MAPPING EUCALYPTUS SPECIES USING WORLDVIEW 3 AND RANDOM FOREST

L. Alonso 1, J. Picos 1, J. Armesto 1*

1 Forestry Engineering School, University of Vigo, A Xunqueira campus, Pontevedra, 36005, Spain (laura.alonso.martinez, jpicos, julia)@uvigo.es

Commission III, WG III/10

KEY WORDS: Eucalyptus nitens, Eucalyptus globulus, High resolution multispectral image, WorldView-3, forestry

ABSTRACT:

Recent advances in remote sensing technologies have allowed the development of new innovative methodologies to obtain geospatial information about Eucalyptus genus distribution. This is an important task for forest stakeholders due to the high presence of this genus in forest plantations worldwide. Therefore, the next step in research should focus on exploring remote sensing possibilities to discern between Eucalyptus species. It would be an important step forward in forest management since different Eucalyptus species present different characteristics and properties that imply different management plans and industrial usages. This study accomplish the classification of E. nitens and E. globulus, the most common Eucalyptus species in the Iberian Peninsula. Worldview-3 images and random forest are used in a forest area placed in Galicia (Northwest of Spain). The differentiation of Eucalyptus species resulted in a producer’s accuracy of 84% and a users’ accuracy of 70% for E. nitens, while for E. globulus accuracy metrics did not reach 70%. The most important bands in the classification were the coastal blue and the blue, followed by the red related ones. The resulting unequal accuracy metrics might be caused by an imbalanced presence of both species in the selected study area. Therefore, further studies might be developed in different locations.

1. INTRODUCTION

Information about natural and productive forests in terms of species abundance, distribution and composition is essential for forest management and for stakeholders to be able to devise and execute sustainable forest management plans. Because of its particular characteristics, such as its rapid growth and great adaptability to different types of soils and climates (Queirós et al., 2020), Eucalyptus is one of the most common tree genera in forest plantations worldwide (FAO, 2005). Eucalyptus trees are planted in biomes as different as tropical and subtropical regions and temperate forests (Queirós et al., 2020). Of the approximate 700 different species that comprise this genus in its native environment (Coppen, 2002) (Australian forests), only a few are used in forestry plantations. The species most commonly encountered in plantations worldwide are E. camaldulensis, E. grandis, E. tereticornis, E. globulus, E. nitens, E. urophylla, E. saligna, E. dunnii, E. pellita and viable hybrids between different pairs of these species (Harwood C., 2011). Given the relevance of Eucalyptus in the forestry sector worldwide it is quite important for forest managers and policy makers to have information about their abundance and distribution. Additionally, nature conservation groups have a special interest in monitoring Eucalyptus populations due to the invasive potential of some Eucalyptus species in specific environments and management conditions (Forstyh et al., 2004; Calviño and van Etten, 2018; MITECO, 2017).

In Europe, Eucalyptus is most commonly planted in Portugal and in the Northwest of Spain (Bruß et al., 2011). In Portugal, for instance, it covers 845,000 ha, around 26% of the continental forest area (ICNF, 2015), and in Galicia, a region in Northwestern Spain, it covers an estimated 287,983,79 ha, accounting for 20% of the total forested area (MITECO, 2011a). In these regions, the most commonly planted species are Eucalyptus globulus and Eucalyptus nitens (MITECO, 2011a). Being able to distinguish between these two species when monitoring Eucalyptus is important since they present different characteristics and properties that imply different management plans and industrial usages. Some differences can be found in their growth rates (Pérez-Cruzado et al., 2011), frost tolerances (Close et al., 2000; Davidson et al., 2004) and their responses and susceptibilities to pathogens (Gonçalves et al., 2019; Smith et al. 2007), growth stress, wounds and pruning (Beadle et al., 2001; Dellorio et al., 2007; Wiseman et al., 2009). Also, they present crucial differences regarding their industrial properties as wood (McKinley et al., 2002) and the moisture contents of their logs (Bow and Lasserre, 2015). In this sense, the consensus is that E. globulus presents greater interest for the paper industry (Antes and Joutsimo, 2015; Pérez et al., 2006; Pérez-Cruzado et al., 2011) while E. nitens is superior from an energy point of view (Pérez et al. 2006). Differences have even been found regarding the C litter accumulation of each species (Pérez-Cruzado et al., 2011).

The monitoring of these forests in Portugal and Spain is done through field sampling recorded in National Forest Inventories (NFI) (MITECO, 2011a; ICNF, 2015). This method is highly time-consuming and therefore it is difficult to maintain updated inventories. For example, in Spain, the most up-to-date official map that is available to the public and contains information about Eucalyptus species’ distribution dates from 2011 (MITECO, 2011b). This low frequency of update can make the mapping and monitoring of areas covered by E. globulus and E. nitens difficult since, in this region, these species’ rotation cycles are short (between 12-15 years) (Tolosana et al., 2017; Arenas et al., 2019).

* Corresponding author
Nowadays, remote sensing is a widely used forest monitoring tool that can help to overcome obstacles, such as the infrequency of field assessment and inventory update. In fact, successful detection of *Eucalyptus* without identification at the species level has already been done in the regions in question using medium resolution satellites such as Sentinel-2 (Forstmaier et al., 2020; Alonso et al., 2021a) and even very high-resolution satellites such as worldview (Alonso et al., 2021b). Therefore, a logical next step in research should focus on exploring the possibilities of discerning between *Eucalyptus* species through remote sensing. Some previous studies have been done in different ecosystems to try to detect differences in the spectral response of different species of *Eucalyptus* using hyperspectral images and field spectrometers. Peerbhay et al. compile several studies where hyperspectral imagery has proven useful to distinguish between *Eucalyptus* species although the accuracy values obtained vary according to the target *Eucalyptus* species in each study (Peerbhay et al., 2013). In particular, regarding the reflectance measurements of different species, difficulties can be encountered in differentiating between some species due to the strong similarities in their spectral reflectances (Kumar, 2007; Dutt, 1999). In their work, Peerbhay et al. (2013) also investigated hyperspectral imagery to distinguish between the most commonly planted South African *Eucalyptus* and they observed that the most important wavebands were located within the visible region so it may be worth investigating the capabilities of high-resolution multispectral sensors for this purpose. However, few studies have been found that address these capabilities. One example is the subsequent study of Peerbhay et al. (2014) that used Worldview-2 to distinguish between *E. grandis*, *E. nitens* and *E. smithii* in South Africa, obtaining user’s and producer’s accuracies of between 80 and 100 percent (Peerbhay et al., 2014). Verma et al. (2019) have also been able to distinguish between 5 *Eucalyptus* species (E. bridgesiana, E. caliginosa, E. blakelyi, E. viminalis, E. melliodora) in an Australian native forest obtaining accuracies of 65% using multispectral high-resolution images (VIS+NIR) and LiDAR data. However, none of these studies examine both *E. globulus* and *E. nitens* which, as mentioned before, are included among the few species of *Eucalyptus* that are commonly used in plantations worldwide. Therefore, this study seeks to accomplish the classification of *E. globulus* and *E. nitens* using high-resolution multispectral images, specifically Worldview-3 data. It is conducted in Galicia, one of the European regions with the most *Eucalyptus* plantations.

2. MATERIALS AND METHODS

2.1 Study area

The study area is located in the Northwest of the Iberian Peninsula, in Galicia, Spain (see Figure 1). A pilot area where both *Eucalyptus* species coexist was selected in this region. It covers a total of 160.2 km².

![Figure 1. Study area. Reference image PNOA (National Aerial Orthophotography Plan, for its initials in Spanish) image (IGN, 2021).](image)

<table>
<thead>
<tr>
<th>Band</th>
<th>Spectral range (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coastal (MS Band 1)</td>
<td>400 - 450</td>
</tr>
<tr>
<td>Blue (MS Band 2)</td>
<td>450 - 510</td>
</tr>
<tr>
<td>Green (MS Band 3)</td>
<td>510 - 580</td>
</tr>
<tr>
<td>Yellow (MS Band 4)</td>
<td>585 - 625</td>
</tr>
<tr>
<td>Red (MS Band 5)</td>
<td>630 - 690</td>
</tr>
<tr>
<td>Red Edge (MS Band 6)</td>
<td>705 - 745</td>
</tr>
<tr>
<td>NIR1 (MS Band 7)</td>
<td>770 - 895</td>
</tr>
<tr>
<td>NIR2 (MS Band 8)</td>
<td>860 - 1040</td>
</tr>
</tbody>
</table>

Table 1. Spectral characteristics of Worldview-3 images.

The detailed observation of the images revealed that the scene contains shadows caused by the tree vegetation height, the satellite azimuth and the sun’s elevation. These shadows can compromise classification results as the spectral response in shadowed areas is greatly altered.
2.3 Field Data

The training polygons and ground truth data regarding the Eucalyptus species in the study area were obtained through the measurement of sample plots and in situ observation of forest stand variables. The location of the plots was established using the latest available Spanish Forest Map, MFE25 (MITECO, 2011b), and through the photointerpretation of PNOA images (IGN, 2021). The plots centers were defined as georeferenced vector points. The position was established in areas that, according to the PNOA images (IGN, 2021) and the MFE, contained the target species in single-species, mature stands. The sample plots for each species were distributed among varying topographical orientations and slopes in order to collect the spectral variability that each species may present due to variations in topographical conditions.

The field work consisted of measuring plot and stand variables. Plot variables included the position of the plot center collected using a GPS device with centimetric precision. The stand parameters were observed in a circular area of a 10m radius. They included: tree species composition, estimated tree height, health status and canopy cover. Photographs and background information were included for each plot to aid in the interpretation of the data.

Once the field work was performed, the data was put into digital format and reviewed. Sample plots were discarded in cases where the forest stands were non-homogeneous in terms of tree species, or where the stand characteristics differed significantly from optimal ground truth data (e.g., leaf bleaching or extremely low canopy cover). A total of 140 Eucalyptus plots were obtained. They were used for training and for verification.

2.4 Methodology

2.4.1 Image pre-processing

Images were first pre-processed to remove shadows. This was done using the shadow detection index (SDI) described by Shahi et al. (2014), which is specific for WorldView images. This index depends on the Blue band and the two NIR bands. The index equation is presented following (equation 1).

\[ SDI = \frac{NIR2-Blue}{NIR2+Blue} \]

2.4.2 Eucalyptus mapping

A general classification was performed to obtain the area covered by Eucalyptus in the study area. The methodology described by Alonso et al. (2021b) was followed. The target classes of the classification were: Eucalyptus, Young Eucalyptus, Conifers, Other Broadleaves, Shrubs, Crops and Pastures, Anthropogenic Areas and Water. The Eucalyptus class refers to Eucalyptus trees with adult leaves. Young Eucalyptus refers to Eucalyptus trees covered mainly by young leaves either because they are in an early stage of development or because they are outgrowth. This differentiation was made in order to avoid intra-specific spectral mixtures in the subsequent step of Eucalyptus species differentiation since the young and adult leaves of each Eucalyptus species show a clear difference in coloration, and therefore exhibit different spectral responses.

Once the target classes were identified, training areas were collected for each class. They were obtained through photointerpretation of orthorectified aerial images obtained from the Spanish cartographical institute (IGN, 2021) (PNOA, for its acronym in Spanish). The PNOA image was obtained in 2020 and it had a spatial resolution of 0.25 m and an RMSE x,y = 0.50 m. Afterwards, the supervised classification was performed by applying the random forest algorithm (Breiman, 2001) to all of the available bands. Finally, a random sample of 240 points was distributed around the study area to obtain a confusion matrix and verify the results obtained.

2.4.3 Eucalyptus species classification

Once the areas covered by adult and young Eucalyptus were mapped, the differentiation between species was performed. This section describes the classification of species in adult stands. The analysis of species in young-leaved Eucalyptus stands was discarded due to the impossibility of finding enough plots in the study area to build consistent training datasets for each species. The first step in performing the species classification in the adult stands was to mask out the Worldview-3 image with those areas covered by the Eucalyptus class according to the general map obtained in the previous section. By doing this, a new raster file was obtained containing the spectral information only of the pixels occupied by adult-leaved Eucalyptus. The species classification was applied to this masked image.

The training dataset was built using the sample plots obtained through field work. A total of 36 field plots were used to define polygons by a manual vectorizing of the areas adjacent to the sample plots. This process yielded a set of 103,561 pixels corresponding to E. nitens and a set of 28,070 pixels corresponding to E. globulus. These training pixels were used to perform a supervised classification of the image using the Random forest algorithm (Breiman, 2001) which was applied using randomForest from the R software library (Liaw and Wiener, 2002). The default configuration parameters were used (Number of trees: 500). As a result, each Eucalyptus pixel was classified according to Eucalyptus species.

A cross-verification process was performed using the sample plots obtained through field work that were not used in the training step. They were 64 and 42 plots for E. nitens and E. globulus, respectively. This plots were independent for the ones used for training. A confusion matrix was built; The User’s Accuracy (UA), Producer’s Accuracy (PA), Overall Accuracy (OA) and Kappa Index were calculated. Each variable’s importance was also calculated, based on its Mean decrease in Gini (69), in order to find out which variables were the most valuable in the random forest prediction.

3. RESULTS

3.1 Image Pre-processing

The index applied allowed for the successful removal of shadows. An example is shown in Figure 2. Figure 2A. shows an amplified area of the original image in false color infrared. Shadowed pixels can be detected, especially on the edges of the plantation. The Figure 2B. shows the result of the shadow index, bright areas correspond to shadows. Figure 2C. shows the image after applying the shadow mask. Areas represented in white are removed shadows whose pixels were assigned NULL values. These NULL values correspond to areas where the SDI had a value above the 95th percentile, estimated at -356.04, of the SDI histogram. As can be seen in the image, large shadowed areas were removed while some small shadows within the tree stands were left unmodified.
3.2 Eucalyptus mapping

An example of the general classification obtained for the study area is shown in Figure 3. It illustrates that young and adult Eucalyptus stands are differentiated from one another, as well as from other tree covers. The cross verification yielded high accuracy metrics (An overall accuracy of 90%). The results of the cross verification are shown in Table 2. In tree-related classes, high accuracy metrics were obtained with UAs and PAs of between 77% and 100%. The exception was the Young Eucalyptus class where the PA decreased to 43%. This might be due, as mentioned before, to the scarcity of young-leaved Eucalyptus stands detected in the study area on the image acquisition date. The Shrub class, which is often confused with the crop and pasture class, also had low accuracy metrics. It should be highlighted that the PA and the UA were above 90% for the Eucalyptus class, which is the target study class.

<table>
<thead>
<tr>
<th>Class</th>
<th>Ground Truth</th>
<th>PA (%)</th>
<th>UA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eucalypts</td>
<td>55</td>
<td>91</td>
<td>98</td>
</tr>
<tr>
<td>Young Eucalypts</td>
<td>14</td>
<td>43</td>
<td>75</td>
</tr>
<tr>
<td>Conifers</td>
<td>15</td>
<td>100</td>
<td>83</td>
</tr>
<tr>
<td>Broadleaves</td>
<td>34</td>
<td>100</td>
<td>77</td>
</tr>
<tr>
<td>Crops and pastures</td>
<td>86</td>
<td>98</td>
<td>91</td>
</tr>
<tr>
<td>Shrub</td>
<td>12</td>
<td>17</td>
<td>67</td>
</tr>
<tr>
<td>Anthropogenic areas</td>
<td>22</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Water</td>
<td>2</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>TOTAL</td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OA (%)</td>
<td>90</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Results of the cross-verification of the general classification. Ground Truth: Number of Ground Truth points; PA: Producer’s Accuracy; UA: User’s Accuracy, OA: Overall Accuracy.

3.3 Eucalyptus species classification

An example of the classification of Eucalyptus species obtained for the study area is shown in Figure 4. The results of the cross verification performed on the supervised classification, which was used to distinguish between *E. nitens* and *E. globulus*, are presented in Table 3. The Overall accuracy obtained was 69%. Higher accuracy metrics were obtained for *E. nitens*, with a UA and a PA of over 70%, than for *E. globulus*, which had a UA and a PA of 65% and 45%, respectively. This great difference in accuracies is reflected in the Kappa Index obtained, which was 0.31.

<table>
<thead>
<tr>
<th>Class</th>
<th>Ground Truth</th>
<th>PA (%)</th>
<th>UA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. nitens</td>
<td>55</td>
<td>75</td>
<td>95</td>
</tr>
<tr>
<td>E. globulus</td>
<td>14</td>
<td>65</td>
<td>45</td>
</tr>
</tbody>
</table>

Figure 4. Example of result of the Eucalyptus species classification. Reference image WorldView image in false color infrared (B8, B5, B3). White correspond to shadowed areas.
Table 3. Confusion matrix of the supervised classification performed to distinguish between \textit{E. nitens} and \textit{E. globulus}.

<table>
<thead>
<tr>
<th>Class/Label</th>
<th>\textit{E. globulus}</th>
<th>\textit{E. nitens}</th>
<th>Total</th>
<th>PA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{E. globulus}</td>
<td>19</td>
<td>23</td>
<td>42</td>
<td>45</td>
</tr>
<tr>
<td>\textit{E. nitens}</td>
<td>10</td>
<td>54</td>
<td>64</td>
<td>84</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>77</td>
<td>106</td>
<td>OA</td>
</tr>
<tr>
<td>UA (%)</td>
<td>65</td>
<td>70</td>
<td>OA</td>
<td>69%</td>
</tr>
</tbody>
</table>

The variable importance results for the random forest prediction are shown in Figure 5. The most important variables for distinguishing between these two species of \textit{Eucalyptus} are the Coastal Blue Band (B01) and the Blue band (B02), followed by the red-related bands: Red Edge, NIR2, Red and NIR1. The least important bands were the Green and Yellow bands.

4. DISCUSSION

4.1 Image Pre-processing

The shadow index was applied successfully in a totally different environment to the one used in the original study by Shahi et al. (2014). Given that the authors had called for further research to confirm the effectiveness of the shadow index in different environments, its effectiveness in this study represents a step forward in terms of the applicability of the index. This case examines a forested environment, where the SDI was useful for removing large shadows; however, small shadows still remain un-removed. This observation is congruent with the observations made by Shahi et al. (2014) in an urban environment.

It is worth noting that in this study, an additional step is introduced to automate the mapping of shadows. This step is the use of the SDI raster histogram to distinguish between shadowed and non-shadowed areas. Other studies rely, as well, on the utilization of a shadow-specific index histogram to discern between shadowed and non-shadowed areas (Mostafa et al., 2019).

4.2 General Classification

The accuracy results of the supervised classification performed to obtain the distribution of the general land cover classes present in the study area were in line with the accuracy results obtained in a similar study by Also et al. This is especially true for the Eucalyptus, Conifer and Broadleaf classes. However, Alonso et al. (2021b) obtained better accuracy results for the Shrub and Young Eucalyptus classes. Such a difference might be due to a greater abundance of these classes in their study area, allowing for a greater amount of training and verification data.

4.3 \textit{Eucalyptus} Classification

In the classification of \textit{E. globulus} and \textit{E. nitens}, the two species yielded quite different accuracy metrics. Lower values were obtained for \textit{E. globulus}. This might be due to a lesser presence of this species in the study area. In fact, in the field work campaign it was noted that \textit{E. nitens} stands were larger and more extensive than the \textit{E. globulus} stands, the repercussion being an imbalance in the amount of training and verification data for the two species. It would therefore be worth performing this study in a different area where the amount of data for each species were more comparable. However, it should be kept in mind that distinguishing between these two species may not be an easy task due to the great similarities in their spectral reflectances, something that has been reported among other species of the genus \textit{Eucalyptus} as well (Kumar, 2007).

The variable importance results are in line with the Peerhbay et al. (2013) study, which indicated that the bands in the spectral range of 393–723 nm were the most useful ones. In fact, in this study, the most important bands were the coastal blue and the blue band, whose spectral ranges go from 400 nm to 510 nm. This might incite the scientific community to continue exploring the possibilities of this use of remote sensing data to monitor \textit{Eucalyptus} species. It may also be worth studying the possibility of combining spectral data with additional information, for example textural features. In this line, Verma et al. (2019) observed that accuracy increases when spectral and textural information is considered in conjunction in order to distinguish between some species of \textit{Eucalyptus} in their native environment.

Finally, it should be mentioned that the results of this study show that \textit{E. nitens} is more abundant than \textit{E. globulus} in the study area. Nonetheless, the latest Spanish NFI report deemed the presence of \textit{E. nitens} as residual and restricted (MITECO, 2011a). A possible interpretation of this discrepancy is the rapid increase in the presence of \textit{E. nitens} in recent years, something which has been mentioned in recent scientific studies (Pérez-Cruzado et al., 2011). This reveals the importance of continuing to develop remote sensing methods and strategies in order to maintain forest species cartography up-to-date, a task which is essential for the design and execution of proper forest management actions.

5. CONCLUSIONS

In this study the potential of WorldView-3 for \textit{Eucalyptus} mapping at the species level is evaluated in a temperate climate study area in Spain. Specifically, we have focused on \textit{Eucalyptus globulus}, one of the most prominent species in forestry in recent decades, and \textit{Eucalyptus nitens}, a species which is gaining relevance. A methodology to automate the removal of shadows from the scene was successfully applied. Subsequently, the Random forest classifier was used to obtain a general map of the different land covers in the study area. Adult stands of \textit{Eucalyptus} were differentiated from young stands, as well as from other common forest covers. The adult stands were successfully mapped with very high accuracy, while the lesser effectiveness of mapping young stands suggests that further research is necessary. This could be done, it seems, in areas where there are abundant young stands of each of the two species.

The differentiation of \textit{Eucalyptus} species resulted in a producer’s accuracy of 84% and a users’ accuracy of 70% for \textit{E. nitens}, while for \textit{E. globulus} accuracy metrics did not reach 70%. Being the results satisfactory enough for the first species, the unequal metrics for the other one suggest that the mapping of these species should be explored in different locations were the imbalance in stand abundance between the classes is less..
significant. Also further research could be done using different satellites (e.g. Sentinel-2) or using additional information (e.g. textural information). Maps at eucalyps species level are essential to track forest policies and plan future management actions.

REFERENCES


Arenas, S., Rodríguez-Soalleiro, R., Díaz-Balteiro, L., 2019. Turno óptimo de *Eucalyptus nitens* en Galicia introduciendo la fiscalidad en el análisis. *XII Congreso de Economía Agraria*.


ICNF (Instituto da Conservação da Natureza e das Florestas ), 2015, 6º Inventario Forestal Nacional. Relatório Final.

Kumar, L., 2007. A comparison of reflectance characteristics of some Australian eucalyptus species based on high spectral resolution data — discriminating using the visible and NIR regions. *Journal of Spatial Science*.


MITECO, 2017, Solicitud de dictamen sobre la posible inclusión de *Eucalyptus camaldulensis*, *E. globulus*, *E. nitens* y cualquier otra especie del género *Eucalyptus* en el Catálogo Español de Espesies Exóticas Invasoras regulado por el R.D. 630/2013, de 2 de agosto y remitida al MAPAMA por el Alcalde del Ayuntamiento de Teo (A Coruña). MITECO.


