



Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

A comparison of regression tree ensembles: Predicting *Sirex noctilio* induced water stress in *Pinus patula* forests of KwaZulu-Natal, South Africa

R. Ismail*, O. Mutanga

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

ARTICLE INFO

Article history:

Received 19 November 2008

Accepted 3 September 2009

Keywords:

Regression trees

Ensembles

Random forest

Sirex noctilio

ABSTRACT

In this study we compared the performance of regression tree ensembles using hyperspectral data. More specifically, we compared the performance of bagging, boosting and random forest to predict *Sirex noctilio* induced water stress in *Pinus patula* trees using nine spectral parameters derived from hyperspectral data. Results from the study show that the random forest ensemble achieved the best overall performance ($R^2 = 0.73$) and that the predictive accuracy of the ensemble was statistically different ($p < 0.001$) from bagging and boosting. Additionally, by using random forest as a wrapper we simplified the modeling process and identified the minimum number ($n = 2$) of spectral parameters that offered the best overall predictive accuracy ($R^2 = 0.76$). The water index and Ratio₉₇₅ had the best ability to assay the water status of *S. noctilio* infested trees thus making it possible to remotely predict and quantify the severity of damage caused by the wasp.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Sirex noctilio is currently the most destructive pest of conifers in South Africa and the wasp is currently causing considerable tree mortality in *Pinus patula* forests located in the southern parts of the country. Recent estimates indicate that 35,000 ha of *P. patula* forest are infested and dying (Hurley et al., 2007). In lieu of the future availability of hyperspectral data in South Africa (van Aardt and Coppin, 2006) there is a keen interest amongst remote sensing researchers to apply novel methods and techniques that will allow for the accurate prediction and quantification of *S. noctilio* infestations.

Regression trees (Breiman et al., 1984) have been widely used for prediction purposes in the remote sensing domain (DeFries et al., 1997; Hansen et al., 2002; Lobell et al., 2007; Michaelson et al., 1994). However, regression trees are very sensitive to small perturbations in the training dataset and have been identified as unstable learners that are prone to overfitting (Breiman, 1996). Simply stated, relatively small changes in the values of the training dataset can lead to significant changes in the selection of variables that are used to create the regression tree (Hastie et al., 2001). Therefore, the instability of regression trees introduces uncertainty in their interpretation and limits their predictive performance (Elith et al., 2008).

Bagging (Breiman, 1996), boosting (Freund and Shapiro, 1996; Friedman, 2002) and random forest (Breiman, 2001) are popular ensembles that have been used to improve the performance of unstable learners (Hamza and Larocque, 2005). As a result of their improved performance these techniques have been applied to a wide variety of remote sensing applications (Gislason et al., 2006; Ham et al., 2005; Lawrence et al., 2004; Lawrence et al., 2006; Pal, 2005). However, to the best of our knowledge, remote sensing applications thus far have focused on using classification trees rather than using regression trees as the base learner.

The question then arises: How would bagging, boosting and random forest perform in regression type applications? Initial research carried out by Breiman (2001) on machine learning datasets revealed that the results were mixed. Random forest always produced better results than conventional bagging while in some of the datasets, a modified version of bagging known as adaptive bagging outperformed random forest. More recently, Prasad et al. (2006) compared random forest and bagging for predicting species distribution under climate change scenarios. Results from the study concluded that random forest and bagging have similar predictive abilities. We are unaware of any studies that compare regression tree ensembles using remotely sensed data. Consequently, the objective of this study was to compare the performance of random forest, bagging and boosting for prediction purposes using remotely sensed data. More specifically, we compared regression tree ensembles for predicting *S. noctilio* induced water stress in *P. patula* trees using several spectral parameters derived from hyperspectral data (Prasad et al., 2006).

* Corresponding author. Tel.: +27 33 347 6695.

E-mail address: riyad.ismail@sappi.com (R. Ismail).

2. Materials and methods

2.1. Spectral reflectance and water content measurements

P. patula foliage was collected from a known *S. noctilio* infested compartment located at the Sappi Pinewoods plantation (centroid 30°4'13.83"E and 29°38'36.06"S) in KwaZulu-Natal, South Africa (Ismail et al., 2008). To facilitate a representative sample, *P. patula* trees were carefully examined with the assistance of experienced foresters and classified into the healthy, green and red stages of infestation (Fig. 1).

Using tree climbers, samples representing each class (healthy = 24, green = 30 and red = 12) were randomly obtained from the upper, middle and lower crowns of selected *P. patula* trees (Ismail et al., 2008). After clipping, each sample (approximately 1 kg) was immediately placed on the ground and spectral measurements were collected. The measurements were taken on a clear sunny day between 10:00 and 14:00 h, using the analytical spectral devices (ASD) Field Spec Pro FR spectroradiometer. The spectroradiometer senses in the 350–2500 nm spectral range. The first sensor measures reflection in wavelengths between 350 nm and 1050 nm with a spectral resolution of 1.4 nm while the second sensor measures reflection between 1000 nm and 2500 nm with a spectral resolution of 2 nm (Analytical Spectral Devices Inc., Boulder, Co.). In accordance with established protocols, the spectroradiometer was mounted on a tripod with a 25° field of view and positioned 0.5 m above each sample at nadir position. Additionally, radiance measurements were converted to

target reflectance using a calibrated white spectralon panel of known spectral characteristics (Analytical Spectral Devices Inc., Boulder, Co.). To control for variation in leaf orientation, 10 spectral reflectance measurements were averaged for each sample and individual samples were rotated 30° between scans (Pontius et al., 2005). Fig. 2 shows the average spectral reflectance of the healthy, green and red stages of infestation.

Once the spectral measurements were completed, foliar samples were immediately sealed in a plastic bag and kept in a cooler at 5 °C. The samples were then transported to the Institute of Commercial Forestry Research (ICFR) laboratory for water content analysis. Following Bowyer and Danson (2004), water content (WC) was calculated as follows:

$$WC(\%) = \frac{FW - DW}{DW} \times 100 \tag{1}$$

where FW is the fresh weight of the sample and DW is the weight of the sample after been oven dried for 24 h at 60 °C.

2.2. Spectral parameters

To minimize external variability and optimize the sensitivity of the spectral response to changes in water content of the foliar samples, several spectral indices, including simple ratios, normalized ratios and three band ratios were calculated from the ASD reflectance measurements. Table 1 shows the various spectral indices that were used in the study. Additionally, we applied continuum removal to water absorption features located at




Stages		Symptoms
Healthy		No signs of <i>Sirex noctilio</i> infestation.
Green		The appearance of resin droplets and the presence of ovipositors on the bark with a dark fungal stain appearing along the cambium. There is minimal needle loss and the canopy appears green and healthy.
Red		Severe chlorosis results in the canopy of the attacked tree changing colour from green to yellow to reddish brown. Larvae are present in the tree. There is very high needle loss and there is a scattering of dead and dying trees in the plantation.

Fig. 1. Description of the healthy, green and red stages of *Sirex noctilio* infestation (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

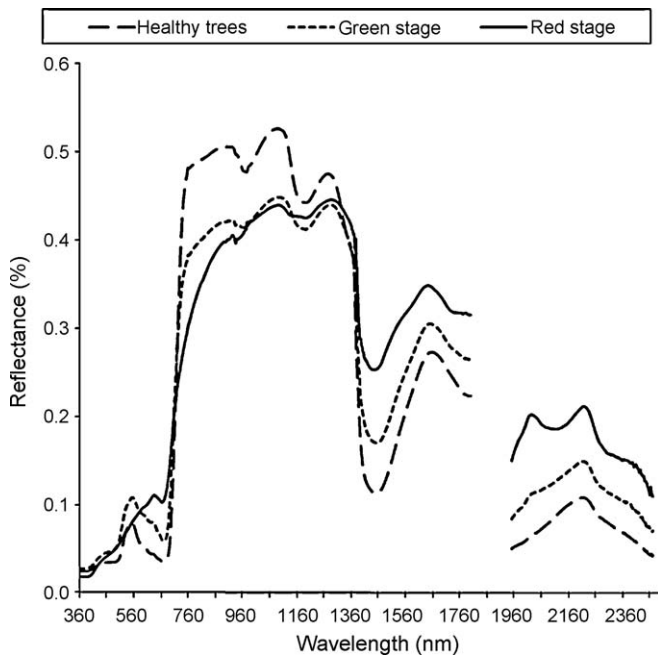


Fig. 2. Spectral reflectance for the healthy, green and red stages of *Sirex noctilio* infestation. Reflectance values between 1800 nm and 1950 nm; and between 2470 nm and 2500 nm displayed a high level of noise and were therefore removed from further analysis.

$R_{920-1120}$ and $R_{1070-1320}$ (Pu et al., 2003). Although previous studies have calculated several parameters from the continuum-removed absorption features (Pu et al., 2003), we only used the band depth (BD), which is computationally efficient and, therefore, more suitable for the practical application of this study (Mutanga and Skidmore, 2004).

2.3. Statistical analysis

We used regression tree ensembles to predict water content as a function of multiple spectral parameters ($n = 9$) using a hold out sample. This was done by repeatedly and randomly ($n = 1000$) dividing the original dataset into training (70%) and test (30%) datasets. For each run, regression tree ensembles developed on the training dataset ($n = 46$) were then used to predict the water content on the test dataset ($n = 20$). The final predictive accuracy used to compare the regression ensembles consisted of an averaged adjusted R^2 value for all the runs carried out. All statistical analysis was carried out using the R package (R

Development Core Team, 2008). The section below briefly describes the regression tree ensembles used in this study.

2.3.1. Bagging, boosting and random forest

Bagging or bootstrap aggregation is a relatively simple idea that uses many bootstrap samples (Efron and Tibshirani, 1993) with replacement from the original dataset and then applies a regression tree to each bootstrap sample. The results from each regression tree are then averaged to obtain the overall prediction. When a bootstrapped sample is drawn, approximately 37% of the dataset is excluded from the sample and the remaining data is replicated to bring the dataset to full size. The excluded one third of the samples is known as the out of bag samples (OOB), while the replicated dataset is known as the in bag samples (Breiman, 1996).

Random forest is similar to bagging but has the additional modification of selecting only a random subset of candidate features ($mtry$) to determine the split at each node of a tree. As each regression tree is maximally grown, it makes predictions on the OOB sample for that particular tree. The prediction error then provides an unbiased assessment of the predictive accuracy, since the OOB sample is not used in the training process. Additionally, random forest provides an internal measure of variable importance using the OOB sample. The variables associated with the OOB sample are randomly permuted and regression trees are grown on the modified dataset. The important measure of each variable is then calculated as the difference in the mean square error between the original OOB predicted dataset and the modified dataset (Breiman, 2001).

While bagging and random forest rely on bootstrapped aggregations of the original training data to generate trees in the ensemble, boosting relies on the results from a previous iteration. Boosting uses a forward stagewise procedure to iteratively fit trees to the training dataset and gradually increases emphasis on poorly modeled observations by the existing collection of trees (Elith et al., 2008). For regression related problems, boosting assumes the form of a functional gradient decent (Friedman, 2002). The boosting algorithm grows the first regression tree to maximally reduce the loss in predictive performance (such as deviance) and the next tree then focuses on the variation in the response (i.e. residuals) that could not be explained by its predecessor. The final model therefore is a linear combination of many trees with the contribution of each tree usually shrunk by a learning rate (lr) to achieve best performance (Elith et al., 2008).

2.3.2. Variable selection

In order to simplify the modeling process we would like to identify the smallest number of spectral parameters that offer the best predictive power and help in the interpretation of the final

Table 1
The various spectral indices that were used in the study.

	Spectral Indices	Formula	Reference
1	Water index	$WI = \frac{\rho_{900}}{\rho_{970}}$	Peñuelas et al. (1997)
2	Normalized difference water index	$NDWI = \frac{\rho_{860} - \rho_{1240}}{\rho_{860} + \rho_{1240}}$	Gao (1996)
3	Normalized difference vegetation index	$NDVI = \frac{\rho_{860} - \rho_{690}}{\rho_{860} + \rho_{690}}$	Rouse et al. (1973)
4	Ratio ₉₇₅	$Ratio_{975} = \frac{2\rho_{960-990}}{\rho_{920-940} + \rho_{1090-1110}}$	Pu et al. (2003)
5	Ratio ₁₂₀₀	$Ratio_{1200} = \frac{2\rho_{1180-1220}}{\rho_{1090-1110} + \rho_{1265-1285}}$	Pu et al. (2003)
6	Moisture stress index	$MSI = \frac{\rho_{1600}}{\rho_{819}}$	Hunt and Rock (1989)
7	Normalised difference infrared index	$NDII = \frac{\rho_{819} - \rho_{1600}}{\rho_{819} + \rho_{1600}}$	Hardinsky et al. (1983)

model. To address this issue we used a wrapper (Kohavi and John, 1997) that searches for the best subset of spectral parameters by using the regression tree ensembles as part of the evaluation process. More specifically, we implemented a backward elimination greedy search function (Guyon and Elisseeff, 2003). The search function commenced with all the spectral parameters ($n = 9$) and then progressively eliminated the least promising spectral parameters. The nested subset of spectral parameters with the lowest root mean square error (RMSE) was then selected. According to Kohavi and John (1997) the resulting subset of spectral parameters should be evaluated on an independent test set that was not used during the variable selection process. For comparative purposes, we evaluated the final subset of spectral parameters using (i) hold out test dataset ($n = 20$), (ii) 10 fold cross validation (CV) and (iii) out of bag samples (OOB).

3. Results

3.1. Model optimization

In an attempt to streamline the model building and evaluation process we optimized certain input parameters for bagging,

boosting and random forest. Using the training dataset ($n = 46$), the optimal input parameters for the ensemble were selected based on the lowest RMSE as calculated by ten fold cross validation (CV).

For random forest, we examined the effect of the number of randomly selected variables ($mtry$) on the prediction error (Hamza and Larocque, 2005). We optimized the $mtry$ value by creating random forest ensembles for all possible $mtry$ values ($n = 9$) and then selecting the optimal $mtry$ value based on the lowest RMSE across all forests. The lowest RMSE (8.19) for the random forest ensemble was obtained when using an $mtry$ value of two. Increasing the $mtry$ value has no additional impact on producing a lower RMSE.

To optimize bagging trees, we varied the number of trees ($nbag$) in the ensemble by adding 25 trees at a time and then recorded the resulting RMSE up to a maximum of 500 trees. The lowest RMSE value (9.35) for bagging was obtained when using 300 trees.

According to Elith et al. (2008) there are two important parameters that need to be optimized for boosting ensembles. The first parameter is the learning rate (lr) which determines the contribution of each tree to the final model, and the second parameter is the tree complexity (tc) which controls whether interactions are fitted. We subsequently identified the number of

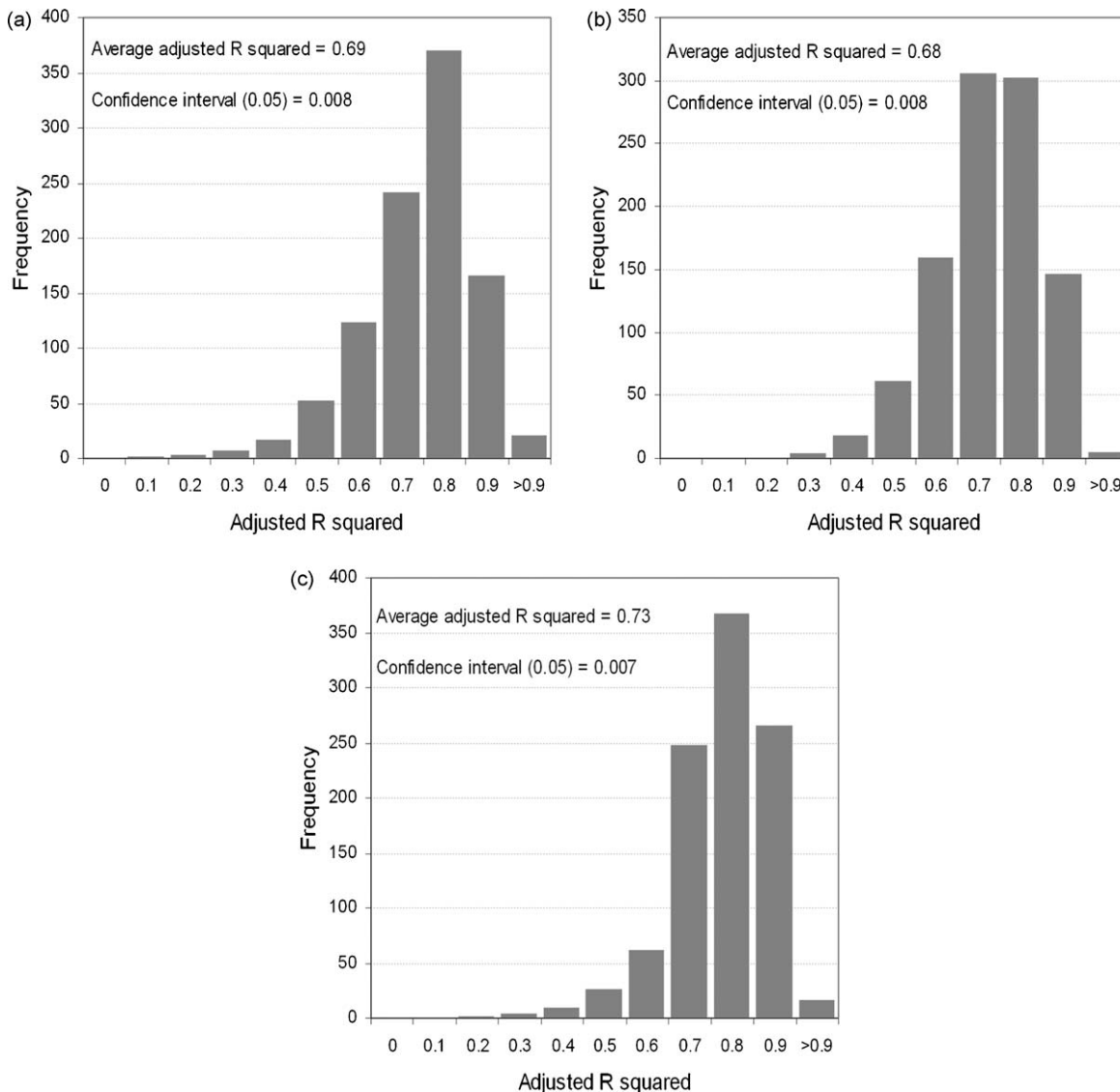


Fig. 3. Histograms showing the frequency of the adjusted R^2 values for the regression tree ensembles used in this study. (a) Shows the distribution of the adjusted R^2 for bagging trees, (b) shows the distribution of the adjusted R^2 for generalized boosting trees and (c) shows the distribution of the adjusted R^2 for random forest.

trees (*n.tree*) that achieved the lowest RMSE for each combination of *tc* (1, 2, 3, 5, 7, and 10) and *lr* (0.1, 0.05, 0.01, 0.005, 0.001 and 0.005). Results indicated that the lowest RMSE (9.57) for boosting was obtained when *tc* = 1, *lr* = 0.01 and *n.tree* = 1000.

3.2. Comparing bagging, boosting and random forest

Fig. 3 shows the histogram of adjusted R^2 values obtained when using the repeated hold out sample ($n = 1000$). There is a narrow confidence interval for all the regression tree ensembles, implying that the methods predicted with high precision (Mutanga et al., 2004). In order to assess whether random forest is significantly better or worse than bagging and boosting, a Bonferroni corrected, one tailed paired *t* test was carried out. Results from paired *t* test indicated that there was a significant difference between random forest and boosting ($t = 6.24, p < 0.001$) and between the random forest and bagging ($t = 8.68, p < 0.001$). However, there was no significant difference ($t = 2.23, p > 0.05$) in the adjusted R^2 values between the boosting and bagging ensembles.

The average performance of all three predictors was relatively close, with the adjusted R^2 values ranging between 0.68 and 0.73 (Table 2). However, random forest produced the best overall performance with an average adjusted R^2 value of 0.73. We also

Table 2

The average adjusted R^2 and RMSE values obtained by bagging, boosting and random forest.

Model	Adjusted R^2	RMSE
Random forest	0.73	8.33
Boosting	0.68	10.27
Bagging	0.69	9.19

calculated the adjusted R^2 value for a single regression tree. As expected, the single regression tree obtained the lowest predictive performance with an adjusted R^2 value of 0.58. Using the random forest ensemble produced a 15% increase in predictive accuracy when compared to single regression trees, a 4% increase in accuracy when compared to bagging and a 5% increase in accuracy when compared to boosting. To check the validity of the comparisons between the regression tree ensembles, we also calculated the RMSE using the hold out samples. The results show that random forest also produced the lowest RMSE (Table 2).

3.3. Variable selection

We used random forest for variable selection since the ensemble produced the best predictive accuracy. However, before

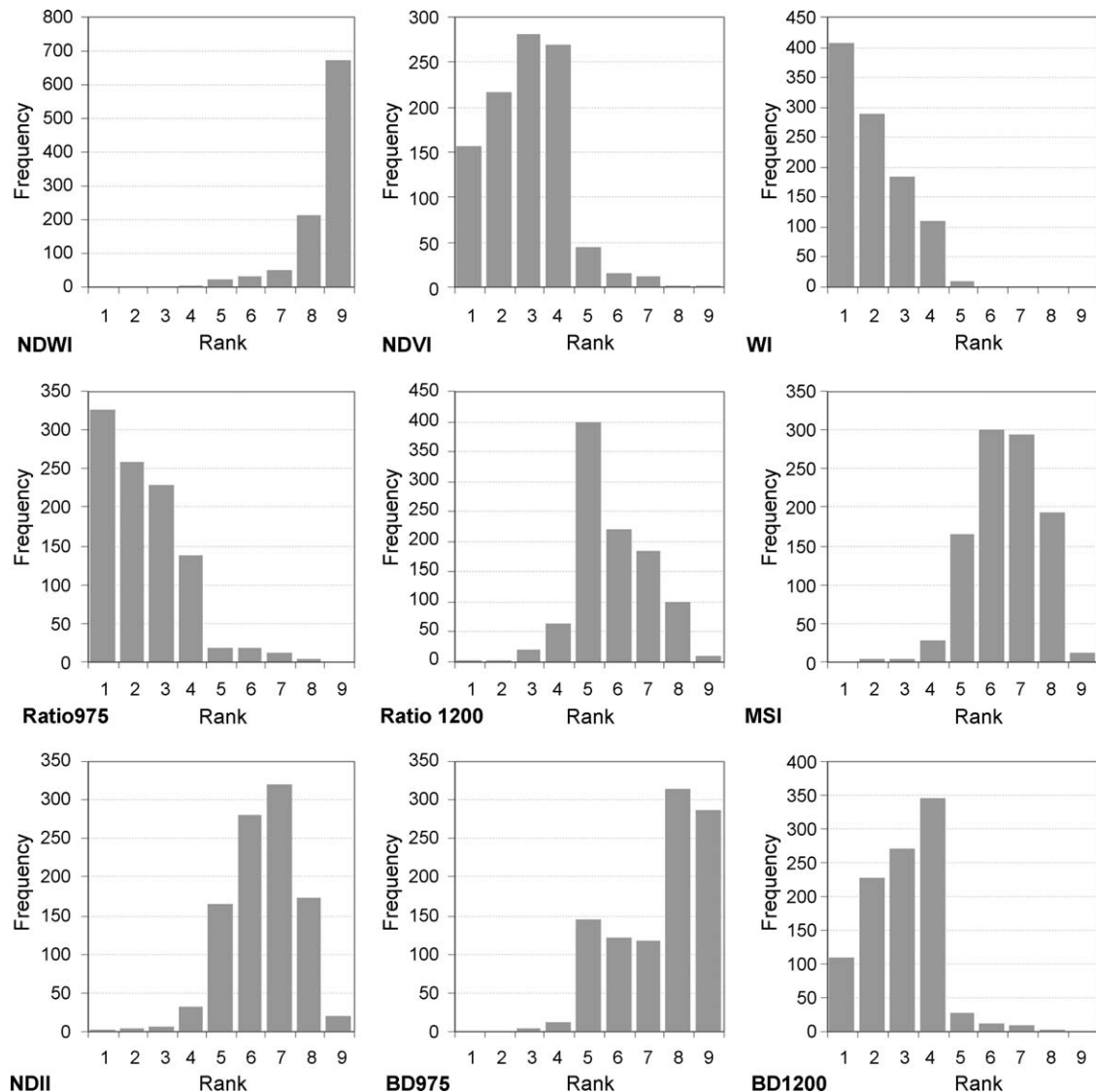


Fig. 4. Histograms showing the ranked variable importance of the spectral parameters used in this study.

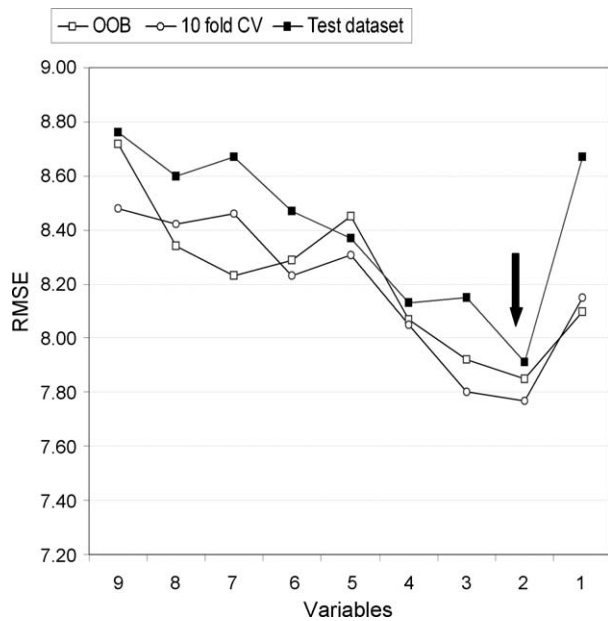


Fig. 5. Variable selection using the backward elimination search function. The resulting RMSE for the OOB sample, 10 fold cross validation and the test dataset ($n = 20$) are shown.

we carried out the variable selection process, we examined the importance of individual spectral parameters in the modeling process. Fig. 4 shows the importance of individual spectral parameters as determined by the random forest OOB sample. The spectral parameters were ranked according to their importance during each of the runs ($n = 1000$) that was carried out during the ensemble comparison phase of the study (Section 3.2). For example, if the spectral parameter had the highest difference in RMSE between the OOB predicted data and the permuted dataset, it was ranked first for that particular run and if the spectral parameter had the second highest difference in RMSE it was ranked second and so on.

To determine if the spectral parameters used in this study were statistically different in terms of their importance in the modeling process, a rank based analysis was performed using Friedman's ANOVA by ranks. The overall test was significant ($p < 0.001$) indicating that the spectral parameters were statistically different. The average rank as calculated by Friedman's ANOVA for the spectral parameters were as follows: WI (2.04), Ratio₉₇₅ (2.38), NDVI (2.94), BD₁₂₀₀ (3.03), Ratio₁₂₀₀ (5.80), NDII (6.46), MSI (6.47), BD₉₇₅ (7.43) and NDWI (8.45). We subsequently used these rankings to determine the sequence in which to eliminate variables using the backward elimination search function.

Fig. 5 shows the results of the variable selection process. As the spectral parameters were progressively eliminated by the backward elimination search function, the RMSE generally decreased, with the lowest RMSE obtained by using only two variables (WI and Ratio₉₇₅). The use of WI and Ratio₉₇₅ produced the lowest RMSE using the hold out test dataset (7.91), 10 fold CV (7.77) and the OOB sample (7.85). We subsequently recalculated the adjusted R^2 value for random forest using WI and Ratio₉₇₅ as input variables. Results indicated that by using WI and Ratio₉₇₅, an adjusted R^2 value of 0.76 is obtained by the random forest ensemble.

4. Discussion

In recent years random forest has gained popularity as an effective classification method in the remote sensing domain (Chan and Paelinckx, 2008; Gislason et al., 2006; Ham et al., 2005; Lawrence et al., 2006; Pal, 2005). Results from this study

additionally confirm that the random forest ensemble is a robust and accurate method for regression type applications as well. In terms of an adjusted R^2 value, random forest produces the best overall performance ($R^2 = 0.73$) and the predictive accuracy of the ensemble is statistically different from bagging and boosting. However, there was no significant difference between bagging and boosting. Similar results were obtained by Hamza and Larocque (2005) when they carried out an empirical comparison of ensemble methods using classification trees.

Besides obtaining the best overall predictive accuracy, using random forest as a wrapper allowed us to simplify the modeling process and identify the minimum number of spectral parameters that offer the best predictive accuracy. Using the backward elimination search function, we only used two spectral parameters while still producing the best overall predictive accuracy ($R^2 = 0.76$). More specifically, results show that WI and the Ratio₉₇₅ indices have the best ability to assay the water content of *S. noctilio* infested trees thus making it possible to remotely quantify the severity of damage caused by the wasp.

The ability of WI and the Ratio₉₇₅ to quantify water content can be explained by the significant variation in water content amongst the healthy, green and red stages. Physiological research has shown that tree mortality due to *S. noctilio* infestation is linked to the combined effects of a toxic mucus and the fungus *Amylostereum areolatum* that is injected into the tree by the female wasp during oviposition (Slippers et al., 2003). The mucus changes the water balance of the tree, thereby creating conditions that are ideal for the growth and spread of the fungus. In turn, the fungus rots and dries the wood, providing a suitable environment for the survival and development of the insect larvae (Slippers et al., 2003).

Additionally Coutts (1970) showed that water content of trees decreased rapidly after only 2–3 weeks of infestation. Thus using spectral indices like WI and the Ratio₉₇₅ which directly measure spectral variance caused by varying plant water status makes it possible to detect *S. noctilio* infestations from an early stage of infestation when the canopy appears green rather than relying on the appearance of reddish-brown foliage which occurs during the later red stage of infestation.

5. Conclusions

The results from this study show that (i) there is a strong link between existing spectral indices (WI and the Ratio₉₇₅) and the water status of *P. patula* foliage thereby improving the chances of remotely detecting *S. noctilio* at a landscape level; (ii) the random forest ensemble provides the best overall predictive accuracy when compared to the boosting and bagging ensembles and (iii) using random forest as part of a wrapper allowed us to simplify the modeling process and identify the minimum number of spectral parameters that offer the best predictive accuracy. Ultimately, this study provides the foundation for the potential upscaling of results to either an airborne or spaceborne platform. This is especially pertinent since it is envisaged that South Africa will soon launch the ZASat-003 satellite that will carry a hyperspectral sensor thus making high spectral resolution data more accessible and available to remote sensing researchers in the country.

Acknowledgements

We thank Sappi for allowing us access to the Pinewoods plantations. The contributions of Marcel Verleur in identifying *Sirex noctilio* infestations are gratefully acknowledged. We thank Wayne Jones for assisting with the sampling of pine needles. Additionally, we appreciate all the computer programming assistance provided by Chris Muncaster. Eric Economon from the Agricultural Research Centre (ARC) of South Africa provided

assistance with the ASD spectroradiometer. Funding for this research was provided by the National Research Foundation (NRF) South Africa.

References

- Bowyer, P., Danson, F.M., 2004. Sensitivity of spectral reflectance to variation in live fuel moisture content at leaf and canopy level. *Remote Sensing of Environment* 92, 297–308.
- Breiman, L., 1996. Bagging predictors. *Machine Learning* 26, 123–140.
- Breiman, L., 2001. Random forests. *Machine Learning* 45, 5–32.
- Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, California.
- Chan, J.C., Paelinckx, 2008. Evaluation of random forest and adaboost tree based ensembles classification and spectral band selection for ecotope mapping airborne hyperspectral imagery. *Remote Sensing of the Environment* 112, 2999–3011.
- Couts, M.P., 1970. The physiological effects of the mucus secretion of *Sirex noctilio* on *Pinus radiata*. *Australian Forest Research* 4, 23–26.
- DeFries, R.S., Hansen, M., Steininger, M., Dubayah, R., Sohlberg, R., Townshend, J., 1997. Subpixel forest cover in Central Africa from multisensor, multitemporal data. *Remote Sensing of Environment* 60, 228–246.
- Efron, B., Tibshirani, R., 1993. *An Introduction To Bootstrapping*. Monographs on Statistics And Applied Probability. Chapman and Hall/CRC, Boca Raton, Florida, New York, p. 436.
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77, 802–813.
- Freund, Y., Shapiro, R.E., 1996. Experiments with a new boosting algorithm. In: *Machine learning proceedings of the 13th international conference*. Morgan-Kaufman, San Francisco, California, pp. 148–156.
- Friedman, J., 2002. Stochastic gradient boosting. *Computational Statistics and Data Analysis* 38, 367–378.
- Gao, B.C., 1996. NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* 58, 257–266.
- Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random Forests for land cover classification. *Pattern Recognition Letters* 27, 294–300.
- Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research* 3, 1157–1182.
- Ham, J., Chen, Y., Crawford, M., Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing* 43.
- Hamza, M., Larocque, D., 2005. An empirical comparison of ensemble methods based on classification trees. *Journal of Computation and Simulation* 75, 629–643.
- Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Sohlberg, R., Dimiceli, C., Carroll, M., 2002. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment* 83, 303–319.
- Hardinsky, M.A., Klemas, V., Smart, M., 1983. The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of *Spartina alterniflora* canopies. *Photogrammetric Engineering and Remote Sensing* 49, 77–83.
- Hastie, T., Tibshirani, R., Friedman, J., 2001. *The Elements Of Statistical Learning: Data Mining, Inference And Prediction*. Springer-Verlag, New York, p. 500.
- Hunt, E.R., Rock, B.N., 1989. Detection of changes in leaf water content using near and middle infrared reflectances. *Remote Sensing of Environment* 30, 43–54.
- Hurley, B., Slippers, B., Wingfield, J., 2007. A comparison of control results for the alien invasive woodwasp, *Sirex noctilio*, in the southern hemisphere. *Agricultural and Forest Entomology* 9, 159–171.
- Ismail, R., Mutanga, O., Ahmed, F., 2008. Discriminating *Sirex noctilio* attack in pine forest plantations in South Africa using high spectral resolution data. In: Kalacska, M., Sanchez-Azofeifa, A. (Eds.), *Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests*. CRC Press, Taylor and Francis, p. 350.
- Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. *Artificial Intelligence Journal* 97, 273–324.
- Lawrence, R.L., Bunn, A., Powell, S., Zambon, M., 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of Environment Remote Sensing of Environment* 90, 331–336.
- Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForests). *Remote Sensing of Environment* 100, 356–362.
- Lobell, D.B., Ortiz-Monasterio, J.L., Asner, G.P., Naylor, R., Falcon, W., 2007. Combining field surveys, remote sensing and regression trees to understand yield variations in an irrigated wheat landscape. *Agronomy Journal* 97, 241–249.
- Michaelson, J., Schimel, D.S., Friedl, M.A., Davis, F.W., Dubayah, R.O., 1994. Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys. *Journal of Vegetation Science* 5, 673–696.
- Mutanga, O., Skidmore, A.K., 2004. Integrating imaging spectroscopy and neural networks to map tropical grass quality in the Kruger National Park, South Africa. *Remote Sensing of Environment* 90, 104–115.
- Mutanga, O., Skidmore, A.K., Prins, H.H.T., 2004. Predicting in situ grass quality in the Kruger National Park, south Africa, using continuum-removed absorption features. *Remote Sensing of Environment* 89, 393–408.
- Pal, M., 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing* 26, 217–222.
- Peñuelas, J., Pinol, J., Ogaya, R., Filella, I., 1997. Estimation of plant water concentration by the reflectance Water Index WI (R900/R970). *International Journal of Remote Sensing* 18, 2869–2875.
- Pontius, J., Hallet, R., Martin, M., 2005. Assessing Hemlock decline using visible and near-infrared spectroscopy: Indices comparison and algorithm development. *Applied Spectroscopy* 59, 836–843.
- Prasad, A., Iverson, L., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199.
- Pu, R., Ge, S., Kelly, N.M., Gong, P., 2003. Spectral absorption features as indicators of water status in coast live oak (*Quercus agrifolia*) leaves. *International Journal of Remote Sensing* 24, 1799–1810.
- R Development Core Team 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. In: *Monitoring vegetation systems in the Great Plains with ERTS Third ERTS Symposium*. NASA SP-351. Goddard Space Flight Center, Washington D.C., pp. 309–317.
- Slippers, B., Coutinho, T.A., Wingfield, B.D., Wingfield, M.J., 2003. A review of the genus *Amylostereum* and its association with woodwasps. *South African Journal of Science* 70–74.
- van Aardt, J., Coppin, P., 2006. Current state and potential of the IS-HS project: Integration of in situ data and hyperspectral remote sensing for plant production modeling. In: Ackerman, P.A., Längin, D.W.A.M.C. (Eds.), *Precision Forestry in Plantations, Semi-Natural And Natural Forests*. Proceedings of the International Precision Forestry Symposium, Stellenbosch University, Stellenbosch, Stellenbosch University.