7 Leaf and canopy radiative transfer modeling

Aim

So far, all our analyses were based on statistical models. These models are able to accurately describe observed relationships between spectral data and thematic information. They do not, however, take into account the physical principles behind these relationships. The radiative transfer models PROSPECT (Jacquemoud & Baret 1990) and 4SAIL (Verhoef 1984) and other members of the PROSPECT/SAIL model family address this issue and model the path of light through a leaf and a vegetation canopy, respectively. In this training, we explore the coupled PROSPECT and SAIL models (= PROSAIL) and use a statistical inversion of the model to map Leaf Mass per Area (LMA [g*cm^{-2}]), a functional plant trait, in a wetland under water stress.

PROSAIL

The PROSAIL model has been designed to simulate the spectral signal of vegetation canopies based on a set of nine parameters that characterize its structure, biochemistry, and background soil (table 7.1). Additional parameters (which we will not consider today) define the illumination and observer geometries. Of the leaf and canopy parameters, some originate from the PROSPECT model and describe the biochemistry and structure of the leaves. The model uses information on contents of pigments, water, dry matter, and proteins as well as leaf structure in combination with the known absorption coefficients to simulate the reflectance spectrum as measured at the leaf. This spectrum is then passed to the SAIL model. Requiring three additional parameters describing the canopy structure and background soil reflectance, the SAIL model is able to simulate the scattering of light within the canopy, resulting in the simulated canopy reflectance spectrum.

The source code of the PROSPECT and SAIL models is provided in various implementations freely on the web (http://teledetection.ipgp.jussieu.fr/prosail/). French researcher Jean-Baptiste Feret provides implementations for R through his github repository (https://jbferet.gitlab.io/prosail/), which we are using today. Please install them or the corresponding Python implementation.

Table 7.1: Parameters used in the PROSAIL model

Parameter	Description	Unit
N	Leaf structure parameter (= # mesophyll layers)	-
CHL	Chlorophyll a & b concentration	µg cm^{-2}
CAR	Carotenoid concentration	µg cm^{-2}
ANT	Anthocyain concentration	µg cm^{-2}
BROWN	Brown pigment content	-
EWT	Equivalent water thickness	cm
LMA	Leaf dry mass per area	g cm^{-2}
PROT	Protein content	g cm^{-2}
CBC	NonProtCarbon-based constituent content	g cm^{-2}
LAI	Leaf area index	-
lidfa, lidfb	Leaf angle distribution (average leaf slope and biomodality)	-

Q6.1: What are the default values for N, CHL, CAR, EWT, LMA, LAI and leaf angle distribution?

As a first application, we explore how a changing LAI affects the canopy spectrum when all other canopy characteristics remain constant. The model output provides several reflectance and absorption components. We use the hemispherical-directional reflectance factor in viewing direction (rdot), which corresponds to the sensor's observation.

```
library (prosail)
wl <- 450:2500 ## simulated wavelength range
lai1 <- PRO4SAIL (lai=1)
plot (wl, lai1$rdot, ylim=c (0, 0.6), col=1, type="l", xlab="Wavelength / nm",
            ylab="Reflectance")
lai2 <- PRO4SAIL (lai=2)
lines (wl, lai2$rdot, col=2)
lai3 <- PRO4SAIL (lai=3)
lines (wl, lai3$rdot, col=3)
lai4 <- PRO4SAIL (lai=4)
lines (wl, lai4$rdot, col=4)
lai5 <- PRO4SAIL (lai=5)
lines (wl, lai5$rdot, col=5)
lai6 <- PRO4SAIL (lai=6)
lines (wl, lai6$rdot, col=6)
lai7 <- PRO4SAIL (lai=7)
lines (wl, lai7$rdot, col=3)
lai0 <- PRO4SAIL (lai=0)
lines (wl, lai0$rdot, col=8)
```

Q7.2: Which parts of the spectrum are changing and which remain stable? Is the relationship between LAI and canopy reflectance strictly linear?

The PRO4SAIL() function claims to accept whole tables of parameters or parameter combinations as data.frame. This functionality eases a systematic exploration of the response to changing parameters. Within the parameter data.frame, each column represents one parameter and is assigned to the corresponding argument via the column name. This allows, for example, for an in-depth analysis of the influence of the chlorophyll concentration within the leaf on canopy reflectance. Unfortunately, the current package version under Ubuntu refuses to keep up with this promise. Maybe it works for your system. I hence run a quick-and-dirty work-around employing a loop. If a parameter table works for your implementation, generate a vector of concentrations to be considered and convert it into a data.frame. The name of the respective column has to correspond to the argument for the chlorophyll concentration in the PROSAIL() function. The parameter list is then passed to the function using the 'Input_PROSPECT' argument. Here, I am taking an inelegant detour (sorry!).


```
CHL <- data.frame ('CHL'= seq (0.3, 48, 0.3))
Cab.specs <- matrix (NA, nrow=nrow(CHL), ncol=2101)
for (i in 1:nrow (CHL))
  Cab.specs[i,] <- PRO4SAIL (CHL = CHL[i,])$rdot
```


We plot these spectra using a color gradient to illustrate the changing chlorophyll concentration.
```
wl <- 400:2500 ## Wavelength vector
greencol <- colorRampPalette (c ("orange4", "green", "darkgreen", "black"))
```

```
plot (wl, Cab.spec[1,], type="n", ylim=c (0, 0.5), xlab="Wavelength / nm",
      ylab="Reflectance", main="Chlorophyll µg/cm2")
for (i in 1:160)
   lines(wl, Cab.spec[i,], col=greencol (160)[i])
legend ("topright", legend=c (1.5, 7.5, 15, 24, 36, 48), lwd=1,
        col=greencol (160)[c (5, 25, 50, 80, 120, 160)], ncol=6, cex=0.6)
```

Q7.3: Explain in 2-3 sentences the influence of chlorophyll on canopy reflectance.

Building a lookup table

Many applications of statistical models as we used them are hampered by the sample size of the available ground truth data. In particular the data required for spectral analyses of ecosystem properties such as functional plant traits or leaf biochemistry are difficult to obtain since they require a vast number of lab analyses of perishable plant samples. In such cases, radiative transfer models offer the opportunity to simulate a large set of spectral data related to known biochemical and structural properties. These simulated data can subsequently be used to train a statistical model. The pairwise sets of biochemical and structural properties on the one hand and corresponding simulated spectral data are known as lookup tables.

As a final exercise in this class, we will use the PROSAIL model to build a large lookup table for the retrieval of Leaf Mass per Area (which corresponds to the LMA parameter and can be further broken down into Proteins and Carbon-based constituents) from spectral data. This lookup table will then be used to train a random forest regression model that we apply onto image data to map LMA across our study site.

LMA is a species-specific functional plant trait that describes the mass per area ratio of a dry leaf. It corresponds to the thickness of plant leaves. Like its inverse Specific Leaf Area LMA is related to the relative growth rate and to the investment of a plant in its leaves. Richness in resources leads in general to decreasing LMA values. Differences in LMA between species as well as among individuals of a single species can thus be used as an indicator for the response of plants to the prevailing environmental conditions (Cornelissen et al. 2003).

First, we generate a parameter table that covers all possible combination of seven PROSAIL parameters (we leave the leaf angle distribution with the default 'spherical', the soil background with a dry soil spectrum that corresponds to the soil condition at the time of image acquisition and do not go into the individual constituents of LMA). The ranges of the seven parameters are assessed from literature. LMA is sampled in detail, the other parameters are included in coarser intervals. The variation of the other parameters is necessary to enable a robust retrieval of LMA for any given pigment and water concentration independent from the canopy structure and background. Ranges and intervals for the parameters are defined as vectors; the expand.grid() function builds a data.frame containing all possible combinations.

```
N <- seq (1.1, 2.3, 0.15)
CHL <- seq (1, 46, 3)
CAR <- seq (1, 25, 2)
BROWN <- seq (0, 1, 1)
EWT <- seq (0.002, 0.024, 0.002)
LAI <- seq (0.5, 7, 0.5)
parameter <- expand.grid (N, CHL, CAR, BROWN, EWT, LAI)
```

Q7.4: How many combinations of the six parameters (not including LMA) are possible for the given ranges and intervals?

Since the number of possible combinations is too large for our processing capabilities, we draw a random and representative statistical sample of this list (*n*=20000). This number is considerably larger than any sample size that we could generate with lab analyses, still covers the multidimensional parameter space rather well, and can be easily processed.

```
parameter <- parameter [sample (1:nrow (parameter))[1:20000],]
```

Now we add LMA to the parameter list.

```
LMA <- seq (0.002, 0.018, length.out=20000)
parameter <- cbind (LMA, parameter)
colnames (parameter) <- c("LMA", "N", "CHL", "CAR", "BROWN", "EWT", "LAI")
```

The order of the combinations is randomized and passed to the PROSAIL model, resulting in a data set that completes the lookup table.

```
parameter <- parameter [sample (1:20000),]
specs <- matrix (NA, nrow=20000, ncol=2101)
for (i in 1:20000){
  specs[i,] <- PRO4SAIL (Input_PROSPECT=parameter[i,],
            rsoil=SpecSOIL$Dry_Soil)$rdot}
```

Image data

The image data were taken by the airborne sensor HyMap operated by the German Aerospace Center (DLR) in 2009 over the wetland Las Tablas de Daimiel, a national park in Central Spain (see Feilhauer et al. 2018). Over the past five decades, human induced activities led to decreasing water levels in the wetland. During the most recent drought period in the 21st century, the wetland was almost dry with no surface water lamina in 2009. In 2010, heavy rainfalls and management actions allowed for a rewetting that stopped the degradation process. In a field campaign simultaneous to the acquisition of the image data, we sampled the vegetation (species composition, structural parameters and biochemistry) and soils (standard morphological and physicochemical parameters) in plots covering the northern part of the park. The HyMap data cover the solar electromagnetic spectrum from 455 nm to 2445 nm in 115 spectral bands with a spatial resolution of 5m x 5 m on the ground. Pre-processing of the data included radiometric, atmospheric, and geometric corrections, resulting in at-surface reflectance values. These data allow to analyze the spatial distribution of LMA across Las Tablas de Daimiel.

```
library (terra)
hymap <- rast ("090812_level1FGMosNorth.img")
plotRGB (hymap, 25, 13, 8, stretch="lin") ## false color composite
plotRGB (hymap, 13, 8, 2, stretch="lin") ## true color
```

Q7.5: Why is an accurate atmospheric correction of the image data inevitable if they are analyzed with the PROSAIL model?

Resampling the spectral library

The spectral resolution of the lookup table does not correspond to the resolution of the HyMap data. We thus have to resample the spectral library to the HyMap resolution. For this purpose, the little homemade function resamp() is applied on the spectra. This function linearly interpolates the spectra and extracts the reflectance values for the center wavelengths of the HyMap bands.

```
hymap.wl <- c(0.455400, 0.469400, 0.484300, 0.499100, 0.513800, 0.528800,
              0.543600, 0.558400, 0.573100, 0.588100, 0.602900, 0.617600,
              0.632000, 0.646500, 0.660900, 0.675500, 0.690000, 0.704500,
              0.718900, 0.733200, 0.747600, 0.761800, 0.775900, 0.790100,
              0.804600, 0.818800, 0.832900, 0.847100, 0.861100, 0.874700,
              0.887800, 0.893000, 0.908500, 0.923900, 0.939400, 0.955200,
              0.970400, 0.985800, 1.001400, 1.016600, 1.031800, 1.046900,
              1.062000, 1.076600, 1.091300, 1.106200, 1.120800, 1.135300,
              1.149700, 1.164100, 1.178600, 1.192800, 1.206900, 1.221000,
              1.235100, 1.249200, 1.263100, 1.277000, 1.290700, 1.304300,
              1.318300, 1.330100, 1.462600, 1.477100, 1.491100, 1.505000,
              1.518800, 1.532600, 1.546300, 1.559800, 1.573200, 1.586400,
              1.599500, 1.612700, 1.625900, 1.638900, 1.651700, 1.664400,
              1.677100, 1.689600, 1.702100, 1.714600, 1.726900, 1.739300,
              1.751500, 1.763600, 1.775600, 1.787600, 1.798100, 2.008300,
              2.027500, 2.046700, 2.065500, 2.084100, 2.102500, 2.120900,
              2.139000, 2.157000, 2.174700, 2.191700, 2.210300, 2.228100,
              2.245600, 2.263400, 2.280400, 2.297400, 2.314400, 2.331400,
              2.348300, 2.365000, 2.381500, 2.397700, 2.414100, 2.430300,
              2.446500) * 1000

resamp <- function (spec, wl, wl.out){ ## define the function
  out <- approx (wl, spec, xout=wl.out)$y
} ## end of the function

specs <- t (apply (specs, 1, resamp, wl=wl, wl.out=hymap.wl))
      colnames (specs) <- paste0 ("nm", hymap.wl)
```

Training a regression model for the statistical inversion

The resampled lookup table is then used to train a random forest regression (RFR) model. The regression version of random forest employs the same principles you have already learned, simply replacing the classification approach by regression models. For training and evaluation, we split the lookup table in a calibration and validation data set.

```
LMA <- parameter[,1]
specs.cal <- specs[1:5000,]
LMA.cal <- LMA[1:5000]
specs.val <- specs[5001:20000,]
LMA.val <- LMA[5001:20000]

library (randomForest)
rfmod <- randomForest (specs.cal, LMA.cal)
rfmod
```

Q7.6: How large is the percentage of variance in the calibration data that could be explained by the random forest model?

Now we use the validation data for an additional evaluation of the performance of the model.

```
pred <- predict (rfmod, specs.val)
```

```
dev.off ()
plot (LMA.val, pred, xlab="Observed / g*cm^-2", ylab="Predicted / g*cm^-2",
      main="LMA")
r2 <- cor (LMA.val, pred)^2
r2
```

Q7.7: Does the explained variance (=R^2*100) in validation differ from the explained variance in calibration?

```
rmse <- sqrt (mean ((LMA.val - pred)^2))
rmse
```

With this RMSE we can assume that the model is sufficiently able to retrieve LMA from the simulated spectral data. In the following, we hope that this assumption also holds true for a transfer to real-world image data.

Q7.8: Build another retrieval model using PLSR. Describe your modeling workflow. Does the model fit differ from the fit gained through RFR? Compare the band importance metrics of the RFR and PLSR. Do they agree on similar spectral bands that are best suited to retrieve LMA from the spectra?

Mapping LMA

To map LMA across the northern parts of Las Tablas de Daimiel, we simply apply the model onto the image data.

```
pixval <- values (hymap) / 10000
colnames (pixval) <- colnames (specs)
LMApix <- predict (rfmod, pixval)
LMAmap <- hymap[[1]]
values (LMAmap) <- LMApix
```

A shapefile with the boundaries of the wetland is used to mask all areas outside the wetland.

```
daimiel <- vect ("Las_Tablas_north.shp")
daimiel <- rasterize (daimiel, LMAmap)
LMAmap <- LMAmap * daimiel
lmacol <- colorRampPalette (c ("black", "orange4", "tan", "beige"))
plot (LMAmap, col=lmacol (100), main="LMA / g*cm^-2")
```

Q7.9: Describe the pattern of the LMA distribution in 2-3 sentences. Can you identify spatial trends across the wetland?

Cross-verification of the prediction

An independent validation of the LMA prediction would be desirable. Unfortunately, I was not smart enough to take LMA samples in the 2009 field campaign (the data were sampled for a different purpose). Still, we can extract LMA values for species growing in monostands that are in good vigor (LMA is changing under resource stress) and compare these predicted values to the values published for these species in trait data bases like the LEDA data base (Kleyer et al. 2008). In the field, we found monostands of *Cladium mariscus* (sawtooth sedge), *Phragmites australis* (reed), and *Scirpus maritimus* (saltmarsh bulrush). LEDA lists for these species the following LMA values:
• *C. mariscus* 0.01340483 g*cm^{-2}
• *P. australis* 0.006321113 g*cm^{-2}
• *S. maritimus* 0.00805153 g*cm^{-2}

Q7.10: Which species has thus the thickest leaves?

The shapefile 'species.shp' contains the locations of a monostand for each of these species. We use this shapefile to extract the predicted LMA values and compare them to the LEDA values.

```
plots <- vect ("species.shp")
points (plots, col=2, pch=19)
lmaplots <- t (zonal (LMAmap, plots, mean))
literature <- c (0.01340483, 0.006321113, 0.00805153)
par (pin=c (3, 3))
plot (literature, lmaplots, xlim=c (0.002, 0.018), ylim=c (0.002, 0.018),
    main="LMA / g*cm^-2")
abline (0, 1) # 1:1 line
text (literature, lmaplots, c ("C. mariscus", "P. australis", "S. maritimus"),
    pos=3, cex=0.7)
```

The predicted value ranges meet more or less the published LMA values for the respective species. In case of S. maritimus, I explain the deviation by the influence of the background soil, which was for this species site rather dark and is not considered in the odel. Still, the LMA map is able to illustrate species-specific differences in LMA. Further analyses are able to reveal changes in LMA due to resource stress. For example, *P. australis,* which is nowadays a dominant species in the park, shows a clear trend in LMA from the green center of the park near the tributary river feeding the wetland to the dry edges with high soil salinity. Multitemporal analyses covering different stages of wetness have the potential to monitor the response of the wetland vegetation to changing environmental conditions.

Q7.11: Read the paper by Darvishzadeh et al. (2011). They use an alternative inversion approach differing from our statistical solution. How does this approach work?

Q7.12: Post your own question for the discussion in Moodle.

References

Cornelissen JHC, Lavorel S, Garnier E, Diaz S, Buchmann N, Gurvich DE, Reich PB, ter Steege H, Morgan HD, van der Heijden MGA, Pausas JG, Poorter H (2003). A handbook of protocols for standardized and easy measurement of plant functional traits worldwide. Australian Journal of Botany 51, 335-380.

Darvishzadeh R, Atzberger C, Skidmore A, Schlerf M (2011). Mapping grassland leaf area index with airborne hyperspectral imagery: a comparison study of statistical approaches and inversion of readiative transfer models. ISPRS Journal of Photogrammetry and Remote Sensing 66, 894-906.

Feilhauer H, Schmid T, Faude U, Sánchez-Carillo S, Cirujano S (2018). Are remotely-sensed traits suitable for ecological analysis? A case study of long-term drought effects on Leaf Mass per Area of wetland vegetation. Ecological Indicators 88, 232-240.

Jacquemoud S, Baret F (1990). PROSPECT: a model of leaf optical properties spectra. Remote Sensing of Environment 34, 75-91.

Kleyer M, Bekker RM, Knevel IC, Bakker JP, Thompson K, Sonnenschein M, Poschlod P, van Groenendael JM, Klimeš L, Klimešová J, Klotz S, Rusch GM, Hermy M, Adriaens D, Boedeltje G, Bossuyt B, Dannemann A, Endels P, Götzenberger L, Hodgson JG, Jackel AK, Kühn I, Kunzmann D, Ozinga WA, Römermann C, Stadler M, Schlegelmilch J, Steendam HJ, Tackenberg O, Wilmann B, Cornelissen JHC, Eriksson O, Garnier E, Peco B

(2008). The LEDA Traitbase: a database of life-history traits of the Northwest European flora. Journal of Ecology 96, 1266–1274.

Verhoef W (1984). Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. Remote Sensing of Environment 16, 125-141.