

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/249970185>

Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models

Article in *International Journal of Digital Earth* · December 2012

DOI: 10.1080/17538947.2012.748848

CITATIONS

133

READS

2,400

2 authors:



Víctor F. Rodríguez-Galiano

Universidad de Sevilla

72 PUBLICATIONS 6,145 CITATIONS

[SEE PROFILE](#)



Mario Chica Rivas

University of Granada

14 PUBLICATIONS 1,439 CITATIONS

[SEE PROFILE](#)

Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models

Victor F. Rodriguez-Galiano^{a*} and Mario Chica-Rivas^b

^a*Departamento de Geodinámica, Universidad de Granada, Granada, Spain;* ^b*Departamento de Análisis Matemático, Universidad de Granada, Granada, Spain*

(Received 3 April 2012; final version received 8 November 2012)

Land cover monitoring using digital Earth data requires robust classification methods that allow the accurate mapping of complex land cover categories. This paper discusses the crucial issues related to the application of different up-to-date machine learning classifiers: classification trees (CT), artificial neural networks (ANN), support vector machines (SVM) and random forest (RF). The analysis of the statistical significance of the differences between the performance of these algorithms, as well as sensitivity to data set size reduction and noise were also analysed. 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 cover categories in south Spain. Overall, statistically similar accuracies of over 91% were obtained for ANN, SVM and RF. However, the findings of this study show differences in the accuracy of the classifiers, being RF the most accurate classifier with a very simple parameterization. SVM, followed by RF, was the most robust classifier to noise and data reduction. Significant differences in their performances were only reached for thresholds of noise and data reduction greater than 20% (noise, SVM) and 25% (noise, RF), and 80% (reduction, SVM) and 50% (reduction, RF), respectively.

Keywords: land cover; remote sensing; classification trees; random forest; neural networks; support vector machines

1. Introduction

The application of classification methods for the mapping and monitoring of land covers is one of the most relevant remote sensing applications. The accuracy of the generated mapping can be assessed quantitatively (Congalton and Green 2009). This accuracy is not only dependent on the classification scene or the data themselves, but it is also strongly bound to the applied classification method. The accuracy with which land cover mapping is carried out may not always satisfy the needs of a given application; hence, it is necessary to develop more sophisticated and accurate new classification methods. Selecting a suitable classification algorithm is essential to obtain an accurate thematic map. It depends, mainly, on the user's experience and knowledge level, but also on the capacity of the algorithm to classify land covers correctly, its operational capacity, interpretability and transparency. However, for practical purposes, choosing a classification algorithm will depend

*Corresponding author. Email: vrgaliano@ugr.es

on software availability and/or ease of use, not only on the goodness of the methodology. On the other hand, the choice of a classification method must consider the method's capacity to handle the possible errors in remote sensing and auxiliary variables (DeFries and Chan 2000; Rogan et al. 2008). The main source of error is the noise incorporated by data into the classification methods. This noise is a consequence of a poor location of the reference data used to train classification algorithms and validate the accuracy of the maps produced (Foody 2002) or of the erroneous interpretation of reference field data or orthophotographs (Rogan et al. 2008).

There is a wide range of algorithms that have been used for the classification of land covers and land use (Lu and Weng 2007): from unsupervised algorithms to supervised parametric such as maximum likelihood (Jensen 2005) or non-parametric algorithms, among which machine learning algorithms can be found. It should be noted that the term 'non-parametric' refers to techniques that do not rely on data belonging to any particular statistical distribution. Several studies demonstrate that machine learning algorithms are more accurate than the traditional classification techniques such as maximum likelihood, especially when the feature space is complex and data present different statistical distributions (Mas and Flores 2008; Na et al. 2010; Mountrakis, Im, and Ogole 2011). The increasing use of this type of methods in remote sensing over recent years is due to different factors (Mas and Flores 2008; Mountrakis, Im, and Ogole 2011; Shao and Lunetta 2012): the ability of these techniques to learn complex patterns, non-linear in many cases; the high generalisation capacity of these algorithms, which makes it possible to apply them on incomplete or noisy databases; the possibility of incorporating *a priori* information; and last, but not least, their independence with respect to the data statistical distribution. This latter characteristic makes it possible to incorporate data from different sensors, auxiliary variables such as those derived from digital terrain models, or even categorical variables (Rogan et al. 2008).

Machine learning is a relatively recent scientific field, which is under continuous development. In the last decades, a large number of classification methods for the generation of thematic maps has emerged (Wilkinson 2005; Tso and Mather 2009). Among those most widely used are classification trees (CT) (Breiman et al. 1984), artificial neural networks (ANN) (Mas and Flores 2008), support vector machines (SVM) (Mountrakis, Im, and Ogole 2011; Yang 2011) and ensembles of classification trees such as Random Forest (RF) (Chan and Paelinckx 2008; Wang, Waske, and Benediktsson 2009; Na et al. 2010; Rodriguez-Galiano et al., "Assessment of the Effectiveness," 2012).

The aim of this study is to assess the suitability of different machine learning algorithms (decision trees, ANN, SVM and RF) for the classification of land covers of the Province of Granada, a complex Mediterranean area with a high number of land covers and a low interclass separability (Rodriguez-Galiano et al., "Random Forest," 2012). These algorithms were specifically chosen because they are being increasingly used in land cover mapping, yet have not been compared with one another exhaustively and also because of the free software availability. The comparative analysis carried out was approached from different perspectives: the mapping accuracy of classifications and noise and training data reduction sensitivity.

2. Theoretical bases of machine learning methods

This section summarises the theoretical bases of classification methods based on machine learning used in this study.

2.1. Decision trees

CT, along with ANN, are the most widely used machine learning algorithms in the classification of remote sensing data (Pal and Mather 2003; Rogan et al. 2008). The increasing use of CT is linked to their simplicity and interpretability, their low computational cost and the possibility of being graphically represented. A CT represents a set of restrictions or conditions, which are hierarchically organised and which are successively applied from a root to a terminal node or leaf of the tree (Breiman et al. 1984). In order to induce the CT from a dataset, an evaluation measure of each of the variables is used to maximise the interclass heterogeneity. There are many approaches to select attributes that can be used for the induction of decision tree models. Some of the most frequent ones are *gain-ratio*, Gini index and Chi-square. Once the evaluation measure has been chosen, the variable from which splits will start is determined (root node). From the root node, the data splitting process in each internal node of a classification rule of the tree is repeated until all examples show the same label or a stop condition previously specified be reached. Once the tree's induction process is finished, pruning is applied with the aim of improving the tree's generalisation capacity by reducing its structural complexity. The number of leaf nodes, the number of internal nodes, or tree-depth can be taken as pruning criteria.

2.2. Random forest

Within the field of meta-classifiers, RF (Breiman 2001) is a classification method that has recently been included in remote sensing (Chan and Paelinckx 2008; Sennie et al. 2008; Guo et al. 2011; Rodriguez-Galiano et al., "Assessment of the Effectiveness," 2012). RF is a meta-classifier that uses decision trees as base classifiers, in which each classifier contributes a vote to assign the most frequent class to the input vector. RF increases decision trees diversity by making them grow from different data subsets created by bagging or bootstrap aggregating (Breiman 1996). CTs that make up RF apply the Gini index from a fixed number of random variables, instead of using the best splitting variables. This can reduce each individual tree's accuracy; however, on the other hand, it also reduces correlation among trees, and hence, the generalisation error of the overall model is lower (Breiman 2001). Another interesting characteristic is that trees generated by RF grow without pruning, which reduces the computational cost.

2.3. Neural networks

From the 1990s to the present, the use of different types of ANN has been frequent in land cover and land use classification (Chen et al. 2009; Elatawneh et al. 2012; Petropoulos, Arvanitis, and Sigrimis 2012). There are many different types of ANN. However, it is not the aim of this study to describe the different types of networks,

which can be consulted from the bibliography (Mas and Flores 2008). This section provides a simplified description of one of the most widely used ANN in remote sensing: the *feed-forward propagation neural network* (Venables and Ripley 2002). An analogy can be drawn between the human brain and ANN, neurons being their main processing units. In ANN, neurons are situated on layers and they are connected in such a way that information flows from the input units, through those units situated on the input layer or hidden layers, to those units on the output layer. The input units distribute the signal to the hidden units of the second layer. These units add the inputs taking different weights into account; they add a constant (the bias) and apply an activation function to the result.

In order for the network to be able to represent any useful function, weights must be adjusted. With that aim, examples consisting of input-output pairs are presented to the network: an input vector and the corresponding output desired for the network. The input vector would be made up of values from the spectral bands of the training areas, and the output vector, of those thematic categories corresponding to the different land covers. The network output is compared to the desired output (category assigned to each training area), and subsequently the weights of connections are modified to reduce this difference. This process is carried out in an iterative way, minimising the overall error, typically calculated as the sum of squared errors, for all the input-output pairs with respect to the network weights, by using non-linear optimization methods (Venables and Ripley 2002).

2.4. Support vector machines

Over the last decade, SVMs have emerged as an alternative to the methods previously described for the classification of remotely sensed data (Pal and Mather 2003; Mountrakis, Im, and Ogole 2011; Petropoulos, Kalaitzidis, and Prasad Vadrevu 2012). SVMs were introduced by Vapnik in the early 1990s (Cortes and Vapnik 1995). SVMs try to find the optimal separating hyperplane interclasses; i.e. the plane for which the interclass separability is the highest. The examples that are on the border of this hyperplane are called support vectors. These examples are the most difficult to classify as they present a lower separability. In the simplest scenario, two classes in a bidimensional space in which data are linearly separable, the optimal hyperplane would be defined by a straight line. However, remotely sensed data are particularly complex, do not only have two dimensions, and the ultimate aim of the classification is not usually to differentiate only two classes, either. SVMs must deal with multiple predictive variables and, sometimes, also with auxiliary variables (Digital Terrain Models) in the classification of satellite data. On the other hand, the interclass separability can be low, with separating non-linear curves (Cortes and Vapnik 1995). In order to solve the problem of the non-linear separation of classes, an approach based on a soft margin that enables the making of some errors was incorporated. In this case, in order to find that hyperplane that makes the minimum number of errors, a regularisation constant (cost) is introduced. This constant controls the trade-off between the complexity of SVMs and the number of non-separable examples. When this latter approach is not feasible, it is necessary to apply non-linear transformation functions called *kernels*, which transform the input space into a Hilbert space of greater dimensionality in which data are linearly separable

(Muller et al. 2001). Some of the most commonly used kernel functions are (Meyer 2001):

$$K_{\text{linear}}(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}' \quad (1)$$

$$K_{\text{polynomial}} = (\gamma \mathbf{x} \mathbf{x}' + \mathbf{r})^\rho \quad (2)$$

$$K_{\text{RBF}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2\right) \quad (3)$$

$$K_{\text{sigmoid}}(\mathbf{x}, \mathbf{x}') = \tanh(\gamma \mathbf{x} \mathbf{x}' + \mathbf{r}) \quad (4)$$

Where \mathbf{x} and \mathbf{x}' are vectors belonging to the input and features spaces, respectively; ρ is the degree of the polynomial; and γ and \mathbf{r} are parameters that makes it possible to maximise the accuracy of SVM. Additionally, SVM considers a regularisation constant (cost or penalty), which allows for the making of some classification errors controlling the trade-off between the complexity of the model and the number of non-separable training areas.

3. Study area and data

3.1. 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, surrounded by the Penibetica mountain range (Figure 1). This area occupies 12,635 km² and elevation ranges from sea level to the Mulhacen Peak (3,482 m) in Sierra Nevada National Park. The climate of Granada is Mediterranean with a continental influence, characterised by hot and dry summers and wet and cold winters. Average annual temperatures range from 18°C on the coast to 10°C in the mountains. Climate ranges from arid to semi-arid; rainfall estimates range from 300 to 500 mm. The study area is composed of a variety of land cover types, mainly including agriculture (46%), with tobacco and corn fields, olive trees, tropical crops and substantial greenhouse production. The remainder of the study area is characterised by the presence of upland conifer forest (18%), shrub grasslands (22%) and oak grove (8%).

3.2. Datasets description and pre-processing

In this study, spring and summer images have been employed (April and August, respectively) in the land cover classification since these images contain most of the phenological variations (Schriever and Congalton 1995). Two Landsat Thematic Mapper-5 scenes of the same area in southeast Spain were captured. The images were acquired on 18 August and 12 April 2004. Image location corresponds with path 200 row 34 of Landsat worldwide reference system, with coordinate centre 0030822 W 372400 N WGS-84. These two dates represent peaks in productivity of the phenological development of the major vegetation types in the area (del Barrio et al. 2010), which is critical for the accurate classification of land cover types. In summer images, annual crops (e.g. tobacco and corn) can be confused with conifer forests and poplar groves. Highly reflective surfaces, such as urban areas, can be confused with bare soils. The inclusion of spring images allows discrimination between annual crops and evergreen natural vegetation. On the other hand, soils,

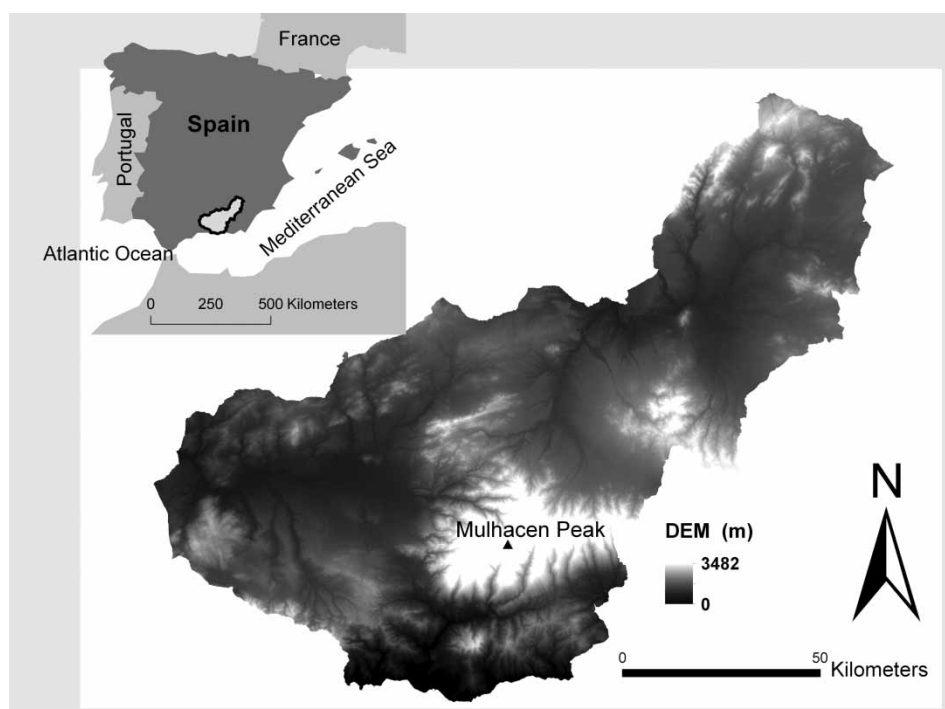


Figure 1. Location of the study area.

which remain bare during drought periods (summer), are usually covered by grass in spring time, which can facilitate differentiation from urban areas (Rodriguez-Galiano et al., “Incorporating the Downscaled Landsat TM,” 2012). The definition of summer and spring is related to the Northern Hemisphere.

Images were corrected independently for geometric offset by using digital orthophotos of 1 m spatial resolution. One hundred fifty points for each image provided a third-order polynomial transformation with less than one-half Landsat Thematic Mapper pixel root mean square error. Nearest neighbour resampling was chosen to preserve the original values of the pixels. The images were converted to radiance values and then into reflectance values using the Envi FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes) module, which includes a radiative transfer code based on MODTRAN4 (MODerate Resolution TRANsmittance) (Kneizys et al. 1996; Berk and Adler-Golden 2002). The reflectance values obtained for both dates were rescaled from 0 to 255 (8 bit reflectance). Images were corrected at surface reflectance, given it is advisable to apply an atmospheric correction prior to the application of the Kauth Thomas transformation (tasseled cap). Reflectance based images are more appropriate for scene to scene analysis (Yarbrough, Easson, and Kuzmaul 2012), which is an important issue in this study as multi-seasonal images are used simultaneously.

The images were enhanced spectrally using the Kauth Thomas linear transformation before being used in classification in order to facilitate both the development of the classification models reducing the volume of spectral information and the interpretation. This way, the multi-seasonal spectral information was condensed to

half of its features, hence reducing the correlation between them. This transformation produced six features: summer brightness, summer greenness, summer wetness, spring brightness and both spring greenness and wetness.

Several topographic variables were included as input variables to each classification: elevation, slope and aspect. These ancillary variables were derived from a 20-m resolution digital elevation model and rescaled to the spatial resolution of the spectral variables (30 m).

3.3. Land cover categories and reference data

The land cover of the Mediterranean can be very complex and challenging to classify (Rodriguez-Galiano et al., “Random Forest,” 2012). Relief complexity and high anthropogenic influence results in a very heterogeneous landscape, which makes it possible that fourteen different thematic categories can be distinguished in the study area (Table 1). The classification scheme was based on Andalusian land cover maps (ALCM) developed in 2003 by the Andalusian Regional Government.

Jensen (2005) proposed that the number of training pixels should at least be equal to ten times the number of variables used in the classification model for a parametric classification approach. However, several studies have shown that non-parametric machine learning algorithms need a larger number of training data in order to attain optimal results (Foody and Arora 1997; Pal and Mather 2003; Pal 2005).

To create an exhaustive database with an optimal size for the training and accuracy assessment it was necessary to resort to auxiliary information due to the retrospective nature of this study. Reference data were obtained from a combination of a set of crop reference sites collected in the summer of 2004 and a stratified random sampling scheme using pre-existing land cover maps (ALCMs). More specifically, the ALCM was reclassified into 14 categories, and 150 sites were sampled randomly from each category. The digital true-colour orthophotos (1:10000), corresponding to the sample sites and acquired during 2004, were then

Table 1. Land cover classification scheme.

Class ID	Class name
1	Urban
2	Poplar grove
3	Conifers
4	Greenhouses
5	Shrublands
6	Olive grove
7	Grasslands
8	Oak grove
9	Herbaceous irrigated
10	Ligneous irrigated
11	Herbaceous dry
12	Bare soils
13	Tropical crops
14	Water

interpreted and 2100 sites were obtained. The ground reference dataset was divided randomly into 2/3 and 1/3 for training and testing, respectively. The number of the training sites per class was kept equal (100 training sites and 50 testing sites per land cover category).

4. Methods

4.1. Classification and algorithm parameterization

To study the performance of the algorithms for land cover classification, nine spectral and auxiliary variables were used. On the one hand, spectral variables consisted of the Kauth Thomas multi-seasonal components of the summer and spring images. On the other hand, the auxiliary variables included in the analysis were elevation, slope and aspect, derived from the Digital Terrain Model (Section 0). All these variables were used in the training of every model. The contribution of the auxiliary and spectral variables in the classification of the study area is discussed in depth in the studies of Rodriguez-Galiano et al., “Incorporating the Downscaled Landsat TM” (2012) and Rodriguez-Galiano and Chica-Olmo (2012).

There are several commercial and open source implementations for machine learning classification. In this study, several packages within the open source software R 2.10.1 (R-Project) were used: ‘rpart’, ‘nnet’, ‘e1071,’ and ‘randomForest’. In order to study the performance of the different machine learning algorithms, it is very important to determine a suitable combination of parameters, which allows generating operative robust classification models with a high generalisation capacity. To assess the optimal value of the different parameters of every method, the classifications derived from all possible parameter combinations were evaluated using the overall accuracy and the kappa index (Congalton and Green 2009). The ‘best’ model was the one with the largest overall accuracy and kappa index.

4.1.1. Decision trees

It is necessary to set a series of parameters for the training of decision trees, such as dissimilarity measure, the depth of the tree, and the minimum number of observations per node. The dissimilarity measure or heterogeneity influences the way in which the algorithm performs data splits in each node. The depth of the tree and the minimum number of observations are parameters linked to the structural complexity of trees: the more the number of levels and the less the number of minimum observations in nodes, the greater the structural complexity of the model. Hence, it is necessary to set these parameters in order to achieve the highest accuracy in the classification of training data, avoiding the creation of complex tree structures, which overfit data and lose generality (Pal and Mather 2003). For this study, CART decision tree models were used (Breiman et al. 1984). For the induction of trees, two different types of dissimilarity measures were considered: Gini index and information gain (Breiman et al. 1984). With the aim of obtaining robust and generalizable models, all possible decision trees were assessed, for depths of tree from 2 to 29, with a minimum number of observations per node between 1 and 50, which resulted in 2700 different classifications.

4.1.2. Random forest

Unlike most methods based on machine learning, RF only needs two parameters to be set for generating a prediction model: the number of classification trees and the number of predictive variables (m), which are used in each node to make decision trees grow (Rodriguez-Galiano et al., "Assessment of the Effectiveness," 2012). Breiman (1996) demonstrated that by increasing the number of trees, the generalisation error always converges; hence, overtraining is not a problem. On the other hand, reducing the number of m brings as a result a reduction in the correlation among trees, which increases the model's accuracy. In order to optimise these parameters, a large number of experiments were carried out using different numbers of trees and split variables. The range of the number of trees was set between 1 and 1000, and the number of splits variables between one and nine, at one interval. This resulted in 9000 different combinations for the classification of the study area.

4.1.3. Neural networks

There are different factors that affect the capacity of ANN to generalise, i.e. to classify new data from the learning carried out with training data. Among these the intrinsic factors to network design can be found: number of nodes and network architecture. Given the number of input and output units determined by predictive variables and the number of categories to be classified, the problem of how to define the most suitable network architecture is related to the nature of the hidden layer. There is no rule for determining the number of hidden layers, but, theoretically, one single hidden layer can represent any Boolean function (Atkinson and Tatnall 1997). In general terms, the higher the number of units of the hidden layer, the greater the network capacity to represent the training data patterns. However, the fact that the hidden layer has a high number of units also produces a loss in the networks' generalisation power (Atkinson and Tatnall 1997; Foody and Arora 1997).

With the aim of making a trade-off between network accuracy and generalisation power, different feed-forward propagation ANN were made using a standard sigmoid transfer function. To this end, ANN of different architectures were trained, made up of a single hidden layer, whose number of units was set between 1 and 20. Likewise, in order to optimise the network training, the range of initial weights assigned by the network was set between the intervals 1 and 1, with increases of 0.02. From these initial values, different weight decay values were considered (between 0.01 and 0.1 at 0.005 intervals). This combination of parameters gave as a result a total of 15,580 different ANN classifications. The optimal value of weights was set by means of least squares to help the optimization process and to avoid over-fitting. Several criteria were adopted to stop the training of the network: a maximum of 10,000 iterations, a relative difference for the entropy fit higher than 0.01, and the reduction of the fit criterion by a factor of at least 1 (Venables and Ripley 2002).

4.1.4. Support vector machines

SVMs need the adjustment of a high number of parameters for their optimization: (1) linear, polynomial, sigmoid and radial basis kernel (RBF) functions; (2) cost; (3) gamma of the kernel function, with the exception of the linear kernel; (4) bias on

the kernel function, only applicable to the polynomial and sigmoid kernels; and, finally, (5) degree of the polynomial, only applicable to the polynomial kernel. The adequate value of these parameters is data specific; therefore it is necessary to optimise them in order to get generalizable models; i.e. these must not overfit or underfit data; therefore they must be accurate (Yang 2011).

In order to assess the impact on the mapping accuracy of each of the abovementioned parameters, a set of 621,000 SVMs were built for different parameter combinations considering all kernel types. For the building of SVM, the cost was fixed between 0.1 and 100, at 0.1 intervals and gamma between 0.05 and 1, at 0.05 intervals. In the case of the polynomial kernel, SVMs were calculated considering the said different parameter combinations for each of the 10 possible degrees of the polynomial (number of variables + 1). The bias, which also held values between 1 and 10, was applied to every single classifier of the sigmoid kernel. However, in the case of the polynomial kernel, the bias was only applied to the degree of the polynomial for which the best results derived from the test were obtained.

5. Results and discussion

5.1. Assessment of the mapping accuracy of the best classifications

The assessment of accuracy was carried out on the basis of confusion matrices, from which the overall accuracy, user's and producer's accuracies and overall and per categories kappa were calculated (Congalton and Green 2009).

The overall accuracy and the kappa index of all classification methods, with the exception of CT were over 0.9. The classification generated by RF was the most accurate, with overall accuracy and kappa coefficients equal to 0.92, followed by the classifications made with SVM (RBF) and ANN, with overall accuracy values equal to 0.92 and 0.91 and kappa indexes of 0.91 and 0.90, respectively. CT produced a mapping accuracy significantly lower than that of the rest of methods, with overall accuracy and kappa coefficients equal to 0.86 and 0.85, respectively (Table 2). These findings are supported by different studies of similar characteristics. Pal (2005) and Waske et al. (2009) compared the performance of RF and SVM and found that they performed equally well. Shao and Lunetta (2012) and Huang, Davis, and Townshend (2002) compared thematic mapping accuracies produced using four different classification algorithms: CT, ANN, SVM and the maximum likelihood classifier (MLC). Their results suggested that the accuracy of SVM-based classifications generally outperformed the other three classification algorithms. On the other hand, Otukey and Blaschke (2010) compared the CT, SVM and MLC algorithms and found CT performed better than SVM. However, in this study, the subset of the satellite bands selected by CT were used for the training of all classifiers, which could overestimate the performance of CT. Duro, Franklin, and Dubé (2012) concluded that no statistical significant differences between CT, RF and SVM classifications were found. However, it should be noted that the default parameters were used in the training of RF, and a very limited combination of parameters were tried for SVM optimization.

Table 2 also shows the values of the producer's and user's accuracies and kappa index of land covers of the study area. The average user's and producer's accuracies and the kappa index per categories of the classifications made by CT, RF, ANN and SVM

Table 2. Summary of the overall and per categories mapping accuracy obtained by the different classification methods.

Class Id	CT			RF			ANN			SVM		
	OA			OA			OA			OA		
	(0.86)		K (0.85)	(0.92)		K (0.92)	(0.91)		K (0.90)	(0.92)		K (0.91)
	PA	UA	K	PA	UA	K	PA	UA	K	PA	UA	K
1	0.74	0.88	0.72	0.88	0.90	0.87	0.86	0.84	0.85	0.94	0.84	0.93
2	0.98	0.98	0.98	1.00	1.00	1.00	0.98	1.00	0.98	1.00	1.00	1.00
3	0.72	0.80	0.70	0.92	0.88	0.91	0.88	0.92	0.87	0.90	0.90	0.89
4	0.96	1.00	0.96	1.00	1.00	1.00	0.96	1.00	0.96	1.00	1.00	1.00
5	0.86	0.66	0.85	0.80	0.83	0.79	0.88	0.86	0.87	0.88	0.81	0.87
6	0.84	0.93	0.83	0.98	0.91	0.98	0.94	0.85	0.93	0.92	0.90	0.91
7	0.80	0.95	0.79	0.84	1.00	0.83	0.86	0.93	0.85	0.86	1.00	0.85
8	0.72	0.67	0.70	0.86	0.84	0.85	0.92	0.84	0.91	0.88	0.86	0.87
9	0.92	0.96	0.91	1.00	0.98	1.00	1	0.91	1.00	0.92	0.96	0.91
10	0.86	0.90	0.85	0.86	0.98	0.85	0.9	0.96	0.89	0.90	0.98	0.89
11	0.94	0.84	0.93	1.00	0.89	1.00	0.98	0.91	0.98	1.00	0.91	1.00
12	0.74	0.66	0.72	0.82	0.84	0.81	0.66	0.77	0.64	0.68	0.83	0.66
13	0.94	0.90	0.94	0.98	0.91	0.98	0.9	0.94	0.89	0.96	0.87	0.96
14	0.98	1.00	0.98	1.00	1.00	1.00	1	1.00	1.00	1.00	1.00	1.00

OA, overall accuracy; K, kappa; PA, producer's accuracy; UA, user's accuracy.

were equal to 0.86, 0.92, 0.91 and 0.92; 0.87, 0.93, 0.91 and 0.92; and 0.85, 0.92, 0.90 and 0.91, respectively. The standard deviations of these parameters were equal to 0.10, 0.08, 0.09 and 0.08; 0.12, 0.07, 0.07 and 0.07; and 0.11, 0.08, 0.09 and 0.09, respectively.

As a consequence of the spectral features of land covers, a shared pattern could be observed in the mapping accuracy of the different categories, regardless of the performance of each algorithm. The best classified vegetation covers were poplars, greenhouses and water. Specifically, RF and SVM classified all the areas corresponding to these vegetation covers correctly. Herbaceous irrigated classified by RF also presented producer's accuracy and kappa values equal to one. Vegetation covers that were most difficult to classify were those with a higher intraclass variability and those with a similar spectral behaviour, such as shrublands, grasslands, bare soils, quercus and urban. Thus, the urban class was classified sometimes as bare soils and vice versa, as both land covers present high reflectivity values. On the other hand, the high reflectivity of soils in the study area may hide the spectral response of patches of scarce vegetation covers such as olive grove and make their classification difficult. Shrubland, quercus and conifer were also classified less accurately, given their high spectral similarity. Those vegetation covers with a similar spectral and seasonal behaviour, i.e. herbaceous dry and grasslands (herbaceous vegetation with a high vigour in springtime), were also mixed up. With relation to the differences in the classification accuracy of land covers in terms of the classification method used, generally, RF and SVM were the methods that classified land covers in our study area best. Conifers, olive grove, bare soils and tropical crops were classified more accurately by RF. SVM improved the classification of the rest of methods for the urban cover. However, ANN classified quercus more accurately.

5.2. Significance of differences in accuracy

The assessment of accuracy is usually carried out on the basis of the calculation and comparison of the kappa coefficients and the proportion of cases classified correctly (overall accuracy) derived from each map. This approximation assumes that the samples used for the calculation of the kappa index (test) are independent, which is not fulfilled in many cases, as the same test set is used in the assessment of each map's accuracy. In those cases in which tests are not independent, the statistical significance of the differences in accuracy of two classified maps can be assessed by means of the McNemar test (Foady 2009). This test considers that the differences in accuracy are statistically significant for a 5% confidence level, i.e. for Z values over 1.96. Together with the comparison of the mapping accuracy carried out in previous section, the significance of the differences in accuracy of the results obtained by the different classification methods was assessed using the McNemar test. Table 3 shows the values of Z calculated from the results of applying the different classification methods, once the optimal parameters related to each classifier were selected. As it can be seen from this table, all classifiers were significantly more accurate than CT and sigmoid SVM. The results of the mapping accuracy derived from RF, ANN, polynomial SVM and radial SVM were equivalent. As far as SVM is concerned, the differences in the mapping accuracy between the linear, polynomial and radial kernels were not statistically significant. However, the differences in accuracy between RF and SVM with a linear kernel type were significant indeed.

5.3. Noise sensitivity

The information needed for the training of classification algorithms is obtained through field works or through the interpretation of aerial photographs. In either case, errors may happen in the labelling of land cover types, above all in those circumstances in which categories are very heterogeneous and the landscape is complex (Rogan et al. 2008). The effect of mislabelling some examples of training information contributes to increase intraclass variability and, therefore, has a direct effect on the accuracy of those maps classified from this information. Machine learning algorithms can learn these 'noisy patterns' to a greater or lesser extent, which results in a decrease in their ability to classify correctly. There are many studies that account for the negative effect noise has in the classification of satellite data by machine learning algorithms (DeFries and Chan 2000; Simard, Saatchi, and De Grandi 2000), although none of them assess the statistical significance of results.

Table 3. Results of the evaluation of the statistical significance (Z) of the differences in kappa coefficients of the thematic maps classified by the different machine learning algorithms.

	CT	RF	ANN	SVM Lin.	SVM Pol.	SVM Rad.
RF	4.96	–	–	–	–	–
ANN	3.63	1.48	–	–	–	–
SVM Lin.	3.62	1.73	0.14	–	–	–
SVM Pol.	3.75	1.6	0	0.26	–	–
SVM Rad.	4.4	0.93	0.7	1.3	1.13	–
SVM Sig.	0.48	4.61	3.64	4.23	4.63	4.65

The robustness of classifiers against noise was assessed through the inclusion of wrongly labelled areas (noise). Figure 2 shows the kappa statistic resulting from the classifications carried out from training subsets with different noise proportions. Every classifier was negatively affected with the addition of wrongly labelled examples. Table 4 shows the assessment of the significance of the differences between the kappa statistics of the classifications carried out from original data and those from data that include noise. RF and SVMs were the most robust classifiers against noise. For noise percentages lower than 20% or 25%, no significant decrease in the mapping accuracy took place. However, in the case of CT and ANN the addition of just 5% of the wrongly labelled cases resulted in a significant decrease in the mapping accuracy, which indicated that both algorithms tended to overfit data. On the other hand, it is worth mentioning that for noise thresholds lower than 40%, RF and SVMs were more accurate than the CT classifier applied to the original data. The greater stability of RF and SVM was due to different reasons. On the one hand, RF only used two-thirds of the training examples in the classification made by each decision tree. Therefore, as the pattern diversity, from which the trees of the classifier learn, increases, the probability of learning noise is lower (Breiman 2001). On the other hand, SVMs use soft margins in the search for a hyperplane that can contain a certain number of errors; therefore they include a constant (cost) that controls the trade-off between the complexity of the model and the number of wrongly classified examples (Cortes and Vapnik 1995). This way, by optimising this constant, noise learning by SVMs can be reduced.

5.4. Sensitivity to the reduction of training areas

The acquisition of reference data that can be used in the training and validation of the classification process is a time-consuming and expensive task. On the one hand, it is necessary for training areas to be as numerous and diverse as possible so that they can represent the whole variety present on each land cover or category (Foody and Arora 1997; Pal and Mather 2003). On the other hand, it is essential to design a sampling that allows reaching an acceptable level of mapping accuracy, which, in

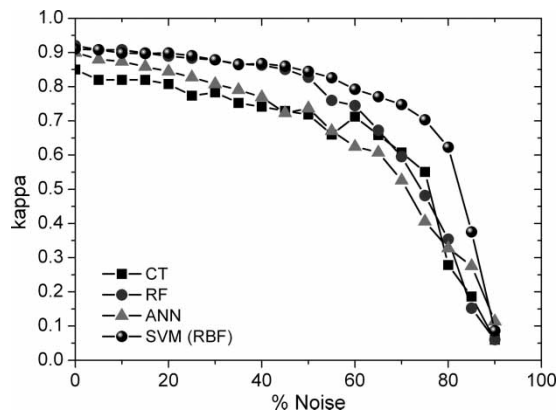


Figure 2. Effect of adding noise in training data on the mapping accuracy. SVM is less noise sensitive than the rest of classifiers, especially for noise proportions over 50%.

Table 4. Z-score values obtained for data classified from training data with different noise proportions with respect to the original results.

% Noise	5	10	15	20	25	30	35	40	45
CT	2.36	2.29	2.32	3.20	4.90	4.24	5.61	6.49	6.73
RF	0.33	0.28	1.15	1.98	2.69	2.78	3.54	3.68	4.30
ANN	1.84	2.01	3.50	4.39	5.18	5.84	7.13	7.54	9.17
SVM (RBF)	0.00	1.22	1.35	1.10	2.20	3.04	4.13	3.83	4.18
% Noise	50	55	60	65	70	75	80	85	90
CT	6.86	8.99	7.49	9.35	10.52	11.99	18.45	20.09	22.07
RF	5.67	8.64	9.10	11.46	13.38	16.04	18.39	21.83	23.37
ANN	8.59	10.94	12.42	13.03	14.63	17.36	18.63	19.65	22.41
SVM (RBF)	5.17	6.20	7.46	14.63	9.34	10.61	12.92	18.29	22.90

turn, can be operative both in economic terms and time. The number of training data is not a problem for the classification of homogeneous categories. However, the data volume necessary to classify land covers with a dense variability is high. The effect of the training dataset size in the accuracy of the classifications carried out by the different machine learning algorithms was assessed by means of the kappa index. Figure 3 and Table 5 show, in a similar way as the previous section, the results of the kappa index and of the McNemar test obtained from the classifications of the different reduced-size data subsets. As it can be seen from Figure 3, the accuracy of the classifications decreased with training data size, although this effect was slighter than that produced by the addition of noise. In absolute terms, SVM, RF and ANN presented a similar pattern. The kappa index remained stable until it reached a reduction threshold of about 50% of the data, although for ANN, from the said threshold, the decreasing rhythm was higher. However, CT is much more sensitive to data size, and accuracy decreases for reduction values of 15%. Table 5 shows the significance of the relative differences in the kappa index. These differences among classifiers were significant from different thresholds, depending on the classifier used.

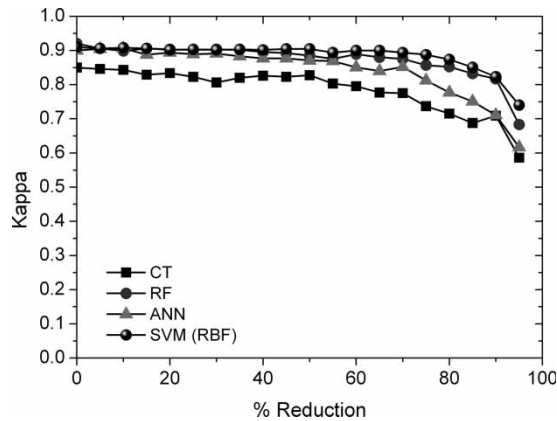


Figure 3. Effect of reducing training data on the mapping accuracy. RF and SVM show a similar behaviour with relation to the reduction of training data. However, the ANN and CT underwent a more noticeable decrease of mapping accuracy, especially for high reduction values. This may mean a higher need for training data of these algorithms.

Table 5. Z-score values obtained for data classified from reduced-size training datasets with respect to the original results.

% Reduction	5	10	15	20	25	30	35	40	45
CT	0.22	0.56	1.95	1.17	2.10	2.71	1.86	1.48	1.69
RF	0.00	1.39	0.00	0.45	0.77	0.58	0.69	1.46	1.62
ANN	0.28	0.46	1.24	0.62	1.18	1.02	1.73	2.03	2.03
SVM (RBF)	0.26	0.00	0.22	0.76	0.77	0.89	0.60	0.78	0.38
% Reduction	50	55	60	65	70	75	80	85	90
CT	1.35	3.11	3.21	3.94	4.31	5.96	6.98	7.54	7.03
RF	2.40	3.09	1.86	2.87	3.16	4.44	4.64	5.66	6.48
ANN	2.32	2.42	3.97	4.13	3.49	5.80	7.69	8.40	9.51
SVM (RBF)	0.35	1.62	0.85	0.93	1.57	1.90	2.99	4.45	5.97

These thresholds for the SVM, RF, ANN and CT classifiers were equal to 80, 50, 40 and 25%, respectively. It is worth mentioning that in the case of data reduction, greater differences in the behaviour of SVM and RF classifiers were appreciated than in the case of noise, which is logical, as RF has an implicit data reduction of 33% (*bagging*). Nonetheless, the proportion of data for the training of RF can be adapted to conditions in which data are scarce.

6. Conclusions

The comparative analysis of algorithms carried out was approached from different perspectives: the mapping accuracy of classifications and, finally, sensitivity to noise and training data reduction.

The assessed classification algorithms have a different difficulty rating in their training. Decision tree-based algorithms involve a lesser difficulty in their training. This applies to both simple decision trees and meta-classifiers based on trees (RF). CT only needs the setting of the measure of heterogeneity, the minimum number of observations in each node and the depth of the trees. RF, by definition, uses a fixed measure of heterogeneity (the Gini index), and the trees which make it up grow up to their maximum depth without pruning. Therefore, it only needs the setting of two parameters: the number of trees that make up the ensemble and the number of random variables considered in the split of each node. However, ANN and SVMs are more complex. ANN need an optimal network architecture design (number of hidden layers and units) and the setting of the optimal cost and weight decay values. Finally, SVMs are based on different kernel types, according to which the combination of parameters to be optimised is different: from the degree of the polynomial, and the cost and gamma parameters of the polynomial kernel, to an only cost parameter for the linear kernel.

The greatest accuracy of classifications was achieved by RF and radial SVMs, with kappa values equal to 0.92 in both cases. ANN also achieved a high level of mapping accuracy (kappa equal to 0.91), although only for a very concrete combination of their adjustment parameters. Lastly, the maximum kappa index derived from the classifications carried out by decision trees was considerably lower than that of the rest of classifications (0.83). Hence, no significant differences in the mapping accuracy obtained by RF, SVM and ANN were observed. It is worth

mentioning that this conclusion can only be applied to the best classification methods obtained from a complex optimization process, since, in general terms, the performance of RF and SVMs was better than that of ANN. Regarding the results of classifications per categories, the choice of classifier resulted in differences in the accuracy of classifications according to the type of vegetation cover. Conifers, olive grove, soils and tropical crops were classified more accurately by RF. SVM improved the classification carried out by the rest of methods for the urban class. Finally, ANN classified quercus more accurately.

The assessed classifiers responded in different ways to the addition of noise and reduction of the number of training areas. The SVMs and RF classifiers were the most robust against noise and presented significant differences in the kappa index for noise values over 20%–25%. On the other hand, ANN and CT underwent significant decreases for values equal to 5%–10%. As regards training data reduction, SVMs were the most robust classifiers, as these only obtained significant differences for reduction thresholds over 80%, followed by RF (50%), ANN (40%) and CT (25%).

Acknowledgements

We are grateful for the financial support given by the Spanish MICINN (Project CGL2010-17629) and Junta de Andalucía (Group RNM122). We thank the reviewers for their constructive criticism.

References

- Atkinson, P., and A. Tatnall. 1997. "Introduction Neural Networks in Remote Sensing." *International Journal of Remote Sensing* 18 (4): 699–709.
- Berk, A., and S. M. Adler-Golden. 2002. "Exploiting Modtran Radiation Transport for Atmospheric Correction: The Flaash Algorithm." Fifth International Conference on Information Fusion, Annapolis, July 8–11.
- Breiman, L. 1996. "Bagging Predictors." *Machine Learning* 24 (2): 123–140.
- Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.
- Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen. 1984. *Classification and Regression Trees*. 1st ed. Belmont, CA: Chapman and Hall/CRC.
- Congalton, R. G., and K. Green. 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. 2nd ed. Boca Raton, FL: CRC Press.
- Cortes, C., and V. Vapnik. 1995. "Support-Vector Networks." *Machine Learning* 20 (3): 273–297.
- Chan, J. C.-W., and D. Paelinckx. 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 (6): 2999–3011.
- Chen, H. W., N. B. Chang, R. F. Yu, and Y. W. Huang. 2009. "Urban Land Use and Land Cover Classification Using the Neural-Fuzzy Inference Approach with Formosat-2 Data." *Journal of Applied Remote Sensing* 3 (1): art. no. 033558.
- Defries, R. S., and J. C.-W. Chan. 2000. "Multiple Criteria for Evaluating Machine Learning Algorithms for Land Cover Classification from Satellite Data." *Remote Sensing of Environment* 74 (3): 503–515.
- del Barrio, G., J. Puigdefabregas, M. E. Sanjuan, M. Stellmes, and A. Ruiz. 2010. "Assessment and Monitoring of Land Condition in the Iberian Peninsula, 1989–2000." *Remote Sensing of Environment* 114 (8): 1817–1832.
- Duro, D. C., S. E. Franklin, and M. G. Dubé. 2012. "A Comparison of Pixel-Based and Object-Based Image Analysis with Selected Machine Learning Algorithms for the Classification of Agricultural Landscapes Using Spot-5 Hrg Imagery." *Remote Sensing of Environment* 118 (0): 259–272.

- Elatawneh, A., C. Kalaitzidis, G. P. Petropoulos, and T. Schneider. 2012. "Evaluation of Diverse Classification Approaches for Land Use/Cover Mapping in a Mediterranean Region Utilizing Hyperion Data." *International Journal of Digital Earth*. doi:10.1080/17538947.2012.671378
- Foody, G. M. 2002. "Status of Land Cover Classification Accuracy Assessment." *Remote Sensing of Environment* 80 (1): 185–201.
- Foody, G. M. 2009. "Sample Size Determination for Image Classification Accuracy Assessment and Comparison." *International Journal of Remote Sensing* 30 (20): 5273–5291.
- Foody, G. M., and M. K. Arora. 1997. "An Evaluation of Some Factors Affecting The Accuracy of Classification by an Artificial Neural Network." *International Journal of Remote Sensing* 18 (4): 799–810.
- Guo, L., N. Chehata, C. Mallet, and S. Boukir. 2011. "Relevance of Airborne Lidar and Multispectral Image Data for Urban Scene Classification Using Random Forests." *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (1): 56–66.
- Huang, C., L. S. Davis, and J. R. G. Townshend. 2002. "An Assessment of Support Vector Machines for Land Cover Classification." *International Journal of Remote Sensing* 23 (4): 725–749.
- Jensen, J. R. 2005. *Introductory Digital Image Processing*. 3rd ed. Upper Saddle River, NJ: Prentice Hall.
- Kneizys, F. X., L. W. Abreu, G. P. Anderson, J. H. Chetwynd, E. P. Shettle, A. Berk, L. S. Bernstein, et al. 1996. *The Modtran 2/3 Report and Lowtran 7 Model*. Hanscom, MA: Phillips Laboratory.
- Lu, D., and Q. Weng. 2007. "A Survey of Image Classification Methods and Techniques for Improving Classification Performance." *International Journal of Remote Sensing* 28 (5): 823–870.
- Mas, J. F., and J. J. Flores. 2008. "The Application of Artificial Neural Networks to the Analysis of Remotely Sensed Data." *International Journal of Remote Sensing* 29 (3): 617–663.
- Meyer, D. 2001. "Support Vector Machines." *R News* 1 (3): 23–26.
- Mountrakis, G., J. Im, and C. Ogole. 2011. "Support Vector Machines in Remote Sensing: A Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (3): 247–259.
- Muller, K. R., S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf. 2001. "An Introduction to Kernel-Based Learning Algorithms." *Neural Networks, IEEE Transactions on* 12 (2): 181–201.
- Na, X. D., S. Q. Zhang, X. F. Li, H. A. Yu, and C. Y. Liu. 2010. "Improved Land Cover Mapping Using Random Forests Combined with Landsat Thematic Mapper Imagery and Ancillary Geographic Data." *Photogrammetric Engineering & Remote Sensing* 76 (7): 833–840.
- Otukei, J. R., and T. Blaschke. 2010. "Land Cover Change Assessment Using Decision Trees, Support Vector Machines and Maximum Likelihood Classification Algorithms." *International Journal of Applied Earth Observation and Geoinformation* 12 (Suppl. 1): S27–S31.
- Pal, M. 2005. "Random Forest Classifier for Remote Sensing Classification." *International Journal of Remote Sensing* 26 (1): 217–222.
- Pal, M., and P. M. Mather. 2003. "An Assessment of the Effectiveness of Decision Tree Methods for Land Cover Classification." *Remote Sensing of Environment* 86 (4): 554–565.
- Petropoulos, G. P., K. Arvanitis, and N. Sigrimis. 2012. "Hyperion Hyperspectral Imagery Analysis Combined with Machine Learning Classifiers for Land Use/Cover Mapping." *Expert Systems with Applications* 39 (3): 3800–3809.
- Petropoulos, G. P., C. Kalaitzidis, and K. Prasad Vadrevu. 2012. "Support Vector Machines and Object-Based Classification for Obtaining Land-Use/Cover Cartography from Hyperion Hyperspectral Imagery." *Computers and Geosciences* 41: 99–107.
- R-Project. 2012. "The R Project for Statistical Computing." R-Project. Accessed December 2012. <http://www.r-project.org>
- Rodriguez-Galiano, V., and M. Chica-Olmo. 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 (1–2): 208–218.
- Rodriguez-Galiano, V. F., M. Chica-Olmo, F. Abarca-Hernandez, P. M. Atkinson, and C. Jeganathan. 2012. "Random Forest Classification of Mediterranean Land Cover Using

- Multi-Seasonal Imagery and Multi-Seasonal Texture.” *Remote Sensing of Environment* 121: 93–107.
- Rodriguez-Galiano, V. F., B. Ghimire, E. Pardo-Iguzquiza, M. Chica-Olmo, and R. Congalton. 2012. “Incorporating the Downscaled Landsat TM Thermal Band in Land-Cover Classification Using Random Forest.” *Photogrammetric Engineering & Remote Sensing* 78 (2): 129–137.
- Rodriguez-Galiano, V. F., B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sánchez. 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., J. Franklin, D. Stow, J. Miller, C. Woodcock, and D. Roberts. 2008. “Mapping Land-Cover Modifications Over Large Areas: A Comparison of Machine Learning Algorithms.” *Remote Sensing of Environment* 112 (5): 2272–2283.
- Schriever, J. R., and R. G. Congalton. 1995. “Evaluating Seasonal Variability as an Aid to Cover-Type Mapping from Landsat Thematic Mapper Data in the Northeast.” *Photogrammetric Engineering and Remote Sensing* 61 (3): 321–327.
- Sesnie, S., P. Gessler, B. Finegan, and S. Thessler. 2008. “Integrating Landsat TM and SRTM-DEM Derived Variables with Decision Trees for Habitat Classification and Change Detection in Complex Neotropical Environments.” *Remote Sensing of Environment* 112 (5): 2145–2159.
- Shao, Y., and R. S. Lunetta. 2012. “Comparison of Support Vector Machine, Neural Network, and Cart Algorithms for the Land-Cover Classification Using Limited Training Data Points.” *ISPRS Journal of Photogrammetry and Remote Sensing* 70 (0): 78–87.
- Simard, M., S. S. Saatchi, and G. De Grandi. 2000. “The Use of Decision Tree and Multiscale Texture for Classification of JERS-1 SAR Data Over Tropical Forest.” *Geoscience and Remote Sensing.* *IEEE Transactions on* 38 (5): 2310–2321.
- Tso, B., and P. M. Mather. 2009. *Classification Methods for Remotely Sensed Data*. 2nd ed. New York: CRC Press.
- Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with s*. 4th ed. New York: Springer.
- Wang, X. L., B. Waske, and J. A. Benediktsson. 2009. “Ensemble Methods for Spectral-Spatial Classification of Urban Hyperspectral Data.” *2009 IEEE International Geoscience and Remote Sensing Symposium* 1–5: 3324–3327.
- Waske, B., J. A. Benediktsson, K. Arnason, and J. R. Sveinsson. 2009. “Mapping of Hyperspectral Aviris Data Using Machine-Learning Algorithms.” *Canadian Journal of Remote Sensing* 35 (Suppl. 1): S106–S116.
- Wilkinson, G. G. 2005. “Results and Implications of a Study of Fifteen Years of Satellite Image Classification Experiments.” *Geoscience and Remote Sensing, IEEE Transactions on* 43 (3): 433–440.
- Yang, X. 2011. “Parameterizing Support Vector Machines for Land Cover Classification.” *Photogrammetric Engineering and Remote Sensing* 77 (1): 27–37.
- Yarbrough, L. D., G. Easson, and J. S. Kuszmaul. 2012. “Proposed Workflow for Improved Kauth–Thomas Transform Derivations.” *Remote Sensing of Environment* 124 (0): 810–818.