DISCRIMINATING THE EARLY STAGES OF *SIREX NOCTILIO* INFESTATION USING RESAMPLED HYMAP DATA

Ismail, R[1], Mutanga, O[1] and Kumar, L[2]*

Department of Geography and Environmental Studies
University of KwaZulu- Natal
Private Bag X01
Scottsville
3209
South Africa

2. Department of Ecosystem Management
School of Environmental Sciences and Natural Resources Management
University of New England
Armidale
NSW 2351
Australia

*Corresponding author
riyad.ismail@sappi.com

ABSTRACT

In this study we evaluated whether the random forest (RF) algorithm can accurately discriminate between healthy trees and the early stages of *Sirex noctilio* infestation using resampled HYMAP data. More specifically, we examined the potential role of three variable selection methods (a filter, the random forest out of bag samples, and a wrapper) to produce the smallest subset of shortwave infrared (SWIR) wavelengths with the lowest misclassification error in a hyperspectral application where (i) the numbers of samples were less than the number of variables and (ii) the class labels have similar spectral characteristics. Using the .632+ bootstrap method, we compared the misclassification error of the RF algorithm against a competing algorithm known as boosting trees (BT). Results showed that the RF algorithm produces slightly better classification results than the BT algorithm for all three variable selection methods. Using wavelengths selected by the wrapper method as input variables into the RF algorithm produces (i) better accuracies than using all the SWIR wavelengths, (ii) the lowest overall misclassification error (6.14%), and (iii) the largest difference in misclassification error (1.4 %) between the RF and BT algorithms. Finally, we evaluated the performance of the RF algorithm when noise is introduced into the class labels or into the reflectance values of the selected wavelengths. Results showed that wavelength noise is less harmful than class noise on the performance of the RF algorithm. Classification accuracies remain below 8% when noise is introduced into the selected wavelengths. The ability of the selected wavelengths located at 1990 nm, 2009 nm, 2028 nm, 2047 nm and 2065 nm to discriminate between healthy and green stage spectra could be explained due to the rapid physiological changes that occur as a result of toxic mucus and a fungus that are injected into the tree during the early stages of *S.noctilio* infestations. Overall the results are encouraging and show that there is a link between the selected SWIR wavelengths and existing physiological knowledge thereby improving the chances of detecting the green stage at a canopy or landscape level.