ISPRS Journal of Photogrammetry and Remote Sensing 67 (2012) 93–104

Contents lists available at SciVerse ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

An assessment of the effectiveness of a random forest classifier for land-cover classification

V.F. Rodriguez-Galiano [a,*], B. Ghimire [b], J. Rogan [b], M. Chica-Olmo [a], J.P. Rigol-Sanchez [c]

[a] *Dept. de Geodinámica, Universidad de Granada, Granada 18071, Spain*
[b] *Graduate School of Geography, Clark University, Worcester, MA, USA*
[c] *Dept. de Geología, Universidad de Jaén, Jaén 23071, Spain*

ARTICLE INFO

Article history:
Received 15 March 2011
Received in revised form 31 October 2011
Accepted 3 November 2011
Available online 1 December 2011

Keywords:
Remote sensing
Machine learning
Classification
Random forest
Land-cover
Landsat Thematic Mapper

ABSTRACT

Land cover monitoring using remotely sensed data requires robust classification methods which allow for the accurate mapping of complex land cover and land use categories. Random forest (RF) is a powerful machine learning classifier that is relatively unknown in land remote sensing and has not been evaluated thoroughly by the remote sensing community compared to more conventional pattern recognition techniques. Key advantages of RF include: their non-parametric nature; high classification accuracy; and capability to determine variable importance. However, the split rules for classification are unknown, therefore RF can be considered to be black box type classifier. RF provides an algorithm for estimating missing values; and flexibility to perform several types of data analysis, including regression, classification, survival analysis, and unsupervised learning.

In this paper, the performance of the RF classifier for land cover classification of a complex area is explored. Evaluation was based on several criteria: mapping accuracy, sensitivity to data set size and noise. Landsat-5 Thematic Mapper data captured in European spring and summer were used with auxiliary variables derived from a digital terrain model to classify 14 different land categories in the south of Spain. Results show that the RF algorithm yields accurate land cover classifications, with 92% overall accuracy and a Kappa index of 0.92. RF is robust to training data reduction and noise because significant differences in kappa values were only observed for data reduction and noise addition values greater than 50 and 20%, respectively. Additionally, variables that RF identified as most important for classifying land cover coincided with expectations. A McNemar test indicates an overall better performance of the random forest model over a single decision tree at the 0.00001 significance level.

1. Introduction

Land-cover mapping and monitoring is one of the major applications of Earth observing satellite sensor data and is essential for the estimation of land cover change. Large scale land cover monitoring is important because human and/or natural land cover modifications affect biophysical and biogeochemical properties of land surfaces (Bala et al., 2007; Betts et al., 2007; Bonan, 2008; Brovkin et al., 2004). Large area monitoring is used to estimate land-cover change and deforestation, perform forest inventory, and determine priority areas for biodiversity conservation (Lambin et al., 2001; Mas et al., 2004; Turner et al., 2007). Additionally, changes in land-cover affect the climate through changes in the composition of carbon dioxide and other greenhouse gases in the atmosphere (Bala et al., 2007; Betts et al., 2007; Bonan, 2008; Brovkin et al., 2004; Fearnside, 2000). Thus, many applications rely on reliable and timely land-cover mapping products over large heterogeneous landscapes.

Increased numbers of satellite sensor images have made it easier to establish land-cover monitoring programs for large area mapping over regular time intervals (Friedl et al., 2002). Operational large area land monitoring programs are well established (Franklin and Wulder, 2002). Unfortunately, there are several limitations related to large area monitoring that need to be resolved. First, large area mapping in complex landscapes is difficult because of abrupt changes in environmental gradients (e.g. moisture, elevation and temperature) and a legacy of past disturbance (Rogan and Miller, 2006). Such heterogeneous landscapes are characterized by land-cover categories that are difficult to separate spectrally due to low inter-class separability and high intra-class variability. Second, large area mapping requires algorithms that can be interpreted

* Corresponding author. Tel.: +34 958 243363; fax: +34 958 248527.
E-mail addresses: vrgaliano@ugr.es (V.F. Rodriguez-Galiano), bghimire@clarku.edu (B. Ghimire), jrogan@clarku.edu (J. Rogan), mchica@ugr.es (M. Chica-Olmo), jprigol@ujaen.es (J.P. Rigol-Sanchez).

0924-2716/$ - see front matter © 2011 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2011.11.002

readily and automated as well as operated easily due to user-defined parameters that are simple to adjust. Third, the choice of a suitable land-cover classification algorithm for large area mapping depends on the ability of the algorithm to handle noisy observations, a complex measurement space, and a small number of training data relative to the size of the study area (DeFries and Chan, 2000; Rogan et al., 2008).

A variety of classification methods have been used to map land cover using remotely sensed data. Classification methods range from unsupervised algorithms such as ISODATA or *K*-means to parametric supervised algorithms such as maximum likelihood (Jensen, 2005); to machine learning algorithms such as artificial neural networks (Mas and Flores, 2008), decision trees (Breiman, 1984), support vector machines (Mountrakis et al., 2011) and ensembles of classifiers (Breiman, 1996). In the last five years, machine learning algorithms have emerged as more accurate and efficient alternatives to conventional parametric algorithms, when faced with large dimensional and complex data spaces and have been used for large area mapping (Hansen et al., 1996; Huang et al., 2002; Rogan et al., 2003). These algorithms are efficient and effective because they do not rely on data distribution assumptions (e.g. normality) and generally have higher accuracy (Foody, 1995; Friedl and Brodley, 1997). However, some machine learning techniques (e.g. neural networks and support vector machines) are complicated due to the large number of parameters that need to be adjusted and are difficult to automate (Atkinson and Tatnall, 1997; Foody, 2004). Additionally these algorithms have a tendency to over-fit the data (Breiman et al., 1984).

An emerging type of machine learning technique which utilizes ensembles of classifications (e.g. neural network ensembles, random forests, bagging and boosting) is receiving highlighted interest (Friedl et al., 1999; Ghimire et al., 2010; Gislason et al., 2006; Hansen and Salamon, 1990; Krogh and Vedelsby, 1995; Sesnie et al., 2008; Steele, 2000). Ensemble learning algorithms use the same base classifier to produce repeated multiple classifications of the same data (Breiman, 2001; Friedl et al., 1999), or use a combination of different base classifiers to generate multiple classifications of the same data or to target different subsets of the data (Mountrakis et al., 2009). The collection of multiple classifiers of the same data are combined using a rule based approach (such as, maximum voting, product, sum, and Bayesian rule), or based on an iterative error minimization technique by reducing the weights for the correctly classified samples (e.g. boosting) (Friedl et al., 1999; Ghimire et al., 2010; Steele, 2000). Ensemble learning techniques have higher accuracy than other machine learning algorithms because the group of classifiers performs more accurately than any single classifier, and utilizes the strengths of the individual group of classifiers while at the same time the classifier weaknesses are circumvented (Ghimire et al., 2010; Kotsiantis and Pintelas, 2004).

An ensemble learning technique called random forests is increasingly being applied in land-cover classification using multispectral and hyperspectral satellite sensor imagery (Chan and Paelinckx, 2008; Ghimire et al., 2010; Lawrence et al., 2006; Pal, 2005; Sesnie et al., 2008), and lidar and radar data (Guo et al., 2011; Latifi et al., 2010; Martinuzzi et al., 2009; Waske and Braun, 2009). However, most studies that have used random forests have focused on relatively small study areas (Pal, 2005; Waske and Braun, 2009), classified few land-cover classes (Gislason et al., 2006; Lawrence et al., 2006; Prasad et al., 2006), or only used single season imagery for classification (Chapman et al., 2010; Ghimire et al., 2010; Ham et al., 2005). Moreover, most studies have not investigated the behavior of the random forest classifier by assessing the influence of training data quality/noise and variations in training data set size on classifier performance (Chan and Paelinckx, 2008; Gislason et al., 2006; Ham et al., 2005; Lawrence et al., 2006; Pal, 2005; Prasad et al., 2006; Sesnie et al., 2008). The objective of this study was to assess the performance of the random forest classifier in a large heterogeneous landscape with diverse land-cover categories using multi-seasonal Landsat, and auxiliary data. The behavior of random forests is assessed by considering multiple criteria related to variations in classifier parameter values, and sensitivity to noise and training size variations. The performance of the RF is also evaluated in comparison to classification trees.

2. Study area

The Province of Granada (GP) is the study area chosen for this project. It is located in the south of Spain on the Mediterranean coast, encircled by the Penibetica mountain range (Fig. 1). This area occupies 12,635 km^2 and elevation ranges from sea level to the

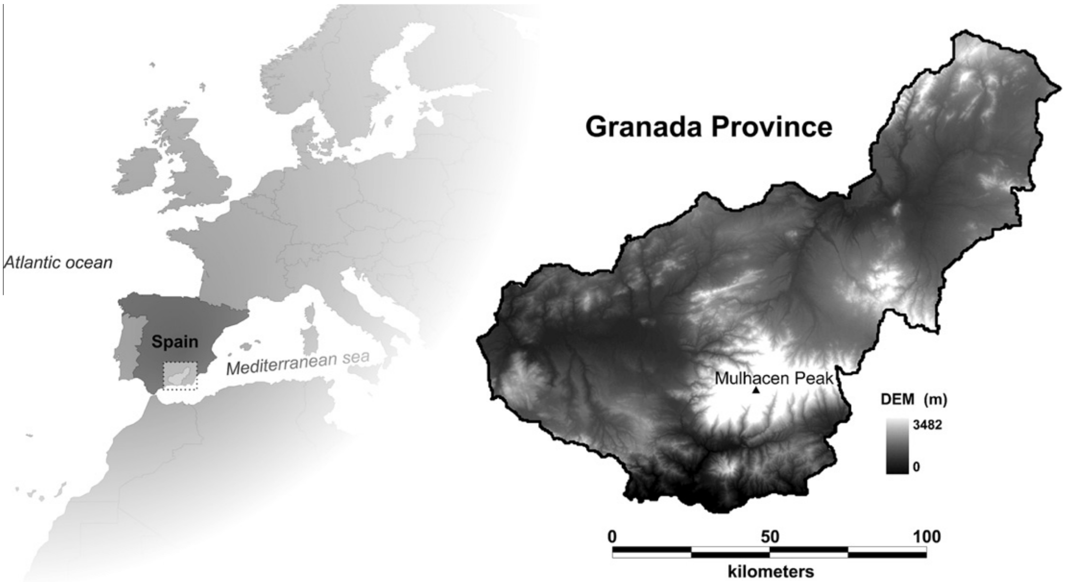

Fig. 1. Location of study area in Spain.