

## DECISION TREES ON LIDAR TO CLASSIFY LAND USES AND COVERS

Jorge Garcia-Gutierrez<sup>a</sup>, Luis Gonçalves-Seco<sup>b</sup>, Jose C. Riquelme-Santos<sup>a</sup>

<sup>a</sup>School of Computer Engineer, University of Seville  
Reina Mercedes s/n, 41012 Seville, Spain  
jgarcia@lsi.us.es, riquelme@us.es

<sup>b</sup>Faculty of Sciences, University of Porto  
Rua Campo Alegre 687, 4169-007 Porto, Portugal  
lgs@fc.up.pt

**KEY WORDS:** LIDAR, point cloud, unsupervised classification, land covers, C4.5, decision trees.

### ABSTRACT:

The area of Huelva, in the South of Spain, is a well-known case of human pressure on the natural environment. In Huelva, National Parks, like Doñana, and industrial and tourist zones coexist in difficult balance. The Regional Ministry of Andalusia is commissioned to assure the preservation of the natural resources in this part of Spain although its cost can be high in time and money. Remote sensing is a very suitable tool to carry out this task and automatic land use and cover detection can be a key factor to reduce costs. In addition, Light Detection and Ranging (LIDAR) has the advantage of being able to create elevation surfaces that are in 3D, while also having information on LIDAR intensity values. Many measures based on its intensity, density and its capacity for describing third dimension have been used previously with other purposes and outstanding results. In this paper, a new approach to identify land cover at high resolution is proposed selecting the most interesting attributes from a set of LIDAR measures. Our approach is based on data mining principles to take advantage on intelligent techniques (attribute selection and C4.5 algorithm decision tree) to classify quickly and efficiently without the need for manipulating multispectral images. Seven types of land cover have been classified in a very interesting zone at the mouth of the River Tinto and Odiel with results of accuracy between 71% and 100%. An overall accuracy of 85% has been reached for a resolution of 4 m<sup>2</sup>.

### 1 INTRODUCTION

Andalusia is the most populated and the second largest region of Spain. It is located in the South and is well-known because the quality of its coasts and even more because of its culture. Tourism has mainly been supporting economically several processes described by the Regional Ministry of Andalusia as modernization. With these processes, the government has tried to progress from an agrarian society helped by important tourism structures to an industrial society in less than 30 years. But these processes have their own dangers. One of the most important is the possibility of environmental damages.

The coast of Huelva is a very clear example in which the dualism between industrial progress and environment protection coexists in small space. On the one hand, large areas dedicated to industrial uses can be found, like for example: refineries, coal deposits... and on the other hand, protected zones like National Park of Doñana, where highly threatened species like Iberian lynx fight to survive. Human pressure on natural environments is a worrying task that has to be solved by Regional Ministries in Spain. Remote Sensing can be a valuable tool to automatize and speed up large area controlling by means of land use classification techniques.

Since its emergence, remote sensing has been used with different purposes related to natural resources. Lately, authors have used remote sensing techniques to monitor species and changes in cities (Gamanya et al., 2009), measure different environmental variables related to gas emission or fire severity in woods (Schneider et al., 2009), detect kinds of special soil (Hughes et al., 2009)... In addition, land covers and uses are studied profusely to manage zones especially interesting from an economic or natural point of view. In this cases, planning and managing play an important role to exploit their resources which can be seen in several important studies (McColl and Aggett, 2007) (Dorigo et al., 2007)...

LIDAR has become an excellent tool to improve remote sensing results. Its capacity for 3D description helps users to overcome traditional limits of remote sensing. It gives the third dimension to distinguish between the floor and the top of the objects using and developing DTM's (Digital Terrain Models). Many approaches have been proposed to DTM creation and a deep study of a set of them and their accuracy can be seen in some very important studies (Sithole and G.Vosselman, 2003). Moreover, laser is not affected by shadows and the problems they produce in traditional image-based remote sensing, though it has to be calibrated like others data sources. All this advantages, beside a progressive descent on costs in opposition to other data sources like hyperspectral images, have made LIDAR be one of the leading technologies in environmental researching.

In accordance with the proved utility of LIDAR, many researchers have used it as a supporting technology for traditional imagery betting on fusion of different sensors to improve results (Schubert et al., 2008) (Arroyo et al., 2008) while others focus their efforts in LIDAR as an only source data (Pascual et al., 2008) (Chust et al., 2008) with excellent results. Each strategy has its own pros and cons. While fusion gives big amounts of data which can produce extra quality classification, it also needs extra work to adapt data from multiple sensors and increases development and testing time. Moreover, some studies show very little improve on LIDAR classification results when fusion with others sensors is used to sort out some kind of tasks (Goetz et al., 2007) (Jensen et al., 2008). Due to these differences, further research is needed in this field.

In recent times, object-oriented techniques have been applied to LIDAR as an only data source to solve several tasks with outstanding results (Antonarakis et al., 2008) (Pascual et al., 2008). These techniques are mainly based on computer vision segmentation using a set of measures from LIDAR data. Then, classification method tries to learn from segmented objects in order to

classify the future data. Despite the fact that results with these approaches are highly interesting, segmentation is not an easy work, and there is all a research line in computer vision dedicated to this kind of problems. Actually, proprietary software like eCognition is usually responsible for the data segmentation and approaches usually work with LIDAR reflectivity as one of the main parameters to segment whole data but it is well-known that intensity is affected by several factors (Hofle and Pfeifer, 2007) like angle of incidence, distance from sensor to object... Other researchers used eCognition working with heights and adjusting segmentation parameters depending on the situation, but this is hardly to automate solution. In this context, a traditional approach based on pixels and working with models resulted from advanced intelligent techniques can be applied with good results. Speaking of that, intelligent techniques from world of data mining (Witten and Frank, 2005) have shown good results in order to solve problems related to LIDAR environment. Data mining algorithms usually extract its potential to classify or predict from machine learning techniques and they can be applied to LIDAR data without much effort. As a result, we can find several approaches based on different techniques like support vector machines (Koetz et al., 2008), neuronal networks (Brzank et al., 2008) (Canty, 2008), clustering (Pascual et al., 2008), nearest-neighbours algorithms (Magnussen et al., 2009), and finally decision trees (Tooke et al., 2008).

With this in mind, this work shows a new application of intelligent techniques in order to extract knowledge that is hidden in LIDAR data to be applied to natural and urban zones. And more specifically to:

- Define a general method based on decision trees to classify LIDAR as an only data source in different land uses and covers.
- Quantify urban and industrial advance in a zone with mixed land uses and covers in order to define a process to monitor the industrial activities and avoid possible damage in natural environments.

## 2 DATA DESCRIPTION

This study is based on LIDAR data provided by REDIAM (Consejería de Medio Ambiente de la Junta de Andalucía, Red de Información Ambiental de Andalucía, n.d.) that belongs to the Regional Ministry of Andalucía. Data were acquire from coastal zones in the provinces of Huelva and Cádiz, as can be seen in Figure 1, between the 23th and 25th of September in 2007 and it was operated at a flight altitude of 1200 m with low angles (< 11 grades) and with a point density of 2 returns/m<sup>2</sup>. The pulses were geo-referenced and validated. The accuracy report indicates an accuracy of 0.5 m. in x-y position and an accuracy of 0.15 m. in z position. In addition, the rest of variables in standard LAS were provided: intensity, angle,... Together with LIDAR data, aerial photography were collected in the same flight. The aerial photography was used to assist in the selection of training and test sets.

The study zone locates in the south of the province of Huelva in the mouth of rivers Tinto and Odiel next to Atlantic Ocean (UTM30; 150960E 4124465N). Close to the city of Huelva, a mix of land covers can be found in which industrial zones, roads and railways, port facilities and natural zones stand out. Vegetation can be divided in three classes. One of them is the scarce trees of genus *eucalyptus* forming high vegetation class. Middle vegetation class is formed by different kinds of Mediterranean shrub that surround roads and urban zones mostly. Dry grass and bare

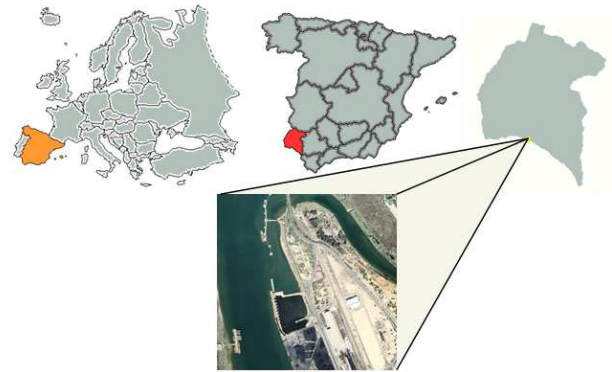


Figure 1: Study site. It locates in Huelva city, between the mouths of the rivers Tinto and Odiel. Andalusia (Spain).

earth is classified as low vegetation. In addition, the primitive land formed by marshlands near the river is another important class for land covers in this ecosystem.

LIDAR data can mainly be exploded depending of three main features: density, intensity and height of the points. A brief study of the different answers by each type of land cover in every characteristic can be useful to figure out the main differences among every class.

**Water** LIDAR does not usually reflect on water. That means plots classified as water will have low density. In addition, the few returns that reflect on water will have a low intensity because a great part of its energy is lost when it tries to go through the water surface. At last, height difference will not be very high because river usually have soft slopes near its mouth.

**Marsh** Marshlands are transition zones between watered terrain and vegetation and urban terrains. They are formed by low shrubs and grass. They are characterized by low heights and a medium/high distribution of intensities.

**Grass and bare earth** They are interior zones with very scarce vegetation or very low vegetation which produces few returns. It has the biggest intensities because of its high reflectivity in comparison with the rest of the land covers. Its height distribution is low but higher than marshland's.

**Middle vegetation** It is formed by bushes with medium height and they are mainly located between roads, trees,... They have a medium level of double and triple returns for every pulse. Intensities are in a medium level depending if they beat trunk or leaves. Their heights are over 1 m.

**High vegetation** High vegetation are mostly trees and big bushes with similar heights as trees. They have the biggest number of returns per pulse and their averaged height is high.

**Roads and railways** This class is formed by the infrastructure made to transport people or materials. It is characterized by low heights and high intensities. In addition, most of pulses produce just one return because of the absence of obstacles.

**Urban zones** The most complex class because of its variety. Intensities vary from minimum to maximum. The same can be applied for heights. This is possible because in this class we can find buildings, rubbish dumps, dock facilities and they are very different from each other.

### 3 METHOD

A widely used method based on machine learning techniques and expert systems has been chosen to carry out the classification: C4.5 decision tree. This technique takes a training set and makes a hierarchical binary tree model. Then, for any new unclassified instance, the system assigns a class based on the previous knowledge. A general view of the chosen whole process can be seen in Figure 2. In our case, a traditional pixel-based strategy has been selected. In the first step, an area size is set and extra data is taken to produce features from raw data. The second step uses expert knowledge to extract the training data. The third step produces a decision tree as a knowledge model after applying C4.5 algorithm. The last step classifies the rest of the data using the previous model.

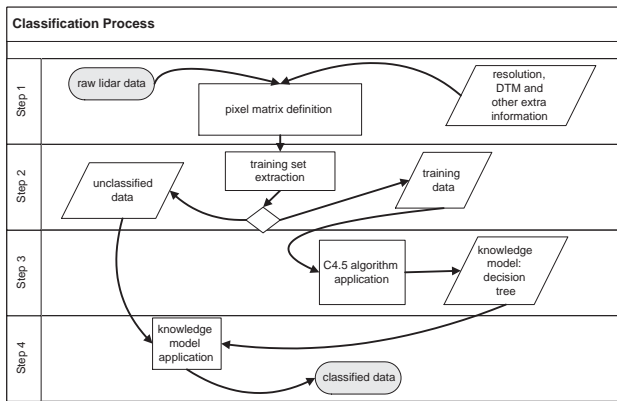


Figure 2: Classification process.

#### 3.1 Area size selection and preprocess

When pixel-oriented land cover classification is wanted to be done, it is necessary to set up the resolution previously. As a initial parameter,  $\epsilon$  will be the area in every pixel. In our case,  $4 m^2$ . Resolution depends on the density of returns directly. In our case of study, over 2 returns per  $m^2$  have been collected. A lower resolution would not have enough points while higher resolution will produce more noise in small classes like roads and railways, which do not usually have bigger widths than 3 or 4 meters.

Collected data during the flight have to be preprocessed to remove noise in order to improve the classification results. Therefore, two sorts of preprocessed was done. First, intensity correction was carried out as can be seen in bibliography (Hofle and Pfeifer, 2007) according to the equation:

$$I(R_s) = I * \frac{R^2}{R_s^2} \quad (1)$$

where I is the original intensity for a return, R is the distance from the laser source to the most furthest return and  $R_s$  is the real distance from the source to the return itself.

Apart from that, escaped returns have to be deleted. For this case, a previous phase was applied to classify data in two clusters depending on its heights. This separation is used to group escaped returns in the highest cluster while the most of returns are in the lowest. After this, a deep analyzing of the cluster and the different objects that cluster was made of, concluded a good value to exclude outliers was 17 m. which is the height reached by returns on port machinery. The returns with higher heights were removed consequently.

#### 3.2 Training set and feature selection

Supervised learning needs classified data previously. This task is very important to obtain good results. Expert knowledge has been applied to classify manually over a 3% of total data. Experts leaned on photographs taken in the same flight as LIDAR data was collected. In cases which photos were not useful, visits on the zone were planned to label the problematic terrain.

After the resolution was set up, a pixel matrix was built. Then, every pixel was labeled as part of training set or not. If it was, a class was assigned to it. Otherwise, it was marked as raw data. In Figure 4, the training data can be seen. After this, a set of measures based on intensity, heights and distribution of the returns was calculated for each pixel. Those measures can be classified as intrapixel or interpixel and they take advantage of different kinds of terrain have their own characteristics that make possible to lay down differences among them visually or morphologically. Separability between classes has to be assured and the more measures you have the best results you can provide. In our case, thirty-three different measures were calculated for every pixel. In addition, it has to be said that LIDAR data was raw so no previous interpolation were done and empty pixels would be used as another source of information.

Table 1 contains the thirty-three different measures used in this study. Most of the measures have been extracted from bibliography (Hudak et al., 2008). Interplot measures have been developed ad hoc and they are an original contribution e.g. relative difference, that is calculated as the absolute difference of all the pixel's measures and its neighbors' divided by the total number of neighbors. In order to simplify the model extraction process, a feature selection method was applied by classification algorithm. In this case, the gain ratio was calculated for every variable and those with the maximum values were selected. Gain ratio selector evaluates the worth of an attribute by measuring the gain ratio with respect to the class:

$$GainR(Class, Var) = \frac{(H(Class) - H(Class|Var))}{H(Var)} \quad (2)$$

where H is the Shannon Entropy value whose definition can be seen in 3.3.

#### 3.3 C4.5 decision tree algorithm

In this work, a classic hierarchical decision tree builder algorithm has been selected: C4.5(Quinlan, 1996). This algorithm is one the most used to build decision trees. C4.5 can handle continuous and discrete attributes, training data with missing attribute and attributes with different costs and it can even prune trees at the end of execution if it's necessary. C4.5 builds decision trees from a set of training data using the concept of information entropy. The training data is a set of already classified samples. Each sample is a vector that represents attributes or features of the sample. Information entropy is a measure of the uncertainty associated with a random variable. The term by itself in this context usually refers to the Shannon entropy, which quantifies, in the sense of an expected value. The information entropy of a discrete random variable X with possible values  $x_1 \dots x_n$  is:

$$H(x) = E(I(x)) \quad (3)$$

Here E is the expected value function, and I(X) is the information content or self-information of X. I(X) is itself a random variable.

Variable	Description	Type
<b>IMIN</b>	Intensity minimum	Intrapixel
<b>IMAX</b>	Intensity maximum	Intrapixel
<b>IMEAN</b>	Intensity mean	Intrapixel
IVAR	Intensity variance	Intrapixel
ISTD	Intensity standard deviation	Intrapixel
IAAA	Intensity average absolute deviation	Intrapixel
IRANGE	Intensity range	Intrapixel
<b>HMIN</b>	Height minimum	Intrapixel
<b>HMAX</b>	Height maximum	Intrapixel
<b>HMEAN</b>	Height mean	Intrapixel
HVAR	Height variance	Intrapixel
HSTD	Height standard deviation	Intrapixel
HAAA	Height average absolute deviation	Intrapixel
HRANGE	Height range	Intrapixel
<b>IKURT</b>	Intensity Kurtosis	Intrapixel
<b>ISKEW</b>	Intensity Skewness	Intrapixel
<b>HKURT</b>	Height Kurtosis	Intrapixel
<b>HSKEW</b>	Height Skewness	Intrapixel
ICV	Intensity coefficient of variation	Intrapixel
HCV	Height coefficient of variation	Intrapixel
SLP	Slope	Interpixel
<b>RDIFF</b>	Relative difference among neighbors	Interpixel
<b>RZDIFF</b>	Elevation difference between first return and last return	Interpixel
PCT1	Percentage 1st returns	Intrapixel
<b>PCT2</b>	Percentage 2nd returns	Intrapixel
PCT3	Percentage 3rd or later returns	Intrapixel
PCT31	Percentage 3rd returns over 1st returns	Intrapixel
PCT21	Percentage 2nd returns over 1st returns	Intrapixel
PCT32	Percentage 3rd returns over 2nd returns	Intrapixel
NOTFIRST	Percentage 2nd or later returns	Intrapixel
EMP	Empty plots surrounding	Interpixel
TPO	Total number of points	Intrapixel
<b>CRR</b>	Canopy relief ratio	Intrapixel

Table 1: Thirty-three candidate predictor variables with ten selected variables indicated in bold

If  $p$  denotes the probability mass function of  $X$  then the information entropy can explicitly be written as:

$$H(x) = \sum p(x)I(x) = - \sum p(x) \log_b p(x) \quad (4)$$

where  $b$  is the base of the logarithm used. Common values of  $b$  are 2.

C4.5 uses the fact that each attribute of the data can be used to make a decision that splits the data into smaller subsets. C4.5 examines the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is the one used to make the decision. The algorithm then recurs on the smaller sublists. In order to assess the quality of the decision tree, data-mining software was used: WEKA (Holmes et al., 1994).

Then a new software developed ad hoc classified the rest of data as the model commanded. In Table 1 the thirteen selected attributes can be seen indicated in bold.

#### 4 RESULTS AND DISCUSSION

After the decision tree was extracted and the unclassified data was processed, a stratified test was built and executed. The test data was taken randomly from the unclassified data and later classified manually by experts. The proportion among the classes was kept in relation to the training data set original proportion. In Table 2, a summary of the confusion matrix, accuracies and kappa statistic can be seen.

Although it is well-known that riparian zones are very hard to be classified, results show a very high global accuracy. These results prove separability between classes in the training set from Huelva just with LIDAR data and they show our approach is very promising and it can provide excellent results. Special success was achieved in most of classes like water, roads and marshlands. The worst results were obtained working with middle vegetation and urban zones.

Apart from that, it has to be said some misclassification appears in docks and port facilities as can be seen in 5. This is due to pixel-oriented approaches try to classify a piece of data. The same problem can be seen in some buildings which do not have roof structure. The problem is structures in the study zone are built with the same sort of terrain that surrounds them, so the algorithm cannot separate properly the structure's inner pixels from neighboring pixels that belong to another class because they share heights and reflectivity so its efficiency is lowered. This is an inherent problem of this kind of approach and it has to be solved in future work. In addition, some zones show some serious noise. This is mainly because of the intensity outliers. As before, intensity is one of the three key parameters in which the classification lean on. So it is very important to delete this interference in order to avoid algorithm is deceived.



Figure 3: Ortophoto of the study zone.

User class \ sample	Water	Marshland	Roads and railways	Low Vegetation	Middle Vegetation	High Vegetation	Urban terrain
Water	32	0	0	0	0	0	0
Marshlands	0	30	0	0	0	0	4
Roads and railways	0	0	28	0	0	0	7
Low Vegetation	0	0	0	14	0	0	1
Middle Vegetation	0	0	0	2	12	0	3
High Vegetation	0	0	0	0	0	3	0
Urban terrain	0	0	2	4	5	1	38
Producer's accuracy	1.0	1.0	0.93	0.7	0.71	0.75	0.73
User's accuracy	1.0	0.88	0.8	0.93	0.71	1.0	0.76
Total accuracy	0.85						
KIA	0.81						

Table 2: Summary for the test set confusion matrix

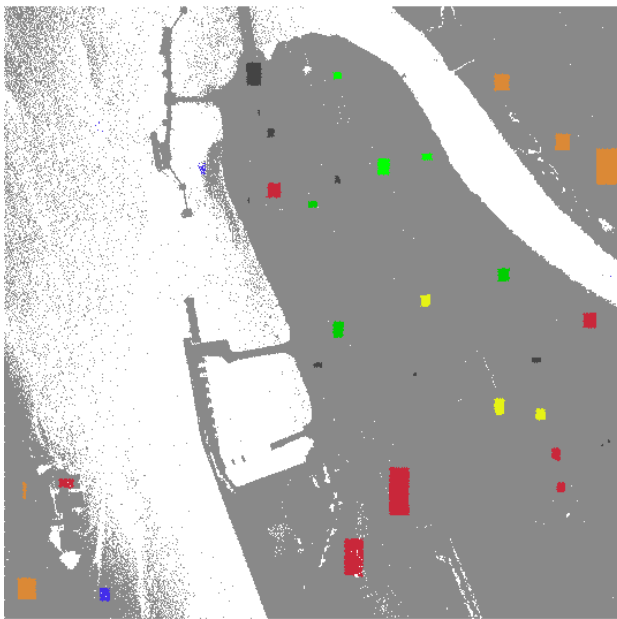


Figure 4: Training set: water in blue, urban zone in red, roads and rails in dark grey, middle vegetation in green, low vegetation or bare soils in yellow, high vegetation in light green, marsh in brown and no training data in light grey.

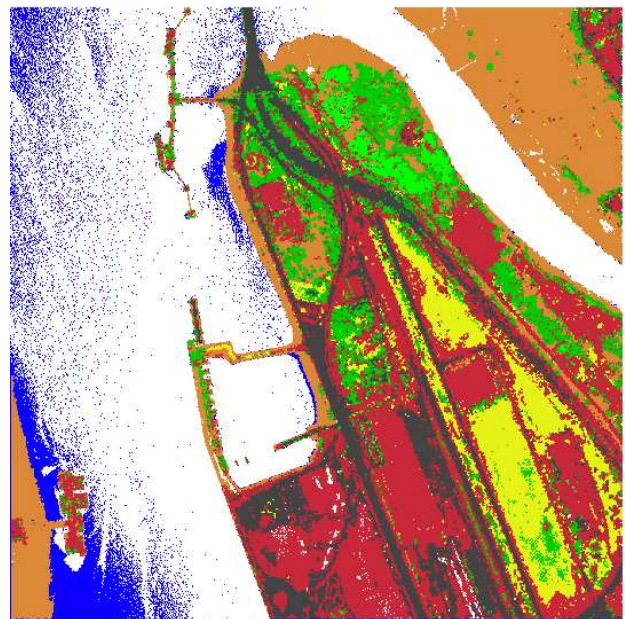


Figure 5: Resulted classification for a resolution of 4 m<sup>2</sup>.

## 5 CONCLUSIONS AND FUTURE WORK

A LIDAR-based approach to classify and monitor land covers from Mediterranean mixed zones has been analyzed in this work. To be precise, a pixel-based approach has been proposed in order to classify raw data into seven different classes. Thus, it has been demonstrated that different kinds of terrain can be differentiated by applying a well-known data mining technique, such as C4.5 algorithm, integrated in a multi-step cascade process of feature extraction and classification and without uses of extra data like multispectral imagery. The accuracy shown is certainly excellent to be a riparian zone and very promising since no extra computation apart is added to the approach, achieving a low computational cost.

Concerning to future work, some problems have been detected. Some of them are inherent to pixel-based approaches and it would be very interesting to apply a metaphase in which for a small quantity of computational cost, a segmentation and object-detection process could deal with data to extract the most difficult structures to be classified. Even so, results are very promising themselves but testing denotes modeling has to be improved to assure optimal

results in future. In addition, outlier detection has been noticed as a very important task because most of misclassification detected has its origin in outliers. Together with this, it is necessary to set up a general method to extract training and test data to be able to achieve quality assessment in future comparisons between several methods and sensors and to validate any work which is a task that very few authors have invested in.

## ACKNOWLEDGEMENTS

We gratefully thank the Regional Ministry of Environment from Andalusia for all the support we have received in the development of this work and especially we thank Irene Carpintero and Juan José Vales for all the time they have invested in us.

## REFERENCES

- Antonarakis, A., Richards, K. and Brasington, J., 2008. Object-based land covers classification using airborne lidar. *Remote Sensing of Environment* 112, pp. 2988–2998.
- Arroyo, L. A., Pascual, C. and Manzanera, J. A., 2008. Fire models and methods to map fuel types: The role of remote sensing. *Forest Ecology and Management* 256, pp. 1239–1252.

- Brzank, A., Heipke, C., Goepfert, J. and Segel, U., 2008. Aspects of generating precise digital terrain models in the wadden sea form lidar-water classification and structure line extraction. *ISPRS journal of Photogrammetry & Remote Sensing* 63, pp. 510–528.
- Canty, M. J., 2008. Boosting a fast neural network for supervised land cover classification. *Computers & Geoscience*.
- Chust, C., Galparsoro, I., Borja, A., Franco, J. and Uriarte, A., 2008. Coastal and estuarine habitat mapping, using lidar height and intensity and multi-spectral imagery. *Estuarine, Coastal and Shelf Science*.
- Consejería de Medio Ambiente de la Junta de Andalucía, Red de Información Ambiental de Andalucía, n.d.
- Dorigo, W. A., Zurita-Milla, R., de Wit, A. J. W., Brazile, J., Singh, R. and Schaepman, M. E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation* 9, pp. 165–193.
- Gamanya, R., Maeyer, P. D. and Dapper, M. D., 2009. Object-oriented change detection for the city of harare, zimbabwe. *Remote Sensing of Environment* 36, pp. 571–588.
- Goetz, S., Steinberg, D., Dubayah, R. and Blair, B., 2007. Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, usa. *Remote Sensing of Environment* 108, pp. 254–263.
- Hofle, B. and Pfeifer, N., 2007. Correction of laser scanning intensity data: Data and model-driven approaches. *ISPRS journal of Photogrammetry & Remote Sensing*.
- Holmes, G., Donkin, A. and Witten, I., 1994. Weka: A machine learning workbench. In: *Proc Second Australia and New Zealand Conference on Intelligent Information Systems*, Brisbane, Australia.
- Hudak, A. T., Crookston, N. L., Evans, J. S., Halls, D. E. and Falkowski, M. J., 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from lidar data. *Remote Sensing of Environment* 112, pp. 2232–2245.
- Hughes, M., Schmidt, J. and Almond, P. C., 2009. Automatic landform stratification and environmental correlation for modelling loess landscapes in north otago, south island, new zealand. *Geoderma* 149, pp. 92–100.
- Jensen, J. L. R., Humes, K. S., Vierling, L. A. and Hudak, A. T., 2008. Discrete return lidar-based prediction of leaf area index in two conifer forests. *Remote Sensing of Environment* 112, pp. 2988–2998.
- Koetz, B., Morsdorf, F., van der Linden, S., Curt, T. and Allgower, B., 2008. Multi-source land cover classification for forest fire management based on imaging spectrometry and lidar data. *Forest Ecology and Management* 256, pp. 263–271.
- Magnussen, S., McRoberts, R. E. and Tomppo, E. O., 2009. Model-based mean square error estimators for k-nearest neighbour predictions and applications using remotely sensed data for forest inventories. *Remote Sensing of Environment* 113, pp. 476–488.
- McColl, C. and Aggett, G., 2007. Land-use forecasting and hydrologic model integration for improved land-use decision support. *Journal of Environmental Management* 84, pp. 497–512.
- Pascual, C., Garcia-Abril, A., Garcia-Montero, L., Martin-Fernandez, S. and Cohen, W., 2008. Object-based semi-automatic approach for forest structure characterization using lidar data in heterogeneous pinus sylvestris stands. *Forest Ecology and Management* 255, pp. 3677–3685.
- Quinlan, J. R., 1996. Improved use of continuous attributes in c4.5. *Journal of Artificial Intelligence Research* 4, pp. 77–90.
- Schneider, J., Grosse, G. and Wagner, D., 2009. Land cover classification of tundra environments in the arctic lena delta based on landsat 7 etm+ data and its application for upscaling of methane emissions. *Remote Sensing of Environment* 113, pp. 380–391.
- Schubert, J. E., Sanders, B. F., Smith, M. J. and Wright, N. G., 2008. Unstructured mesh generation and landcover-based resistance for hydrodynamic modeling of urban flooding. *Advances in Water Resources* 31, pp. 1603–1621.
- Sithole, G. and G.Vosselman, 2003. Comparison of filtering algorithms. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 34, pp. 71–78.
- Tooke, T. R., Coops, N. C., Goodwin, N. and Voogt, J. A., 2008. Extracting urban vegetation characteristics using spectral mixture analysis and decision tree classifications. *Remote Sensing of Environment*.
- Witten, H. and Frank, E., 2005. *Data mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers.