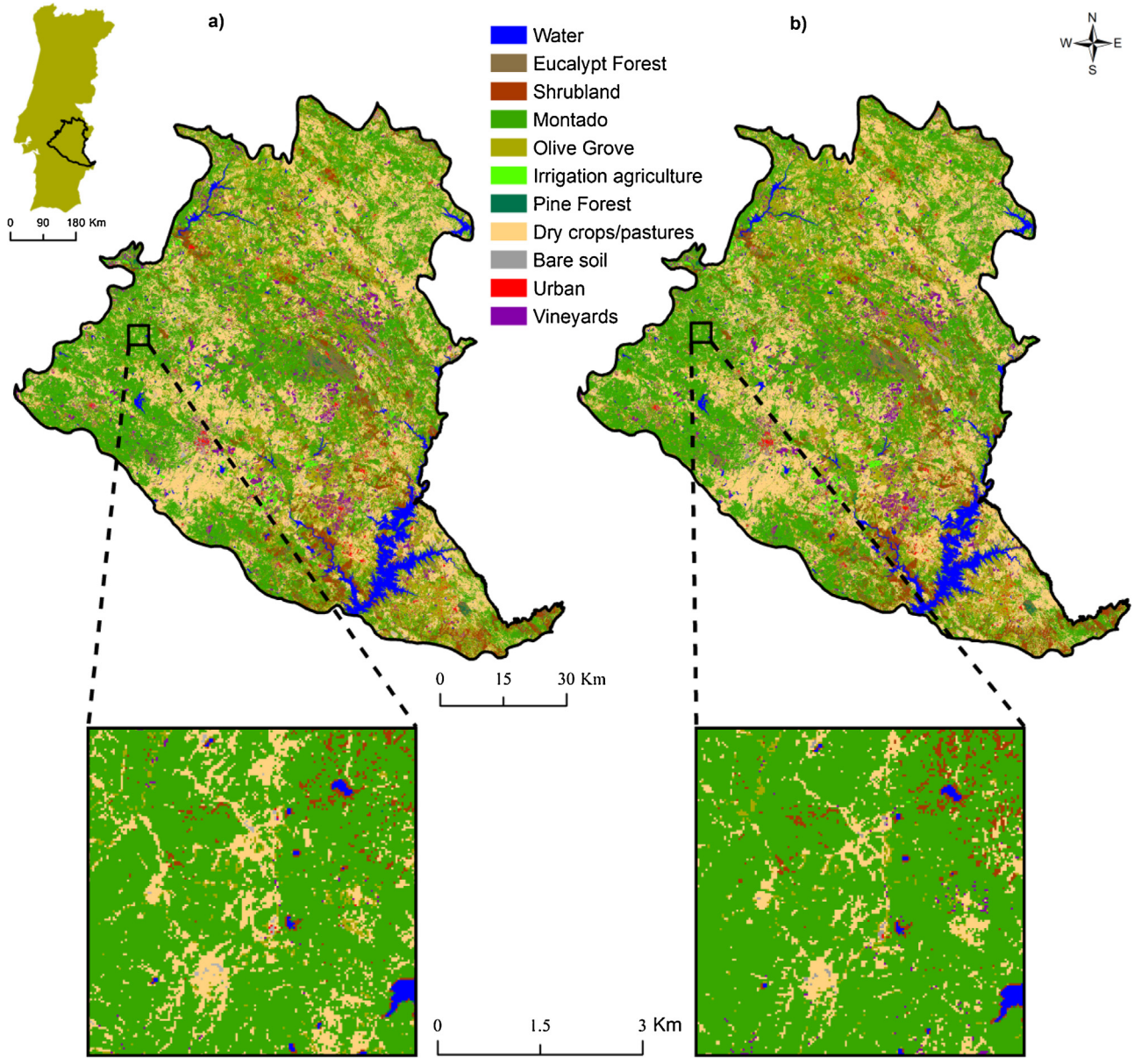

Fig. 5. Comparison of land cover products derived from SGB classification applied to Dataset I (LS8)—on the left—and Dataset II (LS8 + VIs)—on the right.

environments, in which water scarcity, soil background effects, and the predominance of senescent plant species are the main characteristics (*e.g.*, Hill, 2013; Marsett et al., 2006; Sauceda et al., 2008).

2.4. Training and validation data

A dataset including 1903 sample points covering the study area was produced using a stratified approach to ensure a thorough and representative reference dataset made up of eleven land cover categories (Table 2).

All of these points were chosen by photo-interpretation using a set of high-resolution (0.5 m) true-colour orthophotomaps (produced in 2005 by IGP—the Portuguese Geographic Institute), which is a well-established method for collecting reference land cover data (*e.g.*, Rodriguez-Galiano and Chica Olmo, 2012), and also by using the true colour composite of the summer Landsat 8 OLI image. This true colour composite was generated by merging the panchromatic band (band 8) with the multi-spectral bands to produce a pan-sharpened image which increases the level of spatial detail and, therefore, the photo interpretation process. Of the selected 1903 sample points, 174 were labelled with uncertainties during the photo-interpretation process (*montado* = 53, eucalyptus forest = 18, pine forest = 11, shrubland = 24, olive grove = 37, and vineyards = 31). Hence, to guarantee the overall quality of the final dataset of sample points, a field validation was performed in June 2014 to check all the points that presented less certitude in their labelling *via* photointerpretation. From the field verification it was detected that 47 sample points (*montado* = 12; eucalyptus forest = 7, pine forest = 3, shrubland = 9, olive grove = 8, and vineyards = 8) had indeed been mislabelled during the photo-interpretation task. Therefore, all of these points were replaced in the main dataset by new correct samples.

For each sample point, vegetation indices and reflectance values from the 12 bands (6 spring and 6 summer ones) were extracted from both Landsat 8 OLI images. Considering that the available pool of sample points for this research was restrictive (1903), we avoided splitting the dataset into training and validation sets. This was because each sample point was needed for model building, and

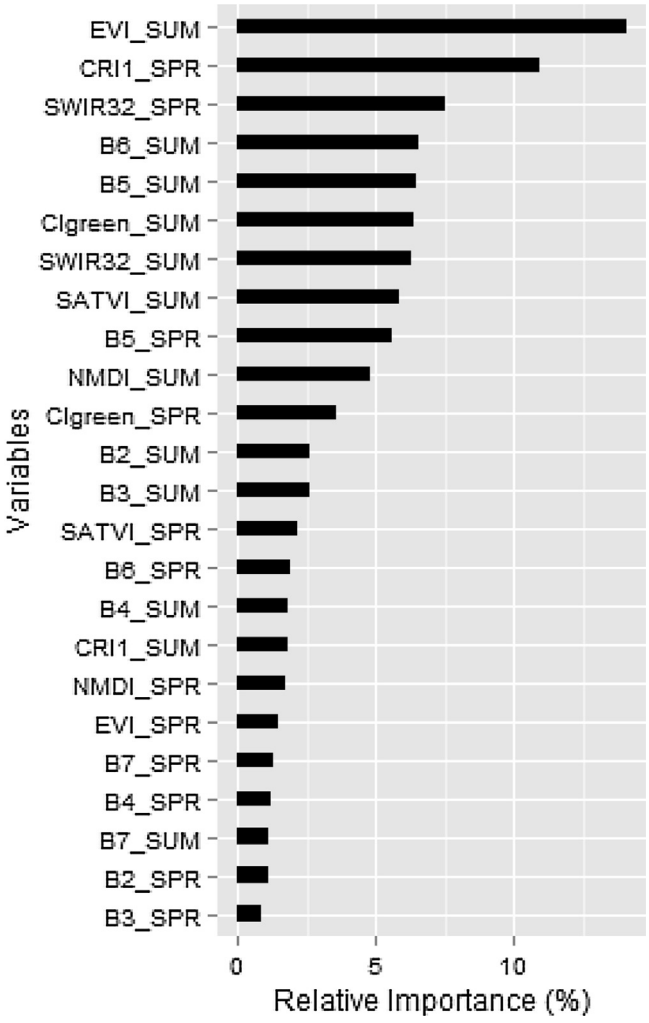

Fig. 6. Relative importance of each one of the 24 variables used in the SGB classification model applied to Dataset II. SUM: summer, SPR: spring.

also because the size of the validation set may have limited utility as a judge of the modelís performance (Kuhn and Johnson, 2013). Therefore, all available sample points were used to tune, train and validate SGB classification models by applying a repeated (three times) 10-fold cross-validation resampling method, which is considered to be an appropriate approach when large amounts of data are not available (Kuhn and Johnson, 2013).

2.5. Image classification using stochastic gradient boosting

SGB is a hybrid machine learning algorithm that combines both the advantages of bagging and boosting procedures (Friedman, 2001, 2002). Specifically, at each SGB iteration, a simple base learner (decision tree) is built using a random sub-sample of the training dataset (without replacement) producing substantial improvements to the model's accuracy. The boosting process in the SGB model is based on the steepest gradient algorithm in order to increase emphasis on misclassified training data that are close to their correct classification, rather than the worst classification data (Lawrence et al., 2004). Several advantages of using the SGB algorithm have been highlighted, including its low level of sensitivity to outlier effects, the ability to deal with inaccurate training and unbalanced datasets, the stochastic characteristic in modelling non-linear relationships, robustness in dealing with interaction effects among predictors, and also the capability to quantify the importance of variables (Friedman, 2001). Moreover, the stochastic component of the SGB algorithm is a powerful element in improving classification accuracy, as well as in reducing the occurrence of over-fitting (Friedman, 2002).

Analytically, the SGB algorithm involves a parameter-tuning process that maximises predictive accuracy. These parameters are: (i) bag fraction (*bf*), which is the random fraction of the training data used to perform each classification tree; (ii) tree complexity (*tc*), which represents the number of splits that should be performed in each tree; (iii) learning rate (*lr*), which determines the contribution of each tree to the growing model and helps to control over-fitting by controlling the gradient steps; and (iv) the number of trees (*nt*) (Elith et al., 2008). To determine the optimal combination of these parameters in the interests of achieving the highest overall model accuracy, a set of SGB models were tested using different values for *bf* (0.50, 0.60, and 0.75), *tc* (3, 5, 9, and 11), *lr* (0.001, 0.01, 0.05, and 0.07), and *nt* (50–1500). Optimal values for SGB-tuning parameters were selected for two different datasets: (i) using solely multi-seasonal Landsat 8 OLI spectral bands (Dataset I: LS8), and (ii) adding vegetation indices to the multi-seasonal Landsat 8 OLI spectral bands (Dataset II: LS8 + VIs). For both Datasets, the optimal bag fraction value was assessed testing $tc = (3, 5, 9, 11)$, $lr = 0.01$, and *nt* ranging from 50 to 1500 trees, in order to reduce the number of candidate models. Therefore, a total of 48 candidate SGB models were evaluated through the *gbm* and *caret* R packages (Kuhn, 2014; Ridgeway, 2013) implemented in the R statistical software (R Development Core Team, 2014).

2.6. Performance assessment

The performance of the SGB algorithm was compared to the K-Nearest Neighbour (KNN) classifier, which is one of the simplest, and yet widely used, instance-based learning algorithms for remote sensing applications (*e.g.*, Budreski et al., 2007; Marçal et al., 2005). KNN classification was performed using the *caret* package and the *knn* method (Kuhn and Johnson, 2013). This method only needs to tune the *k* parameter, the number of neighbours. The optimal value of *k* was determined using the same cross-validation resampling approach applied in the SGB procedure (Section 2.4). Confusion matrices resulting from a 10-fold cross-validation process were used to provide an estimation of SGB (Dataset I and II) and KNN (Dataset I) classifications accuracy. For each matrix, four traditional accuracy assessment measures were calculated: overall classification accuracy (OA), producer's accuracy (PA), user's accuracy (UA), and the Kappa coefficient (*K*) (Congalton and Green, 2009). In addition, to illustrate the superiority of one algorithm over another, the statistically significant difference between KNN and SGB classifications was evaluated using McNemarís test (Foody, 2004). Based on this non-parametric test, two classifications may be considered to be of different accuracy at the 95% level of confidence if $|Z| > 1.96$. This test was also used to assess the significance of differences in adding vegetation indices in the classification process (Dataset I *vs* Dataset II).

To investigate whether separate vegetation indices differ between land cover types, several Kruskal–Wallis and Tukey–Kramer tests (post-hoc pairwise comparisons) were performed. This approach is useful for analysing the degree of dissimilarity between vegetation indices across land cover types. This is, therefore, an indirect way of measuring the importance of these indices for the separability of land cover categories.

3. Results and discussion

3.1. Identifying the optimal SGB parameters

Based on the repeated 10-fold cross-validation results, the highest overall accuracy level was achieved with a bag fraction value of 0.50 for both Dataset I and II. In fact, no major differences were observed between the overall accuracy determined for the three tested bag fractions (Dataset I: $bf_{0.50}$ = 80.70%; $bf_{0.60}$ = 80.59%; $bf_{0.75}$ = 80.64%; Dataset II: $bf_{0.50}$ = 85.60%; $bf_{0.60}$ = 85.48%; $bf_{0.75}$ = 85.53%). As shown in Fig. 2, the tuning procedure determined that the highest overall accuracy levels were achieved when *tc* = 11, *lr* = 0.05, and *nt* = 1200 and when *tc* = 9, *lr* = 0.07, and *nt* = 500 parameter values were used for Dataset I (81.52% and 81.83%, respectively). Regarding Dataset II, the highest overall accuracy levels were obtained when *tc* = 11, *lr* = 0.01, and *nt* = 1050 and when *tc* = 11, *lr* = 0.05, and *nt* = 1050 parameter values were used (85.60% and 85.12%, respectively). However, the overall accuracy levels achieved by both 0.05 (Dataset I and II) and 0.07 (Dataset I) learning rates present a degree of instability (peaks and valleys in the *tc* curves) in the classification accuracies for each tree, which may have important practical implications related to SGB classification performance (Elith et al., 2008). According to Elith et al. (2008) it is preferable to use slower learning rates (*i.e.*, low values of *lr*) to control this instability in the models' accuracy. The results shown in Fig. 2 reveal that a more stable model for Dataset I, which also presented a high overall accuracy level (80.70%) was achieved when *lr* = 0.01 was used. It can therefore be concluded that for the data used in this study, *lr* = 0.01 is the optimal value for this SGB parameter for both Dataset I and II. In summary, the best combination of tuning parameters achieved during the tests, *i.e.*, *bf* = 0.50, *lr* = 0.01, and *tc* = 11, were used to produce the best SGB classification model for Dataset I and II.

3.2. Effectiveness of Landsat 8 OLI multi-spectral imagery for montado ecosystem mapping

The SGB classification performed only using the selected multi-seasonal Landsat 8 OLI spectral bands (Table 3) showed a moderate overall agreement and good accuracy (OA = 80.70%; K = 0.78), demonstrating that this sensor is suitable for generating land cover maps for these Mediterranean ecosystems. Regarding the SGB performance, McNemarís test clearly points to the significant improvements shown by SGB classification compared to the best KNN classification (*Z*-value = 5.28). The results included in Table 4 illustrate the per-class Kappa index and the difference between SGB and KNN classifications. One may observe that the SGB algorithm generally performs better than the KNN classifier, presenting an overall accuracy improvement of 5.80% and an average Kappa increase of 0.06 (SD = 0.04). At the land cover/vegetation class level, the *montado* ecosystem classification derived from the SGB algorithm was reasonably accurate (74.59% and 82.73% of user and producer accuracy, respectively) (Table 3). It can be seen, through the confusion matrix, that some areas of olive groves, dry crops/pastures, and vineyards were classified as *montado*, and vice versa. These errors occurred due to the spatial variability of tree density in the *montado* ecosystem, thereby contributing to a lower inter-class separability between these land cover categories. Some low-density *montado* areas (tree cover between 10% and 30%) were misclassified by having been mapped as dry crops/pastures and vineyards. In these areas, the vegetation cover density is generally sparse and, therefore, the high reflectance of the bare soil may have masked small components reflected from sparse vegetation (Berberoglu et al., 2000; Rodriguez-Galiano et al., 2012). On the other hand, in dense *montado* areas (tree density >50%), classification confusions may occur with olive groves, with certain limitations being present when only the original Landsat 8 OLI bands are used in the classification model.

3.3. Assessing how montado classification accuracy may be improved by adding vegetation indices to the classification scheme

The Kruskal–Wallis tests revealed that each vegetation index value was significantly different between land cover types ($p < 0.001$ for all vegetation indices). Water and irrigation agriculture were intentionally excluded from this analysis due to their high and implicit separability potential when compared to the remaining land cover types. The boxplot graphs (Fig. 3) clearly show the high degree of dissimilarity of the vegetation indices between land cover categories, which is a desirable factor in order to increase their separability (Tolpekin and Stein, 2009). Among the indices, the dissimilarity patterns related to EVI (summer), CRI1 (spring), and CI_{green} (summer) demonstrate their usefulness in differentiating the fractional cover of photosynthetic vegetation (*e.g.*, Hill, 2013). The differentiation of pine forest, eucalyptus forest, *montado*, and vineyards from the remaining categories became evident when the EVI (summer) index was used (Fig. 3a). During the summer season, the contrast between herbaceous layers and tree layers is more obvious than it is in spring because the herbaceous layer is already dry while the tree layer is still green (Carreiras et al., 2006). In such circumstances, NIR, red and blue reflectance values differ significantly between land cover types, where high NIR and generally low red and blue reflectance values can be found in green vegetation (*e.g.*, pine, eucalyptus, *montado* and vineyards) (Jones and Vaughan, 2010). Therefore, the results reported here clearly suggest that the EVI index (which incorporates NIR, red and blue bands) was far more useful for increasing the separability between land cover types in the summer season, in particular for distinguishing herbaceous and shrubland cover from the tree layer. In addition, EVI was also suitable for increasing the separability between olive grove and *montado* areas. Both land cover types constitute one of the most confused pairs when only using the original Landsat 8 OLI bands (Table 3).

The use of leaf pigment indices such as the CRI1 and CI_{green} resulted in an increase in class separability. These vegetation indices were extremely useful, for example, for distinguishing *montado* from vineyards (CRI1) and from olive groves (CI_{green}), as well as for enhancing the differences between eucalyptus and pine forests (Fig. 3b and d). However, neither index was suitable for differentiating urban-crops/pastures and urban-bare soils land cover pairs (Tukey–Kramer pairwise comparisons show $p > 0.05$). To improve the differentiation between such land cover classes, further auxiliary information could be integrated into the classification procedure, such as textural information or digital elevation models (Rodriguez-Galiano and Chica Olmo, 2012). Several authors (*e.g.*, Dash et al., 2007; Sauceda et al., 2008) have reported that leaf pigment content (*e.g.*, chlorophylls and carotenoids) in arid and semi-arid landscapes vary from season to season and between vegetation types, thereby providing valuable information that may enhance land cover spectral differences.

The contrasting behaviour presented by the SWIR32 and SATVI indices (Fig. 3c, e, and f), which are both short-wave infrared-derived indices, emphasises their effectiveness in dealing with non-photosynthetic vegetation, bare soil, and green vegetation fractions (Guerschman et al., 2009; Hill, 2013; Marsett et al., 2006). High SWIR32 index (spring and summer) values were found mainly in urban areas, bare soils, and vineyards. The dissimilarity between these three land cover types was significant ($p < 0.001$). However, the spectral difference between urban areas and bare soils was much higher in spring than in summer (Fig. 3c and f), because bare soils are covered by grass in spring (Rodriguez-Galiano et al., 2012). The SATVI index was useful for distinguishing dry crops/pastures,

montado areas, and shrubland from the other categories by minimising soil spectral signal effects, which constitute a particular feature of this index (Marsett et al., 2006). Finally, the inclusion of the NMDI index in the classification model also proved useful for supporting a more effective differentiation of land cover classes. Given the water scarcity of the Mediterranean region, soil and vegetation moisture are very important factors in determining the spectral behaviour of land cover types. Bare soils may have similar reflectance features to urban areas and similar NIR reflectance to dry crops/pastures (Berberoglu et al., 2000). However, the NMDI index has the capacity to quantify the water content of soils and vegetation, where high values of NMDI indicate the presence of dry bare soils (Wang and Qu, 2007). Sparsely vegetated areas, such as vineyards and some *montado* areas (with tree density 10–30%) were distinguished from bare soils and urban areas when the summer NMDI index was included ($p < 0.001$). As reported by Wang and Qu (2007), using the NMDI index may be effective for detecting significant differences in vegetation moisture. In fact, the differentiation between shrubland and the remaining vegetation classes in this study was evident ($p < 0.001$) given the extremely low NMDI value observed.

Fig. 4 shows the results obtained by applying the SGB algorithm to Dataset I (LS8) and Dataset II (LS + VIs). As may be observed, the highest classification accuracy level was achieved when vegetation indices were used in combination with the multi-seasonal Landsat 8 OLI bands, attaining an overall accuracy level of 85.60%.

Table 5 illustrates the confusion matrix resulting from the 10-fold cross-validation procedure for the accuracy assessment of the SGB model when vegetation indices were included in the classification process. Overall, strong agreement and high levels of accuracy (OA = 85.60% and $K = 0.84$) were observed. For the land cover/vegetation class level, all classes also showed a strong agreement and a good accuracy (UA >76%). Regarding the *montado* ecosystems, the results were highly positive (PA = 86.36%; UA = 80.85%). The combination of vegetation indices with multi-spectral bands provided the best classification performance, with overall accuracy and the Kappa coefficient increasing by 4.90% and 0.06, respectively, when compared with the same accuracy measures obtained by applying SGB to the Landsat 8 OLI multi-spectral bands (Tables 3 and 5).

Integrating vegetation indices into the classification model resulted in significant improvements to accuracy levels for the Dataset II classifications when compared with Dataset I (Tables 3 and 5) (McNemar's test: Z-value = 4.50), leading to differences in the producer accuracy values for olive groves (9.36%), eucalyptus forest (8.07%), and urban areas (7.31%). With regard to user accuracy results, a significant increase was also observed, mainly for dry crops/pastures (8.49%), olive groves (7.65%), and *montado* (6.26%). In addition, the estimated per-class kappa values also demonstrated that these three land cover types were classified better when using Dataset II, showing a kappa value increase of 0.10, 0.09, and 0.08, respectively (Table 6). These results contrast with those obtained by Li et al. (2011), who argued that vegetation indices had a limited role in improving overall classification performance levels. For instance, by performing a visual interpretation of the resulting SGB classification maps derived from both Datasets I and II, it can be seen that in the first case (Dataset I-derived), some *montado* patches were classified as dry crops/pasture areas. Nevertheless, this misclassification did not occur when using Dataset II, which included a combination of selected vegetation indices (Fig. 5a and b).

Fig. 6 shows the relative contribution of the 24 variables used, and it is clear that vegetation indices have the highest relative importance when it comes to differentiating between land cover/vegetation classes. Based on these results, EVI, CRI1, SWIR32, CI_{green}, SATVI, and NMDI ranked in the top 10 most significant variables for classification, with EVI (Summer), CRI1 (Spring), and SWIR32 (Spring) being the three most prominent ones. As reported by Dash et al. (2007), the results obtained in this study support the conclusion that data acquired in the summer have a greater differentiating element than data acquired in the spring. In fact, seven variables from the top 10 of those most significant for classification were sourced from summer data (Fig. 6).

4. Conclusions

The classification of the Landsat 8 OLI imagery combined with vegetation indices as ancillary information was performed using an SGB machine-learning algorithm. Good levels of accuracy were obtained for both LS8 and LS8 + VIs products (80.70% and 85.60%, respectively) by using this methodological approach. It has been demonstrated that SGB has the capability for producing accurate models using remote sensing data covering a complex landscape such as the Mediterranean region. Moreover, the importance that the SGB tuning parameter procedure had on the results obtained in this study confirms the significance of selecting the best combination of parameters, given that the parameterisation process is image data-driven.

The classification accuracy of the thematic maps produced with the use of multi-seasonal Landsat 8 OLI spectral bands increased significantly with the integration of vegetation indices into the classification model. This methodological approach allowed for the most suitable vegetation indices to be determined in order to map *montado* ecosystems more accurately. As a result, for the *montado* category, the highest level of accuracy was obtained using the dataset that combines both multi-spectral bands and vegetation indices, yielding an average accuracy of 83.61% and an increase of 0.09 in Kappa value. The usefulness of vegetation indices as variables to differentiate land cover classification were highlighted throughout the procedure adopted, where EVI (summer), CRI1 (spring), and SWIR32 (spring) vegetation indices were the most useful, accounting for a combined relative importance of 32.53%. It was also shown that this multi-seasonal approach may significantly enhance existing differences in the variations of phenological and spatial features of vegetation classes between spring and summer, and therefore be more useful in discriminating between classes of vegetation. In short, the results provided by this study may be of great relevance to land cover classification in the Mediterranean. It has been shown that the use of vegetation indices as a source of ancillary information may noticeably enhance the biophysical differences between land cover/vegetation categories, and therefore increase their spectral inter-separability.

The lack of accurate *montado* maps describing the ecosystem's spatial-temporal distribution limits the implementation of a comprehensive *montado* monitoring programme. Such information is crucial to supporting decision-making in the implementation and monitoring of land planning and management policies. This study constitutes a first step towards the development of a broader long-term *montado* research line that addresses the need to identify and map the spatial patterns of these ecosystems in southern Portugal over recent decades by using current (Landsat 8 OLI) and previous mission Landsat data (TM and ETM+ sensors).

Acknowledgements

This research was funded by the Portuguese Foundation for Science and Technology (FCT—Fundação para a Ciência e Tecnologia) under the POPH-QREN-Tipologia 4.1 Programme, scholarship number SFRH/BD/77897/2011 to Sérgio Godinho, and Post-Doctoral Research Project with the reference SFRH/BPD/100017/2014 to co-author Artur Gil.

References

Berberoglu, S., Lloyd, C.D., Atkinson, P.M., Curran, P.J., 2000. The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean. Comput. Geosci. 26, 385–396, http://dx.doi.org/10.1016/S0098-3004(99)00119-3.

Berberoglu, S., Curran, P.J., Lloyd, C.D., Atkinson, P.M., 2007. Texture classification of Mediterranean land cover. Int. J. Appl. Earth Obs. 9, 322–334, http://dx.doi.org/10.1016/j.jag.2006.11.004.

Berk, A., Adler-Golden, S., Ratkowski, A., Felde, G., Anderson, G., Hoke, M., Cooley, T., Chetwynd, J., Gardner, J., Matthew, M., 2002. Exploiting MODTRAN radiation transport for atmospheric correction: the FLAASH algorithm. In: In Proceedings of the Fifth International Conference on Information Fusion, Annapolis, MA, USA, 8–11 July 2002, pp. 798–803.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32, http://dx.doi.org/10.1023/A:1010933404324.

Budreski, K.A., Wynne, R.H., Browder, J.O., Campbell, J.B., 2007. Comparison of segment and pixel-based non-parametric land cover classification in the Brazilian Amazon using multitemporal Landsat TM/ETM+ imagery. Photogramm. Eng. Remote Sens. 73, 813–827, http://dx.doi.org/10.14358/PERS.73.7.813.

Bugalho, M., Plieninger, T., Aronson, J., Ellatifi, M., Crespo, D.G., 2009. Open woodlands: a diversity of uses (and overuses). In: Aronson, J., Pereira, J.S., Pausas, J.G. (Eds.), Cork Oak Woodlands on the Edge, Ecology, Adaptive Management, and Restoration. , first ed. Society for Ecological Restoration International, Island Press, Washington D.C, pp. 33–45.

Carrão, H., Gonçalves, P., Caetano, M., 2008. Contribution of multispectral and multitemporal information from MODIS images to land cover classification. Remote Sens. Environ. 112, 986–997, http://dx.doi.org/10.1016/j.rse.2007.07.002.

Carreiras, J.M.B., Pereira, J.M.C., Pereira, J.S., 2006. Estimation of tree canopy cover in evergreen oak woodlands using remote sensing. For. Ecol. Manag. 223, 45–53, http://dx.doi.org/10.1016/j.foreco.2005.10.056.

Chirici, G., Scotti, R., Montaghi, A., Barbati, A., Cartisano, R., Lopez, G., Marchetti, M., McRoberts, R.E., Olsson, H., Corona, P., 2013. Stochastic gradient boosting classification trees for forest fuel types mapping through airborne laser scanning and IRS LISS-III imagery. Int. J. Appl. Earth Obs. Geoinf. 25, 87–97, http://dx.doi.org/10.1016/j.jag.2013.04.006.

Cohen, W.B., Goward, S.N., 2004. Landsat's role in ecological applications of remote sensing. Bioscience 54, 535–545, http://dx.doi.org/10.1641/0006-3568(2004)054[0535].

Congalton, R.G., Green, K., 2009. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, second ed. CRC Press, Boca Raton, Florida.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297, http://dx.doi.org/10.1023/A:1022627411411.

Dash, J., Mathur, A., Foody, G.M., Curran, P.J., Chipman, J.W., Lillesand, T.M., 2007. Land cover classification using multi-temporal MERIS vegetation indices. Int. J. Remote Sens. 28, 1137–1159, http://dx.doi.org/10.1080/01431160600784259.

Doorn, v.A.M., Pinto-Correia, T., 2007. Differences in land cover interpretation in landscapes rich in cover gradients: reflections based on the montado of South Portugal. Agrofor. Syst. 70, 169–183, http://dx.doi.org/10.1007/s10457-007-9055-8.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813, http://dx.doi.org/10.1111/j.1365-2656.2008.01390.x.

Foody, G.M., 2004. Thematic map comparison: evaluating statistical significance of differences in classification accuracy. Photogramm. Eng. Remote Sens. 70 (5), 627–633, http://dx.doi.org/10.14358/PERS.70.5.627.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38, 367–378, http://dx.doi.org/10.1016/S0167-9473(01)00065-2.

Gitelson, A.A., Zur, Y., Chivkunova, O.B., Merzlyak, M.N., 2002. Assessing carotenoid content in plant leaves with reflectance spectroscopy. Photochem. Photobiol. 75, 272–281.

Gitelson, A.A., Gritz, U., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J. Plant Physiol. 160, 271–282, http://dx.doi.org/10.1078/0176-1617-00887.

Godinho, S., et al., 2014. Assessment of environment, land management, and spatial variables on recent changes in montado land cover in southern Portugal. Agrofor. Syst., http://dx.doi.org/10.1007/s10457014-9757-7.

Godinho, S., Gil, A., Guiomar, N., Neves, N., Pinto-Correia, T., 2015. A remote sensing-based approach to estimating montado canopy density using the FCD model: a contribution to identifying HNV farmlands in southern Portugal. Agrofor. Syst., http://dx.doi.org/10.1007/s10457-014-9769-3.

Guerschman, J.P., Hill, M.J., Renzullo, L.J., Barret, D., Marks, A.S., Botha, E.J., 2009. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1Hyperion and MODIS sensors. Remote Sens. Environ. 113, 928–945, http://dx.doi.org/10.1016/j.rse.2009.01.006.

Hill, M.J., 2013. Vegetation index suites as indicators of vegetation state in grassland and savanna: an analysis with simulated SENTINEL 2 data for a North American transect. Remote Sens. Environ. 137, 94–111, http://dx.doi.org/10.1016/j.rse.2013.06.004.

Huang, C., Davis, L.S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. Int. J. Remote Sens. 23, 725–749, http://dx.doi.org/10.1080/01431160110040323.

Huete, A.R., Liu, H.Q., Batchily, K., van Leeuwen, W.J.D., 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote Sens. Environ. 59, 440–451, http://dx.doi.org/10.1016/S0034-4257(96)00112-5.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213, http://dx.doi.org/10.1016/S0034-4257(02)00096-2.

Irons, J.R., Dwyer, J.L., Barsi, J.A., 2012. The next Landsat satellite: the Landsat Data continuity mission. Remote Sens. Environ. 122, 11–21, http://dx.doi.org/10.1016/j.rse.2011.08.026.

Jia, K., Wei, X., Gub, X., Yao, Y., Xie, X., Li, B., 2014. Land cover classification using Landsat 8 Operational Land Imager data in Beijing China. Geocarto Int. 29, 941–951, http://dx.doi.org/10.1080/10106049.2014.894586.

Joffre, R., Lacaze, B., 1993. Estimating tree density in oak savanna-like 'dehesa' of southern Spain from SPOT data. Int. J. Remote Sens. 14, 685–697, http://dx.doi.org/10.1080/01431169308904368.

Joffre, R., Rambal, S., Ratte, J.P., 1999. The dehesa system of southern Spain and Portugal as a natural ecosystem mimic. Agrofor. Syst. 45, 57–79, http://dx.doi.org/10.1023/A:1006259402496.

Jones, H.G., Vaughan, R.A., 2010. Remote Sensing of Vegetation: Principles, Techniques, and Applications. Oxford University Press, 400 pp.

Julien, Y., Sobrino, J.A., Jiménez-Muñoz, J.C., 2011. Land use classification from multitemporal Landsat imagery using the Yearly Land Cover Dynamics (YLCD) method. Int. J. Appl. Earth Obs. Geoinf. 13, 711–720, http://dx.doi.org/10.1016/j.jag.2011.05.008.

Kerekes, J.P., 1994. NDVI sensitivity to atmospheric water vapor as a function of spectral bandwidth. Geoscience and remote sensing symposium: surface and atmospheric remote sensing: technologies. Data Anal. Interpret., 1506–1508, http://dx.doi.org/10.1109/IGARSS.1994.399482.

Kuhn, M., 2014. *caret*: Classification and Regression Training. R package version 6.0–37 (accessed 11.14.) http://CRAN.R-project.org/package=caret.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modelling, 600. Springer, New York, http://dx.doi.org/10.1007/978-1-4614-6849-3.

Lawrence, R., Bunn, A., Powell, S., Zambon, M., 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. Remote Sens. Environ. 90, 331–336, http://dx.doi.org/10.1016/j.rse.2004.01.007.

Li, G., Lu, D., Moran, E., Hetrick, S., 2011. Land-cover classification in a moist tropical region of Brazil with Landsat TM imagery. Int. J. Remote Sens. 32, 8207–8230, http://dx.doi.org/10.1080/01431161.2010.532831.

Li, P., Jiang, L., Fen, Z., 2014. Cross-comparison of vegetation indices derived from Landsat-7 enhanced thematic mapper plus (ETM+) and Landsat-8 operational land imager (OLI) sensors. Remote Sens. 6, 310–329, http://dx.doi.org/10.3390/rs6010310.

Marçal, A.R.S., Borges, J.S., Gomes, J.A., Pinto da Costa, J.F., 2005. Land cover update by supervised classification of segmented ASTER images. Int. J. Remote Sens. 26, 1347–1362, http://dx.doi.org/10.1080/01431160412331291233.

Marsett, R.C., Qi, J.G., Heilman, P., Biedenbender, S.H., Watson, M.C., Amer, S., Weltz, M., Goodrich, D., Marsett, R., 2006. Remote sensing for grassland management in the arid Southwest. Range Ecol. Manag. 59, 530–540, http://dx.doi.org/10.2111/05-201R.1.

Olea, L., San Miguel-Ayanz, A., 2006. The Spanish dehesa: a traditional Mediterranean silvopastoral system linking production and nature conservation. 21st General Meeting of the European Grassland Federation, Badajoz (Spain) (accessed 01.14.) http://www.doctorrange.com/PDF/Dehesa.pdf.

Pinto-Correia, T., 2000. Future development in Portuguese rural areas: how to manage agricultural support for landscape conservation? Landsc. Urban Plan. 50, 95–106, http://dx.doi.org/10.1016/S0169-2046(00)00082-7.

R Development Core Team, 2014. R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Ridgeway, G., 2013. Generalized boosted regression models. Documentation on the R Package 'gbm', version 2.1 (accessed 11.14.) http://cran.r-project.org/web/packages/gbm/gbm.pdf.

Rodriguez-Galiano, V.F., Chica-Olmo, M., 2012. Land cover change analysis of Mediterranean area in Spain using different sources of data: multi-seasonal Landsat images, land surface temperature, digital terrain models and texture. Appl. Geogr. 35, 208–218, http://dx.doi.org/10.1016/j.apgeog.2012.06.014.

Rodriguez-Galiano, V.F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M., Jeganathan, C., 2012. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. Remote Sens. Environ. 121, 93–107, http://dx.doi.org/10.1016/j.rse.2011.12.003.

Rogan, J., Chen, D., 2004. Remote sensing technology for mapping and monitoring land-cover and land-use change. Prog. Plan. 61, 301–325, http://dx.doi.org/10.1016/S0305-9006(03)00066-7.

Roy, D.P., et al., 2014. Landsat-8: science and product vision for terrestrial global change research. Remote Sens. Environ. 145, 154–172, http://dx.doi.org/10.1016/j.rse.2014.02.001.

Sauceda, U.J.I., Gonzalez Rodriguez, H., Ramirez Lozano, R.G., Silva, C.I., Gomez, M.M.V., 2008. Seasonal trends of chlorophylls a and b and carotenoids in native trees and shrubs of Northeastern Mexico. J. Biol. Sci. 8, 258–267, http://dx.doi.org/10.3923/jbs.2008.258.267.

Senf, C., Leitão, P.J., Pflugmacher, D., Linden, S., Hostert, P., 2015. Mapping land cover in complex Mediterranean landscapes using Landsat: improved classification accuracies from integrating multi-seasonal and synthetic imagery. Remote Sens. Environ. 156, 527–536, http://dx.doi.org/10.1016/j.rse.2014.10.018.

Surová, D., Surový, P., Ribeiro, N.A., Pinto-Correia, T., 2011. Integration of landscape preferences to support the multifunctional management of the Montado system. Agrofor. Syst. 82, 225–237, http://dx.doi.org/10.1007/s10457-011-9373-8.

Tolpekin, V.A., Stein, A., 2009. Quantification of the effects of land-cover-class spectral separability on the accuracy of markov-random-field-based superresolution mapping. IEEE Trans. Geosci. Remote Sens. 47, 3283–3297, http://dx.doi.org/10.1109/TGRS.2009.2019126.

Vaz, M., Pereira, J.S., Gazarini, L.C., David, T.S., David, J.S., Rodrigues, A., Maroco, J., Chaves, M.M., 2010. Drought-induced photosynthetic inhibition and autumn recovery in two Mediterranean oak species (*Quercus ilex* and *Quercus suber*). Tree Physiol. 30, 946–956, http://dx.doi.org/10.1093/treephys/tpq044.

Wang, L., Qu, J.J., 2007. NMDI: a normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. Geophys. Res. Lett. 34, 1–5, http://dx.doi.org/10.1029/2007GL031021.

Xie, Y., Sha, Z., Yu, M., 2008. Remote sensing imagery in vegetation mapping: a review. Plant Ecol. 1, 9–23, http://dx.doi.org/10.1093/jpe/rtm005.