Examining region-based methods for land cover classification using stochastic distances

R. G. Negri, L. V. Dutra, S. J. S. Sant'Anna & D. Lu

Taylor & Francis
Taylor & Francis Group

Examining region-based methods for land cover classification using stochastic distances

R. G. Negri [a], L. V. Dutra [b], S. J. S. Sant'Anna[b] and D. Lu [c]

aInstituto de Ciência e Tecnologia, UNESP – Univ. Estadual Paulista, São Paulo, Brazil; bDivisão de Processamento de Imagens, INPE – Inst. Nacional de Pesquisas Espaciais, São Paulo, Brazil; cCenter for Global Change and Earth Observations, MSU – Michigan State University, East Lansing, MI, USA

ABSTRACT

A recent alternative to standard pixel-based classification of remote-sensing data is region-based classification, which has proved to be particularly useful when analysing high-resolution imagery of complex environments, such as urban areas, or when addressing noisy data, such as synthetic aperture radar (SAR) images. First, following certain criteria, the imagery is decomposed into homogeneous regions, and then each region is classified into a class of interest. The usual method for region-based classification involves using stochastic distances, which measure the distances between the pixel distributions inside an unknown region and the representative distributions of each class. The class, which is at the minimum distance from the unknown region distribution, is assigned to the region and this procedure is termed stochastic minimum distance classification (SMDC). This study reports the use of methods derived from the original SMDC, Support Vector Machine (SVM), and graph theory, with the objective of identifying the most robust and accurate classification methods. The equivalent pixel-based versions of region-based analysed methods were included for comparison. A case study near the Tapajós National Forest, in Pará state, Brazil, was investigated using ALOS PALSAR data. This study showed that methods based on the nearest neighbour, derived from SMDC, and SVM, with a specific kernel function, are more accurate and robust than the other analysed methods for region-based classification. Furthermore, pixel-based methods are not indicated to perform the classification of images with a strong presence of noise, such as SAR images.

1. Introduction

The use of region-based classification methods has been increasing, particularly with high-resolution imagery over urban areas, where pixel-based classification normally fails because of the high heterogeneity and complexity of such environments (Liu and Xia 2010; Gigandet et al. 2005; Maillard and Alencar-Silva 2013). Herholz et al. (2014) applied region-based classification to analyse medical imagery. Liu, Wang, and Gong (2014) used this classification approach with light detection and ranging (lidar) data. Region-based

---

CONTACT R. G. Negri ✉ rogerio.negri@ict.unesp.br 📧 Instituto de Ciência e Tecnologia, UNESP – Univ. Estadual Paulista, São José dos Campos, São Paulo, Brazil

classification is also especially useful for radar data, which are normally analysed using pixel-based methods (Freitas et al. 2008; Li et al. 2012a; Li et al. 2012b; Zhang et al. 2013).

Region-based classifiers first aggregate pixels into homogeneous objects using segmentation techniques and then classify the objects individually (Liu and Xia 2010). Typically, classification is performed using a statistical distance between the representative distribution of each class of interest and the pixel distribution inside an unknown region. As presented in Silva et al. (2011), the class at the minimum distance to the unknown region distribution is assigned to the region and this process is known as stochastic minimum distance classification (SMDC). A Gaussian assumption is used for the standard statistical distance definition (Richards and Jia 2005).

Negri, Dutra, and Sant'Anna (2012a) theoretically introduced distinctive ways of using stochastic distances for region-based classification, which have been tested with simulated data. The first study to use the Bhattacharyya kernel function (Kondor and Jebara 2003) and apply Support Vector Machine (SVM) to region-based classification problems was presented in Negri, Dutra, and Sant'Anna (2012b).

Another method that yielded good results for multispectral image classification was proposed by Camps-Valls, Tatyana, and Zhou (2007). This method is based on graph classification and its formalization allows for the use of kernel functions. For these characteristics, it is possible to use the Bhattacharyya kernel function and apply the method proposed in Camps-Valls, Tatyana, and Zhou (2007) for region-based classification, similar to that of Negri, Dutra, and Sant'Anna (2012b).

The present work analyses the methods presented in Silva et al. (2011), Negri, Dutra, and Sant'Anna (2012a), Negri, Dutra, and Sant'Anna (2012b), and Camps-Valls, Tatyana, and Zhou (2007) (with the latter two methods using the Bhattacharyya kernel function) for region-based classification. The equivalent pixel-based versions of the region-based analysed methods were included for comparison.

In this investigation, a practical evaluation of the methods is presented for land use and land cover (LULC) classification using ALOS PALSAR imagery in a study area near the Tapajós National Forest in the western part of Pará state, Brazil. Two classification scenarios with different classes were considered in this study to evaluate the analysed methods.

2. Theoretical background

2.1. Stochastic distances

Stochastic distances were used as discrimination measures. These distances quantify the separability of two sets of information. Probability density functions are used to model the information distribution in each set. The separability of the sets is equivalent to the distance between their probability functions. The Jeffries–Matusita distance (JM) is a stochastic distance that is usually adopted in remote-sensing applications (Richards and Jia 2005):

$$JM(C, D) = \int_{x \in X} \left[\sqrt{f_C(x; \Theta_C)} - \sqrt{f_D(x; \Theta_D)} \right]^2 dx, \qquad (1)$$

where f_C and f_D are probability density functions, with parameters Θ_C and Θ_D, which model the information distribution of the sets C and D; such elements belong to X. Assuming f_C and f_D are Gaussian multivariate distributions, Equation (1) can be reformulated as (Richards and Jia 2005)

$$JM(C, D) = 2\left(1 - e^{-B(C,D)}\right),$$ (2)

where $B(\cdot, \cdot)$ is the Bhattacharyya distance assuming the Gaussian multivariate distribution, defined by

$$B(C, D) = \frac{1}{8}(\mu_C - \mu_D)^{\mathsf{T}}\left(\frac{\sum_C + \sum_D}{2}\right)^{-1}(\mu_C - \mu_D) + \frac{1}{2}\ln\left(\frac{|\sum_C + \sum_D|}{\sqrt{|\sum_C||\sum_D|}}\right),$$ (3)

where μ_Z and $\sum_Z$ are the mean vector and covariance matrix, respectively, estimated for a set Z and T, $|\cdot|$, and $(\cdot)^{-1}$ represent the transpose, determinant, and inverse matrix operations, respectively.

2.2. Stochastic distances and region-based classification

If I is an image defined on a support $S \subset \mathbb{N}^2$ and X is an attribute space, $I(s) = x$ denotes that a pixel $s \in S$ of I has an attribute vector $x \in X$. The region-based classification process consists of associating the class ω_j, $j = 1, \ldots, c$, with a region $R_i \subset S$, $i = 1, \ldots, r$. R_i is a set of pixels s_a, $a = 1, \ldots, \#R_i$, where the attributes of s_a are obtained from $I(s_a)$ and # is the cardinality operator. In this context, the support of I is partitioned into r disjoint regions by a segmentation process. Regions represent sets of spatially connected pixels whose attribute vectors meet a particular uniformity criterion. In the classification process, all pixels of the same region are assigned to a single class.

For a supervised region-based method, it is necessary to construct a set of labelled regions $D = \{(R_i, \omega_j) \in S \times \Omega : i = 1, \ldots, m; j = 1, \ldots, c\}$, where m is the number of training regions. The notation (R_i, ω_j) indicates that R_i is assigned to ω_j.

As mentioned above, SMDC is adopted for region-based classification in Silva et al. (2011). In this case, the pixel distribution in an unlabelled region is used to estimate a probability distribution. This region is then associated with the class with the closest distribution, according to an adopted stochastic distance. The class distributions are modelled based on information from D. Formally, if we let R_i be an unlabelled region and let $M(f_{R_i}, f_{\omega_j})$ be a stochastic distance between the distributions of the attribute vectors of the pixels in R_i and the class ω_j, an assignment (R_i, ω_j) is made when the following rule is satisfied:

$$(R_i, \omega_j) \Leftrightarrow j = \arg\min M(f_{R_i}, f_{\omega_j}),$$
$$j = 1, \ldots, c.$$ (4)

In Equation (4), f_{ω_j} is estimated by considering the attribute vectors of the pixels of all of the labelled regions assigned to ω_j in D. Alternatives to SMDC based on simple changes in $M(\cdot, \cdot)$ were derived in Negri, Dutra, and Sant'Anna (2012a).

The first proposed alternative, called the stochastic minimum averaged distance classifier (SMADC), uses $M_{mean}(\cdot, \cdot)$ instead of $M(\cdot, \cdot)$, which is defined as

$$M_{\text{mean}}(f_{R_i}, f_{\omega_j}) = \frac{1}{t_j} \sum_{l=1}^{t_j} M(f_{R_i}, f_{\omega_j R_l}), \tag{5}$$

where $f_{\omega_j R_l}$ is the probability distribution that models the lth training region assigned to ω_j, which contains t_j training regions in D.

Another alternative is the stochastic nearest neighbour classifier (SNNC), obtained by replacing $M(\cdot, \cdot)$ with $M_{min}(\cdot, \cdot)$, which returns the shortest distance between R_i and one of the training regions assigned to ω_j. $M_{\text{min}}(\cdot, \cdot)$ is defined as

$$M_{\text{min}}(f_{R_i}, f_{\omega_j}) = \min\{M(f_{R_i}, f_{\omega_j R_l}) : l = 1, \ldots, t_j\}. \tag{6}$$

A third alternative is a generalization of Equation (6), which transforms Equation (4) into a stochastic version of the *k*-nearest neighbour (SkNN) when $M(\cdot, \cdot)$ is substituted by $M_{\text{knn}}(\cdot, \cdot)$, and is defined as

$$M_{\text{knn}}(f_{R_i}, f_{\omega_j}) = e^{-h_j(f_{R_i})}, \tag{7}$$

where $h_j(f_{R_i}) = \#\{(\overline{R}, \omega_j) \in V_k(R_i)\}$, such that $V_k(R_i)$ is the set of k training regions close to R_i given a distance $M(\cdot, \cdot)$. Formally,

$$V_k(R_i) = \{(\overline{R}_p, \omega_q) \in D : 0 < M(f_{R_i}, f_{\overline{R}_1}) \leq$$
$$\leq M(f_{R_i}, f_{\overline{R}_2}) \leq \cdots \leq M(f_{R_i}, f_{\overline{R}_k}); p = 1, \ldots, k; \ q = 1, \ldots, c\}. \tag{8}$$

In this formalization, $\overline{R}_p$ represents a new indexing of the k nearest regions of D based on the proximity to R_i.

2.3. *Region-based classification with SVM*

SVM has received great attention in recent years because of its excellent generalization ability, its independence of data distribution, and its robustness with respect to Hughes phenomenon (Bruzzone and Persello 2009).

The adoption of kernel functions, $K : X^2 \rightarrow \mathbb{R}$, is a common strategy to improve SVM classification performance on non-linearly separable patterns. A popular example of a kernel function is the radial basis function (RBF) $K(x_u, x_v) = e^{-\gamma \|x_u - x_v\|^2}$, $\gamma \in \mathbb{R}_+$.

Another use of kernel functions is to generalize the application of SVM in problems where patterns do not have an original vectorial representation, e.g. strings, graphs, and sets of different cardinality. In these cases, measures of distances or similarities between the patterns are used.

The flexibility offered by kernel functions provides a distinct way to apply the SVM to region-based classification problems. For this purpose, we adopt a specific kernel function that considers a set of pixels (i.e. a region on an image) as a single pattern.

The given function $K : X^2 \rightarrow \mathbb{R}$ is a kernel function if K is symmetric and conforms to the Mercer theorem conditions (Cristianini and Shawe-Taylor 2000). However, such a verification may not be trivial; there are alternative ways to develop such a function. For example, the RBF model may be adopted (Schölkopf and Smola 2001):

$$K(x_u, x_v) = g(d(x_u, x_v)), \tag{9}$$

where $d : X^2 \rightarrow \mathbb{R}$ is a distance and $g : \mathbb{R} \rightarrow \mathbb{R}$ is a strictly positive real function, e.g. $g(z) = e^{-z}$.

Based on the model presented in Equation (9), if $g(\cdot)$ is equivalent to the negative exponential function mentioned above and $d(\cdot, \cdot)$ is the Bhattacharyya distance defined in Equation (3), the following kernel function is defined:

$$K(R_u, R_v) = e^{-\gamma B(R_u, R_v)},\tag{10}$$

where R_u and R_v are sets and $\gamma \in \mathbb{R}_+$ is a user-adjusted parameter. Considering the sets R_u and R_v as regions of I, which have known attribute vectors, the use of the kernel function in Equation (10) provides a distinct application of SVM in region-based classification. It should be noted that Equation (10) is similar to the JM stochastic distance defined in Equation (2). The function in Equation (10) was originally proposed in Kondor and Jebara (2003), wherein it is called a Bhattacharyya kernel. In Kim and Park (2009), this kernel is used in artificial neural networks for signal classification. In the present study, the Bhattacharyya kernel is used in the SVM method for region-based classification, and this combination is denoted as SVM-BK.

2.4. *Graph region-based classification*

Graph classification methods are characterized by semi-supervised learning. This learning paradigm is motivated by an insufficient amount of labelled data to adequately train the classifier. In this circumstance, the number of insufficient samples may be minimized by increasing the training set using unlabelled data, which are abundant in most classification problems (Zhu and Goldberg 2009).

In general, graph classification uses an affinity matrix **G**, which is a numerical representation of a graph. In this matrix, the similarities between patterns are represented, whether labelled or not. Formally, let D and $\tilde{D}$ denote labelled and unlabelled data sets, respectively, and let $x_u, x_v \in D \bigcup \tilde{D}$ represent two graph vertices. The value (weight) associated with the edge between such vertices corresponds to a similarity measure g_{uv}, which is an element of **G**. The patterns x_u and x_v tend to be associated with the same class when the value of g_{uv} increases.

According to Zhu and Goldberg (2009), graph methods are based on the 'smoothness assumption', where the pattern labels (classes), in this case the vertices, vary smoothly on the graph. The process of associating a class with an unlabelled vertex depends on the similarity between other neighbour vertices in the graph.

Camps-Valls, Tatyana, and Zhou (2007) presented a graph-based method that allows for the use of kernel functions such as those discussed for SVM. Let $\overline{D} = D \bigcup \tilde{D}$ denote a data set composed of m and r labelled and unlabelled samples (patterns), respectively; the affinity matrix **G** can be determined by

$$\mathbf{G}_{(m+r)\times(m+r)} : g_{uv} = e^{-\gamma\|x_u - x_v\|^2}; u, v = 1, \ldots, m + r; \gamma \in \mathbb{R}_+.\tag{11}$$

Consequently, it is verified that $e^{-\gamma\|x_u - x_v\|^2}$, which corresponds to the RBF kernel, can be substituted by the Bhattacharyya kernel function defined in (10). Under this consideration, and if D is a training region set and $\tilde{D}$ is a set of unlabelled regions drawn from a segmentation of I, as defined in Section 2.2, the method proposed in Camps-Valls,

Tatyana, and Zhou (2007) can address the region-based approach. This method is denoted by GM-BK.

3. Experiments

In this section, we present practical experiments with the objective of comparing the region-based classification methods discussed in Sections 2.2–2.4. Pixel-based equivalent versions of the aforementioned region-based methods were included in the study, which allows the comparison of the methods for both classification approaches.

For this purpose, we considered a case study using an ALOS PALSAR image over an Amazon area. The comparison between methods was made considering the *accuracy of classifiers* and *accuracy of maps* produced by each classification method. In addition to the two types of accuracies considered in the analysis, two distinct classification scenarios were used.

More details of the data used in the experiments are presented in Section 3.1. A complete description of the experiment design is provided in Section 3.2. Results and discussions are presented in Section 3.3.

3.1. *Description of the data*

In this section, we present practical applications of the region-based methods discussed in Sections 2.2–2.4. For this purpose, we used a synthetic aperture radar (SAR) image with HH, HV, and VV polarization acquired on 13 March 2009 by an ALOS PALSAR sensor over a region near the Tapajós National Forest, Pará state, Brazil, whose location is shown in Figure 1(a). The study image, illustrated in Figure 1(b), has 3×3 multi-look processing, is 860×1229 pixels wide, has a 20 m resolution, and covers an area of approximately 289 km^2.

To perform the experiments using the region-based approach, it is necessary, first, to segment the image to be classified. The segmentation was performed using the region-growing method available in the Geographic Information System SPRING (Câmara et al. 1996). The segmentation parameter selection was visually performed. Figure 1(c) represents the contours of the image segmentation.

Among the LULC classes in the study area identified from a fieldwork campaign conducted in September 2009, we considered samples of the following classes: primary forest (PF), regeneration (RE), pasture (PS), bare soil (BS), and three types of agriculture, denoted as Agriculture 1, 2, and 3 (A1, A2, and A3). These agricultural classes differ from each other based on crop type or growing stage.

It is important to mention that the LULC samples identified in the study area were initially randomly divided into two main sets designated for training the methods and testing the classification results. The training set was intentionally defined with approximately double the number of polygons in each class compared to the test set. Unlike traditional studies, where the sample division is performed on the basis of pixel number, the number of polygons was considered to perform the training/test division, where polygons (i.e. regions) were the classification objects for region-based methods, as defined in Section 2.2. The spatial distributions of the samples for scenarios 1 and 2 are represented in Figure 1(d) and (e), respectively, where void polygons were selected
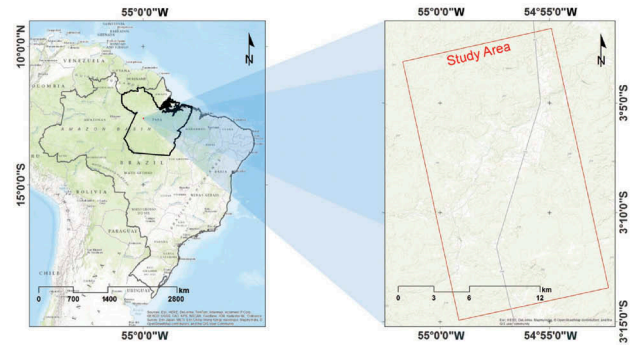
(a) Location of the study area
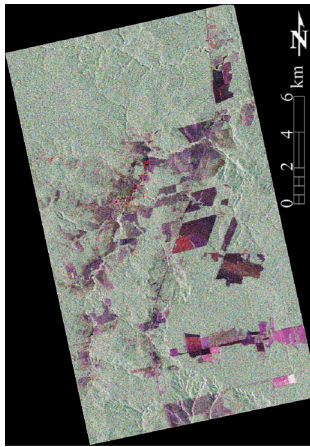
(b) Study image in (HH-HV-VV intensity)RGB composition

(c) Segmentation

PF
RE
BS
PS
A1
A2
A3
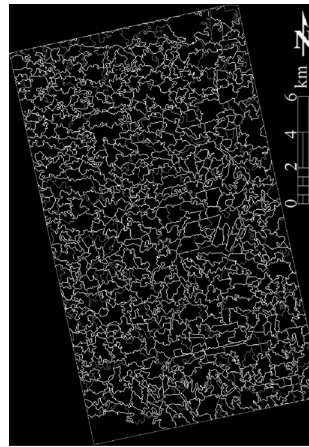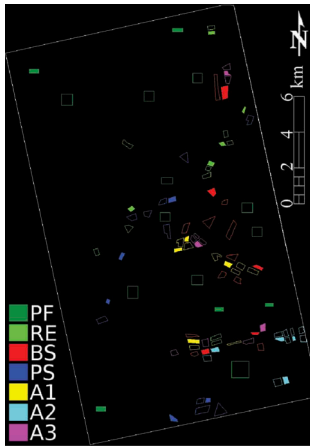
(d) Training and test LULC samples for scenario 1

HB
LB
AA
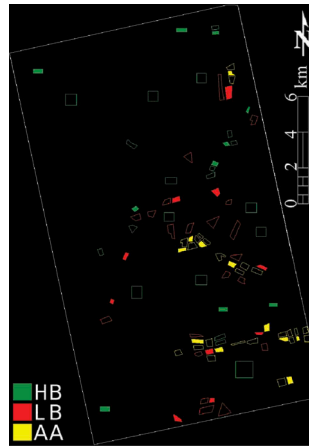
(e) Training and test LULC samples for scenario 2

Figure 1. The PALSAR image, its segmentation, and samples used in the study.

for training and full polygons for testing. To remove spatial correlation between the pixels inside the training and test polygons, a resampling was performed considering a regular grid with a lag of three pixels in both vertical and horizontal directions.

Table 1. Summary of the land-cover classes and scenarios.

LULC Types	Training (void polygons)		Test (solid polygons)		Scenario/Class	
	Pixels	Polygons	Pixels	Polygons	1	2
Primary Forest	9134	10	1563	5	PF	HB
Regeneration	2321	10	835	5	RE	
Bare Soil	3757	11	1586	5	BS	LB
Pasture	3958	10	1025	5	PS	
Agriculture 1	2329	8	975	4	A1	AA
Agriculture 2	2045	8	874	4	A2	
Agriculture 3	1901	8	686	3	A3	

Additionally, Table 1 summarizes for each LULC class the number of pixels and polygons selected for training and testing.

The considered LULC classes were organized into two classification problems, called scenarios. The first scenario was characterized by all seven classes. However, the second scenario was characterized by only three classes, defined by grouping PM and RE as one class (high biomass (HB)), BS and PS as another class (low biomass (LB)), and, finally, the three types of agriculture as a single class (agricultural areas (AA)). The groups of classes that define each scenario are also presented in Table 1. The number of pixels and polygons relative to scenario 2 classes was inferred by summing the relative class numbers from scenario 1.

3.2. *Experiment design*

As previously mentioned, the methods presented in Sections 2.2–2.4 were compared in terms of the accuracy of classifiers and accuracy of maps. The classification methods used were SMDC, SMADC, SNNC, SkNN, SVM-BK, and GM-BK.

In the experiments, the generic stochastic distance M employed to formalize SMDC, SMADC, SNNC, and SkNN on Equations (4), (5), (6), and (8), respectively, was substituted by the JM distance (2) under the Gaussian multivariate assumption. Although a SAR image was adopted in the case study, it is reasonable to use the Gaussian multivariate distribution to model the data once the multi-look process had sufficiently smoothed the speckle noise and, consequently, 'Gaussianized' the data; additionally, a more appropriate distribution in terms of the SAR data is not known.

As already stated, for each region-based method analysed in this study, a corresponding pixel-based classifier was considered. Specifically, the correspondence between such methods is: for SMDC we have the Mahalanobis distance classifier (MDC) (Richards and Jia 2005) as correspondent; similarly, for SMADC we considered the method MADC, which changes M in Equation (5) by the Mahalanobis distance; 1NN and kNN are the respective pixel-wise versions of SNNC and SkNN, but considering M as the Euclidean distance instead of the JM; the correspondent of GM-BK is GM-RBF where the affinity matrix **G** is defined, as in Equation (11); finally, SVM using a RBF kernel, denoted by SVM-RBF, is the correspondent for SVM-BK.

It is worth mentioning that SVM using the RBF kernel for region-based classification, as presented in Liu and Xia (2010), was not included in the comparisons because the superiority of SVM-BK had already been verified in Negri, Dutra, and Sant'Anna (2012b).

To address the problem of multi-class classification on SVM-BK and SVM-RBF, the one-against-all strategy was used.

The accuracy of classifiers was computed following a repetitive training and classification scheme using just the training samples (Section 3.1). For the first step, the training set is randomly divided into two subsets, denoted by A and B. Subset A has approximately 67% of all polygons from the training set and subset B has the remaining 33% of polygons. After defining these subsets, when applicable, the classifier parameter tuning for each method is performed.

For methods for which parameters are to be set (i.e. SkNN, kNN, SVM-BK, SVM-RBF, GM-BK, and GM-RBF), a fine-tuning procedure is performed based on a grid search process with 10-fold cross validation, as discussed in Hsu, Chang, and Lin (2010). The space search considered for the grid search process for each parameter/method is shown in Table 2. This parameter adjustment process has as its objective the guarantee that the conclusions were not impaired by results from methods with ill-tuned parameters. Once the best parameters are found, for those methods, the parameter tuning process applies, and each method is trained using all of the information in subset A and is used to classify the samples from subset B. The results from subset B classification are assessed using the kappa agreement coefficient (Congalton and Green 2009), and the computed measures are stored for future statistical analysis. It is worth mentioning that for GM-BK and GM-RBF, when training or performing the grid search, the labelled data not instantaneously in use were considered as an unlabelled set, denoted as $\tilde{D}$ in Section 2.4.

All steps described, from the selection of subsets A and B to the classification of the accuracy measurement, were performed 50 times. The difference between each execution was just the choice of the distinct A and B subsets. After all 50 repetitions, the average kappa value and its standard deviation for each method were computed. The general classifier accuracy process is illustrated in Figure 2(a). Detailed steps on the classifier parameter tuning are presented in Figure 2(b).

Most of the image classification results from remote-sensing applications are assessed using an accuracy measure, such as the kappa coefficient, and a set of ground truth information. Thus, it is estimated that the accuracy values calculated using the adopted measure from such samples reflect the accuracy of the entire classified image, i.e. the accuracy of the map.

However, several discussions presented in the literature (Pontius and Millones 2011) question this classification assessment approach. As a special case, accuracy values computed using the kappa coefficient may be influenced by the dimensions of the ground truth class sets considered in the study, where bigger sets imply a greater

Table 2. The search space considered in the grid search for SkNN, kNN, SVM-BK, SVM-RBF, GM-BK, and GM-RBF for parameter tuning.

Parameter	Methods	Search Space
C – Penalty	SVM-BK, SVM-RBF	{1; 10; 100; 1000}
γ – kernel flexibility	SVM-BK, SVM-RBF, GM-BK, GM-RBF	{0.25; 0.5; 1.0; 1.25; 2.0; 2.25}
β – graph regularization[1]	GM-BK, GM-RBF	{0.25; 0.5; 0.75; 0.95}
κ – nearest neighbour	SkNN, kNN	{2; 3; 4; 5}

[1]Details about this regularization parameter can be found in Camps-Valls, Tatyana, and Zhou (2007).

(a) General process

(b) Classifier parameter tuning
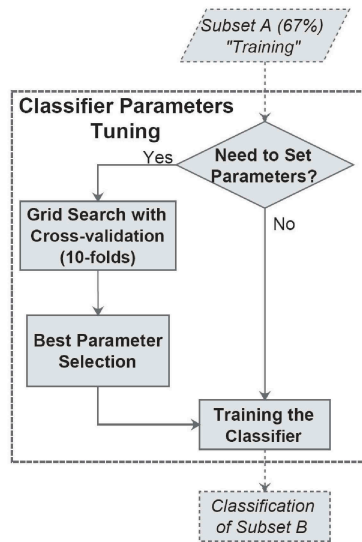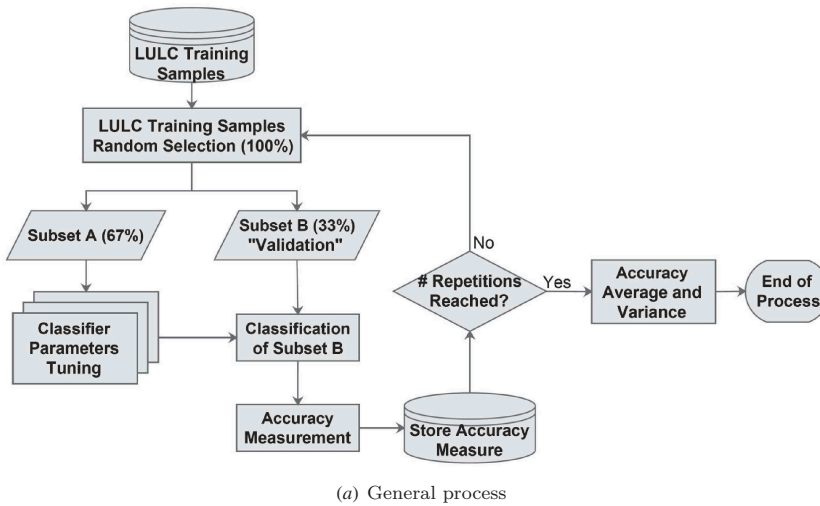
Figure 2. Classifier assessment flow chart.

influence on the accuracy values, although this behaviour is not always desired. Unfortunately, the unavailability of ground truth class sets of the same dimension is to be expected.

In this study, the accuracy of maps is calculated according to a particular scheme presented in Figure 3(a). Initially, all training samples defined in Section 3.1 were used in the training process, similar to the process discussed previously and represented in Figure 2(b), except that all training samples were used and not just a subset. After tuning the parameters and training, each region-based method was applied to classify the study image based on the regions delimited by its segmentation (Figure 1(b) and (c)). Classifications through pixel-wise methods are independent of the segmentation. Again,

(*a*) General process

(*b*) Accuracy analysis
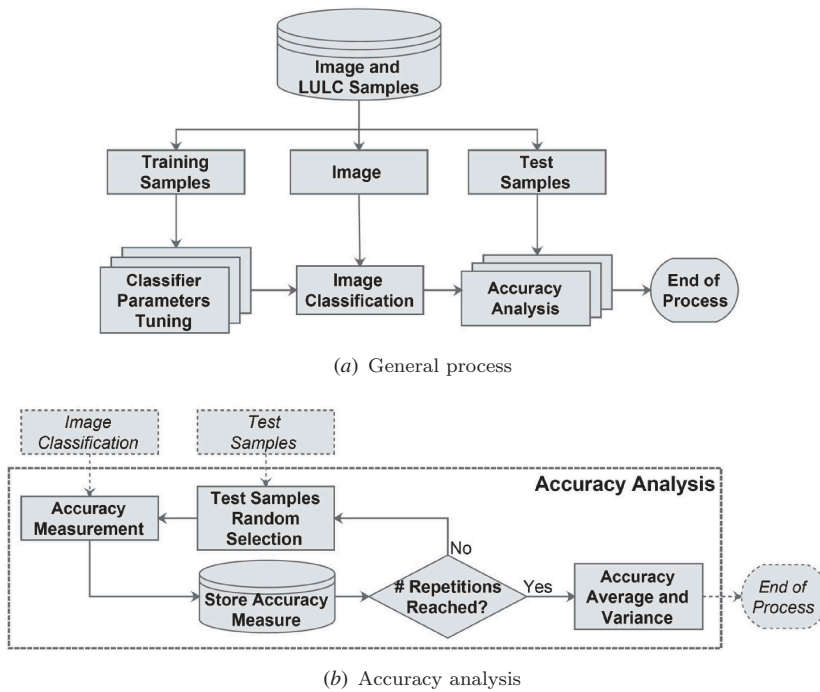
Figure 3. Map assessment flow chart.

it is important to mention in this case that for GM-BK, the unlabelled set $\tilde{D}$ is composed of all regions delimited in the segmentation process. However, computational limitations make impossible considering $\tilde{D}$ in the pixel-wise approach as the set of all unlabelled pixels once to train graph methods is necessary to store and perform matrix operations such as inverse over matrices of (huge) dimension $(D + \tilde{D}) \times (D + \tilde{D})$. To address this limitation, $\tilde{D}$ is composed of pixels sampled from the original image through a regular grid with a lag of 25 pixels in both vertical and horizontal directions.

Once a classification result was obtained from each method, we performed the accuracy analysis process, illustrated in Figure 3(b). This process included three repetitive basic steps: (*i*) random selection of the test sample, (*ii*) computation of the accuracy measures, and (*iii*) storage of the calculated measures. In the first step, we randomly selected 350 pixels from each class. The reason for using 350 pixels was based on the size of the smallest test sample, i.e. A3 (see Table 1); 350 is almost 50%. The remaining quantity increased the possibility of generating test subsets that were at least 50% different from each other. Next, from the defined test subset, we computed the kappa accuracies for each classification result. The accuracies calculated were stored for future statistical analysis. These three steps were performed 50 times.

As a final step, the averages and variances were computed from the respective 50 kappa values stored during the accuracy analysis process. In other words, for each analysed method there was an average performance measure and its deviation according to the accuracies of the classifier and map. In addition to the statistical descriptive values, a bilateral *t*-test was used to compare two population means (Mood and Graybill 1974).

All processing performed were conducted on a computer with an Intel Core *i*5 processor and 8 GB of RAM running the Ubuntu-Linux operating system version 14.04. The Interactive Data Language (IDL) programming language version 7.1 was used to implement the classification methods. The algorithm SVM^{Light} Joachims (1999) was used to train SVM-BK and SVM-RBF.

3.3. *Results and discussion*

Using the data described in Section 3.1 and considering the steps specified in Section 3.2, we obtained the accuracies of the classifiers and the maps. Figure 4(a) and (b) shows these accuracies, respectively. Furthermore, these figures simultaneously present the performances in both scenarios 1 and 2.

As expected, very similar behaviour is observed between the accuracy of the classifiers and the maps (Figure 4(a) and (b)). However, it is worth noting the opposite performance of SNNC, SkNN, GM-BK, and SVM-BK, comparing its accuracies for classifier and map in both scenarios 1 and 2. Such variations are introduced because of the different training and test data sets used in each case. Furthermore, while the accuracy of the classifiers is computed from different data sets for training and test, the accuracy
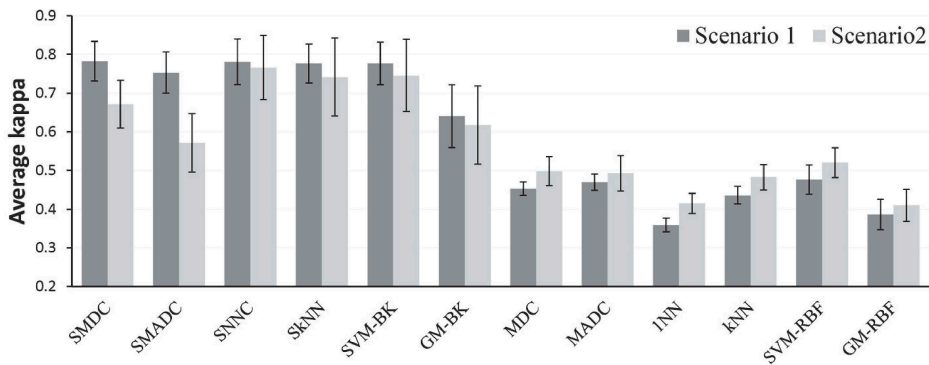
(*a*) Accuracy of the classifiers; error bars represent ±1 standard deviation
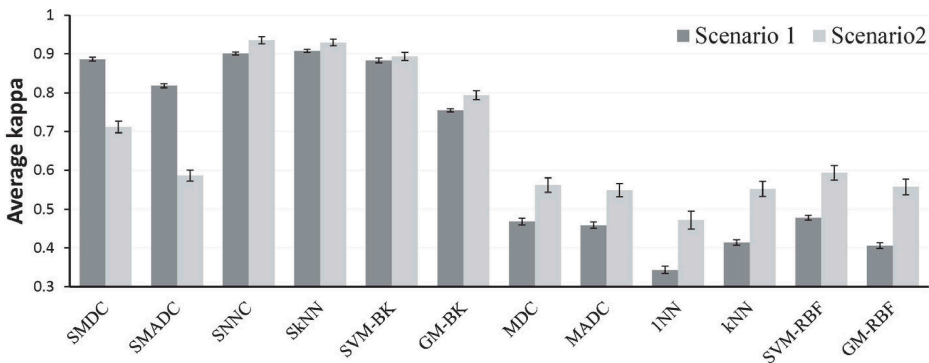
(*b*) Accuracy of the maps; error bars represent ±1 standard deviation

Figure 4. A graphical comparison between classifiers and classification accuracies.

of the maps comes from just one classification result. In other words, the accuracy of the maps (single results) achieved a higher value compared to the average accuracy of the classifiers (50 results).

Observing just the accuracy of the classifiers for scenario 1 (Figure 4(a)), it is possible to note that all region-based methods achieved better performance compared to pixel-based methods. A *t*-test to compare the accuracy of classifiers, at a 95% confidence level, allowed us to verify the statistical equality between all region-based methods, in addition to the statistical equality between all pixel-based methods and the dominant statistical difference between region- and pixel-based methods. We note that GM-BK achieved lower accuracy compared with other region-based methods. Additionally, GM-BK was statistically equivalent to SVM-RBF, where the latter had a higher accuracy value among the observed pixel-based methods.

Conversely, with respect to the accuracy of the classifiers for scenario 2, we observed an evident performance drop for the SMDC and SMADC methods. Furthermore, the decrease in accuracies associated with the increase in deviations of SMADC and GM-BK, with opposition to the increase in the accuracy of all pixel-based methods for scenario 2, led to statistical equality between the methods. The increase of kappa values for scenario 2 might have been influenced by the reduced quantity of classes. It is worth noting that SNNC, SkNN, SVM-BK, and GM-BK present similar behaviours for both scenarios 1 and 2.

The *p*-values for the *t*-tests that compare the analysed methods for scenario 1 with respect to the accuracy of the classifiers are shown in the upper triangular matrix in Table 3. The lower triangular matrix in Table 3 presents the corresponding *p*-values for scenario 2.

Focusing the discussion on the accuracy of the maps, we can observe that the highest accuracies for scenario 1 were obtained with classifications produced by SMDC, SNNC, SkNN, and SVM-BK. In contrast, lower accuracies were obtained with 1NN, kNN, and GM-RBF.

From the results of scenario 2, SNNC, SkNN, and SVM-BK were found to have similar accuracy levels to those obtained in scenario 1; however, SMDC and SMADC showed a lower level of performance. Even given their low performance, GM-BK achieved more accurate values compared to those for scenario 1. Table 4 presents the *p*-values from a *t*-test applied to compare the accuracy of the maps illustrated in Figure 4(b) for scenarios

Table 3. *p*-values from a bilateral *t*-test to compare the classifier accuracies achieved by the analysed methods in scenarios 1 (upper triangular matrix) and 2 (lower triangular matrix).

	SMDC	SMADC	SNNC	SkNN	SVM-BK	GM-BK	MDC	MADC	1NN	kNN	SVM-RBF	GM-RBF
SMDC		0.690	0.981	0.936	0.938	0.139	0	0	0	0	0	0
SMADC	0.305		0.728	0.748	0.759	0.245	0	0	0	0	0	0
SNNC	0.361	0.082		0.959	0.960	0.162	0	0	0	0	0	0
SkNN	0.556	0.177	0.850		0.999	0.154	0	0	0	0	0	0
SVM-BK	0.508	0.146	0.870	0.975		0.165	0	0	0	0	0	0
GM-BK	0.648	0.714	0.256	0.386	0.351		0.024	0.042	0.001	0.016	0.068	0.005
MDC	0.017	0.387	0.003	0.024	0.014	0.269		0.545	0	0.541	0.569	0.121
MADC	0.020	0.375	0.004	0.025	0.015	0.261	0.928		0	0.274	0.867	0.063
1NN	0	0.051	0	0.002	0.001	0.053	0.070	0.140		0.007	0.005	0.527
kNN	0.007	0.283	0.001	0.015	0.008	0.205	0.757	0.857	0.107		0.3d52	0.277
SVM-RBF	0.038	0.547	0.007	0.040	0.025	0.368	0.683	0.648	0.024	0.459		0.098
GM-RBF	0	0.062	0	0.002	0.001	0.058	0.116	0.180	0.915	0.170	0.052	

Table 4. *p*-values from a bilateral *t*-test to compare the map accuracies achieved by the analysed methods in scenarios 1 (upper triangular matrix) and 2 (lower triangular matrix) .

	SMDC	SMADC	SNNC	SkNN	SVM-BK	GM-BK	MDC	MADC	1NN	kNN	SVM-RBF	GM-RBF
SMDC		0	0.023	0.001	0	0	0	0	0	0	0	0
SMADC	0		0	0	0.528	0	0	0	0	0	0	0
SNNC	0	0		0.221	0	0	0	0	0	0	0	0
SkNN	0	0	0.642		0	0	0	0	0	0	0	0
SVM-BK	0	0	0	0		0	0	0	0	0	0	0
GM-BK	0	0	0	0	0		0	0	0	0	0	0
MDC	0	0.300	0	0	0	0		0.424	0	0	0.367	0
MADC	0	0.090	0	0	0	0	0.596		0	0	0.065	0
1NN	0	0	0	0	0	0	0.002	0.007		0	0	0
kNN	0	0.155	0	0	0	0	0.710	0.896	0.007		0	0.428
SVM-RBF	0	0.751	0	0	0	0	0.231	0.076	0	0.124		0
GM-RBF	0	0.241	0	0	0	0	0.863	0.744	0.005	0.851	0.188	

1 and 2. As represented in Table 3, the upper and lower triangular matrices inside the table contain the computed *p*-values for scenarios 1 and 2, respectively.

A confidence level of 95% affirms that almost all results are statistically different from each other, except for the classifications obtained by SNNC and SkNN in scenario 1. Adopting the same level of confidence, the scenario 2 classifications provided by SNNC and SkNN are still statistically equal; furthermore, SMADC is statistically equal to most of the analysed pixel-based methods.

In general, region-based methods are statistically superior to pixel-based methods. Although a similar tendency is shown in Figure 4(b), whereas most of the pixel-based results are different from each other in scenario 1, we have the reverse in scenario 2. It is worth noting that 1NN achieved a lower accuracy level.

The LULC classification maps for scenarios 1 and 2, for which the accuracy of the maps (Figure 4(b)) was computed, are illustrated in Figures 5 and 6, respectively.
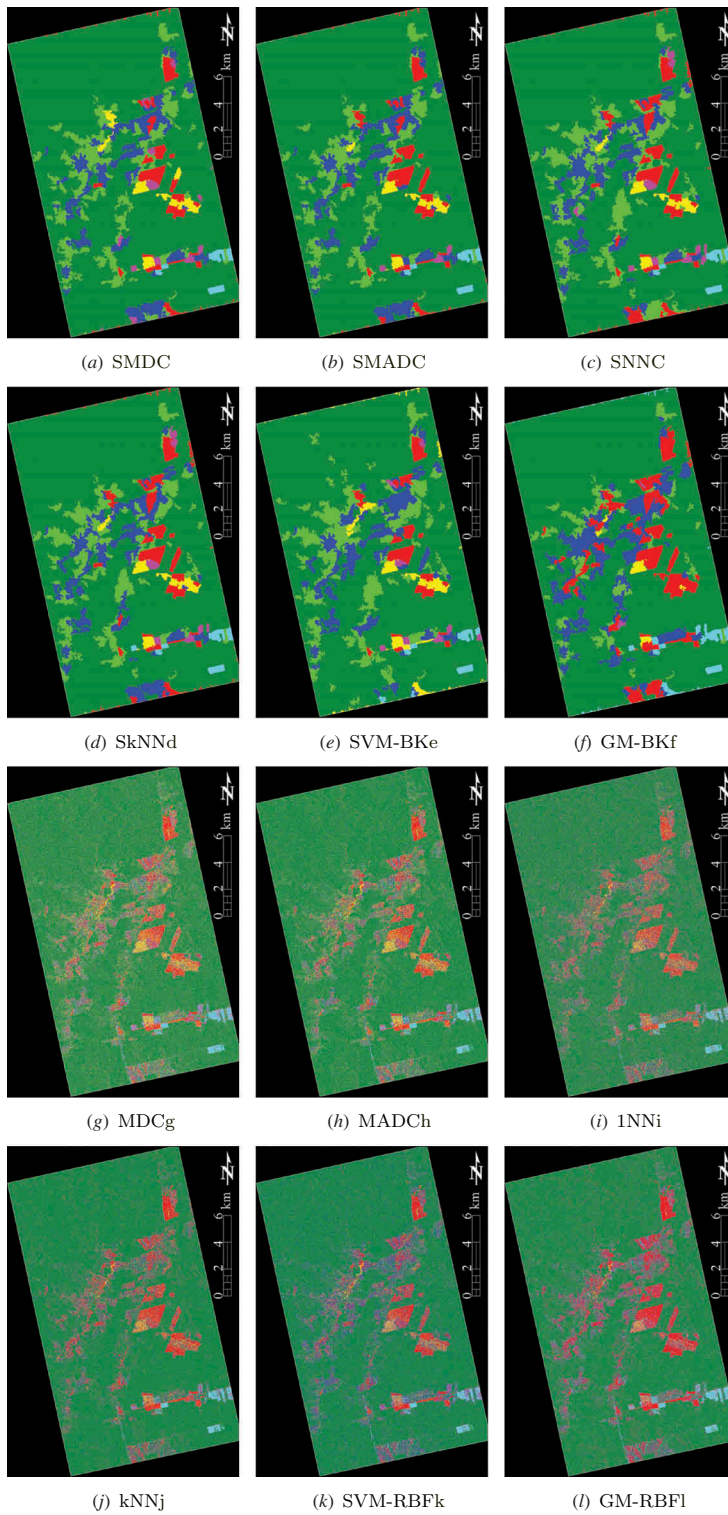
With respect to the performance of region-based classification for scenario 1, there was similarity between SMDC, SNNC, and SkNN, which yielded the highest accuracy values. Satisfactory results were achieved by SMADC and SVM-BK. Loss of classification precision was noted when SMADC misclassified A3, SVM-BK misclassified PS, and GM-BK misclassified the RE and PS areas.

For pixel-based classifications, mainly as a function of the extremely noisy data, poor results were achieved, where A1, A3, PS, and RE classes were frequently confused with each other.

For scenario 2, the highest performances were achieved by the SNNC, SkNN, and SVM-BK methods. Among these methods, we observed that the main divergence occurred when SVM-BK classified the areas at the bottom of the study image. GM-BK did not perform well the classification of LB and AA classes. As expected, because of the data type, the pixel-based results were noisy.
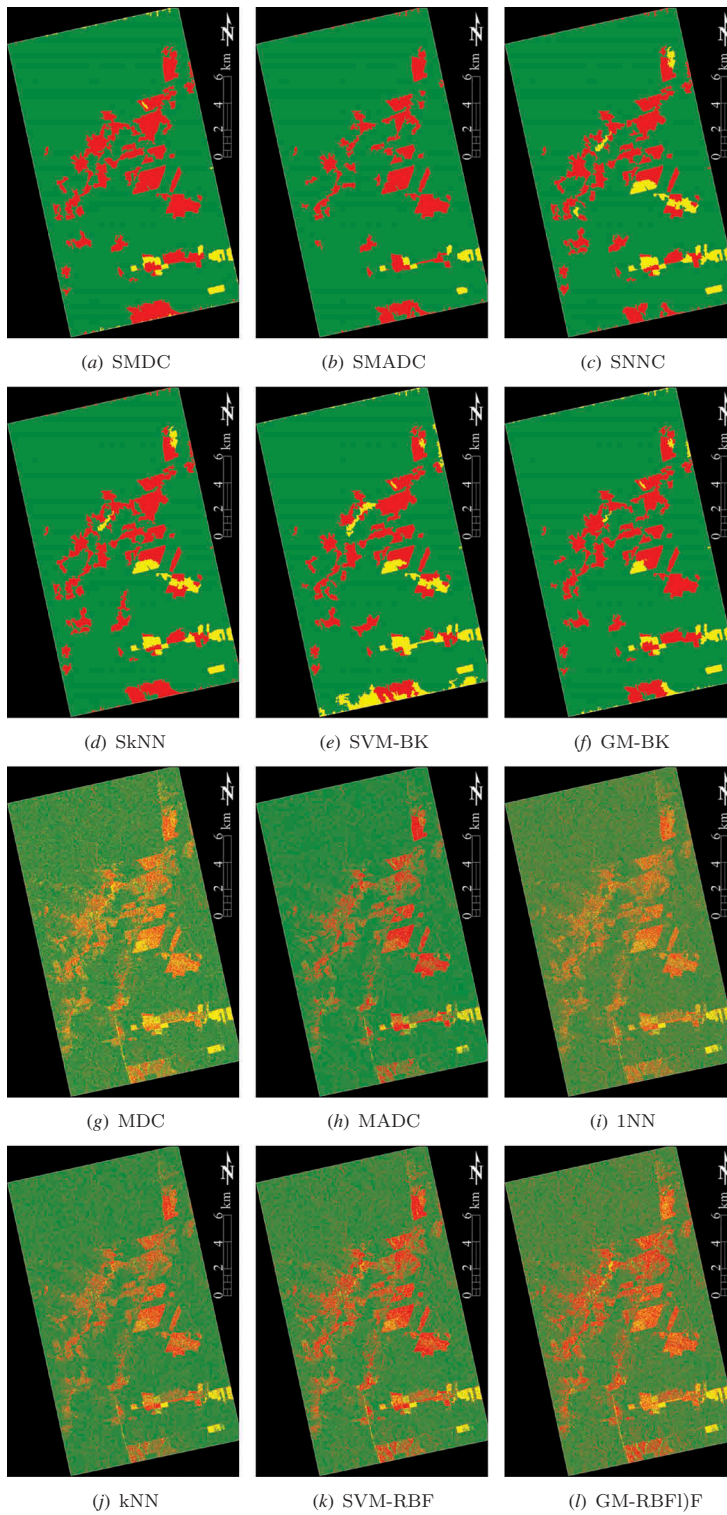
Given the presented results, we can conclude that SMDC and SMADC are less robust than SNNC, SkNN, and SVM-BK because the accuracies achieved in scenario 2 differ considerably compared with scenario 1.

Recapping the theoretical formulation (Section 2.2), SMDC used the information from all training samples (regions) of a given class to estimate a single statistical distribution, which modelled the elements of this class. The classification of an unlabelled region is made based on the most similar class distribution. It is reasonable that the more similar

(a) SMDC   (b) SMADC   (c) SNNC

(d) SkNNd   (e) SVM-BKe   (f) GM-BKf

(g) MDCg   (h) MADCh   (i) 1NNi

(j) kNNj   (k) SVM-RBFk   (l) GM-RBFl

Figure 5. Classification results obtained for scenario 1.

Figure 6. Classification results obtained for scenario 2.

the statistical distributions of the classes involved in the classification problem, the greater were the chances of incorrect classification occurring. Computing the distance between the statistical distributions of the classes in each scenario, modelled from the training samples, while it was found that the greater distance had a value of 1:83 (between PF and BS) for the first scenario, in the second scenario, the greater value was 0:87 (between HB and AA). Based on this fact, we can explain that the union of classes from scenario 1, to define scenario 2, produced classes with similar statistical distributions, which in turn decreased the robustness of the method.

A justification of the SMADC behaviour follows a similar idea. SMADC considers the statistical distances between an unlabelled region and each training sample (region) of a given class. After computing these distances, a 'smoothed' distance value was calculated averaging all distance values previously computed for the initially considered class (Equation (5)). This smoothing introduced possible misclassifications in scenario 2, when the samples of each class in this scenario were most dissimilar compared to the samples of each class in scenario 1. This fact was proved after computing intra-class sample distances, where, whereas the highest value of 0:80 (between samples of A3) was observed in scenario 1, the highest value in scenario 2 was 1:85 (between samples of AA).

On the other hand, SNNC, SkNN, SVM-BK, and GM-BK methods were not affected by this factor. The decision rule of SNNC and SkNN did not depend on the smallest averaged stochastic distances, but on the smallest distance value found between some training samples to the unlabelled region. Focusing on the SVM-BK method, the weighting of training samples, defined through the Lagrangian coefficients (see Theodoridis and Koutroumbas (2008)), to construct the decision rule, prevents problems similar to those presented by SMDC and SMADC.

Despite the robustness of GM-BK on the analysed scenarios, i.e. without presenting sharp drops as SMDC and SMADC, their low accuracy might have derived from their formulation based on the 'smoothing assumption'. This semi-supervised concept established that the pattern labels varied smoothly on the feature space, and for this purpose, the information from unlabelled samples was considered. This assumption may be helpful when the quantity of samples for adequate training is insufficient, but harmful in some cases. According to the present results, there are indications that the use of unlabelled samples affected decision rule estimation.

Regarding the pixel-based methods, it can be observed the existence of a general behaviour independent of the scenario and way of accuracy analysis (i.e. accuracy of the classifier or map), most of the methods (MDC, MADC, kNN, SVM + RBF, and GM + RBF) tend to present similar results. It is worth noting that the highest accuracy values are presented by SVM, in accordance with the previous results discussed in Mountrakis, Im, and Ogole (2011). On the other hand, lower accuracy values are associated with the 1NN method. Such behaviour occurs in function of the decision rule of the method (assign a class to a non-labelled pattern according to the most similar/closer labelled pattern) that becomes inadequate for noisy data (presence of speckle).

In general, the results obtained through the pixel-based approach were overwhelmingly lower than those achieved with the region-based approach. Three factors that explain the differences between these approaches are the adopted image, the pattern type, and how such patterns are treated by the different approaches. Although multi-

look processing performed on the study image had minimized the speckle presence, we could see (Figure 1(b)) that this noise was not eliminated. The high variability introduced by this noise on the pixel's attributes prevented pixel-based methods from performing homogeneous classification of the targets, as illustrated in Figures 5(g)–(l) and 6(g)–(l). On the other hand, region-based methods had higher accuracy because the classification pattern was not a single pixel isolated from its spatial context. Furthermore, the use of stochastic distances improved the distinction between the regions because not only the attributes of pixels inside the region were considered, but also the variability (texture) of such pixels. It is worth recalling that the experiment design included procedures to decorrelate information from the training and test samples, and then the hypothesis that the existence of correlation between samples had favoured region-based methods was discarded.

4. Conclusion and perspectives

In this article, we investigated different image classification methods using the region-based approach. Additionally, the pixel-based approach was included in the analysis for comparison. A case study with respect to LULC on an Amazon area using an ALOS PALSAR image was conducted to compare the methods. For the case study, two classification problems were addressed: the so-called scenarios. While the first scenario dealt with a more specific classification problem composed of seven classes, the other scenario treated three classes, defining a grouping of classes from scenario 1. In addition to the different scenarios, the methods were compared in two distinct assessment ways: the accuracy of the classifiers and the accuracy of the methods. Whereas the accuracy of the classifiers consists in examining the generalization ability of the method, through a repetitive and distinct process of 'training-classification', the accuracy of the maps aims to assess a particular classification result using the kappa coefficient, however avoiding the fact that different quantities of test samples in different classes may influence the accuracy values of such coefficients.

The results showed that nearest neighbour-based classifiers (SNNC and SkNN), independent of the scenario, had better performances than minimum distance-based methods (SMDC and SMADC). Classifiers such as SNNC and SkNN are not available in any commercial or free software; however, their implementation is simple, and they are a good alternative for region-based methods.

The SVM method using the Bhattacharyya kernel function (SVM-BK) provided good results for region-based classification, independent of the scenario. Conversely, SVM with the RBF kernel function (SVM-RBF) used for pixel-based classification yielded inferior performance. This result showed the importance of region-based classification for problems such as the ALOS PALSAR case study discussed in this work.

Moreover, this study presented the use of the graph-based method proposed in Camps-Valls, Tatyana, and Zhou (2007), adapted to the region-based approach using the Bhattacharyya kernel function (GM-BK). The performance of this method in this case study was inferior compared with that of SNNC, SkNN, and SVM using the same kernel function.

Finally, the results indicated that the pixel-based approach was strongly influenced by noise present in the image. Furthermore, the region-based approach could satisfactorily

deal with the presence of noise because of the manner in which the patterns (regions) are compared (i.e. using stochastic distances).

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

R. G. Negri 🔟 http://orcid.org/0000-0002-4808-2362
L. V. Dutra 🔟 http://orcid.org/0000-0002-7757-039X
D. Lu 🔟 http://orcid.org/0000-0003-4767-5710

References

Bruzzone, L., and C. Persello. 2009. "A Novel Context-Sensitive Semisupervised SVM Classifier Robust to Mislabeled Training Samples." *IEEE Transactions on Geoscience and Remote Sensing* 47 (7): 2142–2154. doi:10.1109/TGRS.2008.2011983.

Câmara, G., R. C. M. Souza, F. M. Ii, U. Freitas, and J. Garrido. 1996. "Spring: Integrating Remote Sensing And Gis By Object-oriented Data Modelling." *Computers & Graphics* 20: 395–403. doi:10.1016/0097-8493(96)00008-8.

Camps-Valls, G., V. B. Tatyana, and D. Zhou. 2007. "Semi-Supervised Graph-Based Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 45: 3044–3054. doi:10.1109/TGRS.2007.895416.

Congalton, R. G., and K. Green. 2009. *Assessing the Accuracy of Remotely Sensed Data*. Boca Raton: CRC Press.

Cristianini, N., and J. Shawe-Taylor. 2000. *An Introduction to Support Vector Machines: And Other Kernel-Based Learning Methods*. New York, NY: Cambridge University Press.

Freitas, C. C., L. Soler, S. J. S. Sant'Anna, L. V. Dutra, J. R. Santos, J. C. Mura, and A. H. Correia. 2008. "Land Use and Land Cover Mapping in the Brazilian Amazon Using Polarimetric AirborneP-Band SAR Data." *IEEE Transactions on Geoscience and Remote Sensing* 46 (10): 2956–2970. doi:10.1109/TGRS.2008.2000630.

Gigandet, X., M. B. Cuadra, A. Pointet, R. Cammoun, L. and Caloz, and J. Thiran. 2005. "Region-Based Satellite Image Classification: Method and Validation." IEEE International Conference on Image Processing, Genova, September 11–14, 3832–3835.

Herholz, K., R. Evans, J. Anton-Rodriguez, R. Hinz, and J. C. Matthews. 2014. "The Effect of 18f-Orbetapir Dose Reduction on Region-Based Classification of Cortical Amyloid Deposition." *European Journal of Nuclear Medicine and Molecular Imaging* 41 (11): 2144–2149. http://dx.doi.org/10.1007/s00259-014-2842-3.

Hsu, C. W., C. C. Chang, and C. J. Lin. 2010. *A Practical Guide to Support Vector Classification*. Tech. rep. Tawain. http://www.csie.ntu.edu.tw~cjlin/papers/guide/guide.pdf

Joachims, T. 1999. *Making Large-Scale Support Vector Machine Learning Practical*, 169–184. Cambridge: MIT Press. Advances in Kernel Methods.

Kim, J.-Y., and D.-C. Park. 2009. "Application of Bhattacharyya Kernel-Based Centroid Neural Network to the Classification of Audio Signals." Proceedings of the 2009 International Joint Conference on Neural Networks, Atlanta, Georgia, USA, 2948–2952.

Kondor, R., and T. Jebara. 2003. "A Kernel between Sets of Vectors." International Conference on Machine Learning, Washington, DC, August 21–24.

Li, G., D. Lu, E. Moran, L. V. Dutra, and M. Batistella. 2012a. "A Comparative Analysis of ALOS PALSAR L-Band and RADARSAT-2 C-Band Data for Land-Cover Classification in A Tropical Moist Region." *ISPRS Journal of Photogrammetry and Remote Sensing* 70: 26–38. doi:10.1016/j.isprsjprs.2012.03.010.

Li, G., D. Lu, E. Moran, and S. J. S. Sant'Anna. 2012b. "Comparative Analysis of Classification Algorithms and Multiple Sensor Data for Land Use/Land Cover Classification in the Brazilian Amazon." *Journal of Applied Remote Sensing* 6 (1): 061706–061706. doi:10.1117/1.JRS.6.061706.

Liu, K., Y. Wang, and H. Gong. 2014. "Classification of Lidar Data Based on Region Segmentation and Decision Tree." *Proceedings of SPIE 9262, Lidar Remote Sensing for Environmental Monitoring XIV, 926213*, November 26. doi:10.1117/12.2069203.

Liu, D., and F. Xia. 2010. "Assessing Object-Based Classification: Advantages and Limitations." *Remote Sensing Letters* 1 (4): 187–194. doi:10.1080/01431161003743173.

Maillard, P., and T. Alencar-Silva. 2013. "A Method for Delineating Riparian Forests Using Region-Based Image Classification and Depth-To-Water Analysis." *International Journal of Remote Sensing* 34 (22): 7991–8010. doi:10.1080/01431161.2013.827847.

Mood, A. M., and F. A. Graybill. 1974. *Introduction to the Theory of Statistics*. 3rd ed. Singapore: McGraw-Hill.

Mountrakis, G., J. Im, and C. Ogole. 2011. "Support Vector Machines in Remote Sensing: A review." *ISPRS Journal of Photogrammetry and Remote Sensing Society* 66 (3): 247–259. doi:10.1016/j.isprsjprs.2010.11.001.

Negri, R. G., L. V. Dutra, and S. J. S. Sant'Anna. 2012a. "Stochastic Approaches of Minimum Distance Method for Region Based Classification." *Lecture Notes in Computer Science* 7441: 797–804.

Negri, R. G., L. V. Dutra, and S. J. S. Sant'Anna. 2012b. Support Vector Machine and Bhattacharrya Kernel Function for Region Based Classification. Proceedings International Geoscience and Remote Sensing Symposium, Munich, July 22–27, 5422–5425. IEEE.

Pontius, R. G., and M. Millones. 2011. "Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment." *International Journal of Remote Sensing* 32 (15): 4407–4429. doi:10.1080/01431161.2011.552923.

Richards, J. A., and X. Jia. 2005. *Remote Sensing Digital Image Analysis: An Introduction*. New York: Springer.

Scholkopf, B., and A. J. Smola. 2001. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press.

Silva, W. B., L. O. Pereira, S. J. S. Sant'Anna, C. C. Freitas, R. J. P. S. Guimarães, and A. C. Frery. 2011. "Land Cover Discrimination at Brazilian Amazon Using Region Based Classifier and Stochastic Distance." 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, July 24–29, 2900–2903.

Theodoridis, S., and K. Koutroumbas. 2008. *Pattern Recognition*. 4th ed. Academic Press.

Zhang, B., G. Ma, Z. Zhang, and Q. Qin. 2013. "Region-Based Classification by Combining MS Segmentation and MRF for POLSAR Images." *Journal of Systems Engineering and Electronics* 24 (3): 400–409. doi:10.1109/JSEE.2013.00048.

Zhu, X., and A. B. Goldberg. 2009. *Introduction to Semi-Supervised Learning*. Morgan & Claypool Publishers.