South African Journal of Botany

Landsat-8 based coastal ecosystem mapping in South Africa using random forest classification in Google Earth Engine

Mariel Bessinger [a,b,] Melanie Lück-Vogel [b,c,] Andrew Skowno [a,] Ferozah Conrad [a]

[a] South African National Biodiversity Institute, Kirstenbosch Research Centre, Cape Town, South Africa

[b] Department of Geography & Environmental Studies, Stellenbosch University, Stellenbosch, South Africa

[c] Coastal Systems Research Group, Council for Scientific and Industrial Research CSIR, Stellenbosch, South Africa

Abstract

Coastlines worldwide are home to an increasing number of people and are subject to many pressures. This, combined with natural dynamics and hazards, often results in the degradation of coastal and marine ecosystems and infrastructure. Therefore, it is necessary to adopt effective management strategies to ensure sustainable use of coastal ecosystems, which requires up-to-date data on the extent of coastal ecosystems. This research aimed to create a coastal ecosystem land cover map for South Africa using the random forest algorithm to classify Landsat 8 imagery. Processing was done using the Google Earth Engine platform. A total of 522 Landsat 8 images were called to create a median image for classification. The impact of the number of trees, the number of variables per split, and variable selection on overall classification accuracy and Kappa values were evaluated. This was done by increasing the number of trees from 100 to 500 with increments of 100, setting the number of variables per split to three, four or five, and reducing the number of input variables from the original 18 variables, to the 10 most important variables, to the 5 most important variables, based on variable importance scores. Results suggest that the number of input variables used in the model had a greater impact on accuracy than the number of trees used, or the number of variables used per split. The average overall accuracy was 82.28%, with values ranging between 75.33% and 86.70%, while the average Kappa was 0.8068 and values ranged between 0.7310 and 0.8550. The model with the highest overall accuracy was the model using all input variables, 500 trees, and three variables per split. A major challenge was the misclassification of certain vegetation classes due to the complex successional mosaic they form, causing mixed signals and generally lower classification accuracy. Despite model limitations, results were satisfactory and have shown that coastal land cover classification and monitoring could be aided by the rapid classification of Landsat 8 imagery in Google Earth Engine using the random forest algorithm.