EuroCoppice STSM Report in University of Eastern Finland – Joensuu.

By Ángela Blázquez Casado, föra forest technologies, Spain

Modelling methodologies focussed on different machine learning using Rstudio for stand classification and productivity

1. INTRODUCTION

Spain has suffered from severe deforestation and intense exploitation of remaining forests for centuries, mainly for the use of fuelwood, charcoal, timber and pastures in the rural economy. Nevertheless, during the 20th century this trend has changed and, particularly in the second half of the century, reforestation plans and the flow of population towards the urban environment have contributed to the increment of forested areas and the abandonment of traditional forestry practices. As a result, the no-management forestry in the last 4-5 decades has led to increasing stocking and density of forests all over Spain (Montero and Serrada 2013).

The case of the Laurisilva forests in the Canary Islands is a good example of this scenario. The Laurisilva is a typical subtropical cloud forest in areas with mild temperatures, which occupies, in the Canary Islands, the north-facing slopes of the western islands. Evergreen species of the genera *Laurus*, *Ilex*, *Ocotea*, *Arbutus*, etc. are typical of this ecosystem. Coppicing has been the traditional method of management for the exploitation of fuelwood, charcoal and small pieces of timber for agricultural activities and construction (Arozena et al., 2015). Nevertheless, in the last decades the society has shifted its preferences on the ecosystem services provided by the Laurisilva coppice forest from the traditional abovementioned forest resources to others like ecological, recreational, soil protection, etc.

Even though a plethora of studies focused on various aspects of the Laurisilva, such as taxonomy (Llorent-Martínez et al., 2015), forest-climate interaction (Aboal et al 2013) or landscape dynamics (Arozena and Panareda 2013) can be found, the productivity of this ecosystem has not been sufficiently assessed. The low commercial value of the ecosystem services provided by the Laurisilva coppice can justify this lack of knowledge, but modern technologies like remote sensors can help to significantly reduce the cost of estimating this transcendent forest variable.

Particularly, forest applications of LiDAR technologies have been widely developed during the last decade (Wulder et al., 2013). The generalization of LiDAR information has brought the possibility of having more than one consecutive flights. Based on the same principles as traditional interval or permanent inventory plots, consecutive LiDAR flights allow us assessing important variables for forest management and planning such as productivity and growth. Another important asset of having at least two consecutive flights is to control resources through monitoring harvesting activities carried out during the inter-flight season.

One of the main objectives of the passive remote sensors has been to assess how forest systems change through time. Sentinel is currently the next-generation Earth observation mission, and provides free, full and open data for the user community. Sentinel-2 mission is particularly suitable to oversee changes in vegetation, as it monitors at high spatial, temporal and, especially, spectral resolution (Mas et al., 2016).

The aim of the conduced research at the University of Eastern Finland (UEF) is to give an operationally valid tool for discriminating different kind of typologies in the Laurisilva coppice ecosystem, based on LiDAR and passive remote sensing data, to get their productivity according to the distinctive characteristics of each typology.

2. MATERIAL AND METHODOLOGIES

2.1. Study area

The study area is located in a 1500-ha coppice forest, in the northwestern sector of the Tenerife island (Figure 1), between 200-1300 m.a.s.l. The minimum mean temperature in the coldest month is 6.8°C and the maximum mean temperature in the hottest month is 26.6°C. The mean annual precipitation is 511 mm.

Laurisilva forest ecosystem is highly valuable in economic terms but also considering ecosystem services. It is characterised by a wide variety of species, trees such as willows, but all of them has regrowth capacity. Consequently, Tenerife's evergreen forest has been extensively exploited as a coppice. Dominant species usually are Laurus, Erica, Ilex, Prunus, Myrica and Viburnum genre. The dominance of a given species depends on site conditions (Arevalo et al., 2008). Therefore, Laurisilva coppice forest generally have thin stems with mean canopy height of 10.5 m.

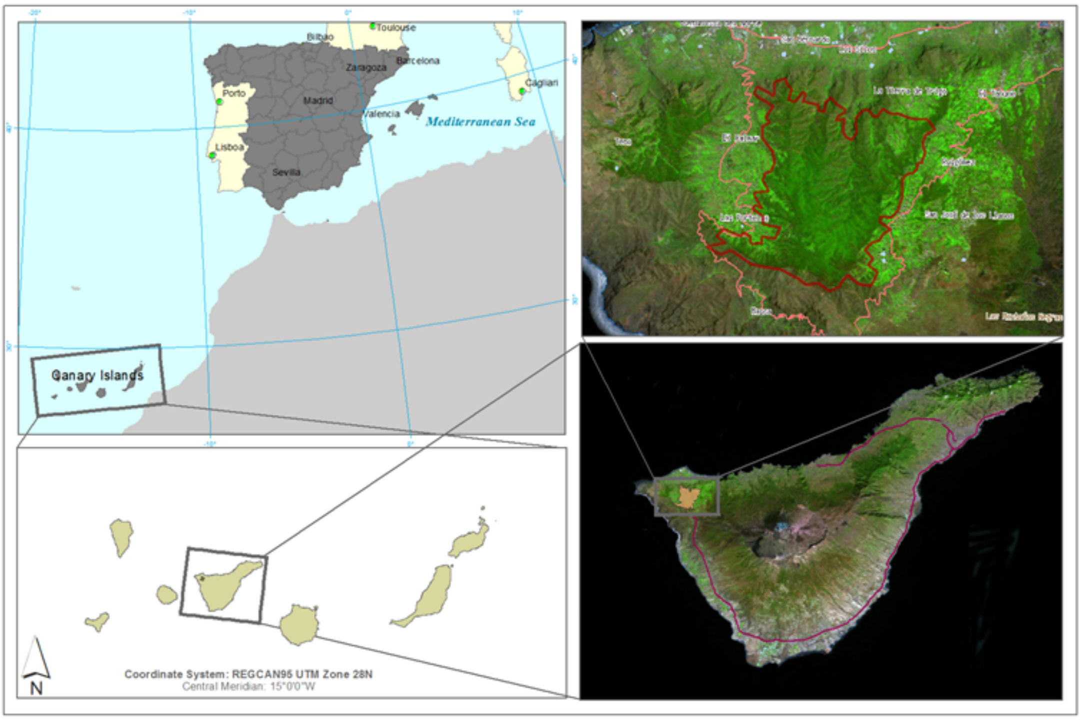

Figure 1: Map location of the study area

2.2. LiDAR DATA

Canary Islands have been scanned with a LiDAR sensor in 2009, 2012 and 2014. The average nominal point density was 0.5 pulses m^{-2}. In terms of measurement accuracy and according to flight specifications, the precision in determining X and Y information was 0.6 m while for the case of Z (height above-ground) the value was 0.20 m.

LiDAR data of the study area were provided in digital files of 2x2 km extension. Point clouds have been captured by LiDAR sensors and automatically classified and coloured taking RGB orthophotos as reference with a resolution of 25 cm. In our case, the official reference geodetic system is REGCAN95 and UTM projection in the 28N zone.

In the three datasets (one per year), LiDAR data were processed as follows: the first echoes of the pulses were used to create the canopy height model (CHM) using a 5-m raster cell as calculation unit. The last return, echoes classified as ground, were used to create the digital terrain model (CHM) raster dataset of the same cell size. The difference between CHM and DTM Z coordinate resulted in raster of above-ground information, which can be used to estimate forest characteristics (e.g. Maltamo et al., 2014). The boundary of each segment was used as a mask to compute a long array of LiDAR-derived metrics. As a result, the same LiDAR metrics were obtained for the three consecutives datasets.

Afterwards, the Digital Terrain Model (DTM) was generated with a resolution of 1m. Next, taking the generated DTM as the zero Z coordinate, the Digital Tree Height Model (DTHM) was obtained with a pixel size of 1 meter. Then, normalized LiDAR statistics were estimated on a 5x5 m raster, which will be used as independent variables in the classification model.

2.3. MULTISPECTRAL IMAGERY DATA SOURCES

Sentinel-2 is a recent sensor launched on 23 June 2015 by the European Space Agency (ESA), which have a global coverage. The promising utility of Sentinel in large-scale forest inventory project is related to the fine spatial resolution (20-10 m depending on the band) and its short interval of time between measurements for a given area (10 days). Particularly, Sentinel-2A is characterized by high temporal resolution (10 days) and variable spatial resolution, depending on the bands. Blue, green, red and near-infrared have 10-m resolution and 20-m in three red bands, one near-infrared and two more bands in the short-wave infrared. Moreover, three additional bands for atmospheric correction are collected at 60-m resolution.

The image for the study area was captured in December 2015. The season is optimal for our study given that it has no clouds and the target species are evergreen. Consequently, deciduous trees have very low reflectance in this image. The Sentinel image, previously orthorectified and georeferenced, was provided by the Government of the Canary Islands. We used the imagery data to calculate several vegetation indexes paralleling preceding studies (Table 1). The indexes aimed to measure vegetation photosynthetic activity, leading to predict biomass abundance. We included all spectral indices as potential predictors in the classification model.

Table 1: Vegetation index estimated as independent variables in the model.

Index	**Acronym**	**Equation**	**Reference**
Simple Ratio Index	SR	B08/B04	Jordan (1969)
Soil Adjusted Vegetation Index	SAVI	1.5 * (B08 - B04) / (B08+B04+0.5)	Huete (1988)
Red Edge – Normalized Difference Water Index	RE-NDWI	(B03-B05)/(B03+B05)	McFeeters (1996)
Red edge – Normalized Difference Vegetation Index	RE-NDVI	(B08-B06) / (B08+B06)	Gitelson and Merzlyak (1994)
Pigment Specific Simple Ratio	PSSR	B08/B04	Blackburn (1998b)
Normalized Difference Water Index	NDWI	(B03-B08)/(B03+B08)	Gao (1996)
Normalized Difference Vegetation Index	NDVI	(B08-B04)/(B08+B04)	Rouse et al. (1974)

Normalized Difference Vegetation Index 705	NDVI705	(B06-B05)/(B06+B05)	Monori et al. 2013
Normalized Differenced Infrared Index	NDII	(B08-B11)/(B08+B11)	Hardisky et al. (1983)
Normalized Differenced Index	NDI45	(B05-B04)/(B05+B04)	Yemefack et al. (2006)
Normalized Differenced built-up Index	NDBI	(B11-B08)/(B11+B08)	Zha et al. (2003)
Normalized Burn Ratio	NBR	(B08-B12) / (B08+B12)	Roy et al. (2006)
Moisture stress index	MSI	B11/B08	Rock and Vogelmann (1985)
Modified Normalized Differenced Water Index	MNDWI	(B03-B11)/(B03+B11)	Han-Qiu (2005)
Green-red Vegetation Index	GRVI	(B03-B04)/(B03+B04)	Motohka et al. (2010)
Green Normalized Difference Vegetation Index	GNDVI	(B08-B03)/(B08+B03)	Gitelson et al. (1996)
Enhanced Vegetation Index	EVI	2.5*(B08 - B04) / (B08 + 6*B04 - 7.5*B02 + 1)	Liu and Huete (1995)
Modified Soil Adjusted Vegetation Index 2	MSAVI2	(B08 + 1) - 0.5 * sqrt((2 * B08 - 1) ^ 2 + 8 * B04))	Laosuwan and Uttaruk (2014)

2.4. SEGMENTATION PROCESS

To define object segmentation in the study area, we executed an OBIA (Object Based Image Analysis) which created an image-object through the aggregation of pixels by image segmentation (vectorization of image data). It is a basic process to generate homogeneous objects based on a segmentation algorithm called *meanshift* from the open access software *Orfeo Tool Box* (OTB) (OTB Development Team, 2017). The two parameters which we worked with are: (i) the spatial resolution and (ii) range domains, which is the allowable spectral range within each segment for each band.

2.5. DATA ANALYSIS

Mean annual height increment (dependent variable) was calculated as the difference between the DTHM in 2012 and 2009, divided by three.

DTHM in 2014 has been initially discarded owing to the high proportion of *nodata* values. The mean annual height growth of each segment was classified in one of 5 increment categories (very low, low, medium, high and very high), based on the standard deviation and the interquartile range of the dependent variable. Break points are settled at mean+1, +2, -1 and -2 standard deviations or, in the case of the interquartile range, in such a way that every class has the same number of segments.

A classification model was built using *randomForest* package in R statistic software (Liaw and Wiener 2002) to predict the height increment class of the whole study area. As one objective of this work is to build a model to predict mean annual height increment using vegetation indicators derived from satellite information when no field data is available, the training dataset consisted of a small sample (26.33%) randomly selected from the objects. Thus, the model would be of use when only one LiDAR flight is available, combined with satellite data.

As a consequence of the notable number of candidate variables in the classification model, a previous step was necessary, drawing up a shortlist in order to remove the variables with redundant information. *VSURF* package (Genuer et al., 2016) was used to select variables that maximize the accuracy of the classification model.

3. RESULTS

3.1. HEIGHT GROWTH MAPS

Firstly, height growth maps were created as the difference between the three digital tree height models divided by the number of years in each temporal horizon. Figure 2 shows height vegetation and height growth for the three temporal horizons.

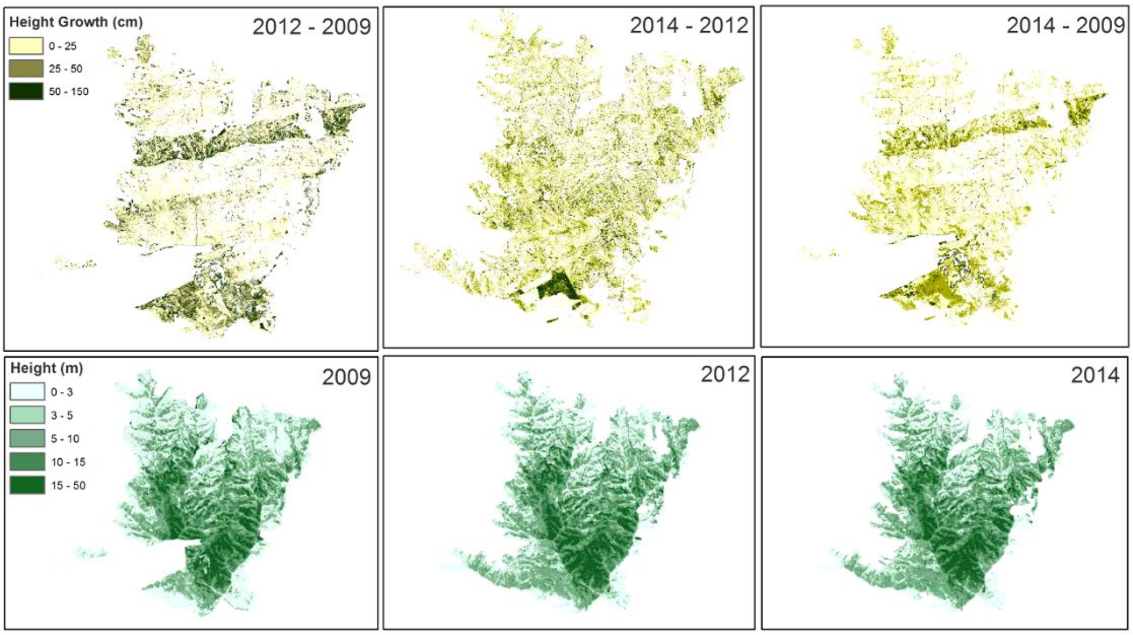

Figure 2: Height and height growth vegetation

3.2. SEGMENTATION

The result of the segmentation process was a shapefile with the study area segmented by homogeneous objects (figure 3).

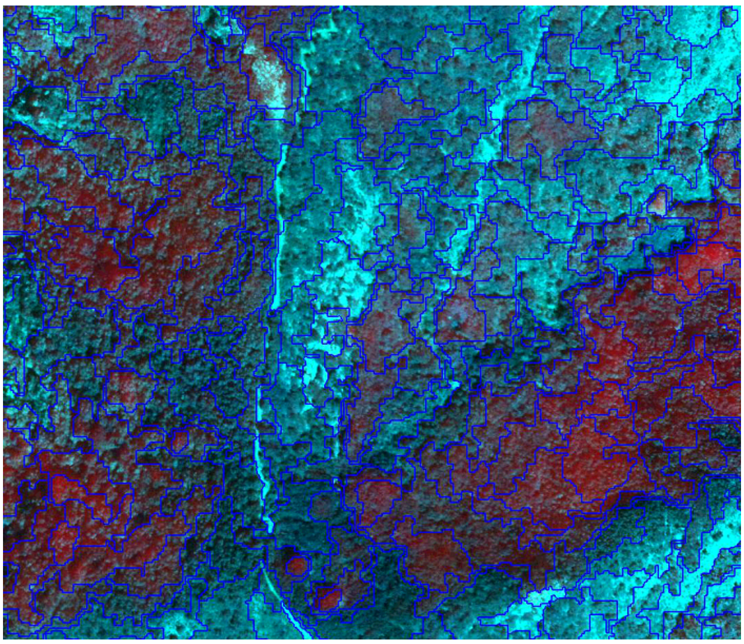

Figure 3: Segmentation area

3.3. RANDOM FOREST

Classification results based on global error of the model (OOB-error) is showed in the next table. Omission and commission errors estimated by the model were referred to percentage of cases. Table 2 shows that the OOB-error obtained by standard deviation classes is 29.39%, omission error was 53.11%, and commission error was 50.68%. Figure 4 show the variables selected by the model.

Table 2: Confusion matrix by Standard deviation classification.

	1	**2**	**3**	**4**	**5**	**Class.error**
1	4315	692	42	11	29	0.1520927
2	923	1817	177	36	36	0.3921044
3	87	299	114	28	34	0.7971530
4	31	67	42	7	30	0.9604520
5	40	78	25	19	296	0.3537118

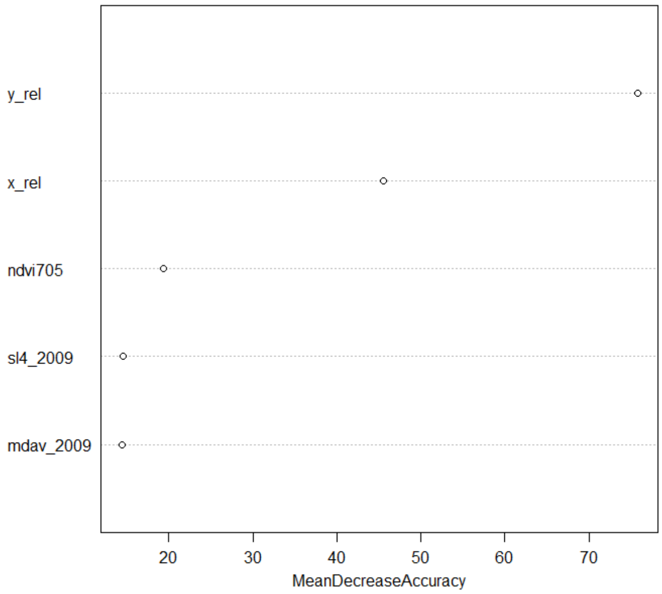

Figure 4: Variables selected by the model with Standard desviation classify.

Table 3 shows that the OOB-error obtained by Interquartile classify is 54.95%, omission error was 54.94%, and commission error was 55.18%. Figure 5 show the variables selected by the model.

Table 3: Confusion matrix by Interquartile classification.

	1	**2**	**3**	**4**	**5**	**Class.error**
1	1087	704	321	126	82	0.5314655
2	709	788	543	180	92	0.6591696
3	355	513	801	485	185	0.6575460
4	137	163	488	1017	518	0.5622040

5	75	83	157	466	1539	0.3366379

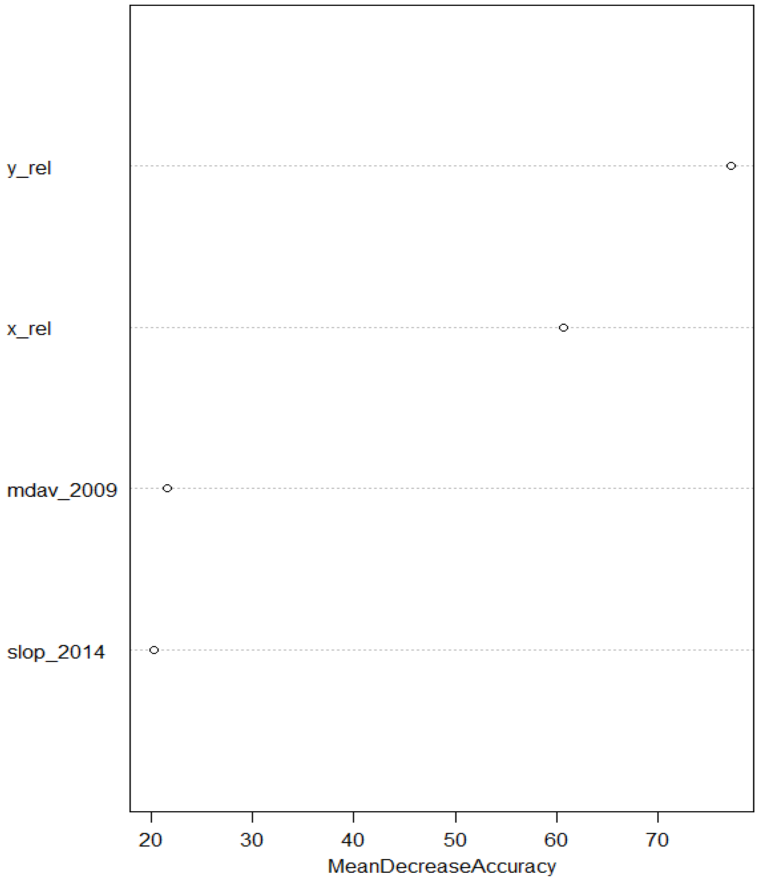

Figure 5: Variables selected by the model with Interquartile classify.

4. DISCUSSION

The main conclusion addressed through this report is the accomplishment of the STSM, since a classification tool based on growth has been developed. Nevertheless, the work done showed the evidence that further research must be carried out in order to obtain more robust models and, subsequently, enhance the classification of laurisilva coppice forests.

Some issues that may help lead to better performance of the models were found during the work. First of all, there was no field data that could support the classification with 'training areas'. Having field data would probably be transformed in more accurate models.

Secondly, only one of the spectral variables is included in the model (NDVI705). On one hand, this index capitalizes on the sensitivity of the vegetation red edge to slight changes in canopy foliage content, gap fraction, and senescence. On the other hand, we can hypothesize that the studied laurisilva coppice forests do not present enough spectral variability and, consequently, most of these variables do not provide significant information to the classification.

Classification of laurisilva coppice seems likely to be importantly driven by climatic variables. Availability of better resolution climatic variables could help with the results, particularly in rugged landscapes like the study area.

Finally, we conjecture that a simpler classification (e.g. 3 categories) of the mean annual height increment would lead to a better performance of the model. In fact, productivity is not usually a key management parameter in this type of ecosystems, hence a less accurate prediction of the variable would not jeopardize the usefulness of the results.

Thus, the future trends for this study may be based on field surveys as training samples and probably, the expansion of the study area with the aim of obtaining a wider variability of the forests.

ACKNOWLEDGEMENTS

This study wouldn't be possible without the support of the host Institution (University of Eastern Finland). Specially to Dr. Blas Mola-Yudego and Adrián Pascual. Lastly, I want to thank COST EuroCoppice (COST-STSM-FP1301-36761), Science and Innovation Spanish Ministry provided A.B.C. with through a 'Doctorado Industrial' contract (DI-14-06953) and the home institution föra forest technologies without which this study would not have been possible.

5. BIBLIOGRAPHY

Aboal J. R., Regalado C. M., Ritter A., Gómez L. A., Fernández A. B. (2013). Interceptación de lluvia y niebla en bosques de laurisilva y pinar de las Islas Canarias. La interceptación de la lluvia por la vegetación en España. Fundación Instituto Euromediterráneo del Agua, Murcia Google Scholar.

Arevalo J. R., Peraza M. D., Alvarez C., Bermudez A., Delgado J. D., Gallardo A., Fernández-Palacios J. M. (2008). Laurel forest recovery during 20 years in an abandoned firebreak in Tenerife, Canary Islands. acta ecologica, 33(1), 1-9.

Arozena M. E., Panareda J. M., Rivero B., Toledo, P. R (2015). La herencia del carboneo en el paisaje de la laurisilva canaria. Análisis espacial y representación geográfica: innovación y aplicación: 1535-1542. Universidad de Zaragoza-AGE. ISBN: 978-84-92522-95-8

Arozena M. E., Panareda J. M. (2013). Forest transition and biogeographic meaning of the current laurel forest landscape in Canary Islands, Spain. Physical Geography, 34(3), 211-235.

Blackburn G.A. (1998b) Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. International Journal of Remote Sensing 19, 657–675. doi.org/10.1080/014311698215919

Gao B.C. (1996). NDWI – a normalized difference water index for remote sensing of vegetation liquid water from space Remote Sens. Environ., 58 (1996), pp. 257–266

Genuer R., Poggi J.M., Tuleau-Malot C. (2016). VSURF: Variable Selection Using Random Forests. R package version 1.0.3. https://CRAN.R-project.org/package=VSURF

Gitelson A., Merzlyak M.N. (1994). Spectral reflectance changes associated with autumn senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral features and relation to chlorophyll estimation. J. Plant Physiol., 143 (1994), pp. 286–292

Gitelson A., Kaufman Y.J., Merzlyak M.N. (1996) Use of a green channel in remote sensing of global vegetation from EOS-MODIS Remote Sens. Environ., 58 (1996), pp. 289–298

Han-Qiu X. U. (2005). A Study on Information Extraction of Water Body with the Modified Normalized Difference Water Index (MNDWI)[J]. Journal of Remote Sensing, 5, 589-595.

Hardisky M.A., Klemas V., Smart R.M. (1983) The influences of soil salinity, growth form, and leaf moisture on the spectral reflectance of Spartina alterniflora canopies Photogrammetric Engineering and Remote Sensing, 49 (1983), pp. 77–83

Huete A. R. (1988) a soil-adjusted vegetation index (SAVI). Remote sensing of environment. 25, 295-309

Jordan C.F. (1969) Derivation of leaf-area index from quality of light on the forest floor. Ecology, 50 (1969), pp. 663–666

Laosuwan T., Uttaruk P. (2014). Estimating Tree Biomass via Remote Sensing, MSAVI 2, and Fractional Cover Model. IETE Technical Review, 31(5), 362-368.

Liaw A., Wiener M. (2002). Classification and Regression by randomForest. R News 2(3), 18-22

Liu, H. Q., Huete, A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. IEEE Transactions on Geoscience and Remote Sensing, 33(2), 457-465.

Llorent-Martínez, E. J., Gouveia, S., Castilho, P. C. (2015). Analysis of phenolic compounds in leaves from endemic trees from Madeira Island. A contribution to the chemotaxonomy of Laurisilva forest species. Industrial Crops and Products, 64, 135-151.

McFeeters S.K., (1996) The use of the normalized difference water index (NDWI) in the delineation of open water features. Int J Remote Sens 17(7):1425–1432

Moroni M., Lupo E., Marra E., Cenedese A. (2013). Hyperspectral image analysis in environmental monitoring: setup of a new tunable filter plataform. Procedia Environmental Sciences 19 (2013) 885 – 894. doi: 10.1016/j.proenv.2013.06.098

Montero G., Serrada R., (2013) La situación de los bosques y el sector forestal en España - ISFE 2013. Edit. Sociedad Española de Ciencias Forestales. Lourizán (Pontevedra)

Motohka T., Nasahara K. N., Oguma H., Tsuchida S., (2010). Applicability of green-red vegetation index for remote sensing of vegetation phenology. Remote Sensing, 2(10), 2369-2387.

OTB Develoment Team (2017). The ORFEO Tool Box Software Guide. Updated for OTB-5.10.0

Rock B.N., Vogelmann J.E., (1985) Field and airborne spectral characterization of suspected acid deposition damage in red spruce (Picea rubens) from Vermont Machine Processing of Remotely Sensed Data, Purdue Research Foundation (1985), p. 71

Rouse Jr J., Haas R. H., Schell J. A., Deering D. W. (1974 o 1973) Monitoring vegetation systems in the great plains with ETRS, In: S.C. Freden, E.P. Mercanti, M. Beker (eds). Third Earth Resources Technology Satellite-1 Symposium. Volume I: Technocal Presentations, NASA SP-351, NASA, Washington, D C., pp. 309-317

Roy D. P., Boschetti, L., Trigg, S. N., (2006). Remote sensing of fire severity: assessing the performance of the normalized burn ratio. IEEE Geoscience and Remote Sensing Letters, 3(1), 112-116.

Wulder M. A., Coops N. C., Hudak A. T., Morsdorf F., Nelson R., Newnham, G., Vastaranta, M., (2013). Status and prospects for LiDAR remote sensing of forested ecosystems. Canadian Journal of Remote Sensing, 39(s1), S1-S5.

Yemefack M., Bijker W., De Jong S. M., (2006). Investigating relationships between Landsat-7 ETM+ data and spatial segregation of LULC types under shifting agriculture in southern Cameroon. International Journal of Applied Earth Observation and Geoinformation, 8(2), 96-112.

Zha Y., Gao J., Ni S., (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. International Journal of Remote Sensing 24 (3), 583–659.