Example application of multi-biome PLSR model to estimate LMA for NEON domains in the eastern U.S.

Shawn Serbin

2019-09-10

Overview

This is an R Markdown Notebook to illustrate how to apply the multi-biome LMA PLSR model to leaf reflectance spectra.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

```
#--------------------------------------------------------------------------------------------------#
# get all required libraries
list.of.packages <- c("readr","scales","plotrix","httr","devtools")  # packages needed for script
# check for dependencies and install if needed
new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)
```

```
# load libraries needed for script
library(readr)    # readr - read_csv function to pull data from EcoSIS
library(plotrix)  # plotCI - to generate obsvered vs predicted plot with CIs
library(scales)   # alpha() - for applying a transparency to data points
```

```
##
## Attaching package: 'scales'
## The following object is masked from 'package:plotrix':
##
##     rescale
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
library(devtools)
```

```
## Loading required package: usethis
```

```
library(httr)
```

```
# define function to grab PLSR model from GitHub
#devtools::source_gist("gist.github.com/christophergandrud/4466237")
source_GitHubData <-function(url, sep = ",", header = TRUE) {
  require(httr)
  request <- GET(url)
  stop_for_status(request)
  handle <- textConnection(content(request, as = 'text'))
  on.exit(close(handle))
  read.table(handle, sep = sep, header = header)
}
#--------------------------------------------------------------------------------------------------#
```

```r
#--------------------------------------------------------------------------------------------------#
### PLSR Coefficients - Grab from GitHub
git_repo <- "https://raw.githubusercontent.com/serbinsh/SSerbin_etal_2019_NewPhytologist/master/"
print("**** Downloading PLSR coefficients ****")
```

```
## [1] "**** Downloading PLSR coefficients ****"
```

```r
githubURL <- paste0(git_repo,"SSerbin_multibiome_lma_plsr_model/sqrt_LMA_gDW_m2_PLSR_Coefficients_10comp
LeafLMA.plsr.coeffs <- source_GitHubData(githubURL)
rm(githubURL)
githubURL <- paste0(git_repo,"SSerbin_multibiome_lma_plsr_model/sqrt_LMA_gDW_m2_Jackkife_PLSR_Coefficien
LeafLMA.plsr.jk.coeffs <- source_GitHubData(githubURL)
#--------------------------------------------------------------------------------------------------#
```

NEON dataset

URL: https://ecosis.org/#result/5617da17-c925-49fb-b395-45a51291bd2d

---

```r
#--------------------------------------------------------------------------------------------------#
### Grab data
print("**** Downloading Ecosis data ****")
```

```
## [1] "**** Downloading Ecosis data ****"
```

```r
ecosis_id <- "5617da17-c925-49fb-b395-45a51291bd2d"
ecosis_file <- sprintf(
  "https://ecosis.org/package/export?package_id=%s&filters=&metadata=true",
  ecosis_id
)

message("Downloading data...")
```

```
## Downloading data...
```

```r
neon_data <- read_csv(ecosis_file)
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Affiliation = col_character(),
##   `Common Name` = col_character(),
##   Domain = col_character(),
##   Functional_type = col_character(),
##   `Latin Genus` = col_character(),
##   `Latin Species` = col_character(),
##   PI = col_character(),
##   Project = col_character(),
##   Sample_ID = col_character(),
```

```
##   `USDA Symbol` = col_character()
## )
## See spec(...) for full column specifications.
```

```
message("NEON data download complete!")
```

```
## NEON data download complete!
```

```
#--------------------------------------------------------------------------------------------------#
```

```
#--------------------------------------------------------------------------------------------------#
## Concatenate data
Start.wave <- 500
End.wave <- 2400
wv <- seq(Start.wave,End.wave,1)

neon_data_spec_subset <- neon_data[, which(names(neon_data) %in% wv)]
names(neon_data_spec_subset) <- c(paste0("Wave_",wv))
neon_lma_data <- data.frame(SampleID=neon_data$Sample_ID, NEON_Domain=neon_data$Domain,
                            PFT=neon_data$Functional_type, USDA_Species_Code=neon_data$`USDA Symbol`,
                            LMA_gDW_m2=neon_data$LMA)
neon_lma_spec_data <- data.frame(neon_lma_data,neon_data_spec_subset)
## cleanup
rm(neon_data_spec_subset,neon_lma_data)
#--------------------------------------------------------------------------------------------------#
```

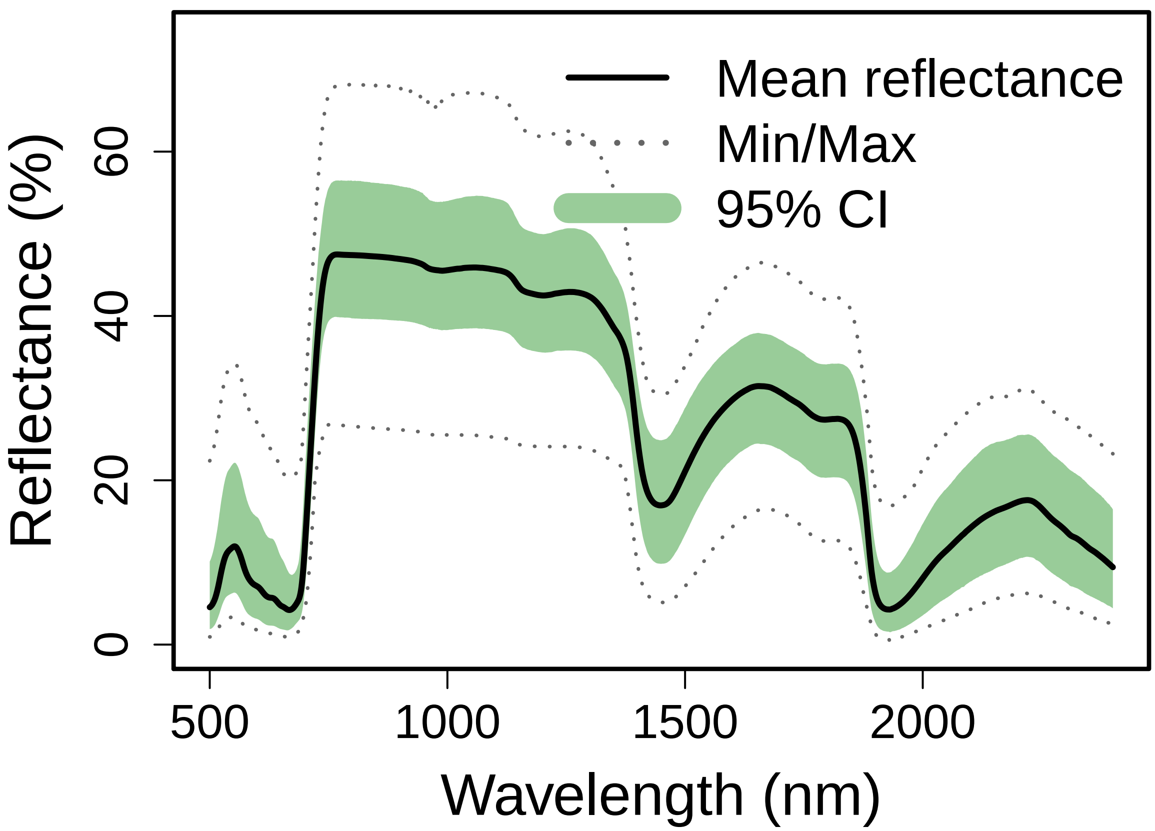
```
#--------------------------------------------------------------------------------------------------#
## Plot data
waves <- paste0("Wave_",wv)
cexaxis <- 1.5
cexlab <- 1.8
ylim <- 74
ylim2 <- 80

mean_spec <- colMeans(neon_lma_spec_data[,which(names(neon_lma_spec_data) %in% waves)])
spectra_quantiles <- apply(neon_lma_spec_data[,which(names(neon_lma_spec_data) %in% waves)],
                           2,quantile,na.rm=T,probs=c(0,0.025,0.05,0.5,0.95,0.975,1))

print("**** Plotting Ecosis data. Writing to scratch space ****")
```

```
## [1] "**** Plotting Ecosis data. Writing to scratch space ****"
```

```
png(file=file.path("~",wd,'NEON_leaf_spectra_summary_plot.png'),height=3000,
    width=3900, res=340)
par(mfrow=c(1,1), mar=c(4.5,5.7,0.3,0.4), oma=c(0.3,0.9,0.3,0.1)) # B, L, T, R
plot(wv,mean_spec*100,ylim=c(0,ylim),cex=0.00001, col="white",xlab="Wavelength (nm)",
     ylab="Reflectance (%)",cex.axis=cexaxis, cex.lab=cexlab)
polygon(c(wv ,rev(wv)),c(spectra_quantiles[6,]*100, rev(spectra_quantiles[2,]*100)),
        col="#99CC99",border=NA)
lines(wv,mean_spec*100,lwd=3, lty=1, col="black")
lines(wv,spectra_quantiles[1,]*100,lwd=1.85, lty=3, col="grey40")
lines(wv,spectra_quantiles[7,]*100,lwd=1.85, lty=3, col="grey40")
legend("topright",legend=c("Mean reflectance","Min/Max", "95% CI"),lty=c(1,3,1),
       lwd=c(3,3,15),col=c("black","grey40","#99CC99"),bty="n", cex=1.7)
box(lwd=2.2)
dev.off()
```

```
## pdf
##   2
```

```r
par(mfrow=c(1,1), mar=c(4.5,5.7,0.3,0.4), oma=c(0.3,0.9,0.3,0.1)) # B, L, T, R
plot(wv,mean_spec*100,ylim=c(0,ylim),cex=0.00001, col="white",xlab="Wavelength (nm)",
     ylab="Reflectance (%)",cex.axis=cexaxis, cex.lab=cexlab)
polygon(c(wv ,rev(wv)),c(spectra_quantiles[6,]*100, rev(spectra_quantiles[2,]*100)),
        col="#99CC99",border=NA)
lines(wv,mean_spec*100,lwd=3, lty=1, col="black")
lines(wv,spectra_quantiles[1,]*100,lwd=1.85, lty=3, col="grey40")
lines(wv,spectra_quantiles[7,]*100,lwd=1.85, lty=3, col="grey40")
legend("topright",legend=c("Mean reflectance","Min/Max", "95% CI"),lty=c(1,3,1),
       lwd=c(3,3,15),col=c("black","grey40","#99CC99"),bty="n", cex=1.7)
box(lwd=2.2)
```




```r
#--------------------------------------------------------------------------------------------------#
```

```r
#--------------------------------------------------------------------------------------------------#
print("**** Applying PLSR model to estimate LMA from spectral observations ****")
```

```
## [1] "**** Applying PLSR model to estimate LMA from spectral observations ****"
```

```r
# setup model
dims <- dim(LeafLMA.plsr.coeffs)
LeafLMA.plsr.intercept <- LeafLMA.plsr.coeffs[1,]
LeafLMA.plsr.coeffs <- data.frame(LeafLMA.plsr.coeffs[2:dims[1],])
names(LeafLMA.plsr.coeffs) <- c("wavelength","coefs")
LeafLMA.plsr.coeffs.vec <- as.vector(LeafLMA.plsr.coeffs[,2])

# estimate LMA
```

```r
sub_spec <- as.matrix(droplevels(neon_lma_spec_data[,which(names(neon_lma_spec_data) %in% waves)]))
temp <- as.matrix(sub_spec) %*% LeafLMA.plsr.coeffs.vec  # Updated: Using matrix mult.
leafLMA <- data.frame(rowSums(temp))+LeafLMA.plsr.intercept[,2]
leafLMA <- leafLMA[,1]^2  # convert to standard LMA units from sqrt(LMA)
names(leafLMA) <- "FS_PLSR_LMA_gDW_m2"

# organize output
'%ni%' <- Negate('%in%')
LeafLMA.PLSR.dataset <- data.frame(neon_lma_spec_data[,which(names(neon_lma_spec_data) %ni% waves)],
                                   FS_PLSR_LMA_gDW_m2=leafLMA)

# Derive LMA estimate uncertainties
print("**** Deriving uncertainty estimates ****")
```

```
## [1] "**** Deriving uncertainty estimates ****"
```

```r
dims <- dim(LeafLMA.plsr.jk.coeffs)
intercepts <- LeafLMA.plsr.jk.coeffs[,2]
jk.leaf.lma.est <- array(data=NA,dim=c(dim(sub_spec)[1],dims[1]))
for (i in 1:length(intercepts)){
  coefs <- unlist(as.vector(LeafLMA.plsr.jk.coeffs[i,3:dims[2]]))
  temp <- sub_spec %*% coefs
  values <- data.frame(rowSums(temp))+intercepts[i]
  jk.leaf.lma.est[,i] <- values[,1]^2
  rm(temp)
}

jk.leaf.lma.est.quant <- apply(jk.leaf.lma.est,1,quantile,probs=c(0.025,0.975))
jk.leaf.lma.est.quant2 <- data.frame(t(jk.leaf.lma.est.quant))
names(jk.leaf.lma.est.quant2) <- c("FS_PLSR_Leaf_LMA_L5","FS_PLSR_Leaf_LMA_U95")
jk.leaf.lma.est.sd <- apply(jk.leaf.lma.est,1,sd)
names(jk.leaf.lma.est.sd) <- "FS_PLSR_Leaf_LMA_Sdev"

## Combine into final dataset
stats <- data.frame(jk.leaf.lma.est.sd,jk.leaf.lma.est.quant2)
names(stats) <- c("FS_PLSR_Leaf_LMA_Sdev","FS_PLSR_Leaf_LMA_L5","FS_PLSR_Leaf_LMA_U95")
LeafLMA.PLSR.dataset.out <- data.frame(LeafLMA.PLSR.dataset,stats,
                                       residual=(LeafLMA.PLSR.dataset$FS_PLSR_LMA_gDW_m2-
                                                   LeafLMA.PLSR.dataset$LMA_gDW_m2))

# output results
write.csv(x = LeafLMA.PLSR.dataset.out,
          file = file.path("~",wd,
                           "Angers_Lopex_PLSR_estimated_LMA_data.csv"),
          row.names = F)

# calculate error stats
rmse <- sqrt(mean(LeafLMA.PLSR.dataset.out$residual^2))
# calculate fit stats
reg <- lm(LeafLMA.PLSR.dataset.out$FS_PLSR_LMA_gDW_m2~
            LeafLMA.PLSR.dataset.out$LMA_gDW_m2)
summary(reg)
```

```
##
```

```
## Call:
## lm(formula = LeafLMA.PLSR.dataset.out$FS_PLSR_LMA_gDW_m2 ~ LeafLMA.PLSR.dataset.out$LMA_gDW_m2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -110.443   -6.950   -0.378    6.599   79.587
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         17.000708   0.400041    42.5   <2e-16
## LeafLMA.PLSR.dataset.out$LMA_gDW_m2  0.851236   0.005042   168.8   <2e-16
##
## (Intercept)                         ***
## LeafLMA.PLSR.dataset.out$LMA_gDW_m2 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.88 on 6310 degrees of freedom
## Multiple R-squared:  0.8188, Adjusted R-squared:  0.8187
## F-statistic: 2.851e+04 on 1 and 6310 DF,  p-value: < 2.2e-16
```

```
#--------------------------------------------------------------------------------------------------#
```

```
#--------------------------------------------------------------------------------------------------#
## Plot up results
ptcex <- 1.8
cexaxis <- 1.3
cexlab <- 1.8

print("**** Plotting Lopex/Angers LMA validation plot. Writing to scratch space ****")
```

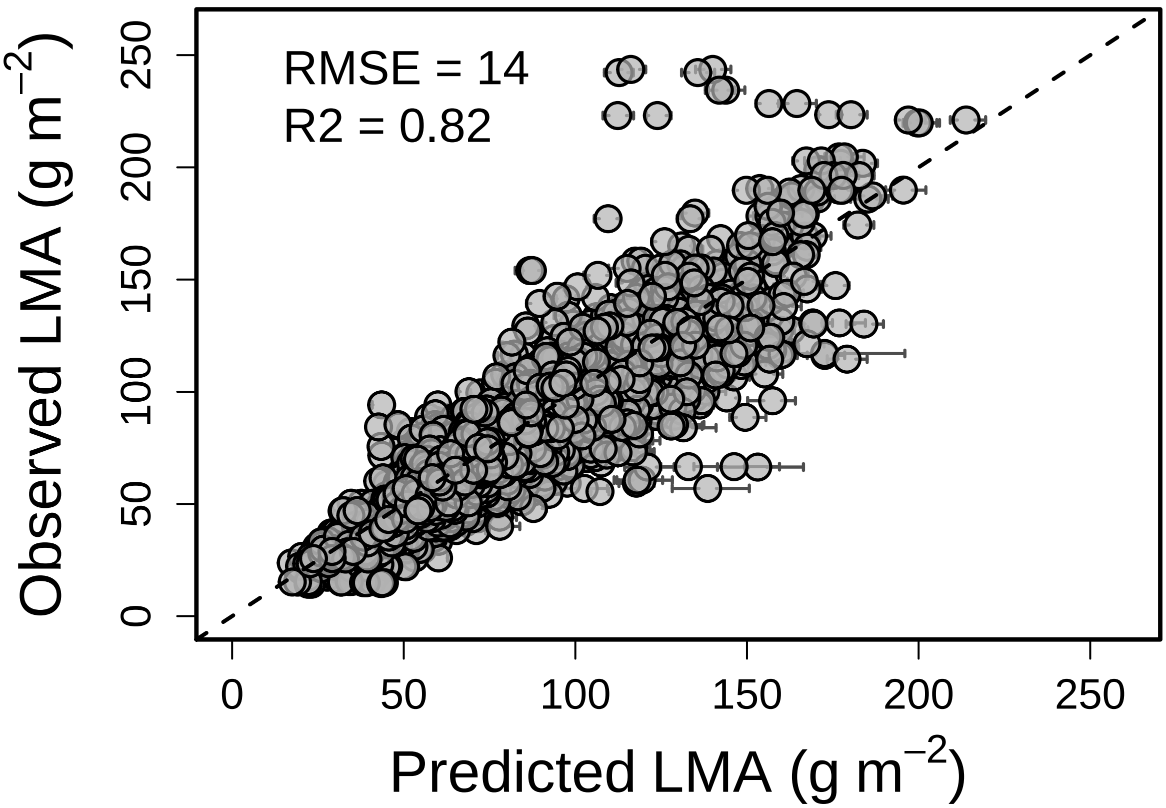
```
## [1] "**** Plotting Lopex/Angers LMA validation plot. Writing to scratch space ****"
```

```
png(file=file.path("~",wd,'NEON_LMA_validation_plot.png'),height=3000,
    width=3900, res=340)
par(mfrow=c(1,1), mar=c(4.5,5.4,1,1), oma=c(0.3,0.9,0.3,0.1)) # B, L, T, R
plotCI(LeafLMA.PLSR.dataset.out$FS_PLSR_LMA_gDW_m2,LeafLMA.PLSR.dataset.out$LMA_gDW_m2,
       li=LeafLMA.PLSR.dataset.out$FS_PLSR_Leaf_LMA_L5,gap=0.009,sfrac=0.004,lwd=1.6,
       ui=LeafLMA.PLSR.dataset.out$FS_PLSR_Leaf_LMA_U95,err="x",pch=21,col="black",
       pt.bg=alpha("grey70",0.7),scol="grey30",xlim=c(0,260),cex=ptcex,
       ylim=c(0,260),xlab="",
       ylab=expression(paste("Observed LMA (",g~m^{-2},")")),main="",
       cex.axis=cexaxis,cex.lab=cexlab)
mtext(side = 1, text = expression(paste(Predicted~LMA," (",g~m^{-2},")")), line = 3.5,
      cex=cexlab)
abline(0,1,lty=2,lw=2)
legend("topleft",legend = c(paste0("RMSE = ",round(rmse)),
                            paste0("R2 = ",round(summary(reg)$r.squared,2))), bty="n", cex=1.5)
box(lwd=2.2)
dev.off()
```

```
## pdf
##   2
```

```
par(mfrow=c(1,1), mar=c(4.5,5.4,1,1), oma=c(0.3,0.9,0.3,0.1)) # B, L, T, R
plotCI(LeafLMA.PLSR.dataset.out$FS_PLSR_LMA_gDW_m2,LeafLMA.PLSR.dataset.out$LMA_gDW_m2,
```

```r
        li=LeafLMA.PLSR.dataset.out$FS_PLSR_Leaf_LMA_L5,gap=0.009,sfrac=0.004,lwd=1.6,
        ui=LeafLMA.PLSR.dataset.out$FS_PLSR_Leaf_LMA_U95,err="x",pch=21,col="black",
        pt.bg=alpha("grey70",0.7),scol="grey30",xlim=c(0,260),cex=ptcex,
        ylim=c(0,260),xlab="",
        ylab=expression(paste("Observed LMA (",g~m^{-2},")")),main="",
        cex.axis=cexaxis,cex.lab=cexlab)
mtext(side = 1, text = expression(paste(Predicted~LMA," (",g~m^{-2},")")), line = 3.5,
       cex=cexlab)
abline(0,1,lty=2,lw=2)
legend("topleft",legend = c(paste0("RMSE = ",round(rmse)),
                            paste0("R2 = ",round(summary(reg)$r.squared,2))), bty="n", cex=1.5)
box(lwd=2.2)
```




```r
#----------------------------------------------------------------------------------------------------#
```

```r
#----------------------------------------------------------------------------------------------------#
rm(list=ls(all=TRUE))    # clear workspace
### EOF
```