TOWARDS A MAP OF THE EUROPEAN TREE COVER BASED ON SENTINEL-2

Thor-Bjørn Ottosen[1], Geoffrey Petch[1], Mary Hanson[1] and Carsten Ambelas Skjøth[1]

[1]Institute of Science and the Environment, University of Worcester, Worcester, UK

ABSTRACT

Many areas of science and policy depend on knowledge of the tree cover in Europe. Sentinel-2 is a new (launched in 2015) satellite with a higher spatial resolution compared to previous satellites. In the present study a new algorithm for mapping tree cover from Sentinel-2 is developed, an analysis of which bands should be used for tree cover mapping is made, the accuracy of the mapping is assessed, and the tree cover from the present approach is compared with previous estimates. Firstly, the feasibility of the present algorithm is demonstrated. Secondly, it is shown that only ten band combinations have good performance in four selected Sentinel-2 tiles and that the bands 3, 5, 6, 12 appear in most combinations. Thirdly, the accuracy is assessed to be high, and lastly it is shown that the relative difference between the tree cover of the present study and the tree cover of previous studies is between -14% and 68%.

Index Terms— *Algorithm design, spectral bands, broadleaved trees, coniferous trees, Intel DAAL.*

1. INTRODUCTION

The European tree cover is important for many areas of science and planning such as climate [1], atmospheric composition [2, 3] and socio-economic values [4]. Until recently, the best estimate of the European tree cover has been $1.47 \cdot 10^6 km^2 \pm 0.02 \cdot 10^6 km^2$ based on either a combination of MODIS images and LISS-3 images [5] or Landsat images [6] with a spatial resolution of 25 m – 30 m. In old cultural landscapes, like the UK and other European countries, trees are often located in small linear features or groups and along roads and rivers, thus making them difficult to detect using satellites such as Landsat. It is thus likely that the actual tree cover is higher than previously thought and that woodland is more widely distributed.

With the launch of Sentinel-2, a new tool has become available with a high spatial resolution and a band combination specifically designed for vegetation studies. However, a number of methodological questions remain to be answered before a European tree map can be produced from Sentinel-2. In this paper we explore a methodology for tree cover mapping from Sentinel-2, we analyse the performance of all 8100 potential band combinations for tree mapping, assess the accuracy of the methodology and compare with previous estimates for four Sentinel-2 tiles.

1. METHODS

1.1. Mapping tree cover

The tree mapping algorithm consists of a number of steps carried out sequentially:

1. Clouds and other artifacts are removed using the accompanying masks.
2. All bands are resampled to 10 m × 10 m.
3. All bands are normalized by mean centering and division with the standard deviation. This is done to normalize the weight of the individual bands.
4. A K-means unsupervised classification is done using Intel Data Analytics Abstraction Library (DAAL) with the number of classes set to 25.
5. The NDVI values of the pixels corresponding to forests in Corine Land Cover (CLC) [7] are extracted from the image.
6. Pixels from the entire image are subsequently removed if their NDVI value is less than μ-2σ of the distribution of values extracted in step 5. In this way, non-vegetation pixels are efficiently removed.
7. The dominating classes for respectively broadleaved and coniferous forests are labelled using CLC by K-means clustering the clusters from the first clustering into two classes (dominating and non-dominating). This is done iteratively starting from the largest polygons within the Sentinel-2 tile proceeding to the smallest polygon until convergence. Convergence is defined when the largest change in a class is smaller than 1%.

1.2. Analysis of band combinations

To analyze whether all spectral bands were needed for this algorithm, and if not, which spectral bands provided the best performance, the algorithm was applied to all 8100 possible band combinations of size 3 to 13 on four Sentinel-2 tiles covering selected areas of Northern Europe. These tiles are named 30UWC, 30VUH, 32NVH and 33VUC. The kappa-coefficient for the wall-to-wall comparison with the respective national forest inventory was subsequently calculated and the performance of each band combination

doi: 10.2760/383579

ranked by summing the kappa coefficient over the four images.

1.3. Accuracy assessment

The accuracy assessment of the present study was performed on tile 30UWC.

1.3.1. Sampling design

1000 pixels were selected across the image using stratified random sampling [8]. The pixels were stratified into *broadleaved trees*, *coniferous trees* and *no trees* to prevent the non-forest category dominating the results.

1.3.2. Response design

The primary land use class of each pixel was subsequently manually determined using Google Earth. The interpreter did not have access to the classified map during the manual classification to avoid biasing the classification (blind interpretation). To enhance consistency among interpreters, a written guide to the classification procedure was produced and 99 points were classified by all interpreters. As well as *broadleaved trees, coniferous trees* and *non-forest*, the interpreter had opportunity to classify a pixel as *unclassified* and *unclassified trees.* Data points that were classified to be in the last two categories were subsequently excluded from the analysis.

2. PRELIMINARY RESULTS AND DISCUSSION

2.1. Mapping tree cover

Figure 1 shows an example of a result of the labelling procedure of the 25 classes produced by the unsupervised classification algorithm. The blue bars are the areas marked as broadleaved forest in CLC and the orange bars are the areas marked as coniferous forest in CLC. In this case, classes 18 and 22 dominate broadleaved forests, whereas classes 8 and 12 dominate coniferous forests. It is likewise evident, that these four classes are the dominant ones and therefore the ones labelled as forests since a complete separation cannot be achieved as the CLC land cover by definition will miss small woodland areas outside forests and miss small bare patches inside forest regions.

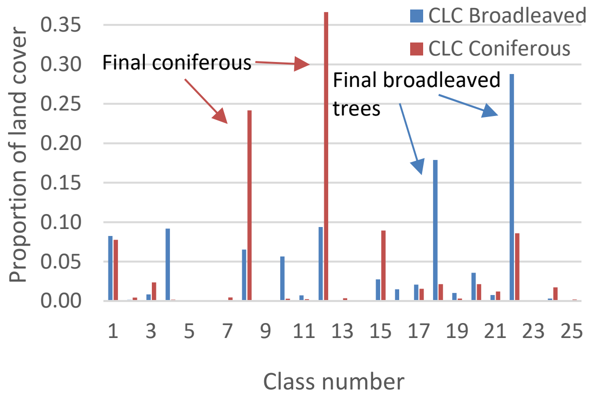

Figure 1 Example of a result of the labelling procedure.

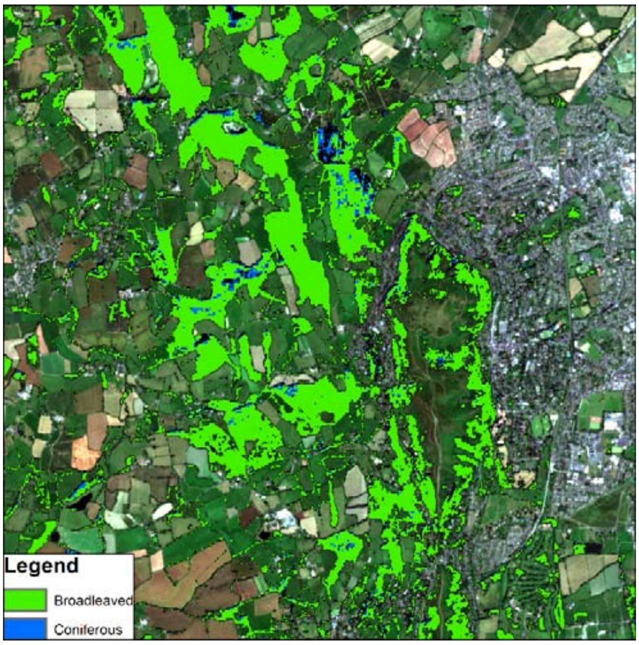

Figure 2 Example of a tree map on the border between Herefordshire and Worcestershire, UK (tile 30UWC) superimposed on a RGB-image from Sentinel-2. The scale is 1:75,000.

An example of a classification result is shown in Figure 2 superimposed on a RGB-image from Sentinel-2. The area is dominated by broadleaved forests with only small patches of coniferous forests. It is evident from this figure, that the algorithm allows the mapping of trees with a high degree of detail. This is seen from the many small features in the left side of the image and the trees mapped in the urban area in the right side of the image. These small features are by definition not a part of the CLC land cover or the more detailed UK tree cover map. Furthermore the map also contains a separation into coniferous and broadleaved trees.

2.2. Analysis of band combinations

Table 1 shows the ten band combinations that appear among the top-5% combinations in all four images. It is evident that bands 2, 3, 6 and 12 appear in many of the combinations. Band 2 is the blue band (496.6 nm), band 3 is the green band (560.0 nm), band 6 is a red-edge band (740.2 nm), and band 12 is a short-wave infrared band (2202.4 nm). Using USGS Spectral Characteristics Viewer it can be seen that these bands are particularly suitable to separate different types of vegetation.

Table 1 Band combinations appearing in top-1500 in all four images and their corresponding kappa-sum.

$\sum \kappa_i$	Combination
2.755	1, 2, 3, 4, 5, 7, 9, 12
2.803	2, 3, 6, 12
2.774	1, 3, 5, 6, 12
2.771	1, 2, 3, 5, 6, 12
2.800	1, 3, 4, 5, 6, 11, 12
2.797	2, 4, 5, 6, 12
2.799	3, 5, 6, 11, 12
2.797	2, 3, 4, 5, 6, 11
2.757	3, 4, 5, 6, 7, 9, 11, 12

Out of these combinations, the combination 2, 3, 6 and 12 was chosen, since it has a good performance in all four images.

2.3. Accuracy assessment

Table 2 Confusion matrix for the satellite map for tile 30UWC versus google earth. 0 is non-forest, 1 is broadleaved forest, 2 is coniferous forest, n=941.

	Classified data			Producer's Acc. (%)
Reference data	0	1	2	
0	303	67	18	0.78
1	23	228	182	0.53
2	3	10	107	0.89
User's Acc. (%)	0.92	0.75	0.35	0.68

The results of the accuracy assessment are shown in Table 2. Compared to the commonly used success criteria of 85% accuracy, an overall accuracy of 68% is low. The algorithm is especially challenged in separating broadleaved trees from coniferous trees. This is natural, since these two classes are spectrally much more similar compared to the non-forest class. This shows that the map tends to overestimate the cover of coniferous forest. Part of this is the result of shadows on forest edges, forest roads etc. that gives the trees a darker color. Future work should aim at reducing this effect. A group of points are classified as forests, and referenced as non-forest. This is largely caused by green fields being spectrally similar to forests. In the future we will implement a temporal averaging procedure over several images, which we hope will reduce this effect. The accuracy for the map (area × accuracy) is 0.76 or 0.90 depending on whether the user's accuracy or the producer's accuracy is used. Averaging the two numbers gives 83%, which is very close to the commonly used limit. The map accuracy is 88% if classified as only *trees* and *no trees*. Considering that the accuracy assessment is done on 10 m × 10 m resolution, which is higher than previous maps, this result must be considered acceptable.

2.4. Comparison with previous datasets

Table 3 compares the amount of tree cover in the dataset of [5] with the tree cover mapped from Sentinel-2 for four tiles in Northern Europe. As can be seen, the relative difference is between -14% and 68% with Sentinel-2 often estimating a larger tree cover. Generally, this looks like a realistic result. Widespread clouds, where the accompanying cloud mask does not remove some of them, influence 32VNH. It is thus plausible that this value is an overestimation. That Sentinel-2 has the lower of the two numbers in 33VUC is likewise an effect of high clouds that have not been removed by the cloud mask. This demonstrates the ability of Sentinel-2 to detect smaller groups of trees compared to previous satellites. It is expected that using several images over the same area will either completely remove or at least reduce the effect of clouds, thus obtaining more accurate estimates of the tree cover in cultural landscapes over very large areas like Europe.

Table 3 Comparison between the tree cover in our map and the tree cover in the map of [5].

Tile code:	Location:	Our map (km^2)	Previous maps (km^2)
30UWC	Worcester, UK	1148	909
30VUH	Scotland	2204	1407
32VNH	West Denmark	2387	1178
33VUC	East Denmark	2175	2515

3. CONCLUSION AND OUTLOOK

The present study developed an algorithm for automatic tree cover mapping based on Sentinel-2 and demonstrated that this is a feasible approach to tree cover mapping, analyzed the performance of 8100 band combinations, assessed the accuracy, and demonstrated that Sentinel-2 can contribute to more accurate tree cover maps of Europe, in particular by identifying smaller woodlands that are important in some areas. In the future, this algorithm will be extrapolated to the entire European domain, to obtain a complete estimate of the European tree cover divided into broadleaved and coniferous trees.

doi: 10.2760/383579

4. ACKOWLEDGEMENTS

The present study has been supported by the BBSRC funded project *New approaches for the early detection of tree health pests and pathogens, projectID: BB/L012286/1.*

5. REFERENCES

[1] Bonan, G.B., "Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests." *Science* (80-.). 320, 1444 LP-1449, 2008.

[2] Kesselmeier, J., Staudt, M., "Biogenic Volatile Organic Compounds (VOC): An Overview on Emission, Physiology and Ecology." *J. Atmos. Chem.* 33, pp. 23–88, 1999.

[3] Pauling, A., Rotach, M.W., Gehrig, R., Clot, B., "A method to derive vegetation distribution maps for pollen dispersion models using birch as an example". *Int J Biometeorol* 56, pp. 949–958, 2012.

[4] FAO, *Global Forest Resources Assessment 2015*, FAO Forestry, 2015.

[5] Kempeneers, P., Sedano, F., Seebach, L., Strobl, P., San-Miguel-Ayanz, J., "Data fusion of different spatial resolution remote sensing images applied to forest-type mapping." *IEEE Trans. Geosci. Remote Sens.* 49, pp. 4977–4986, 2011.

[6] Hansen, M.C., Potapov, P. V, Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S. V, Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G.,. "High-Resolution Global Maps of 21st-Century Forest Cover Change." *Science* (80-.). 342, 850 LP-853, 2013.

[7] EEA, *CORINE land cover technical guide – Addendum 2000*, EEA, Copenhagen, 2000.

[8] Stehman, S. V., "Sampling designs for accuracy assessment of land cover". *Int. J. Remote Sens*. 30, 5243–5272. 2009.