Geoinformation analysis for Conservation: experiences with Open Source Software in the Natural Parks of Barcelona

Agustin Lobo[1,2], Francesc Ara[2,3], Francesc Baró[2,4], Carlos Camino[5,6]

[1]Institut de Ciències de la Terra "Jaume Almera (CSIC), Barcelona, Spain, Agustin.Lobo@ija.csic.es
[2]Institut de Ciència i Tecnologia Ambiental, Universitat Autònoma de Barcelona, Spain
[3]Current address: Geodata Sistemes S.L., Barcelona, Spain
[4]Current address: Lavola S.A., Barcelona, Spain
[5]Centre Tecnològic Forestal de Catalunya, Solsona, Spain
[6]Current address: Centro de Investigaciones Atmosféricas de Izaña (AEMET), Santa Cruz de Tenerife, Spain

Abstract

This contribution describes some applications of geospatial analysis in which OSS has been prevalent, with the purpose of presenting examples of operational use and discussing current limitations. The examples are extracted from projects that we have carried out in Natural Parks of the Province of Barcelona, including use of calibrated satellite vegetation index to monitor recovery after fire, multicriteria analysis to rank and select abandoned fields for restoration, generalization of vegetation maps using a double-criteria hierarchical clustering and processing of infrared imagery acquired by a miniUAV.

1. Introduction

Open Source Software (OSS) is becoming increasingly popular for the analysis of geodata in Ecology, and Environmental and Biological Conservation. OSS is much more than just a cheap solution for disciplines that are most often under the stress of scarce funding, it has become the basis for the development of a community in which cooperative support is spiced with exciting and fruitful debate. The interface between users and developers of OSS is becoming wider and fuzzier, providing an excellent environment for scientific and technical advance. This presentation describes some applications of geospatial analysis in which OSS has been prevalent, with the purpose of presenting examples of operational use and discussing current limitations.

2. Description of the experiences

2.1 Multi-annual analysis of a fire scar

Monitoring both the natural regrowth of vegetation and the impact of remediation actions after a wild fire is important for management of protected areas. We monitored regrowth of vegetation in the scare of a wild fire occurred on 8/2003 in the PNSTLL through calibrated Perpendicular Vegetation Index (PVI) using SPOT images acquired on 29/08/2003, 31/07/2004 and 21/09/2005. SPOT-HRG images include 1 panchromatic band (490-710 nm at 5 m resolution) and 4 multi-spectral bands at 10 m resolution: green (500-590 nm), red (610-680 nm), near infrared (780-890 nm) and short-wave infrared (1580-1750 nm). PVI was preferred because is less affected by soil color than NDVI and thus more appropriate to estimate vegetation cover. SPOT images were previously orthorectified using an implementation of the algorithm of Toutin (2004) in PCI-Orthoengine, and atmospherically corrected using ATCOR-3 (Richter 1998, 2009). A Digital Elevation Model (DEM) provided by the Cartographic Services of the *Diputació de Barcelona* was used for terrain analysis.

We used QGIS as the basic GIS interactive tool, running GRASS through its specific plugin when required. GRASS was used to calculate layers of slope and aspect, as well as to perform map calculations (r.mapcalc).

We extracted multi-spectral image values for a stratified random sample of pixels of the 2003 image located inside and outside the fire scar, including information of the type of cover (unaffected vegetation, less-affected vegetation in "islands" within the burned area, bare soil), height, slope and aspect, and saved as an ESRI shape set. These data were imported into R (*rgdal* package, Keitt *et al.* 2010), where the soil line for the PVI computation was estimated through a plot of near infrared (NIR) *vs* red (RED) values, using the aforementioned information on type of land cover to verify the consistency of the data (Fig. 1). Parameters of the soil line were then used to calculate the PVI layer through the *r.mapcalc* module of GRASS.

In order to obtain our own calibration between PVI and vegetation cover, we acquired low-height color and NIR photography of the area with a tethered balloon *(www.altafoto.com)*, and interactively estimated vegetation cover for a grid matching the SPOT pixels. We then applied the coefficients of the calibration to the PVI layers in order to generate layers of the fraction of pixel area that is covered by vegetation in summer in successive years (Fig. 2).
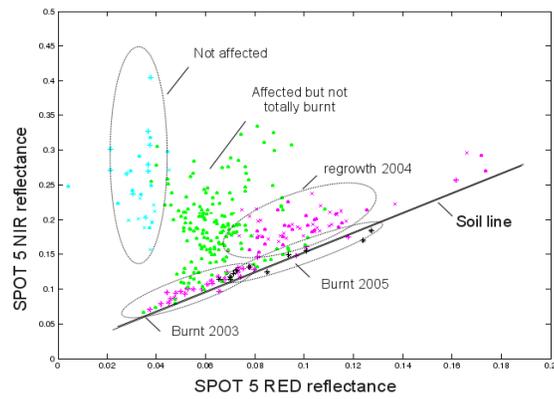
Figure 1. Plot of sampled plixels on the NIR vs RED plane
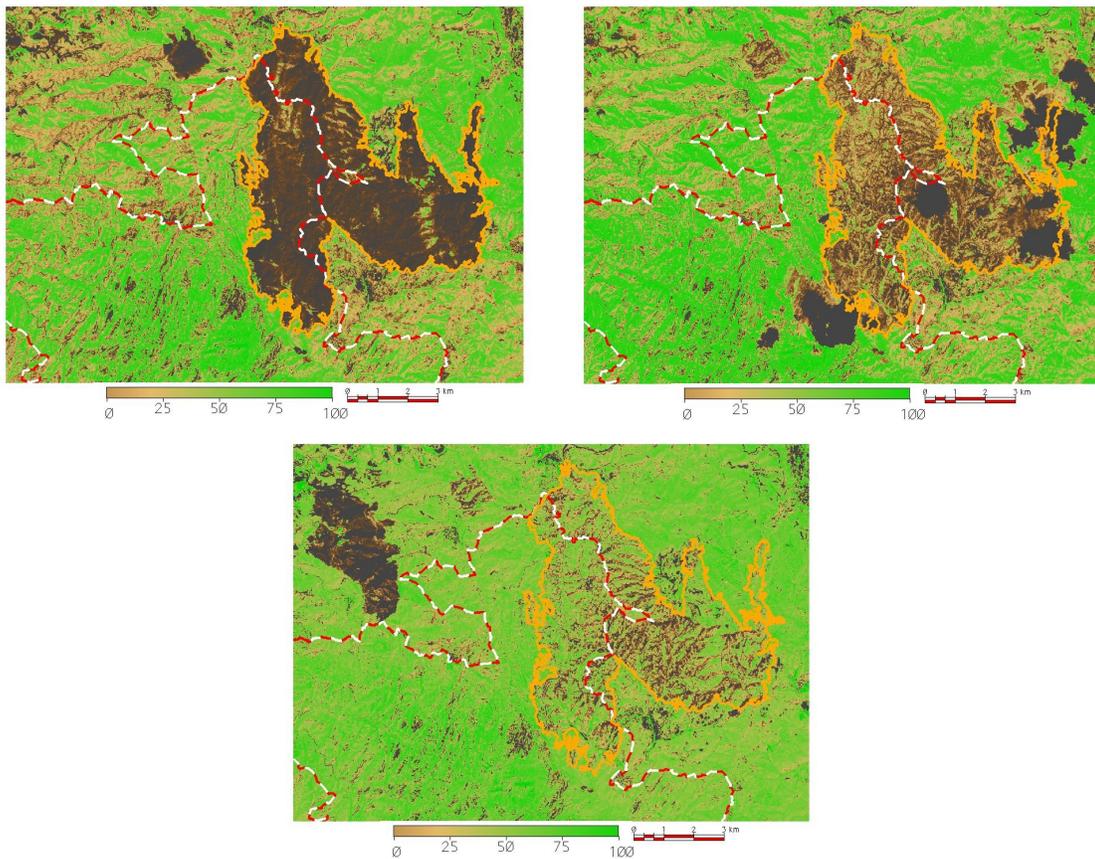
Figure 2. Maps of the fraction of green vegetation cover in 29/08/2003, 31/07/2004 and 21/09/2005

2.2 Multi-criteria decision analysis (MCDA) to rank and select abandoned fields for restoration

Rather than being continuous forest landscapes, the aim of Natural Parks of the Province of Barcelona is to remain as attractive and scenic landscape mosaics resulting from the interaction of rural human activity and natural environment. Abandonment of crop fields and severe decrease of extensive livestock in the last decades because of low monetary revenues, have become a threat to the diversity of the mosaic. Parl managers are now interested on restoration of some abandoned crop fields. Considering economic constraints, abandoned fields must be ranked in terms of their appropriateness for restoration. As maintaining the mosaic is one of the priorities, we included a metric of the impact of the restoration of each candidate field in the mosaic along with other field attributes and applied a multi-criteria analysis for ranking the candidates.

After digitizing both active and abandoned crop fields in the Park and surrounding area from current and historic (1956) aerial photography, our metric was based on (i) calculating a raster layer of distances from a given cell to the nearest active crop field (Fig. 3); (ii) calculating a second layer of distances for which the candidate field was considered as active; (iii) calculating the difference between both distance layers; (iv) summing up all values of the difference raster layer as the Index of Mosaic Impact (IMI) for the given candidate. This procedure is based on the fact that restoring a field that is close (far) to fields that are actually active has an small (large) impact in the distances to the nearest active crop field (Fig. 4). Digitized vector layers of crop fields were converted to raster using GRASS and imported to R, where all steps for calculating the Index of Mosaic Impact were programmed.

We did not consider opening any new crop field, but just the active ones in 1956 were included as candidates. Therefore, we were assuming that the pattern in 1956 was close to the optimum mosaic and/or that transport and associated economic reasons make crop fields other than the ones present in 1956 inviable. Note that while this a reasonable assumption in our geographic region, the situation in other parts of the world can be totally different.

In addition to computing IMI, we selected the following criteria from the total of 50 attributes included in a Geodatabase (MDB format) of active and abandoned crop fields: Fire risk, Erosion factor, Quality as agricultural habitat for birds, Quality as forest habitat for birds, Average conservation importance for birds, Quality as habitat for rabbit. We applied three different methods for ranking: rule-based and MCDA by weighted linear average and MCDA by Ideal Goal. Appropriate functions were programmed in R for the different steps involved in MCDA. Results from all three methods were very similar for the top 10 candidates in terms of priority. See Lobo *et al*. (2007) and Baró (2007) for details.
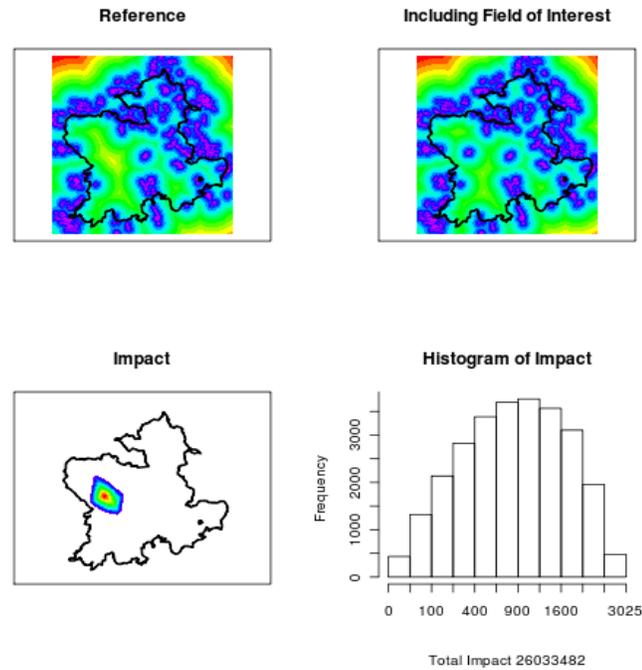
Figure 3. Example of the calculation of the Index of Mosaic Impact (IMI)

2.3 Analysis of satellite imagery for the generalization and error detection in detailed vegetation maps generated by photo-interpretation

While satellite imagery can be used to develop land cover and vegetation maps that are acceptable in many applications, a much finer (and involved) cartography has been produced for Catalonia from digitization and photo-interpretation of real and color infrared aerial photography at scales 1:25000 and 1:5000 (Ibañez and Burriel, 2007; CERBIV, 2004). We developed the method *Double Criteria Agglomerative Clustering* to generalize a very detailed and complex vegetation map (350 classes) of the Montseny Natural Park (Panereda *et al.*, 2007) into categories that keep coherence both from a physiognomic point of view (thanks to the multi-spectral properties of vegetation types as recorded in a SPOT-5 satellite image), and from a vegetation point of view (thanks to vegetation properties as defined by the hierarchical legend).

A table of multi-spectral properties of vegetation types in the SPOT-5 image were calculated in a per-polygon basis with GRASS and used to build a statistical dissimilarity matrix (distance of Jeffries-Matutsita (JM); Richards and Jia, 1999) in R. As this metric is a distance between multi-gaussian distributions, care must be taken not to include erroneously attributed polygons in the parameterization because variances easily become inflated. Thus, we developed a method to automatically filter out suspect polygons based on functions from R package *mvoutlier* (Gschwandtner and Filzmoser 2009).

Figure 4. Example of polygons (red labels) that have been
filtered out as outliers in a given vegetation class

Figure 5. Example of a typical polygon of Fagus sylvatica forest (Fs, left)
and another one that had been detected as outlier (Fs, right)

We built an equivalent dissimilarity table considering the conceptual differences between vegetation classes. Semiquantitative (rank) values in this matrix were set by taking advantage of the hierarchical structure (6 levels) of the legend. Two given classes that would split in the lowest level of the hierarchy, would be considered fairly similar and their dissimilarity set to "1", while two given classes that would be different from the top level would have their similarity set to "6". Note that conceptual dissimilarities are not arithmetically operated but just logically compared. We modified this matrix by

including some additional restrictions, such as that any two classes including, respectively, evergreen and deciduous species would always have a dissimilarity higher than "1".

Once both dissimilarity matrices were calculated, our clustering algorithm would proceed by selecting the two most similar classes in terms of image properties, but merging only takes place if the conceptual dissimilarity is lower than "3"; otherwise, the next two most similar classes in terms of image properties are considered. If merging actually occurs, the dissimilarity matrix is updated and the process iterated. The process stops when no more merges are possible.

After applying this method, we reduce the number of classes from 114 to 18 for the 67% of the area of study. Adding a number of trivial aggregations based on the vegetation properties only, 73% of the study area is simplified from 167 to 25 classes. As a side effect, the method identifies 10% of the polygons as outliers in terms of the image properties of each class, which implies either land cover change between the dates of the spatial support used for generating the original map and that of the satellite image, or errors in the production of the original map. More details can be found in Lobo and Camino (2009).

2.4 Campaign planning and initial processing of low-height aerial imagery acquired with a miniUAV

Satellite remote sensing imagery is used for forestry, conservation and environmental applications, but insufficient spatial resolution, and, in particular, unavailability of images at the precise timing required for a given application, often prevent achieving a fully operational stage. Airborne remote sensing has the advantage of custom selection of sensors, resolution and timing, but most often its price prevents using it as a routine technique for the mentioned fields. Some small Unmanned Aerial Vehicles (miniUAV) might provide a "third way" solution as low-cost techniques for acquiring remotely sensed information, albeit at the expense of lower quality instrumentation and instability.

We conducted a series of aerial campaigns during 2009 in the Montseny National Park (Catalonia, NE Spain) to test a system based on a remotely-controlled mini-UAV (ATMOS-3) equipped with a low-cost color infra-red camera (VEGCAM-PRO) designed and operated by CATUAV. For each campaign, flight lines were planned by editing with QGIS and exported to R. We created a set of R programs that, based on the input flight lines and the optical characteristics of the system, produce, for a given height and interval, a set of rectangles matching the planned footprints of the aerial images, and their centroids. Centroids and rectangles were exported as shapefiles and displayed in QGIS, where the appropriateness of the plan was checked. Centroids were exported as way points (GPSBabel) to the GPS format of the navigation system of the miniUAV and used to control the campaign (Fig. 6).

Each campaign resulted into a set of digital images and a GPS track, without independent angular

information as the system lacks an inertial measurement unit. We made a second set of R programs that selects, for each image, the GPS reading that is closest in time, thus providing its geographic position. Orientation and pitch angle, as well as velocity and acceleration are estimated from the selected and previous GPS records in the track. After reading terrain height from a DEM, and assuming a 0 yaw angle, the program creates a set of rectangles that provide a first estimate of the actual footprint of the images (Fig. 6). Another R program writes a world file for each image that can be then displayed in QGIS. Our elementary processing, based on the data provided by the track of a simple (non geodetic level) GPS only, is sufficient for an initial evaluation of the campaign, but most applications require further, more sophisticated and involved photogrametric processing which offers sub-metric accuracy and mosaic generation.

We were able to cover 70% of the 1915 ha that were initially planned, in 7 different sites, having only 2 failed campaigns: one because of technical problems and another due to meteorological conditions. Therefore, from a flying point of view, the system proves to be operational albeit quite dependent upon relatively calm wind conditions. Our present efforts are focused on improving the colorimetric processing of the images, which currently presents severe problems whenever light conditions are note close to optimal.
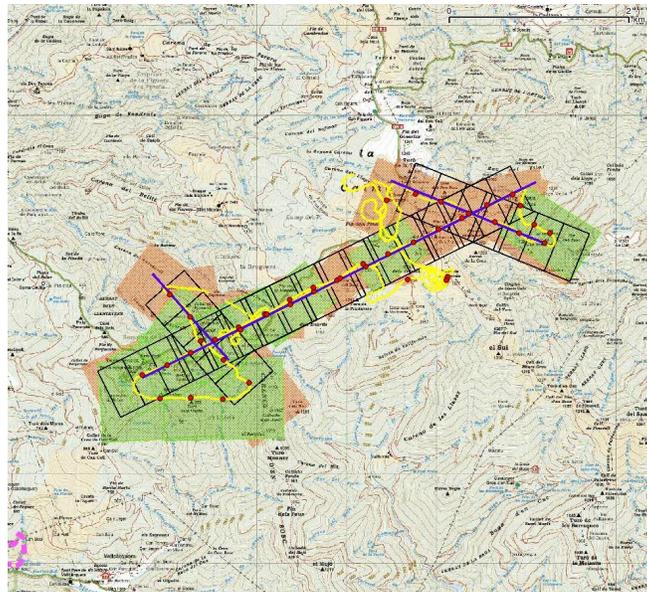
Figure 6. Example of a campaign conducted by the minuUAV ATCOR-3. The blue lines and black rectangles are, respectively, the planned flight lines and image footprints (700 m agl). The yellow line and red dots represent, respectively, the GPS track and positions of image acquisitions. The orange and green areas represent the mosaic of estimated footprints of image acquisitions.

3. Brief evaluation of the use of OSS in these experiences

The mentioned experiences were undertaken from 2003 to 2009, a relatively long period of time if considering the currently fast development of OSS. This fact, in addition to our own progress on detecting, learning and testing OSS tools, implies that some of the currently available tools were actually not used or were used from nowadays outdated versions. Therefore, some of the following comments might not reflect the latest implementations.

Quantum GIS (QGIS) has developed almost during the same period of time, to become a very user-friendly platform for the display of geographic information, with its very convenient feature of wide format support for reading and writing. Our use of QGIS has steadily increased from just being an interface for GRASS to become our most common viewing tool. Nevertheless, QGIS tools for raster display are still insufficient and support for SID and ECW formats does not come "out of the box" but actually require involved post-installation steps. This is an important drawback, as virtually all of the extensive digital aerial orthoimagery of Catalunya is provided in SID format and many popular commercial maps are in ECW. Raster operations were conducted with GRASS, as only recently some raster tools have been added to QGIS through use of GDAL utilities.

A set of tools that has considerably improved in QGIS is the geometric Editing of vectors. Our digitizing work was actually conducted with commercial software because of some initial troubles with earlier versions of QGIS, but digitization work can be certainly done with OSS nowadays.

Use of a Geodatabase in MDB format was an unfortunate requirement set by the agency funding the work on ranking abandoned fields for restoration *(Àrea d'Espais Naturals, Diputació de Barcelona)*. We suspect that an implementation with POSTGIS or SPATIALite would be considerably more powerful, but we cannot offer any comparison yet. This pressure set by a funder on delivering a given proprietary format is not frequent in scientific work, though. Actually, the facilities for auditing and reproducing results as well as the cooperative support that is inherent to most OSS projects are increasingly recognized in scientific communities. Nevertheless, we also faced other difficulties, such as insufficient training of students on OSS tools during their academic studies, and, in some cases, insufficient motivation to learn OSS solutions because the students initially believe that (i) the effort does not pay the advantages, and (ii) future contractors will want them to use commercial GIS.

Virtually all of our (spatial or not) data analysis was done in R (R Development Core Team, 2010), for which we made extensive use of spatial packages, in particular *spgrass6* (which links active GRASS sessions to R), *rgdal* (for reading and writing files in standard spatial formats to and from R), *sp* and *maptools*. R was also used as scripting language, thus actually becoming the central and fundamental piece of our toolbox. Because of RAM problems, dealing with large raster layers (and images are always large raster layers) in R has been a problem until recently. The common approach

(and the one used in most of these experiences) was to perform all heavy raster processing in GRASS, often under the control of R through *spgrass6*. In this way, R scripts were actually controlling the flow of GRASS processes. Nevertheless, the modern addition of package *raster* (Hijmans and van Etten, 2010) to R is bringing this approach close to an end. Package *raster* can work as an interface to the large raster files and makes it possible to operate with them as raster objects in R because the actual data remain in the disk. R can thus deal with part of the spatial processing, in addition to its natural capabilities for spatial and non-spatial statistical analysis.

A recent tool that facilitates the link between the realms of GIS and data analysis is the QGIS plugin *manageR* (Farmer, 2010), which opens a convenient R GUI from within QGIS and makes it possible for the user to transfer (currently vector) layers in the Table of Contents of QGIS to R and *viceversa*. Links between R and GIS packages can actually be extended beyond the transfer of spatial objects as provided by utilities like manageR: R could provide extended facilities for statistically sounded symbolization of vector and raster layers.

All described examples required the combined use of proprietary and OS software. In particular, atmospheric and advanced geometric correction methods are still not included in general-purpose OS GIS packages. This is because such a processing is quite dependent on specific characteristics of the instrument and mission that acquired the imagery, but dedicated toolboxes such as CNES Orfeo (see presentations by T. Feuvrier and E. Christophe), OSSIM (see presentation by M. Lucas) and ESA-funded BEAM project (http://www.brockmann-consult.de/cms/web/beam/welcome) are bringing this sophisticated processing to OSS users.

References

Baró, F., 2007, 'Proposta de planificació del mosaic agroforestal del Parc Natural de Sant Llorenç del Munt. Una priorització dels espais oberts agraris a partir d'una integració SIG- Avaluació Multicriteri ', *Ms. Thesis, Universitat Autònoma de Barcelona*

CERBIV, 2004, 'Manual d'interpretació de la llegenda de la cartografia CHC50' <*http://mediambient.gencat.cat/cat/el_medi/habitats/habitats_documents.htm#mc*>

Farmer, C., 2010, *'manageR',* <http://www.ftools.ca/plugins.html>

Gschwandtner, M. & Filzmoser, P. 2009, *'Multivariate outlier detection based on robust methods',* <*http://www.statistik.tuwien.ac.at/public/filz/*>

Hijmans,R.J. & Jacob van Etten, J, 2010, *'raster: Geographic analysis and modeling with raster data',* R package version 1.1.2/r948. <*http://R-Forge.R-project.org/projects/raster/* >

Ibàñez, J. & Burriel, J.A., 2010, 'Ocupació del sòl a Catalunya: 3a edició del Mapa de Cobertes (MCSC-3)', <*http://www.creaf.uab.es/mcsc/resu_res/CaracteristiquesResultatsMCSC3.pdf* >

Keitt,T. H., Bivand, R., Pebesma, E. & Rowlingson , B. 2010. *'rgdal: Bindings for the Geospatial Data Abstraction Library'* , R package version 0.6-27. <http://CRAN.R-project.org/package=rgdal >

Lewin-Koh,N.J. & Bivand,R (contributions by Edzer J. Pebesma, Eric Archer, Adrian Baddeley, Hans Bibiko, Stphane Dray, David Forrest, Michael Friendly, Patrick Giraudoux, Duncan Golicher, Virgilio Gomez Rubio, Patrick Hausmann, Thomas Jagger, Sebastian P. Luque, Don MacQueen, Andrew Niccolai, Tom Short and Ben Stabler), 2010, 'maptools: Tools for reading and handling spatial objects', R package version 0.7-34. <http://CRAN.R-project.org/package=maptools >

Lobo, A., Vega, C. & Baró, F. 2007, 'Selecció d'espais oberts a recuperar i propostes d'ús: una aproximació per anàlisi espacial i multicriteri'. *Reports on Environmental Sciences (RECERCAT, Dipòsit de la Recerca de Catalunya)* <http://hdl.handle.net/2072/40712>

Lobo, A. & Camino, C,. 2009, 'Soporte geoespacial para modelos de distribución de especies (I): Generalización del mapa de vegetación del Montseny apoyada en información de imágenes de satélite '. *Reports on Environmental Sciences (RECERCAT, Dipòsit de la Recerca de Catalunya),* <***http://hdl.handle.net/2072/83514***>

Panareda, J.M., Salvà, M. & Nuet, J. 2007 , 'Cartografia de l'ús del sòl i de la vegetació, a escala 1:5.000 del Parc Natural del Montseny . Llegenda del mapa', *Servei de Parcs Naturals, Diputació de Barcelona*

Pebesma, E.J. & Bivand, R. 2005, *'Classes and methods for spatial data in R',* R News 5 (2), http://cran.r-project.org/doc/Rnews/.

R Development Core Team, 2010, *'R: A Language and Environment for Statistical Computing',* <http://www.R-project.org>

Richards, J.A. & Jia, X., 1999*, 'Remote Sensing Digital Analysis. An Introduction'* 3rd Edition, Springer, Berlin

Richter, R. 1998, 'Correction of satellite imagery over mountainous terrain', *Applied Optics*, vol. 37, pp. 4004-4015.

Richter, R. 2009, 'Atmospheric / Topographic Correction for Satellite Imagery" , *DLR report DLR-IB 565-01/09, Wessling, Germany,* <http://www.rese.ch/pdf/atcor4_manual.pdf>

Toutin, T. 2004, 'Geometric processing of remote sensing images: models, algorithms and methods', *International Journal of Remote Sensing*, vol. 25, pp. 1893 – 1924

Acknowledgements