Etosha Heights Random Forest Vegetation Summaries

## Overview

The following tables summarize a series of random forest models to predict vegetation communities across the Etosha Heights Private Reserve. The overarching idea is that seasonal timeseries of satellite observations are better suited to mapping fine-scale vegetation communities than a single image because the vegetation communities will respond differently to seasonal variation in precipitation and other environmental drivers. The following analyses utilize a stack of 25 Planet SuperDove images spanning the period from August 2023 to July 2024. Each image includes 8 spectral bands at 3 meter spatial resolution.

## Predicting Vegetation Communities with NDVI

To begin, we’ll predict all 13 vegetation community classes using NDVI since that is the metric that captures the greatest degree of variability between the different vegetation communities across the growing season. The table below clearly indicates that prediction accuracy increases with the number of NDVI images. Using the full stack of 25 images yields approximately 75% accuracy (note the Kappa is a similar metric that takes into account the likelihood of achieving an accurate prediction due to chance).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Accuracy and Kappa statistics for Random Forest models predicted using varying numbers of NDVI images. n indicates the number of images used for predictions. Images were selected as evenly spaced sequences from the entire seasonal time series.   | Accuracy | Kappa | n | | --- | --- | --- | | 0.7506702 | 0.7163359 | 25 | | 0.6769437 | 0.6316600 | 13 | | 0.6112601 | 0.5579729 | 9 | | 0.5536193 | 0.4920064 | 7 | | 0.4316354 | 0.3563273 | 5 | | 0.4865952 | 0.4186970 | 5 | | 0.4061662 | 0.3297571 | 4 | | 0.3699732 | 0.2838907 | 4 | | 0.3016086 | 0.2145559 | 3 | | 0.2801609 | 0.1901861 | 3 | | 0.2815013 | 0.1932255 | 3 | | 0.2908847 | 0.1982764 | 3 | |

## Predicting Vegetation Communities with Multispectral Images from a Single Date

As a next step we can predict vegetation communities using individual 8-band multispectral images. We’ll do this for each of the 25 images in the time series to see if a particular seasonal timing yields better predictions. As the table below demonstrates, accuracy ranges from 53% to 67%, and while the higher accuracy generally occurs early in the rainy season during green-up, there is not an extremely strong seasonal trend. We can also compare this to the NDVI results above in Table 1, and note that using 8 multispectral bands from a single date yields prediction accuracy similar to 9 NDVI images distributed across the study period. This suggest that the number of predictor variables might be more important that capturing seasonal variability, or more specifically, the 8 multispectral bands from a single date capture as much spatial heterogeneity as a similar number of single-band NDVI images distributed across the annual cycle.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Accuracy and Kappa statistics for Random Forest models predicted using individual multispectral images. file indicates the image date (yyyymmdd).   | Accuracy | Kappa | file | | --- | --- | --- | | 0.5563003 | 0.5003642 | 20230818 | | 0.6139410 | 0.5632700 | 20230930 | | 0.6085791 | 0.5557495 | 20231102 | | 0.5683646 | 0.5125222 | 20231128 | | 0.6742627 | 0.6314132 | 20231207 | | 0.5683646 | 0.5111223 | 20231213 | | 0.6273458 | 0.5771432 | 20231220 | | 0.6434316 | 0.5959653 | 20231229 | | 0.5911528 | 0.5398870 | 20240102 | | 0.5495979 | 0.4889400 | 20240112 | | 0.5750670 | 0.5164996 | 20240118 | | 0.6126005 | 0.5625992 | 20240121 | | 0.6032172 | 0.5547023 | 20240210 | | 0.5697051 | 0.5154060 | 20240216 | | 0.5804290 | 0.5270304 | 20240302 | | 0.5764075 | 0.5198752 | 20240315 | | 0.5576408 | 0.4986508 | 20240405 | | 0.5603217 | 0.5052020 | 20240410 | | 0.5924933 | 0.5408738 | 20240418 | | 0.5924933 | 0.5403453 | 20240430 | | 0.6058981 | 0.5543164 | 20240507 | | 0.6179625 | 0.5689732 | 20240519 | | 0.5536193 | 0.4954433 | 20240530 | | 0.6112601 | 0.5604123 | 20240621 | | 0.5321716 | 0.4707628 | 20240712 | |

## Predicting Vegetation Communities with Multispectral Images from Multiple Dates

To start we can predict land cover with an increasing number of multispectral images to examine how adding more data improves predictions. As the table below shows, accuracy increases as multiple images are used for prediction, however the magnitude of increase declines as successive images are added. For example going from one to two images increases accuracy by roughly 10%, whereas going from 5 to 6 predictor images (40 to 48 bands) only increases prediction accuracy by less than 1%. The overall accuracy of ~86% is quite good.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: Accuracy and Kappa statistics for Random Forest models predicted using multiple multispectral images. n indicates the number of multispectral bands used for the prediction. Since each images has 8 bands, n/8 is equal to the number of images. For this analysis we use the top 6 images from Table 2 above in descending order (i.e. n = 8 used the most accurate image, n = 16 used the two most accurate images, etc…)   | Accuracy | Kappa | n | | --- | --- | --- | | 0.6742627 | 0.6314132 | 8 | | 0.7721180 | 0.7413863 | 16 | | 0.8150134 | 0.7908628 | 24 | | 0.8378016 | 0.8166786 | 32 | | 0.8632708 | 0.8454219 | 40 | | 0.8713137 | 0.8545614 | 48 | |

As a lest step we can examine accuracy with 5 multispectral images that are randomly selected from our timeseries stack (as opposed to selecting the five best individual images). This will be a useful way of knowing whether we can simply use any five images, or if the timing somehow matters.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4: Accuracy and Kappa statistics for Random Forest models predicted using 5 multispectral images selected at Random. n indicates the number of multispectral bands used for the prediction. Since each images has 8 bands, n/8 is equal to the number of images.   | Accuracy | Kappa | n | | --- | --- | --- | | 0.8445040 | 0.8241481 | 40 | | 0.8780161 | 0.8619805 | 40 | | 0.8699732 | 0.8533624 | 40 | | 0.8123324 | 0.7868937 | 40 | | 0.8190349 | 0.7949101 | 40 | |

## Summary

Overall these results are quite good. Our highest accuracy values are typical to those observed for broad land cover classification tasks that often include classes that are much more distinct that the savanna vegetation communities that we’re predicting here. The pragmatic utility of our analysis is that it provides some guidelines for managers and researchers wanting to make fine-scale habitat maps. The results indicate that with five random images you can make a vegetation community/habitat map that is quite good. This seems like a good hook for this paper in my view. I might do a bit more tinkering to see if precipitation timing impacts model accuracy for the individual images, but on the whole this seems like a good near-final endpoint.