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

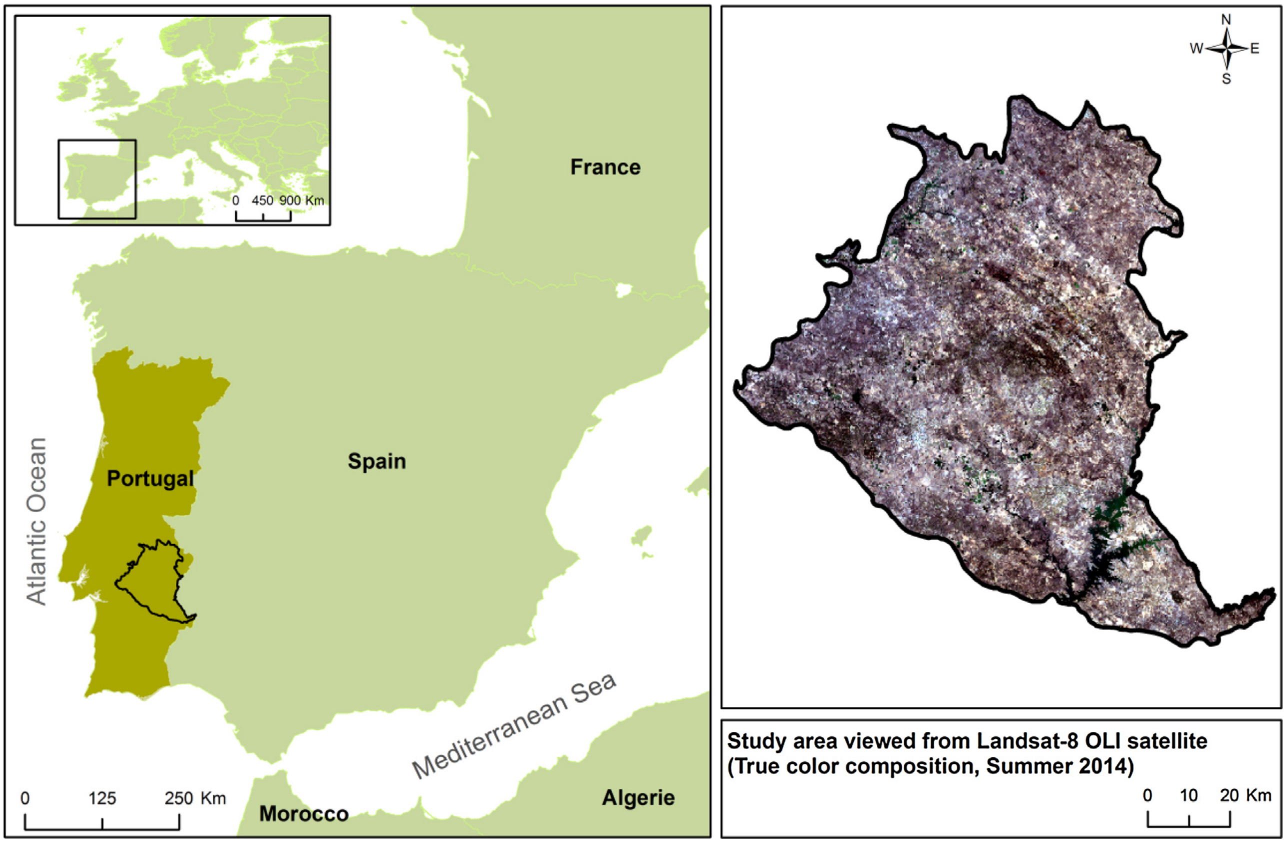

Fig. 1. Study area.

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

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