

## ASSESSING THE SIGNIFICANCE OF HYPERION SPECTRAL BANDS IN FOREST CLASSIFICATION

G. J. Newnham<sup>a</sup>, D. Lazaridis<sup>b</sup>, N. C. Sims<sup>a</sup>, A. P. Robinson<sup>b</sup>, D. S. Culvenor<sup>a</sup>

<sup>a</sup> CSIRO Division of Land and Water and Sustainable Agriculture Flagship, Clayton South, Victoria, Australia

<sup>b</sup> Department of Mathematics and Statistics, University of Melbourne, Parkville, Victoria, Australia

**KEYWORDS:** Forest, Classification, Hyperspectral, Ensemble, Decision Tree, Random Forests

### ABSTRACT:

The classification of vegetation in hyperspectral image scenes presents some challenges due to high band autocorrelations and problems dealing with many predictor variables. The Random Forests classification method is based on an ensemble of decision trees and attempts to address these issues by dealing with only a subset of image bands in each node of each decision tree. Random Forests has previously been used for classification of vegetation using hyperspectral data. However, the variable importance measure that is a by-product of the technique has largely been ignored. In this study we investigate the spectral qualities of variable importance in the classification of forest and non-forest in a single Hyperion scene. The spectral importance curve showed broad bands of importance over wavelength regions known to be significant in biochemical absorption.

### 1. INTRODUCTION

Certain biological and statistical challenges can inhibit the successful use of hyperspectral data for mapping forest extent. Absorption by plant materials *in vivo* generally occur as broad wavelength bands leading to auto-correlation in vegetation reflectance spectra. In addition, many statistical modelling methods have a tendency to over-fit to noise in cases with many predictor variables (Bajcsy and Groves, 2004). Consequently, classification accuracy may be highest when only a small subset of predictor variables is used (Hughes, 1968).

The ensemble decision tree approach described as Random Forests (Breiman, 2001) is suited to addressing these challenges and has been shown to be superior to linear, quadratic and penalised discriminant analysis when using hyperspectral satellite data (Everingham et al., 2007; Sluiter and Pebesma, 2010). Random Forests models also generate a measure of variable importance. High variable importance has been used for selecting narrow bands (Chan and Paelinckx, 2008) and spectral indices (Ismail and Mutanga, 2010) for inclusion in refined classification models. However, the spectral characteristics of variable importance have not been fully explored.

We consider variable importance for a classification of forests and non-forests based on a Hyperion image over high value forest site in Tasmania. Spectral characteristics of the importance curve are compared to known absorption and reflectance characteristics of leaf biochemicals.

### 2. METHODS

The Hyperion scene used in this study was captured on the 13<sup>th</sup> of March 2010 over the Warra Long Term Ecological Research (LTER) site in southern Tasmania (Brown et al., 2001). The image was 88km in the along track direction and included mainly forested land in the south, while grassland and pasture dominated in the north. Pre-processing was performed using the methods described by (Datt et al., 2003) and then registered to a orthorectified mosaic of Landsat Thematic Mapper images

produced as part of the Australian National Carbon Accounting System (Furby, 2002).

A Tasmanian Government state-wide vegetation map was used for training and validation of the classification models. The map is based on aerial photo interpretation and field validation, and includes 154 classes as described by Harris and Kitchener (2005). These classes were aggregated into generic forest and non-forest classes and a raster map created on the same grid as the Hyperion image.

First, we applied the implementation of Random Forests by Liaw and Wiener (2002) to discriminate forest from non-forest classes in the Hyperion image. For each class, 10000 pixels were selected at random as the training set. In each model run, 1000 decision trees were generated. Classification accuracy was assessed across the entire Hyperion scene. The wavelength regions that best discriminate forest from non-forest classes were inferred from the variable importance spectrum. These wavelengths were then compared to published biochemical absorption features to examine which parameters of forest biochemistry may be contributing to the spectral separation of forested from non-forested areas.

### 3. RESULTS

The classifications of the Hyperion image were assessed in terms of overall accuracy and the Kappa statistic (Cohen, 1960). These are summarised in Table 1. Training accuracy was comparable to other published results. Interestingly, when the model was applied to all pixels in the Hyperion scene, the overall accuracy was maintained and the kappa statistic increased slightly. This is not a large increase, but does indicate the stability of the model when applied outside the original data on which it was built.

The significance of Hyperion spectral bands in discriminating the forest and non-forest classes were assessed using the measure of variable importance produced using the Random Forests method. The plot of variable importance as a function of wavelength showed strong auto-correlation, with dominant peaks in significant biochemical absorption regions.

Training Data		Complete Image	
Overall Accuracy	Kappa	Overall Accuracy	Kappa
83	0.65	83	0.67

Table 1. Summary of the accuracy achieved for the forest/non-forest classification

Features in the spectral importance curve include a sharp peak at 1720nm, which sits between two broad liquid water absorption bands. This wavelength region is known to be sensitive to absorption by cellulose and lignin content (Fourty et al., 1996; Gao and Goetz, 1995). Secondary peaks appear in the visible green at 539nm and on the red edge at 701nm. There is also a smaller peak at 640nm in the strongly chlorophyll absorbing red wavelength region.

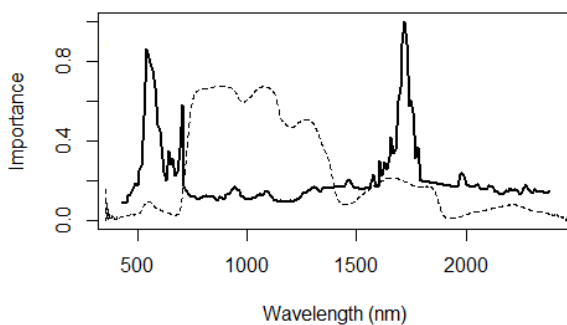


Figure 1. Variable importance (solid line) for the Random Forests classification model and a Eucalyptus leaf reflectance spectrum measured in the laboratory (dashed line).

#### 4. CONCLUSIONS

The accuracy of forest and non-forest classification achieved here was comparable with those reported in previous studies using Random Forests for the classification of vegetation using hyperspectral imagery (Chan and Paelinckx, 2008; Sluiter and Pebesma, 2010). Variable importance highlighted spectrally broad features that have previously been associated with biochemical absorption. For example, the importance feature at 1720nm, which is thought to be associated with cellulose and lignin absorption is the dominant feature. Since the image was collected just after the summer season, this importance may be due to the presence of dead material within non-forest (grassland and pasture) areas of the image.

While importance measures are an interesting diagnostic which may help us to understand the key biophysical characteristics of forests that allow their discrimination within a satellite image scene, they also allow the investigation of appropriate broad band data types for operational monitoring of forests as an ongoing exercise. This is a key focus for our further research in this area.

#### REFERENCES

- Bajcsy, P. and Groves, P., 2004. Methodology for hyperspectral band selection. *Photogrammetric Engineering and Remote Sensing*, 70(7): 793-802.
- Breiman, L., 2001. Random forests. *Machine Learning*, 45(1): 5-32.
- Brown, M., Elliott, H. and Hickey, J., 2001. An overview of the Warra long-term ecological research site. *Tasforests*, 13: 1-8.
- Chan, J.C.W. and Paelinckx, D., 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112(6): 2999-3011.
- Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1): 37-46.
- Datt, B., McVicar, T.R., Van Niel, T.G., Jupp, D.L.B. and Pearlman, J.S., 2003. Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6): 1246-1259.
- Everingham, Y., Lowe, K.H., Donald, D., Coomans, D. and Markley, J., 2007. Advanced satellite imagery to classify sugarcane crop characteristics. *Agronomy for Sustainable Development*, 27: 111-117.
- Fourty, T., Baret, F., Jacquemoud, S., Schmuck, G. and Verdebout, J., 1996. Leaf optical properties with explicit description of its biochemical composition: direct and inverse problems. *Remote Sensing of Environment*, 56(2): 104-117.
- Furby, S., 2002. Land cover change: specification for remote sensing analysis - technical report no. 9. CSIRO Mathematical and Information Sciences, Australian Greenhouse Office.
- Gao, B.C. and Goetz, A.F.H., 1995. Retrieval of Equivalent Water Thickness and Information Related to Biochemical-Components of Vegetation Canopies from Aviris Data. *Remote Sensing of Environment*, 52(3): 155-162.
- Harris, S. and Kitchener, A., 2005. From Forest to Fjaeldmark: Descriptions of Tasmania's Vegetation. Department of Primary Industries, Water and Environment. Australian Government Publishing Service, Canberra.
- Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1): 55-63.
- Ismail, R. and Mutanga, O., 2010. A comparison of regression tree ensembles: Predicting Sirex noctilio induced water stress in Pinus patula forests of KwaZulu-Natal, South Africa. *International Journal of Applied Earth Observation and Geoinformation*, 12: S45-S51.
- Liaw, A. and Wiener, M., 2002. Classification and Regression by randomForest. *R news*, 2(3): 18-22.

Sluiter, R. and Pebesma, E.J., 2010. Comparing techniques for vegetation classification using multi- and hyperspectral images and ancillary environmental data. *International Journal of Remote Sensing*, 31(23): 6143-6161.

#### **ACKNOWLEDGEMENTS**

This project was conducted as part of the Group on Earth Observation (GEO) Forest Carbon Tracking (FCT) initiative. The authors would like to thank Dr. Alex Held and Dr Peter Caccetta from the CSIRO and Prof. Kim Lowell from the Cooperative Research Centre for Spatial Information for their support of the project.