Hermosilla, T.; Ruiz Fernández, LÁ.; Recio Recio, JA.; Fernández-Sarría, A. (2012). Analysis of parcel-based image classification methods for monitoring the activities of the Land Bank of Galicia (Spain). Applied Geomatics. 4(4):245-255. doi:10.1007/s12518-012-0087-z.

Analysis of parcel-based image classification methods for monitoring the activities of the Land Bank of Galicia (Spain)

T. Hermosilla [a], J.M. Díaz-Manso [b], L.A. Ruiz [a], J.A. Recio [a], A. Fernández-Sarría [a], P. Ferradáns-Nogueira [b]

[a] GeoEnvironmental Cartography and Remote Sensing Research Group.
Universitat Politècnica de València.
Camino de Vera s/n, 46022 Valencia, Spain

[b] Sociedade para o Desenvolvemento Comarcal de Galicia.
Estrada Santiago-Noia, km. 1. A Barcia s/n. Santiago de Compostela. Spain.

Corresponding author:
Txomin Hermosilla
txohergo@topo.upv.es
Telephone: 00 34 963877000 ext. 75576

ABSTRACT:
The abandonment of agricultural plots entails a low economic productivity of the land and a higher vulnerability to wildfires and degradation of affected areas. In this sense, the local government of Galicia is promoting new methodologies based on high resolution images in order to classify the territory in basic and generic land uses. This procedure will be used to control the sustainable management of plots belonging to the Land Bank. In this sense, this paper presents an application study for maintaining and updating land-use/land-cover geospatial databases using parcel-oriented classification. The test is performed over two geographic areas of Galicia, in the northwest of Spain. In this region, forest and shrublands in mountain environments are very heterogeneous, with many private unproductive plots, some of which are in a high state of abandonment. The dataset is made of high spatial resolution multispectral imagery, cadastral cartography, employed to define the image objects (plots), and field samples, which were used to define evaluation and training samples. A set of descriptive features is computed quantifying different properties of the objects, i.e. spectral, texture, structural, and geometrical. Additionally, it is tested the effect on the classification and updating processes of the historical land use as a descriptive feature. Three different classification methodologies are analyzed: linear discriminant analysis, decision trees and support vector machine. The overall accuracies of the classifications obtained are always above 90%, and support vector machine method is proved to provide the best performance. Forest and shrublands areas are especially undefined, so the discrimination between these two classes is low. The results enable to conclude that the use of automatic parcel-oriented classification techniques for updating tasks of land-use/land-cover geospatial databases, is effective in the areas tested, particularly when broad and well defined classes are required.

Keywords: Object-based classification, change detection, high-resolution imagery, mapping, agriculture.

1. Introduction

An efficient territory management and monitoring requires the regular use of land use/land cover (LU/LC) geospatial databases. The accuracy and the reliability of these databases is a crucial issue for decision-making tasks. The high dynamism of some geographic areas and the need of periodical updating of the information contained in the geospatial databases entail a high economic cost, which makes difficult to update the information with the appropriate frequency. In this sense, digital image classification techniques contribute to automate the processes of LU/LC geospatial database updating and could substantially reduce the field visits and the costs at production level, enabling to improve the updating frequency (El Kady and Mack, 1992; Dadhwal et al., 2002). Early studies were focused on pixel based analysis of the images (Homer et al., 2004). However, higher applicability of digital image processing techniques has been reached using object-based approaches, which require the definition of objects to divide the territory. An object is defined as a group of pixels with common characteristics created by means of a determined segmentation criterion (Blaschke, 2010). The segmentation method employed is key in the descriptive feature extraction process because the resultant objects will differ depending on the algorithm and parameters selected. Parcel-based image classification is a particular object-based classification case that uses cartographic limits to create the objects. This is a suitable methodology for anthropogenic environments such as rural areas, allowing to easily integrate remotely sensed data derived information with LU/LC geospatial databases. This technique has been widely employed in several works (Arikan, 2004; De Wit and Clevers, 2004; Walter, 2004; Peled and Gilichinsky, 2004; Ruiz et al, 2009). Although first approaches used medium spatial resolution multispectral imagery acquired with Landsat (Petit and Lambin, 2002) or ASTER (Perveen et al., 2008) platforms, or combining these images with other data such as SPOT (Cohen and Shoshany, 2000; Ormeci et al., 2010) or radar (Dupas, 2000, Del Frate et al., 2008), sensors with higher spectral and particularly spatial capabilities, such as IKONOS (Peled and Gilichinsky, 2010), QuickBird (Walsh et al., 2008), or airborne cameras (Walter, 2004; Tansey et al., 2009; Zaragozí et al., 2011) permitted to achieve more detailed analysis. Additionally, the integration with three-dimensional information acquired with aerial laser scanner systems increases the accuracy of the results (Walter, 2005; Hermosilla et al., 2010).

The periodic updating and quality maintenance of the information contained in LU/LC geospatial databases may support the control of the appearance of neglected lands. Abandoned lands reduce the economic productivity and increase the vulnerability to wildfires and degradation in the affected areas. In this sense, this paper presents the results of a preliminary study regarding to the suitability of the employment of parcel-based image classification techniques for LU/LC geospatial database updating in Galicia (Northwest of Spain). In this region, forest and shrublands in mountain environments are very heterogeneous, presenting many private unproductive plots, some of them in a high state of abandonment (Díaz-Manso and Ferradáns-Nogueira, 2011). The regional government is promoting new methodologies based in high resolution images in order to classify the territory in basic and generic land uses, with the goal of creating and maintaining a LU/LC geospatial database. This database will allow for planning specific actions for a sustainable management of neglected plots, including the creation of a Land Bank of Galicia (*Banco de Terras de Galicia*). This study aims to define and evaluate a productive methodology based on parcel-based classification of high-resolution images, for updating a generic LU/LC geospatial database. Object-based classification applied to update agricultural and forest plots can be focused on different thematic levels. In this case, generic classes would provide operative information for discriminating between productive and unproductive plots. Three different classification methods are compared: decision trees, linear discriminant analysis and support vector machine. Additionally, the effect on the classification and updating process of the historical land-use contained in the geospatial database is tested.

2. Material and methods

2.1. Study areas and data

The study has been performed on two local administrative areas (*comarcas*) of Galicia: Baixo Miño and A Limia (see Figure 1). The first one is located in the Atlantic coast of the province of Pontevedra, and is mainly covered by forest, agricultural crops and vineyards. The administrative area of A Limia is located in the province of Ourense and presents large areas of agricultural crops, forest and shrublands.

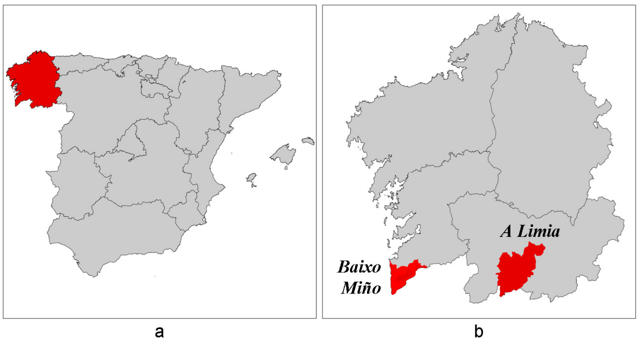

Fig. 1 Location of Galicia in Spain (a) and Location of the analyzed administrative regions (*comarcas*) in Galicia (b)

The images employed were available through the Spanish National Plan of Aerial Orthophotography (PNOA). These images have a spatial resolution of 0.25 m/pixel and 4 spectral bands: red (R), green (G), blue (B) and near infrared (NIR). The images of A Limia were acquired between May and July of 2007, and those of Baixo Miño during the same months of 2008. The imagery was already ortho-rectified and geo-referenced, panchromatic and multispectral bands fused, mosaicking and radiometric adjustments applied, as a part of the PNOA programme. After preliminary tests, and considering the classes used in this study, 0.5 m/pixel spatial resolution proved as good performance as 0.25 m/pixel. Consequently images were resampled by using bilinear interpolation in order to facilitate and to improve the operativeness of the descriptive feature extraction process.

Cartographic boundaries to define the final objects (plots) were obtained from the National Land Parcel Information System (SIGPAC), a geospatial database oriented to agriculture management. The plots represent a continuous area of land within a plot for a single agricultural use, being the total number of plots 468,721 in A Limia and 255,347 in Baixo Miño. Additionally, field sampling segments collected in the same date that the images employed for each region were available. These segments have square shape with side sizes of 350 or 500 meters. Figure 2 shows the distribution of field samples in both areas. In order to confer coherence to the automatic feature extraction process, which requires a minimum plot surface, plots with a surface lower than 60 m^2 were rejected from the tests. Besides, SIGPAC plots with very large dimensions were also excluded (representing images with more than 9,000,000 pixels), due to the RAM memory limitations for processing the per plot feature extraction algorithms.

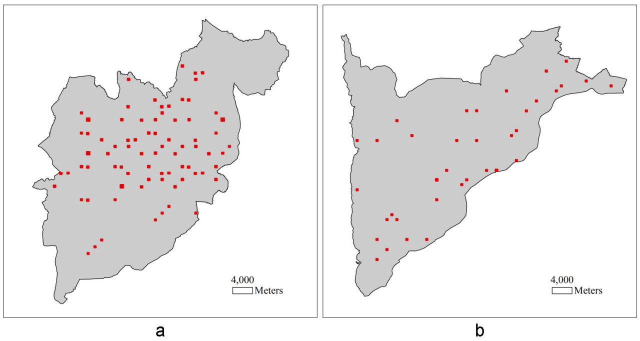

Fig. 2 Distribution of field sampling data in A Limia (a) and Baixo Miño(b)

A number of generic classes for both regions were defined: *Buildings*, *Forest*, *Shrublands* and *Arable and crop lands*. Due to the definition accuracy presented by the classes regarding to roads and rivers in the SIGPAC, these classes were transferred directly from the geospatial database. *Arable and crop lands* class includes also pastures, being built up by aggregation of three sub-classes, differencing the vegetation level of a plot: no vegetation, medium vegetation and cultivated field. Besides, some additional classes were defined in order to adapt the legend to the particular conditions of each region. Thus, a *Water* layer class was defined in A Limia to classify new flooded areas not registered in the SIGPAC database. In Baixo Miño, the additional *Vineyards* and *Greenhouse* classes were defined. Examples of the defined classes are shown in Figure 3. In the region of A Limia, most of the training samples were selected from the register of field segments available from the SIGPAC project, being adequate in number and spatial distribution (Figure 2a). Since the sampling polygons did not coincide with the SIGPAC plots limits (Figure 4), the assignation of samples to each class was manually done. Additional samples were added by photointerpretation in order to avoid a low representation of some classes, particularly *Water* layers, *Forest* and *Shrublands*. As in the region of Baixo Miño, the number of field samples was substantially lower (see Figure 2.b), the training samples were mainly selected using photointerpretation techniques, and using the field registers as ancillary data.

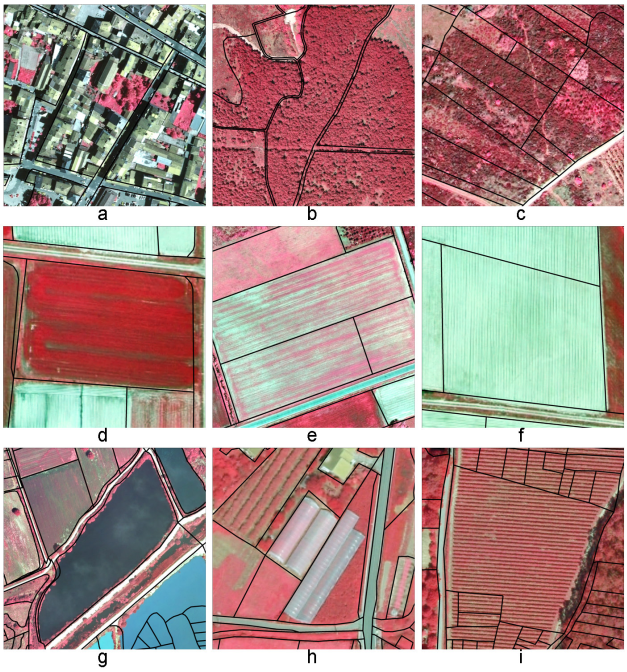

Fig. 3 Examples of the considered classes: Buildings (a); *Forest* (b); *Shrublands* (c); *Arable and crop lands*: cultivated field (d), medium vegetation (e) and no vegetation (f); *Water* (g); *Greenhouse* (h); and *Vineyards* (i)

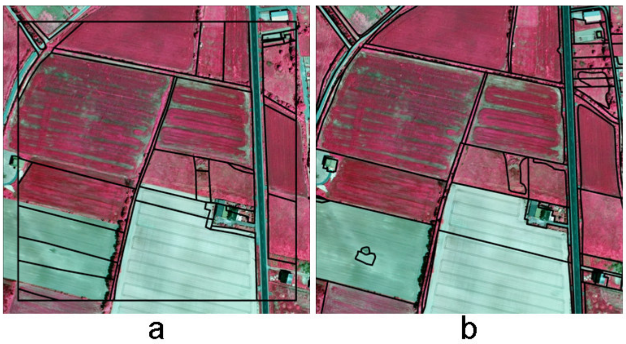

Fig 4 Example of the geometric differences between field sampling polygons inside a segment (a) and SIGPAC plots (b)

2.2. Descriptive feature extraction

The use of efficient features is essential for accurate classification. At this point, every plot was independently processed to extract descriptive features characterising the current land use. The features were computed using the object-based description software FETEX 2.0 (Ruiz et al., 2011). Four feature categories were defined: spectral, textural, structural and shape features. Additionally, the effect of the use of ancillary data by considering the historical land-use was tested.

Spectral features provide information about the spectral response of objects on the visible and near infrared regions of the electromagnetic spectrum, which is related to land cover types, state of vegetation, soil composition, construction materials, etc. These features are particularly useful in the characterization of spectrally homogeneous objects, such as herbaceous crops or fallow fields. Mean and standard deviation were computed as features from bands NIR, R, G, and also from the Normalized Difference Vegetation Index (NDVI).

Texture features inform about the spatial distribution of the intensity values in the image, being useful to quantify properties such as heterogeneity, contrast or uniformity related to each object (Ruiz et al., 2004). These properties are clearly related to the LU/LC inside an object. For every object, the following features proposed by Haralick et al. (1973) based on the grey level co-occurence matrix (GLCM) were computed: contrast, uniformity, entropy, variance, covariance, inverse difference moment, and correlation. This information was completed with the kurtosis and skewness, which depict the first order distribution of values of the histogram of an object, and the mean and the standard deviation of the *edgeness factor* for each plot (Laws, 1985). The edgeness factor represents the density of edges in a neighbourhood. These features were computed from the red band.

Structural features were derived form the semivariogram. The semivariogram curve quantifies the spatial associations of the values of a variable, and measures the degree of spatial correlation between different pixels in an image. This is a particularly suitable tool in the characterization of regular patterns. The multidirectional semivariogram representing each object is obtained by computing the mean of the semivariograms calculated in six directions, ranging from 0º to 150º with a step of 30º (Balaguer-Beser et al., 2011). Afterwards, each semivariogram curve is filtered using a Gaussian kernel with a stencil of 3 positions, in order to smooth its shape and to reduce experimental fluctuations. Several structural descriptive features were computed considering the singular points of the semivariogram, such as the first maximum, the first minimum, the second maximum, etc., all being described in detail in Balaguer et al. (2010).

Shape features inform about the complexity in the shape of the objects. They contribute to differentiate polygons with specific shapes. Several standard features were extracted for each object: compactness, shape index, fractal dimension, area and perimeter.

Finally, the historical land use, contained in the SIGPAC geospatial database, was included as a qualitative descriptive feature in order to evaluate its effect in the classification and change detection steps.

2.3. Classification methods

In order to optimize the process and to analyze the performance of different classification methods, three classification tests were carried out over the data sets from the two areas, based on decision trees, linear discriminant analysis (LDA), and support vector machine (SVM).

2.3.1. Decision trees

A decision tree is a set of organized conditions in a hierarchical structure, in such a way that the class assigned to an object can be determined following the conditions that are fulfilled from the tree roots (the initial data set) to any of its leaves (the assigned class). The algorithm employed in this study is the C5.0 (Quinlan, 1993), widely used to deduce decision trees for classifying images (Zhang and Liu, 2005).

The process of building a decision tree begins by dividing the collection of training samples using mutually exclusive conditions. Each of these sample subgroups is iteratively divided until the newly generated subgroups are homogeneous, i.e., all the elements in a subgroup belong to the same class. These algorithms are based on searching partitions to obtain purer data subgroups, which are less mixed than the previous group where these come from. For each possible division of the initial data group, the impurity degree of the new subgroups is computed, and the condition which gives the lower impurity degree is chosen. This is iterated until the division of the original data into homogeneous subgroups is carried out by using the gain ratio as splitting criterion. This criterion employs information theory to estimate the size of the sub-trees for each possible attribute and selects the attribute with the largest expected information gain, that is, the attribute that will result in the smallest expected size of the sub-trees.

Objects were classified using 10 decision trees, by means of the boosting multi-classifier method, which allows for increasing the accuracy of the classifier. The methodology followed by the boosting to build the multi-classifier is based on the assignment of weights to training samples. After each tree construction, the vector of weights is adjusted according to the model performance. In this way, samples erroneously classified increase their weights, whereas the weights of correctly classified samples decrease. Thus, the model obtained in the next iteration will give more relevance to the samples erroneously classified in the previous step (Hernandez-Orallo et al., 2004). After the construction of the decision tree set, a class is assigned to each object considering the estimated error in the construction of each tree.

2.3.2. Linear discriminant analysis

LDA is a multivariate statistical technique that has been described as similar to multiple regression analysis where the dependent variable is categorical, and the independent variables are continuous and are used to determine the class to which the objects belong (Huberty, 1994; Everitt and Dunn, 2001). The objective is to find the linear relationships between the continuous variables that better discriminate between the defined groups.

The criterion to decide the variables entering into the definition of the discriminant function is to minimize the value of the Wilk's Lambda that is defined as the quotient between the variance inside the groups (classes) and the overall variance. The overall variance of a variable can be expressed as the sum of the variance inside the groups and the variance between groups. Thus, the value of the quotient defined by Wilks' Lambda represents the proportion of the overall variability due to the differences occurred inside the groups. According to this interpretation, a variable is chosen when it minimizes the Wilks' Lambda value, i.e., the observed differences between the values of the discriminant functions are caused by the differences between groups, not by internal differences.

After defining the acceptance criterion, the variable selection is done stepwise. Firstly, the most discriminative variable is selected by independently assessing each variable by using an *F* test, which in this case is a function of Wilks' Lambda value (Λ) as is described in the expression:

$$F = \frac{(n-g)}{g-1} \cdot \left(\frac{1-\Lambda}{\Lambda} \right)$$

where *n* is the number of elements (objects) and *g* the number of groups (classes). The fist variable selected is the one minimising the Wilks' Lambda value, i.e., the one providing the highest and significant *F* value. In the second step, the selected variable is combined with each of the remaining variables, choosing the most discriminative (highest significant value of *F*). In this case, as a variable is added into a model were there are already other features, the F value is computed according to the expression:

$$F = \frac{(n-g-s)}{g-1} \cdot \left(\frac{1-\frac{\Lambda_{s+1}}{\Lambda_s}}{\frac{\Lambda_{s+1}}{\Lambda_s}} \right)$$

where *s* is the number of variables in the model, and Λ_s and Λ_{s+1} are Wilks' Lambda values before and after the addition of the new variable, respectively. *F* represents the increment in the discrimination after the addition of the new variable, with respect to the total value obtained with the variables previously included in the model. Subsequent variables are selected in a similar manner. The variable addition process stops when no variable out of the model fulfil the acceptance criterion.

2.3.3. Support vector machine

The support vector machine is a set of supervised learning methods used for classification and regression analysis developed by Vapnik (1998). They model non-linear class boundaries in high dimensional feature spaces by means of kernel functions and regularization. For classification, according to van der Linden et al. (2010), SVM delineate two classes with an optimal separating hyperplane to the training data in the multidimensional feature space. A good separation is achieved by the hyperplane with the largest distance to the nearest training data points of any class. For linearly not separable classes, the original finite-dimensional space is mapped into a much higher-dimensional space by a kernel function, where the new data distribution enables a better fitting of a linear hyperplane. The parameterization of a support vector classifier requires the selection of the kernel function parameter and a regularization parameter. To define the optimal separating hyperplane is necessary to solve a quadratic optimization problem. As a result, one vector of weights is obtained for each training data vector. Only the training data vectors with non-zero weight are needed to define the optimal separating hyperplane, i.e. the support vectors. In remote sensing applications, the Gaussian radial basis function kernel proved to be effective:

$$K(x, x_i) = \exp(-g|x - x_i|^2)$$

where *g* is the width of the Gaussian kernel function. A regularization parameter (C) controls the trade-off between the maximization of the margin between the training data vectors and the margin errors; limiting the influence of individual training data vectors. To search for adequate values of *g* and C a two-dimensional grid search with internal validation was done.

SVM is only directly applicable for two-class problems. To be applied in multi-class problems is necessary to reduce the multiclass problem into multiple binary classification problems. The one-versus-one approach was used. In this approach, several SVM classifiers are trained for separating all possible class pairs. To assign the multi-class decision, the values of the decision functions of the support vector classifier were transformed into binary probabilities and class probabilities were estimated by pairwise coupling. Then, the class with the highest probability was selected. These procedures were carried out with the *imageSVM* software (Rabe at al., 2010).

2.4. Accuracy assessment

The accuracy of the different classification tests was assessed using leave-one-out cross-validation, in order to maximize the efficiency in the number of samples. This method uses a single observation from the original sample set as validation data, and the remaining observations as training data, iterating until each observation in the sample set is used once as validation data. From the confusion matrix, the user's and producer's accuracies per class were computed, that respectively measure the commission and omission errors (Congalton, 1991). In addition, a specific confusion index was defined to quantify the particular confusion between a pair of classes. This index is computed as the sum of their mutual errors divided by the total objects from that pair of classes. Confusion index value ranges from 0 (absence of per-class-pair errors) to 1 (all the objects of both classes considered are misclassified).

In a standard process of LU/LC geospatial database updating, the class assigned to a parcel in the classification process is compared to the land use contained in the original database. As the land uses defined in the database were different to the employed in the classification, they were grouped to produce a direct correspondence between land use types, allowing for their comparison and detection of possible changes. The differences between them register the potential LU/LC changes produced in the territory, but also the errors produced in the classification. In the updating process, correctly classified cases can be divided in two categories: coincidences and detected changes. Coincidences occur when equal land use is assigned in the classification, reference data and database. A detected change occurs when the classification land use is correctly assigned meanwhile the land use appearing in the database is wrong. The sum of the percentage of coincidences and detected changes is equal to the overall accuracy of the classification. Updating errors can be divided in two respective categories: detectable and undetectable errors. A detectable error is produced when a mistaken land use is assigned in the classification and differs to the one which is contained in the database. An undetectable error occurs when the land use assigned in the classification process and the one contained in the database are the same but incorrect. The accumulation of detected changes and detectable errors composes the number of plots to review in the updating process (Recio et al., 2011). The effect of the addition of historical land-use as a feature in the classification was tested using decision trees as classifier, since this method can directly handle discrete variables as descriptive features, without requiring the use of other strategies such as dummy variable definition.

3. Results and discussion

3.1. Classification analysis

The classification overall accuracy values obtained considering the different methods are shown in Table 1. In both study areas the use of SVM provided the highest values, meanwhile decision trees and LDA methods reached similar overall accuracies.

Table 1 Overall accuracies obtained using different classification methods

Method	A Limia	Baixo Miño
Decision trees	91.4	89.2
LDA	88.2	91.6
SVM	95.0	96.3

The analysis of the per-class user's and producer's accuracies shows that in A Limia (Figure 5) and in Baixo Miño (Figure 6) LDA and decision trees classification methods, respectively, produced unbalanced values for certain classes. Generally, in both study areas, the highest error rate occurred between the classes *Arable and crop lands* with *Shrublands*, and *Shrublands* with *Forest* due to the similar spectral and textural responses of *Shrublands* and *Forest*. Additionally, in Baixo Miño main confusions were produced between *Greenhouse* with *Buildings* and *Arable and crop lands* classes. The use of SVM noticeably mitigate the errors given in the classes, and enable to reach user's and producer's accuracies higher than 90% in most of the classes.

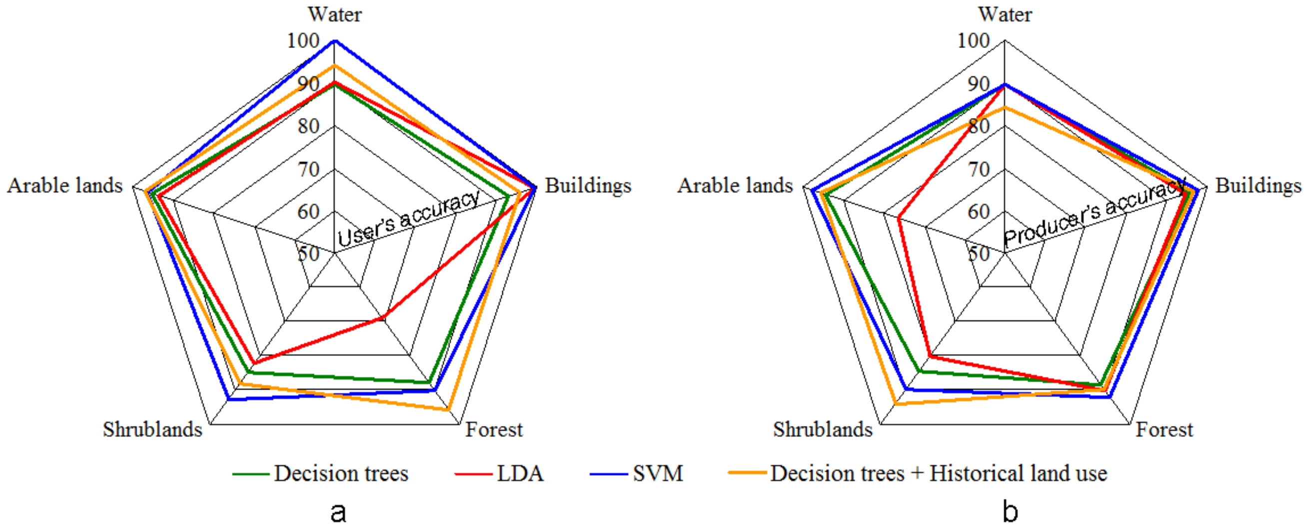

Fig. 5 Comparison of A Limia per-class user's (a) and producer's (b) accuracies using different classification methods

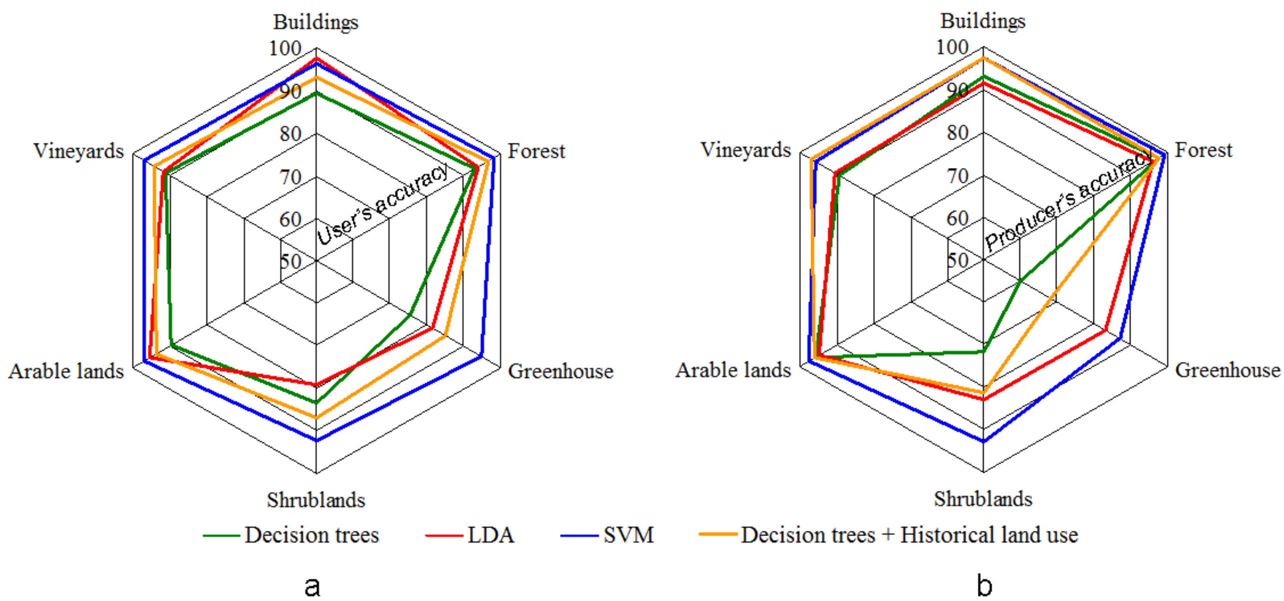

Fig 6 Comparison of Baixo Miño per-class user's (a) and producer's (b) accuracies using different classification methods

Many of the errors detected especially occurred between classes *Shrublands* and *Forest*. Since there was a particular interest to discriminate between these two classes in order to improve the detection of neglected plots, a deeper analysis of these errors was done by means of the mutual confusion index. The values reached employing the different classification methods for the classes *Shrublands* and *Forest* are shown in Figure 7. In both study areas the error pattern was similar, existing a higher confusion rate when LDA method was used. With this methodology, the error rate between both classes was especially remarkable in the study area of A Limia. Decision trees and, especially, SVM markedly produced a decrease of this confusion. Figure 8 shows details of the thematic representation of the classification in both study areas.

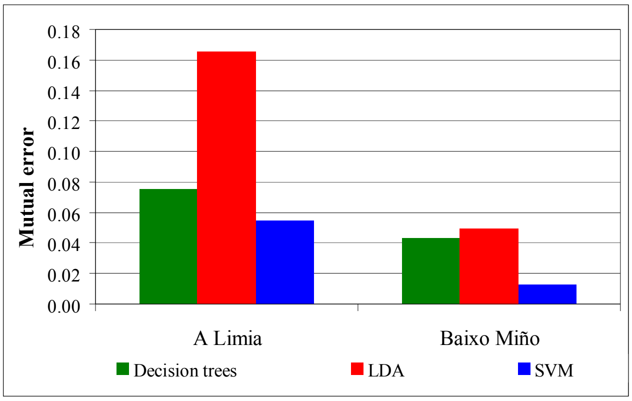

Fig 7 Comparison of mutual errors for the classes *Shrublands* and *Forest* obtained using different classification methods

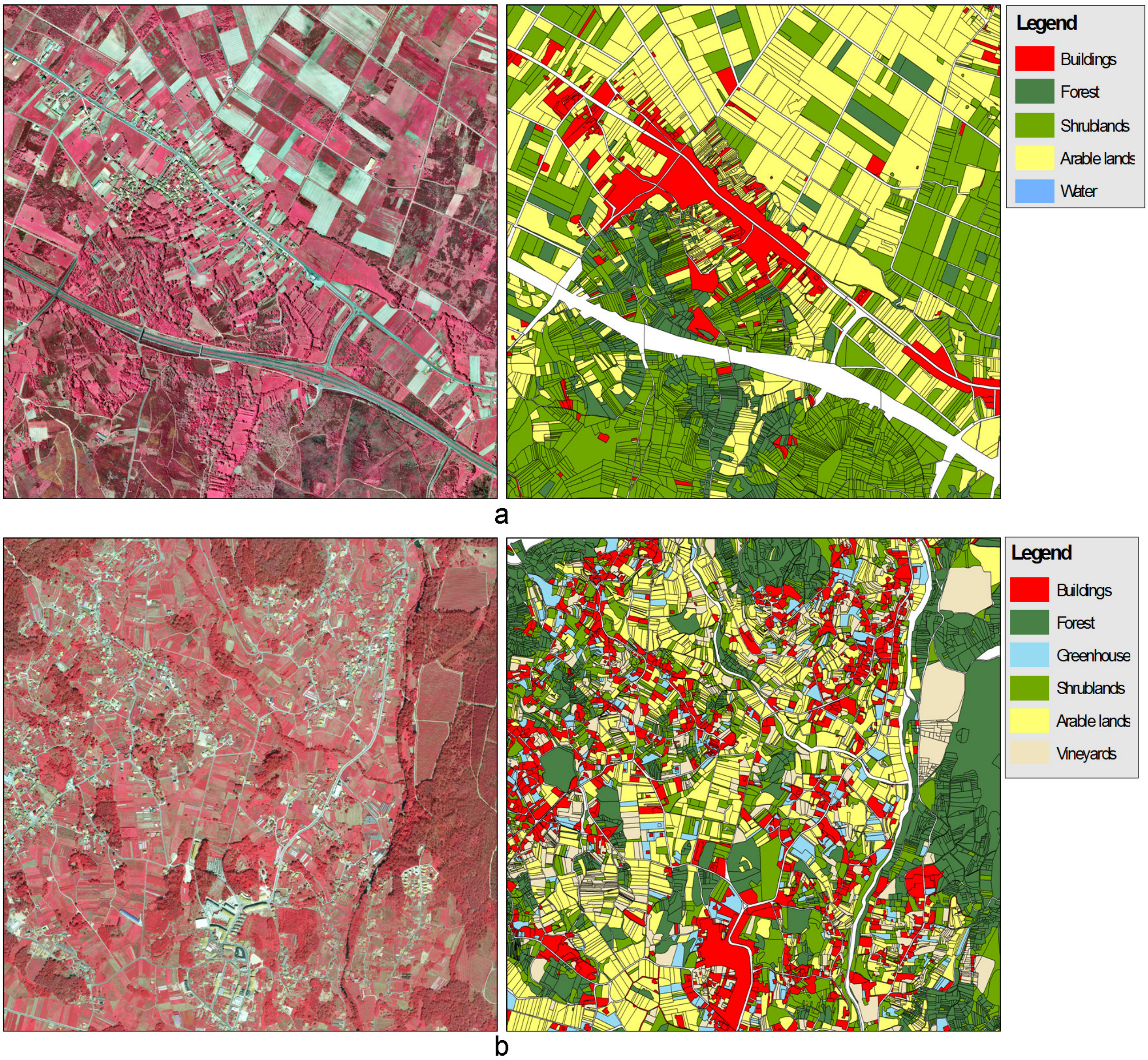

Fig. 8 Details of the thematic representation of the classifications in A Limia (a) and Baixo Miño (b)

In general, a fact that introduce errors in the classification process in both areas of study was related with the size and shape of the plots. Some polygons, particularly in Baixo Miño, presented extremely long and narrow shapes

(Figure 9.a) or very small area, making difficult the extraction of meaningful descriptive features. In contrast, plots with large dimensions normally presented mixed land uses (Figure 9.b). A possible alternative to face this problem is to apply segmentation algorithms and then classify the generated sub-objects. Similarly, since the analyzed regions were mainly rural areas, some built-up zones were contained in plots mixed with vegetation (Figure 9.c). In this sense, the introduction of a post-processing step to filter or control improbable changes could reduce this type of errors, improving the classification accuracy of the classes involved. Other problems were related to the visual interpretation of the classes. In the training sample selection process, partially done through photointerpretation, the visual discrimination between forest and shrubland was sometimes difficult (Figure 9.d). Besides, the class *Shrublands* presents a significant internal heterogeneity in the area of Baixo Miño. Finally, since the regions of study were significantly large, those LC/LU classes having a very low representation were not considered in the classification legend.

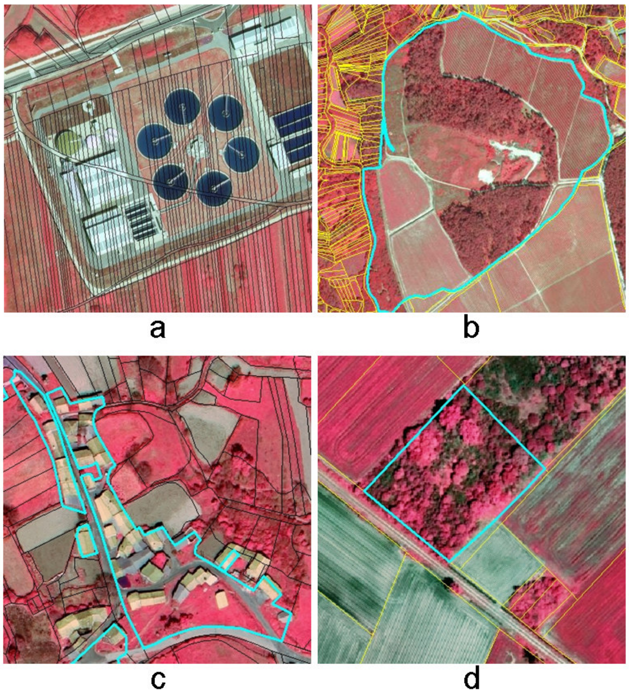

Fig. 9 Examples of cases that difficult a correct classification: narrow plots (a); mixed land-use (b); mixed built-up plots (c); mixed forest and shrublands (d)

3.2. Effect of using historical land-use as a descriptive feature

When the historical land use contained in the SIGPAC geospatial database is added as descriptive feature in the classification, the overall accuracy values experiment an increase of about 3% in both study areas (see Table 2). Consequently, the producer's and user's accuracies of the classes presenting higher confusion degree increased. The highest number of errors was still produced between *Arable and crop lands* and *Shrublands*, and *Shrublands* and *Forest*.

Table 2 Comparison of overall accuracies with and without considering the historical land-use as a descriptive feature

Features	A Limia	Baixo Miño
Image derived	91.4	89.2
Image derived + Historical land use	94.3	92.9

The specific distribution of cases produced in the updating process for the study area of A Limia is shown in Figure 10. Assuming a thematic updating of land use contained in the SIGPAC, the detected changes without historical land-use was 5% versus 3.3% considering this information as a feature. Detectable errors without using historical land-use represented 6.2% against 1.4% when it was used. The addition of the historical land-use produced that the number of plots to be revised, i.e. possible detected changes (sum of detected changes and detectable errors) was reduced to 4.7%, against 11.2% without using it, but also the proportion of detected changes was also notably reduced. In addition, the historical land-use consideration produced that the undetectable errors of change detection process were practically doubled (from 2.4% to 4.3%). This is caused by the significance acquired by the historical land use feature in the rule creation process, due to the presence of a large number of training samples with same historical and actual land uses. This leads to the classification of a significant number of plots with changes attending to their historical land use. Consequently, these parcels become undetectable errors since no change is detected when comparing the classification result to the outdated database used to extract the old use as descriptive feature. This means that even when the use of the historical land use as descriptive feature produces an actual improvement of the classification accuracy, in a LU/LC geospatial database updating process the undetectable errors are noticeably increased.

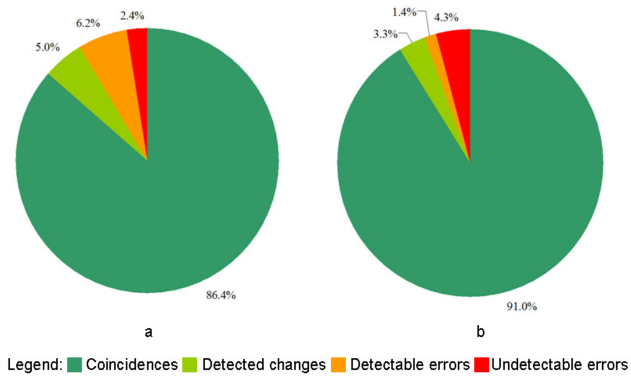

Fig. 10 Distribution of coincidences, changes and errors in the study area of A Limia without considering (a) and considering (b) the historical land use contained in the SIGPAC geospatial database as descriptive feature

4. Conclusions

A methodology for generic LU/LC geospatial database creation and updating based on a plot-based classification approach using high resolution multispectral images is presented and analysed. The classification process is based on the combination of several descriptive features derived from images, plot shape and historical land-use information. The analysis has been focused on detecting neglected use of agricultural and forest plots, in order to control and monitor the correct use of the plots. The *Land Bank of Galicia* intermediates between owners and potential users for a productive reutilization of the land. This information will be useful to define cost-effective control procedures that minimise field visit tasks and allowing for a systematic monitoring of the territory.

The highest overall accuracy values were reached using SVM classification method. This methodology enabled to obtain the most balanced per-class accuracies. Main classification errors were given in the discrimination between forest and shrublands areas, because of the complexity and mix of these cover types in the landscape. In fact, some problems were found to differentiate between these classes by means of photointerpretation.

The addition of the historical land use contained in the LU/LC geospatial database to be updated as a descriptive feature produced and improvement of the classification accuracy, reducing the errors, but also significantly increasing the undetectable errors, which is not advisable for the processes of LULC geospatial database updating. In general terms, the results obtained in this study showed a high capability of the object-based image classification techniques as a supporting tool for updating and managing LC/LU information.

Acknowledgements

The authors appreciate the collaboration and support provided by *Xunta de Galicia*, *Sociedade para o Desenvolvemento Comarcal de Galícia* and *Banco de Terras de Galicia*. The financial support provided by the Spanish *Ministerio de Ciencia e Innovación* in the framework of the projects CGL2010-19591/BTE and CGL2009-14220 is also acknowledged.

References

Arikan, M (2004) Parcel-based crop mapping through multi-temporal masking classification of landsat 7 images in Karacabey, Turkey. Int Arch Photogramm Remote Sens Spat Inf Sci 35:1085-1090.

Balaguer A, Ruiz LA, Hermosilla T, Recio JA (2010) Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. Comput Geosci 36(2):231-240

Balaguer-Besser, A, Hermosilla, T, Recio, JA, Ruiz, LA (2011) Semivariogram calculation optimization for object-oriented image classification. Modelling in Science Education and Learning 4(7):91-104.

Blaschke, T (2010) Object based image analysis for remote sensing. ISPRS J. Photogramm. 65(1):2-16

Cohen Y, Shoshany M (2000) Integration of remote sensing, GIS and expert knowledge in national knowledge-based crop recognition in Mediterranean environment. Int Arch Photogramm Remote Sens 33(Part B7):280-286

Congalton R (1991) A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ 37(1):35–46

Dadhwal, V.K., Singh, R.P., Dutta, S., Parihar, J.S., 2002. Remote sensing based crop inventory: A review of Indian experience. Trop Ecol 43(1):107-122.

De Wit AJW, Clevers JGPW (2004) Efficiency and accuracy of per-field classification for operational crop mapping. Int J Remote Sens 25:4091-4112

Del Frate F, Pacifici F, Solimini D (2008) Monitoring Urban Land Cover in Rome, Italy, and Its Changes by Single-Polarization Multitemporal SAR Images. IEEE J Sel Top Appl Earth Obs Remote Sens 1:87-97

Díaz-Manso, JM, Ferradáns-Nogueira, P (2011) Modelo de uso actual da terra. In: Cobelle-Rico, EJ, Diaz-Manso, JM, Crecente-Maseda, R, Martínez-Rivas, EM (eds) Mercado e Mobilidade de Terras en Galícia, 1st edn. Servizo de Publicacións e Intercambio Científico, Santiago de Compostela, Spain, pp 31-44

Dupas CA (2000) SAR and LANDSAT TM image fusion for land cover classification in the Brazilian Atlantic Forest Domain. Int Arch Photogramm Remote Sens XXXIII(Part B1):96-103

El Kady, M, Mack, CB (1992) Remote sensing for crop inventory of Egypt's old agricultural lands. Int Arch Photogramm Remote Sens 29:176-185

Everitt BS, Dunn G (2001) Applied multivariate data analysis. 2nd ed. Edward Arnold, London

Haralick RM, Shanmugam K, Dinstein I (1973) Texture features for image classification. IEEE Transact Syst Man Cybern 3(6):610-622

Hermosilla T, Almonacid J, Fernández-Sarría A, Ruiz LA, Recio JA (2010) Combining features extracted from imagery and lidar data for object-oriented classification of forest areas. Int Arch Photogramm Remote Sens Spat Inf Sci 38(4/C7)

Hernández Orallo J, Ramírez Quintana MJ, Ferri Ramírez C (2004). Introducción a la minería de datos. Pearson Educación S.A., Madrid

Homer C, Huang C, Yang L, Wylie B, Coan M (2004) Development of a 2001 National Land-Cover Database for the United States. Photogramm Eng Remote Sens 70:829-840

Huberty CJ (1994) Applied Discriminant Analysis. Wiley, New York

Laws KI (1985) Goal-directed texture image segmentation. Appl Artif Intel II, SPIE 548:19-26

Ormeci, C., Alganci, U., and Sertel, E., Identification of Crop Areas Using SPOT – 5 Data, FIG Congress 2010 Facing the Challenges – Building the Capacity, Sydney, Australia, 11-16 April 2010

Peled A, Gilichinsky M (2010) Knowledge-Based Classification of Land Cover for the Quality Assessement of GIS Database. Int Arch Photogramm Remote Sens Spat Inf Sci 38:217-222

Peled A, Gilichinsky M (2004) GIS-driven analyses of remotely sensed data for quality assessment of existing land cover classification. Int Arch Photogramm Remote Sens Spat Inf Sci 35

Perveen, F., Nagasawa, R., Ali, S., Husnain, 2008. Evaluation of ASTER spectral bands for agricultural land cover mapping using pixel-based and object-based classification approaches. Int Arch Photogramm Remote Sens Spat Inf Sci 37(4-C1)

Petit CC, Lambin EF (2002) Impact of data integration technique on historical land-use/land-cover change: Comparing historical maps with remote sensing data in the Belgian Ardennes. Landsc Ecol 17:117-132

Quinlan JR (1993) C4.5: Programs for machine learning. Morgan Kaufmann Publishing, San Francisco

Rabe A, van der Linden S, Hostert P (2010) imageSVM, Version 2.1, software available at www.hu-geomatics.de

Recio JA, Hermosilla T, Ruiz LA, Fernández-Sarría A (2011) Historical land use as a feature for image classification. Photogramm Eng Remote Sens 77(4):377-387

Ruiz LA, Fernández-Sarría A, Recio JA (2004) Texture feature extraction for classification of remote sensing data using wavelet decomposition: A comparative study. Int Arch Photogramm Remote Sens Spat Inf Sci 35(B4):1109-1115

Ruiz LA, Recio JA, Hermosilla T, Fdez. Sarriá A (2009) Identification of Agricultural and Land Cover Database Changes Using Object-oriented Classification Techniques. 33rd International Symposium on Remote Sensing of Environment, May 4 - 8, Stresa (Italy)

Ruiz LA, Recio JA, Fernández-Sarría A, Hermosilla T (2011) A feature extraction software tool for agricultural object-based image analysis. Comput Electron Agric 76(4):284-296

Tansey, K, Chambers, I, Anstee, A, Denniss, A, Lamb, A (2009) Object-oriented classification of very high resolution airborne imagery for the extraction of hedgerows and field margin cover in agricultural areas. Appl Geogr 29(2): 145-157

van der Linden S, Rabe A, Wirth F, Suess S, Okujeni A, Hostert P (2010) imageSVM regression, Application Manual: imageSVM version 2.1. Humboldt-Universität zu Berlin, Germany

Vapnik VN (1998) Statistical Learning Theory. Wiley, New York

Walter V (2005) Object-based evaluation of lidar and multiespectral data for automatic change detection in GIS databases. Geo-Inf Syst 18:10-15

Walter V (2004) Object-based classification of remote sensing data for change detection. ISPRS J Photogramm Remote Sens 58:225-238

Walsh, S J, McCleary, A L, Mena, C F, Shao, Y, Tuttle, J P, Gonzalez, A, Atkinson, R (2008) QuickBird and Hyperion data analysis of an invasive plant species in the Galapagos Islands of Ecuador: implications for control and land use management. Remote Sens Environ 112(5):1927-1941

Zaragozí, BM, Navarro, JT, Ramón, A (2011) A study of drivers for agricultural land abandonment using GIS and Data Mining techniques. Eighth International Conference on Ecosystems and Sustainable Development, April 13 – 15, Alicante, Spain, pp.363-374.

Zhang S, Liu X (2005) Realization of Data Mining Model for Expert Classification Using Multi-Scale Spatial Data. Int Arch Photogramm Remote Sens Spat Inf Sci 26(4/W6):107-111