

# Forest species mapping using airborne hyperspectral APEX data

## Abstract

The accurate mapping of forest species is a very important task in relation to the increasing need to better understand the role of the forest ecosystem within environmental dynamics. The objective of this paper is the investigation of the potential of a multi-temporal hyperspectral dataset for the production of a thematic map of the dominant species in the Forêt de Hardt (France). Hyperspectral data were collected in June and September 2013 using the Airborne Prism EXperiment (APEX) sensor, covering the visible, near-infrared and shortwave infrared spectral regions with a spatial resolution of 3 m by 3 m. The map was realized by means of a maximum likelihood supervised classification. The classification was first performed separately on images from June and September and then on the two images together. Class discrimination was performed using as input 3 spectral indices computed as ratios between red edge bands and a blue band for each image. The map was validated using a testing set selected on the basis of a random stratified sampling scheme. Results showed that the algorithm performances improved from an overall accuracy of 59.5% and 48% (for the June and September images, respectively) to an overall accuracy of 74.4%, with the producer's accuracy ranging from 60% to 86% and user's accuracy ranging from 61% to 90%, when both images (June and September) were combined. This study demonstrates that the use of multi-temporal high-resolution images acquired in two different vegetation development stages (i.e., 17 June 2013 and 4 September 2013) allows accurate (overall accuracy 74.4%) local-scale thematic products to be obtained in an operational way.

## Keywords

Vegetation map • Hyperspectral • Aerial • Supervised classification • Multi-temporal dataset • Forest ecosystem

© University of Warsaw – Faculty of Geography and Regional Studies

**Giulia Tagliabue<sup>1</sup>, Cinzia Panigada<sup>1\*</sup>, Roberto Colombo<sup>1</sup>, Francesco Fava<sup>1</sup>, Chiara Cilia<sup>1</sup>, Frédéric Baret<sup>2</sup>, Kristin Vreys<sup>3</sup>, Koen Meuleman<sup>3</sup>, Micol Rossini<sup>1</sup>**

<sup>1</sup>Remote Sensing of Environmental Dynamics Laboratory (LTDA), Department of Sciences and Technologies for Environment and Landscape (DISAT), University of Milano-Bicocca, Italy

e-mail: g.tagliabue9@campus.unimib.it, cinzia.panigada@unimib.it, roberto.colombo@unimib.it, francesco.fava@unimib.it, chiara.cilia@gmail.com, micol.rossini@unimib.it

<sup>2</sup>Institut National de la Recherche Agronomique (INRA), France

e-mail: baret@avignon.inra.fr

<sup>3</sup>VITO Vlaamse Instelling voor Technologisch Onderzoek, Belgium

e-mail: kristin.vreys@vito.be, koen.meuleman@vito.be

Received: 10 June 2015

Accepted: 15 December 2015

## Introduction

Forests play a key role in local and global scale environmental dynamics; therefore, the temporal and spatial monitoring of their functioning is critical for the purpose of efficient and sustainable management (Linke et al., 2006). In this context, remote sensing is a very useful tool compared with traditional surveying techniques, as it allows the production of large scale digital thematic maps in a fast and reasonably accurate way (Franklin 2001, Panigada et al. 2010). In particular, the composition and distribution of forest ecosystems is a fundamental factor within the carbon, nitrogen and water biogeochemical cycles. For this reason, the accurate mapping of forest species is a very important task. In the last few years, several studies have been carried out using different remote sensing sensors, evaluating the potentialities of both passive sensors (i.e., multispectral and hyperspectral sensors) and active systems (i.e., Light Detection and Ranging (LiDAR) and Synthetic Aperture Radar (SAR) systems).

In the case of a sparse canopy, such as urban vegetation (Alonzo et al. 2014) and Savannah (Cho et al. 2012, Colgan et al. 2012), or a dense canopy characterized by species with different heights or dominated by definable growth stages (Jones et al. 2010, Dalponte et al. 2008, Dalponte et al. 2012, Kempeneers et al. 2014), best results in

species classification were obtained by integrating hyperspectral data with the information about tree height and canopy structure supplied by LiDAR data. Conversely, in mixed forests with closed canopy and a single dominant layer, Ghosh et al. (2014) evidenced that the canopy height provided by LiDAR data did not significantly affect the mapping accuracy. In these conditions, the use of hyperspectral data alone showed good results using either a traditional algorithm such as Maximum Likelihood (ML) applied on selected indices, in order to reduce spectral dimensionality (Boschetti et al. 2007, Pandey et al. 2014), or more sophisticated approaches such as vector machines (Marcinkowska et al. 2014), with higher accuracy level for a higher spatial resolution (Clark et al. 2005, Dalponte et al. 2013, Baldeck et al. 2015).

The objective of this study was to investigate the operational use of Airborne Prism EXperiment (APEX) images for tree species mapping in a complex mixed forest ecosystem. APEX images were acquired on the Forêt de Hardt (Mulhouse, France) in June and September 2013, with a spatial resolution of 3 m. Given the fact that image segmentation to detect single tree crowns was not feasible in a closed canopy forest at this spatial resolution, pixel based classification algorithms were investigated. The possibility

to improve the spectral discrimination of species through the combination of the spectral information derived from images acquired during two different phenological development stages (i.e., growing season and incipient senescence) was evaluated and results were discussed.

### Remotely sensed data collection and processing

Remotely sensed data were acquired with the hyperspectral airborne push broom imaging spectrometer Airborne Prism Experiment (APEX), developed by a Swiss-Belgian consortium on behalf of the European Space Agency (ESA). APEX covers a wide spectral range from the visible to the shortwave infrared region of the spectrum (380-2500 nm) with 288 narrow bands, which allows a detailed characterization of the spectral response of different surfaces.

Data were acquired over the Forêt de Hardt on June 17, 2013 and September 4, 2013 at 12:30±30 local solar time (LST), with APEX flying at a height of about 6000 m, which results in a 3 m ground spatial resolution. To cover the entire study area, 7 flight lines with 30% lateral overlap were planned with a track in the principal solar plane (195.3°), in order to minimize shadow effects.

Radiometric calibration of the APEX data was performed using dedicated software developed and maintained by the APEX consortium, using as input the calibration cubes generated from data collected before the flight season on the APEX Calibration Home Base (Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, Germany) and the dark current data collected in-flight. Following the standard radiometric calibration, a vicarious calibration was applied to the spectral regions suffering from slightly higher calibration uncertainties, i.e., the region around 1030 nm and the end of the SWIR (1300-2500 nm). This was based on four artificial targets (two black, one grey and one white) made of PVC-coated canvas material ('Odyssey' trademark material, from Kayospruce Ltd., UK) and some 'pseudo-invariant' features: concrete, asphalt with different brightness and pit material. Their radiance and reflectance were measured with an ASD spectrometer simultaneously with the APEX overflights.

Spectral misregistrations were detected for every across track pixel, following a spectrum-matching technique (Gao et al. 2004), which was applied on the continuum removed at sensor radiance across selected atmospheric absorption features.

Geometric processing of the APEX data was performed by means of direct georeferencing, using a dedicated software developed by VITO (Vlaamse Instelling voor Technologisch Onderzoek). Direct georeferencing requires knowledge of the position and look direction of the sensor, and of the elevation, to adjust for the topographic relief.

The position and orientation of each image scan line was obtained from the post-processed GPS (Global Positioning System) and IMU (Inertial Measurement Unit) data. Angular offsets in X, Y and Z directions between the APEX and IMU frame were determined through a boresight calibration and accounted for in the boresight misalignment angles. The elevation was obtained from a DTM (Digital Terrain Model) of the Mulhouse region at 25 m horizontal resolution. From this data, the geographic location and the sun-viewing geometry was calculated for every image pixel. The resulting absolute geolocation accuracy was found to be at sub-pixel level.

The data were projected to UTM zone 32N, WGS84, with a spatial resolution of 3 m by 3 m.

The USGS (United States Geological Survey) GCTP (General Cartographic Transformation Package) was used for the coordinate projection, while the nearest neighbour method was applied for the spatial resampling.

### Species classification method

The forest species map was realized on the North portion of the Forêt de Hardt, where the ground surveys were conducted. A spectral endmember set was selected in order to train the algorithm to identify and assign all the pixels of the image to the proper class, as required by a supervised classification approach.

The tree species present in the study area were identified by forest experts through a ground survey conducted over an extensive area. The forest experts visually evaluated the species composition in 42 elementary sampling units (ESU) of 20 x 20 m, considering all the species representing more than 5% of the leaf area of the ESU. The centre of the ESU was geo-located by means of the high precision global positioning system Trimble Geo-XT (Trimble, California, USA). Out of the dominant species found, five – hornbeam (*Carpinus betulus* L.), two species of oak (*Quercus petraea* (Matt.) Liebl., *Quercus robur* L.), linden (*Tilia* L.) and pine (*Pinus* L.) – were identified as the most common. Together, these forest species represent an average 93% of each ESU. Therefore, they were accounted for in the production of the forest species map. Other species were not taken into account, due to the fact that they were identified in only a few ESU (e.g. maple - *Acer campestre* L., *Acer platanoides* L.) or were mainly found as single trees (e.g. larch - *Larix decidua* Mill.).

The endmember set was selected on the APEX images by integrating the visual interpretation of high resolution orthophotos and ground surveys. It included a total of 400 pixels distributed between the four forest species. For each species, homogeneous polygons composed of ten to twelve pure vegetation pixels – selected on two to three crowns, based on the crown dimension – were defined within several ESUs (Table 1). An example of crown pixel selection is depicted in Figure 1.

Each polygon was considered as a separate spectral class during the classification process in order to account for the intrinsic spectral variability of the species over the study area. The spectral classes belonging to the same species were then merged in the post-classification process. *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. were considered as the same classification unit, due to their spectral similarity. An additional class, "shadow", which accounts for the plant inter-crown shadows, was included in the spectral endmember set.

Hence, the spectral endmember set was used to train a ML algorithm to distinguish the different forest species. Spectral indices (i.e., APEX band ratios) in the *red edge* portion of the spectrum (680-750 nm) were used as input, in order to avoid information redundancy, which could decrease the classification accuracy (Hughes 1968). The spectral indices selected were those successfully used by Boschetti et al. (2007). Three indices expressed as band ratios between APEX bands 45 (659.2 nm), 58 (701.6 nm)

**Table 1. Characteristics of the endmember set: for each class, the number of ESU in which pixels to assemble the spectral endmember set were selected, and the number of pixels per ESU collected, are reported**

Forest species	Classification unit	N of ESU	N of pixels per ESU
<i>Carpinus betulus</i> L.	Hornbeam	12	10
<i>Quercus robur</i> L., <i>Quercus petraea</i> (Matt.) Liebl.	Oak	4	10
<i>Tilia</i> L.	Linden	2	12
<i>Pinus</i> L.	Pine	6	10
-	Shadow	3	50

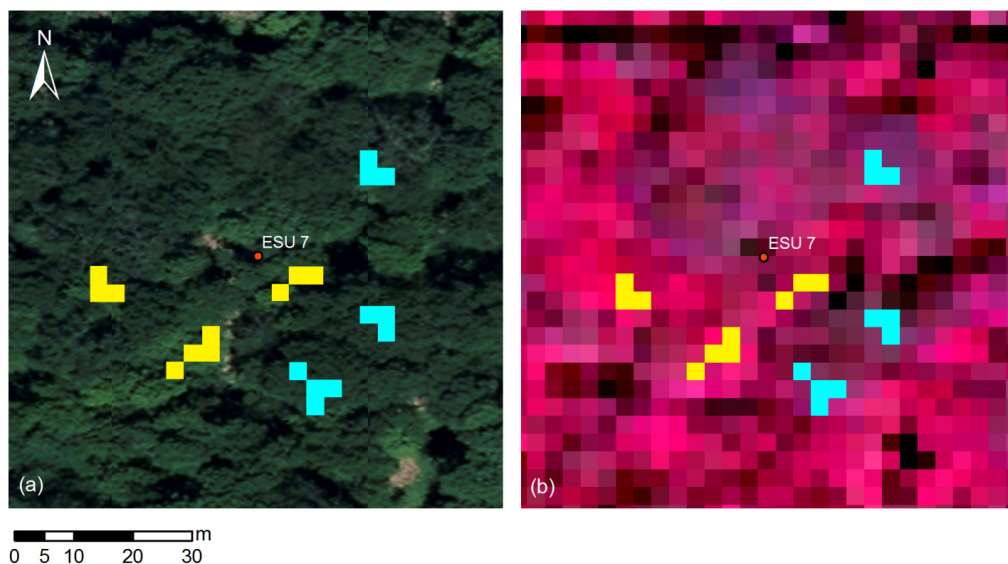


Figure 1. Example of the spectral endmember collection for the training of the ML algorithm: pure crown pixel of hornbeam (yellow) and oak (cyan) identified on (a) a high resolution orthophoto and (b) an APEX false colour image (APEX bands 92, 42, 16)

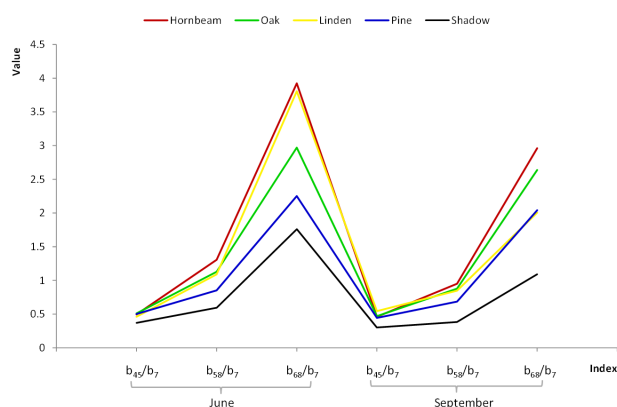


Figure 2. Spectral endmember set used to train the ML algorithm in order to classify the different vegetation species

and 68 (740.2 nm) and band 7 (480.8 nm) were calculated on June and September APEX radiance images. The multi-temporal endmember set used to train the ML algorithm is shown in Figure 2.

The indices selected were located in the *red edge* because spectral differences in this region mirror differences in photosynthetic pigments and canopy greenness that vary between species and within their phenological cycle (Horler et al. 1983, Filella & Penuelas 1994, Hu et al. 2008, Zarco-Tejada et al. 1999). The ratio with the blue band is useful in correcting for atmospheric effects.

Since APEX images were acquired in two different vegetation phenological development stages, the classification was first performed on June and September images separately and then on the joint images in order to assess the improvement on the classification accuracy of the multi-temporal information. The maps obtained were slightly filtered in order to improve the clarity of the thematic product; isolated pixels were identified by the *sieve class* function (ENVI 5.0, ITT Visual Information Solution).

Next, a majority analysis based on a 3 x 3 kernel was used to assign these spurious pixels to the major class.

Finally, the maps were validated using a testing set composed of 250 pixels selected on the basis of a random stratified sampling scheme, taking into account the unequal distribution of the classes. The number of testing pixels for each class reflected the size of that class: 90 testing pixels were randomly selected for hornbeam, 100 for oak, 20 for linden, 20 for pine and 20 for shadow. The testing pixels were labelled according to the visual interpretation of the high resolution orthophotos and used as ground truth in the validation process. The classification performances were then evaluated by computing the Overall Accuracy (OA), the Producer's Accuracy (PA) and the User's Accuracy (UA) of the map. OA is calculated as the ratio between the sum of pixels that are classified in the correct class and the total number of testing pixels. PA refers to the probability that the classifier has labelled a pixel as belonging to a class given that the ground truth is that class. It is computed for each class as the ratio between the number of correctly classified pixels and the total number of testing pixels for that class. UA refers to the probability that a pixel belongs to a class given that the classifier has labelled it as belonging to that class. It is calculated for each class as the ratio between the number of correctly classified pixels and the total number of pixels that were assigned to that class.

**Results**

The classification process enabled the production of thematic maps of the most common forest species in the Forêt de Hardt. The classification accuracy was assessed on the maps obtained using as input: (i) only the image acquired in June, (ii) only the image acquired in September, and (iii) the June and September images together. The results of the classification performed with only one image as input showed better results for the June image (OA=59.5%) than the September image (OA=48%). This can be explained by the fact that the vegetation in June is in its maximum seasonal development stage and the illumination conditions are better (i.e., maximum solar elevation, which minimizes shadow effects). However, the June image did not allow an accurate

Table 2. Confusion matrix obtained by crossing the reference data with the classification result. In each cell, the value is expressed as number of pixels (top) and percentage (bottom). The last line and column represent Producer's Accuracy (PA) and User's Accuracy (UA) respectively

		Testing set						
		Hornbeam	Oak	Linden	Pine	Shadow	Tot	UA
APEX classification	Hornbeam	78 70.91%	9 12.68%	3 15.79%	0 0%	0 0%	90 36%	78/90 86.67%
	Oak	30 27.27%	61 85.92%	3 15.79%	1 5%	5 16.67%	100 40%	61/100 61%
	Linden	2 1.82%	1 1.41%	13 68.42%	1 5%	3 10%	20 8%	13/20 65%
	Pine	0 0%	0 0%	0 0%	16 80%	4 13.33%	20 8%	16/20 80%
	Shadow	0 0%	0 0%	0 0%	2 10%	18 60%	20 8%	18/20 90%
	Tot	110 100%	71 100%	19 100%	20 100%	30 100%	250 100%	
	PA	78/110 70.91%	61/71 85.92%	13/19 68.42%	16/20 80%	18/30 60%		OA=74.4% k=0.6374

classification of linden, which showed high commission errors with hornbeam (UA=36%), while in September the incoming senescence allowed a better discrimination of linden from hornbeam (UA=53%). The use of both images improved the classification result significantly (OA=74.4%), confirming that the use of multi-temporal images supplies additional information, leading to better spectral class discrimination. PA and UA also showed an improvement for all classes, with PA ranging from 60% to 86% and UA ranging from 61% to 90%, except for the pine PA, which decreased from 85% to 80%. Again, this can be explained by the lighting conditions, which were not optimal in September, with shadow effects affecting the conifer canopy more than the deciduous.

The confusion matrix obtained from crossing the ground truth data with the results of the multi-temporal classification is shown in Table 2. An example of the best classification result obtained using the multi-temporal information is depicted in Figure 3. The map shows a clear predominance of the hornbeam-oak association, which is the climax community of forest at these latitudes, while the distribution of pine and linden is more restricted.

In previous studies, higher accuracy levels were achieved by using higher spatial resolution (1 m or less), with OA close to 90% for pixel-based classifications, and up to 95% when individual tree classifications were conducted (Clark et al. 2005, Dalponte et al. 2013, Baldeck et al. 2015). OAs consistent with our results have been shown in previous studies with lower spatial resolution ( $\geq 3$  m). In a similar mixed forest ecosystem, Boschetti et al. (2007) evidenced that the classes with lower accuracy are those with higher fragmentation, such as the typical association of *Quercus*

*robur* L. and *Carpinus betulus* L. This is confirmed by our results: the commission error between oak and hornbeam classes is in fact higher compared to the error between other classes (i.e., pine and linden). Marcinkowska et al. (2014) used APEX images with a spatial resolution of up to 2 m to map vegetation communities rather than single tree species, confirming that, at this spatial resolution, it is worth scaling up at community level more than at tree species level in order to improve vegetation mapping accuracy. Despite this, the use of multi-temporal images allowed a reliable thematic product to be obtained, the quality of which can also be appreciated visually (Figure 4).

**Conclusions**

In this paper, APEX images at 3 m resolution were used to map forest species in the Forêt de Hardt (Mulhouse, France), a mixed forest ecosystem. The use of an operational algorithm, the ML, was improved by the selection of spectral bands combined in ratio indices sensitive to plant pigment content. It was demonstrated that the combination of multi-temporal images in the classification process led to a good accuracy level (OA=74.4%) in forest species mapping. On the other hand, with only one APEX overpass and a spatial resolution of 3 m, the accuracy level of species classification was significantly lower (OA<60%). As a matter of fact, the availability of two images – the first acquired when the vegetation was in its maximum development stage and the second when senescence was incoming – allowed the detection of variations in the spectral response linked to species-specific phenological development, which significantly improved the map accuracy. However, it is stressed that a flight later in the season would be suggested in

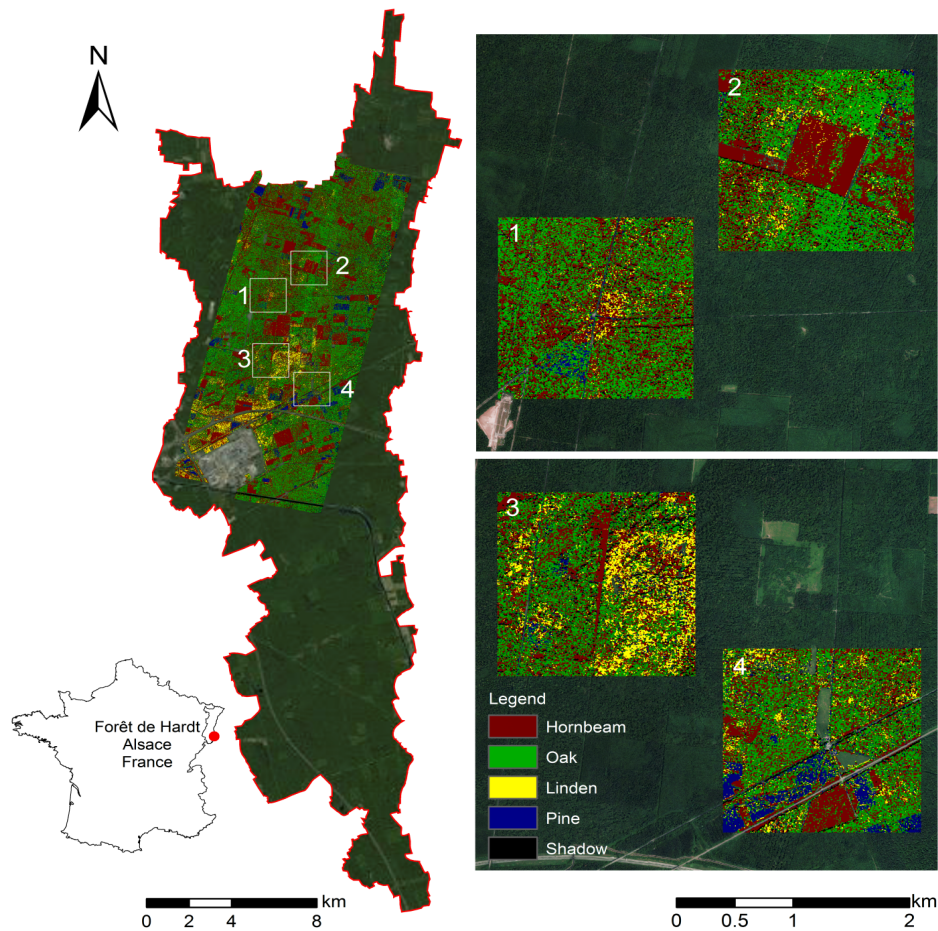


Figure 3. Example of the thematic map that represents the four most common dominant species in the Forêt de Hardt

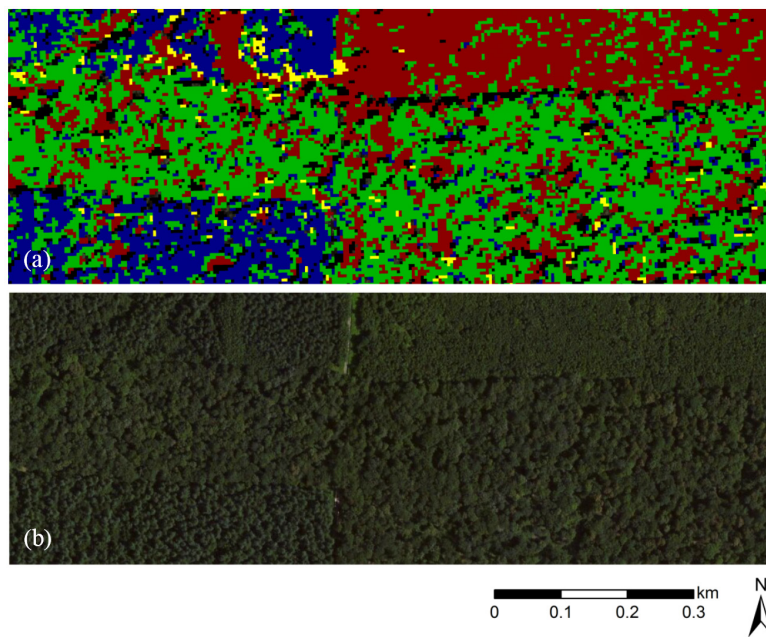


Figure 4. Detail of the thematic product obtained. (a) Forest species map; (b) High resolution orthophoto of the same area

order to emphasise the spectral differences between oak and hornbeam. The use of multi-temporal information was made possible by APEX's high absolute geo-location accuracy (i.e., sub-pixel level), which ensured a good overlapping of the images acquired at different times. This confirmed the high quality of APEX data and its valuable use in forest applications. In order to improve the classification accuracy through the mapping of individual trees, a higher spatial resolution (about 1 m) is suggested.

## References

- Alonzo, M, Bookhagen, B & Roberts, DA 2014, 'Urban tree species mapping using hyperspectral and LiDAR data fusion', *Remote Sensing of Environment*, vol. 148, pp. 70–83.
- Baldeck, CA, Asner, GP, Martin, RE, Anderson, CB, Knapp, DE, Kellner, JR & Wright, SJ 2015, 'Operational Tree Species Mapping in a Diverse Tropical Forest with Airborne Imaging Spectroscopy', *PLOS ONE*, vol. 10, no. 7, pp. 1–21.
- Boschetti, M, Boschetti, L, Oliveri, S, Casati, L & Canova, I 2007, 'Tree species mapping with airborne hyperspectral MIVIS data: the Ticino Park study case', *International Journal of Remote Sensing*, vol. 28, no. 6, pp. 1251–1261.
- Cho, MA, Mathieu, R, Asner, GP, Naidoo, L, van Aardt, J, Ramoelo, A, Debba, P, Wessels, K, Main, R, Smit, IPJ & Erasmus, B 2012, 'Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system', *Remote Sensing of Environment*, vol. 125, pp. 214–226.
- Clark, ML, Roberts, DA & Clark, DB 2005, 'Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales', *Remote Sensing of Environment*, vol. 96, pp. 375–398.
- Colgan, M, Baldeck, C, Féret, JB & Asner, GP 2012, 'Mapping Savanna Tree Species at Ecosystem Scales Using Support Vector Machine Classification and BRDF Correction on Airborne Hyperspectral and LiDAR Data', *Remote Sensing*, vol. 4, no. 11, pp. 3462–3480.
- Dalponte, M, Bruzzone, L & Gianelle, D 2008, 'Fusion of Hyperspectral and LiDAR Remote Sensing Data for Classification of Complex Forest Areas', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1416–1427.
- Dalponte, M, Bruzzone, L & Gianelle, D 2012, 'Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data', *Remote Sensing of Environment*, vol. 123, pp. 258–270.
- Dalponte, M, Orka, HO, Gobakken, T, Gianelle, D & Naesset, E 2013, 'Tree Species Classification in Boreal Forests With Hyperspectral Data', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2632–2645.
- Filella, I & Penuelas, J 1994, 'The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status', *International Journal of Remote Sensing*, vol. 15, no. 7, pp. 1459–1470.
- Franklin, SE 2001, *Remote Sensing for Sustainable Forest Management*, CRC Press, Boca Raton.
- Gao, BC, Montes, MJ & Davis, CO 2004, 'Refinement of wavelength calibrations of hyperspectral imaging data using a spectrum-matching technique', *Remote Sensing of Environment*, vol. 90, no. 4, pp. 424–433.
- Ghosh, A, Fassnacht, FE, Joshi, PK & Koch, B 2014, 'A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales', *International Journal of Applied Earth Observation and Geoinformation*, vol. 26, pp. 49–63.
- Horler, DNH, Dockray, M & Barber, J 1983, 'The red edge of plant leaf reflectance', *International Journal of Remote Sensing*, vol. 4, no. 2, pp. 273–288.
- Hu, B, Miller, JR, Zarco-Tejada, PJ, Freemantle, J & Zwick, H 2008, 'Boreal forest mapping at the BOREAS study area using seasonal optical indices sensitive to plant pigment content', *Canadian Journal of Remote Sensing*, vol. 34, pp. 158–171.
- Hughes, G 1968, 'On the mean accuracy of statistical pattern recognizers', *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63.
- Jones, TG, Coops, NC, & Sharma, T 2010, 'Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada', *Remote Sensing of Environment*, vol. 114, no. 12, pp. 2841–2852.
- Kempeneers, P, Van Coillie, F, Liao, W, Devriendt, F & Vandekerkhove, K 2014, 'Tree species mapping by combining hyperspectral with LiDAR data', *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Québec.
- Linke, J, Betts, MG, Lavigne, MB & Franklin, SE 2006, 'Introduction: structure, function and change of forest landscapes' in *Understanding forest disturbance and spatial pattern: Remote sensing and GIS approaches*, eds. M Wulder & SE Franklin, Taylor & Francis Group, Abingdon, pp. 1–29.
- Marcinkowska, A, Zagajewski, B, Ochtyra, A, Jarościńska, A, Raczko, E, Kupková, L, Stych, P & Meuleman, K 2014, 'Mapping vegetation communities of the Karkonosze National Park using APEX hyperspectral data and Support Vector Machines', *Miscellanea Geographica*, vol. 18, no. 2, pp. 23–29.
- Pandey, PC, Tate, NJ & Balzter, H 2014, 'Mapping Tree Species in Coastal Portugal Using Statistically Segmented Principal Component Analysis and Other Methods', *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4434–4441.
- Panigada, C, Rossini, M, Busetto, L, Meroni, M, Fava, F & Colombo, R 2010, 'Chlorophyll concentration mapping with MIVIS data to assess crown discoloration in the Ticino Park oak forest', *International Journal of Remote Sensing*, vol. 31, no. 12, pp. 3307–3332.
- Zarco-Tejada, PJ & Miller, JR 1999, 'Land cover mapping at BOREAS using red edge spectral parameters from CASI imagery', *Journal of Geophysical Research*, vol. 104, no. D22, pp. 27921–27933.