Estimating Kelp cover from LANDSAT imagery

caperadd.com/news/estimating-kelp-cover-from-landsat-imagery/

Dylan Irion

Monitoring the distribution of kelp

We love kelp here at Cape RADD, if you haven't gotten that by now then I don't know what you're reading. I try to use the word kelp in a sentence at least once a day and I am pretty sure I've seen Jessie feed some to Mike before. Any way, kelp was previously thought to be an important sentinel species whose decline would be the start of the bad things to come from climate change, motivating the need for monitoring the distribution and size of kelp forests off the coast of California. This view is being challenged but doesn't diminish the importance of kelp in the ecosystem, and a need for monitoring the health of this species. This post will be the first in many R Tutorials introducing some of the analytical methods we use to conduct our research at Cape RADD. But first, some background.

The LANDSAT Missions

The LANDSAT series of satellites have orbited our planet for 40+ years, collecting the longest global time-series of multispectral data for the Earth's surface. Each LANDSAT satellite maintains a sun-synchronous near polar orbit of the Earth, passing the same place once every 16 days. This data is used in studies of agriculture, forestry, geology, land cover mapping, resource management, and water and coastal environments. In this exercise, we will be using Level 2 data from the LANDSAT 8 mission. This data comes preprocessed into measurements of surface reflectance across 7 Bands; 4 visible spectra, and 3 infrared

False colour infrared LANDSAT8 Operational Land Imager and Thermal Infrared Sensor Scene over Cape Town.

spectra. Most LANDSAT products are free to order using the USGS EarthExplorer. This tutorial is going to assume you have downloaded some LANDSAT imagery, we'll be using row 84 from path 175 from the 2014-07-14 flyover for this demo. Ideally we want a scene with as little cloud cover as possible. I have clipped this scene to exclude some of the open ocean that we aren't interest in.
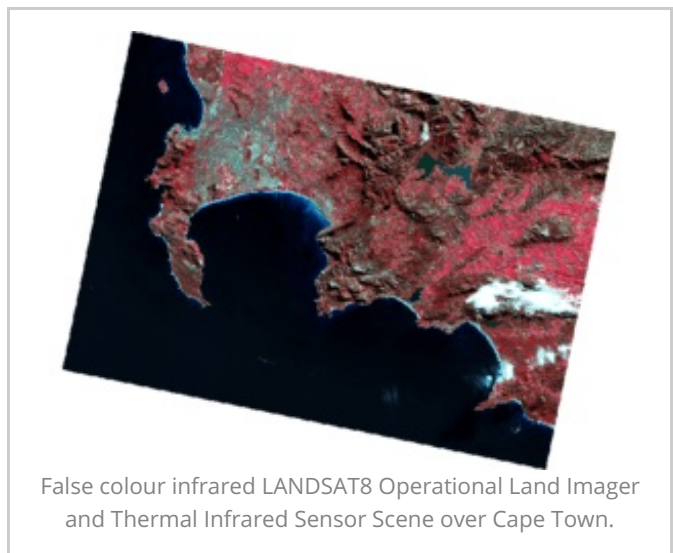
Multiple Endmember Spectral Mixture Analysis (MESMA)

Plants are green because their chloroplasts contain a pigment that absorbs blue and red light, reflecting the remaining visible spectra (green) back into our eyes. Scientists use these same properties to remotely sense different vegetation and land cover types. This is part of the concept behind a measure called Normalised Difference Vegetation Index (NDVI). We are going to use this information to differentiate kelp from water in a spectral analysis. Spectral unmixing decomposes an image, pixel-by-pixel, into two or more 'endmembers' representing pure cover types. MESMA improves upon traditional spectral mixture analysis by allowing each endmember class (e.g. land, or kelp) to be represented by more than one endmember. This allows the unmixing to account for variability in surface reflectance, such as that created by waves, surface glint, and topsoil runoff in the coastal environment. The RStoolbox package in R was recently updated to include a function for running a MESMA which we will be using to recreate the analysis performed in this paper.

> Cavanaugh KC, Siegel DA, Reed DC, Dennison PE (2011) Environmental controls of giant-kelp biomass in the Santa Barbara Channel, California. Mar Ecol Prog Ser 429:1-17.
> https://doi.org/10.3354/meps09141

The first step is to install and load the necessary packages for manipulating spatial data and performing the spectral analysis.

```
install.packages("raster", "rgdal", "RStoolbox")
library(RStoolbox)
library(raster)
library(rgdal)
```

Next you'll need to unzip and read in the LANDSAT imagery that you downloaded. Each band is a separate *.tif* file.

```
#load bands
r20140714<-brick(list("/home/dylan/Documents/Projects/Cape RADD/Kelp
LANDSAT/data/LANDSAT/LC08_L1TP_175084_20140714_20170421_01_T1_sr_band1.tif",
          "/home/dylan/Documents/Projects/Cape RADD/Kelp
LANDSAT/data/LANDSAT/LC08_L1TP_175084_20140714_20170421_01_T1_sr_band2.tif",
          "/home/dylan/Documents/Projects/Cape RADD/Kelp
LANDSAT/data/LANDSAT/LC08_L1TP_175084_20140714_20170421_01_T1_sr_band3.tif",
          "/home/dylan/Documents/Projects/Cape RADD/Kelp
LANDSAT/data/LANDSAT/LC08_L1TP_175084_20140714_20170421_01_T1_sr_band4.tif",
          "/home/dylan/Documents/Projects/Cape RADD/Kelp
LANDSAT/data/LANDSAT/LC08_L1TP_175084_20140714_20170421_01_T1_sr_band5.tif",
          "/home/dylan/Documents/Projects/Cape RADD/Kelp
LANDSAT/data/LANDSAT/LC08_L1TP_175084_20140714_20170421_01_T1_sr_band6.tif",
          "/home/dylan/Documents/Projects/Cape RADD/Kelp
LANDSAT/data/LANDSAT/LC08_L1TP_175084_20140714_20170421_01_T1_sr_band7.tif"))
```

Our first step in the process is to find some representative endmembers for both kelp and water spectra. Cavanaugh et al. (2011) went with one endmember for kelp, and 30 for water, due to the dynamic nature of water in the coastal environment, using a method described in the paper

below.

> Dennison PE, Roberts DA (2003) Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. Remote Sens Environ. 87:123-135. https://doi.org/10.1016/S0034-4257(03)00135-4
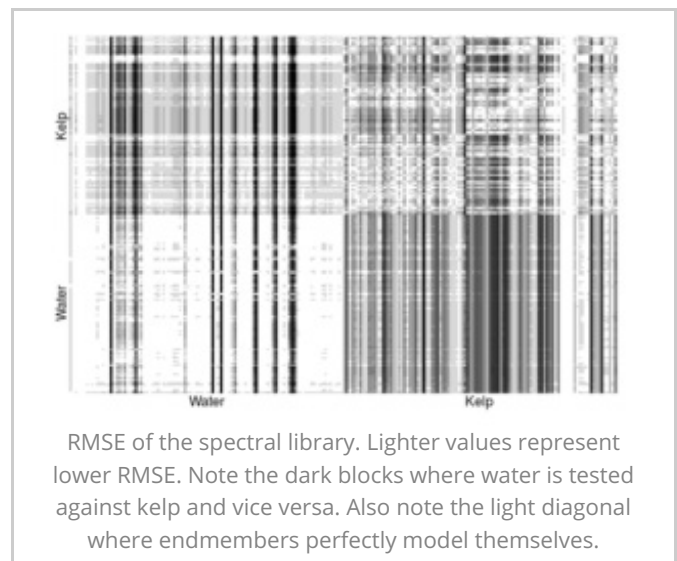
The method involves identifying reference polygons of areas dominated by at least 75% of one cover type, and occupying an area greater than 60mX60m so at least one pixel from the LANDSAT image can fall inside of this polygon. We used a GIS to find polygons of known land cover but ran into difficulty getting R to extract pixels under these polygons within a reasonable amount of time. In the end we identified 100 points in the image that represented known cover types by inspecting the visible spectra bands of the image and an additional aerial photography layer of the area. You can download the necessary shapefiles here, but note that they will only work for this particular LANDSAT scene.

```
#load reference points
water<-readOGR("/home/dylan/Documents/Projects/Cape RADD/Kelp LANDSAT/data/shp/",layer="water_pts")
kelp<-readOGR("/home/dylan/Documents/Projects/Cape RADD/Kelp LANDSAT/data/shp/",layer="kelp_pts")
```

We can now then use these points to extract the underlying pixels from the LANDSAT image to use in our spectral library. I additionally create a theoretical 'shade' endmember with no reflectance. The *rownames* function just assigns a unique name so I can keep track of the cell that each endmember refers to.

```
#extract spectral library (candidate endmembers, 100 each)
water_em<-r20140714[water]
rownames(water_em)<-paste("water",extract(r20140714,water,cellnumbers=T)[,"cells"],sep=".")
kelp_em<-r20140714[kelp]
rownames(kelp_em)<-paste("kelp",extract(r20140714,kelp,cellnumbers=T)[,"cells"],sep=".")
shade_em<-matrix(c(-2000,-2000,-2000,-2000,-2000,-2000,-2000),nrow=1,dimnames=list("shade",c(colnames(water_em))))
em<-rbind(water_em,kelp_em,shade_em)
```

In order to find the most representative endmember for each class we will now run a series of two-endmember MESMAs on each pixel of our spectral library. Each MESMA will contain one of the other candidate endmembers and our theoretical shade endmember. The idea here is to find the endmember that best represents all the other endmembers in its class. We will assess this by calculating the root-mean-square error (RMSE) for the candidate endmember and selecting the one with the lowest RMSE. This step is a little time-consuming. A for loop cycles

RMSE of the spectral library. Lighter values represent lower RMSE. Note the dark blocks where water is tested against kelp and vice versa. Also note the light diagonal where endmembers perfectly model themselves.

through each endmember, and another
nested for loop cycles through all the remaining endmembers and stores the result of the
MESMA in a matrix called *rmse*.

```
#for each spectrum (candidate endmember) run n-1 two endmember models (next endmember + shade
endmember)
rmse<-matrix(numeric(),nrow=nrow(em)-1,ncol=nrow(em)-1,dimnames=list(head(rownames(em),-
1),head(rownames(em),-1)))
for(i in 1:(nrow(em)-1)) {
  for(j in 1:(nrow(em)-1)) {
    rmse[i,j]<-getValues(subset(mesma(brick(array(em[i,],dim=c(1,1,7)))),em[c(j,nrow(em)),]),"RMSE"))
  }
}
```

Cavanaugh et al. (2011) then take the lowest 30 scoring water endmembers and the lowest
single kelp endmember for their full image classification.

```
#water (lowest 30 RMSE)
names(sort(colMeans(rmse[1:100,1:100],na.rm=T))[1:30])

#kelp (single lowest RMSE)
names(which.min(colMeans(rmse[101:200,101:200],na.rm=T)))
```

For the sake of speed we will use only the best 10 scoring water endmembers. The next line
stores these in a matrix called *em_final* and names them according to the class they were pulled
from using a regular expression that simply strips everything after and including the first period
in the original row names that I generated earlier.

```
em_final<-
em[c(names(which.min(colMeans(rmse[101:200,101:200],na.rm=T))),names(sort(colMeans(rmse[1:100,1:100],na.rm=T))
[1:10])),]
rownames(em_final)<-gsub("\\..*","",rownames(em_final))
```

As a last preliminary step, I'd also like to mask out the land and clouds from our analysis. The
LANDSAT files you downloaded contain a Quality Assessment band that does its best to identify
clouds and other features that may be important to an analysis. You can read more about what
each value of this band means here. Pixels with a high confidence of being clouds will have a
value of 480 or 992. We'll use this information to set the values in our scene to NA. This code
iterates through each band of the scene and sets pixels to NA if they are though to be clouds, or
retains the original value otherwise. I've also done this with a land mask layer, but I will leave
that as an exercise for you.

```
qa<-brick(readGDAL("/home/dylan/Documents/Projects/Cape RADD/Kelp
LANDSAT/data/LANDSAT/LC08_L1TP_175084_20140714_20170421_01_T1_pixel_qa.tif"))
for(i in 1:7) {
  values(r20140714[[i]])<-
ifelse(getValues(qa$band1)==480|getValues(qa$band1)==992,NA,getValues(r20140714[[i]]))
}
```

We're now ready to run our classification routine. This will consist of a series of two-endmember MESMAs using our kelp endmember, and one of the water endmembers and then for each pixel, selecting the model with the lowest RMSE. In Cavanaugh et al. (2011) this meant 30 models, but we're going to keep things relatively quick with just 10 water endmembers for a total of 10 models.

```
#model scene iteratively, replacing pixels if a model scores a lower RMSE at each step
for(i in 1:10) {
  if(i==1) {
    scene<-mesma(r20140714,em_final[c(1,i+1),],filename="/home/dylan/Documents/Projects/Cape RADD/Kelp
LANDSAT/data/r20140714_mesma.tif",format="GTiff",options="INTERLEAVE=BAND",overwrite=T)
  } else {
    candidate<-mesma(r20140714,em_final[c(1,i+1),])
    scene<-writeRaster(((scene$RMSE>candidate$RMSE)*candidate)+
((candidate$RMSE>scene$RMSE)*scene),filename="/home/dylan/Documents/Projects/Cape RADD/Kelp
LANDSAT/data/r20140714_mesma.tif",format="GTiff",options="INTERLEAVE=BAND",overwrite=T)
    names(scene)<-names(candidate)
    rm(candidate)
  }
  gc(verbose=F)
}
```

Some Remarks

OK, so we can clearly see that this isn't perfect, or at least I can. We detected a lot of kelp along stretches of sandy beaches that I know should not contain much living kelp. There is a small chance we are detecting dead kelp that has washed ashore, but I am sceptical. Having said that, our spectral library was created from just one image where in practice we would be using the full range of scenes to create a robust set of candidate endmembers. We could also stand to be more systematic in the creation of our spectral library, rather than just picking 100 points that I liked from a map. We can actually see that a lot of our candidate

A zoomed in view of the Cape Peninsula, highlighting areas where the MESMA has predicted the presence of kelp

kelp endmembers were not great at modelling their class-mates (the kelp-kelp section in the second figure should be more uniformly light, I would also like to see darker inter-class blocks). We also had to make a few deviations from the methods referenced in the papers due to using different software, and for the sake of time. All in all however, we did succeed in laying out a protocol with some ability to semi-autonomously identify kelp, and there is something to be said for that! We could use this to come up with an estimate of area covered, and monitor how that changes with time, which is precisely what we aim to do!
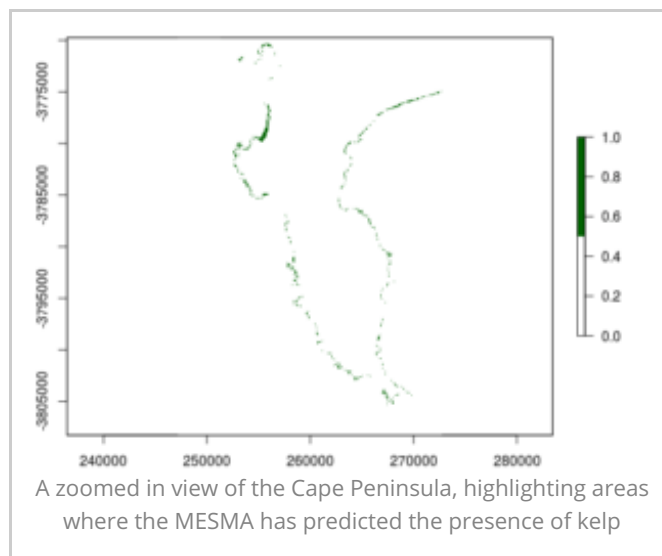
The kelp forests in Cape Town exist very close to shore and appear to be significantly smaller than those off the coast of California, which makes this strategy all the more difficult. Future plans at Cape RADD include timing our research dives with satellite passes in order to record areas of known kelp-cover for use in our endmember selection. There is also the potential to source alternative, higher resolution, multi-spectral data sources from aerial photography. It should make for an exciting and fun project.