Example application of Serbin et al 2014 foliar N PLSR model to existing data in EcoSIS

Shawn Serbin

2019-11-12

Overview

This is an R Markdown Notebook to illustrate how to apply the foliar N 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)
library(knitr)
```

```
# 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)
}
```

```
# not in
`%notin%` <- Negate(`%in%`)
#-----------------------------------------------------------------------------------------------------#
```

```
#-----------------------------------------------------------------------------------------------------#
### Set working directory (scratch space)
output_dir <- file.path("~",'scratch/')
if (! file.exists(output_dir)) dir.create(file.path(output_dir),recursive=TRUE, showWarnings = FALSE)
#setwd(file.path("~",wd)) # set working directory
opts_knit$set(root.dir = file.path(output_dir))
getwd()  # check wd
```

```
## [1] "/Users/shawnserbin/Data/GitHub/SSerbin_etal_2014_EcoApps/vignettes"
```

```
print(output_dir)
```

```
## [1] "~/scratch/"
```

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

```
#-----------------------------------------------------------------------------------------------------#
### PLSR Coefficients - Grab from GitHub
print("**** Downloading PLSR coefficients ****")
```

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

```
git_repo <- "https://raw.githubusercontent.com/serbinsh/SSerbin_etal_2014_EcoApps/master/"
githubURL <- paste0(git_repo,"PLSR_model_coefficients/leaf_Nitrogen/FFT_Leaf_Nitrogen_PLSR_Coefficients_
LeafN.plsr.coeffs <- source_GitHubData(githubURL)
rm(githubURL)
githubURL <- paste0(git_repo,"PLSR_model_coefficients/leaf_Nitrogen/FFT_Leaf_Nitrogen_Jackkife_PLSR_Coe
LeafN.plsr.jk.coeffs <- source_GitHubData(githubURL)
rm(githubURL)
#-----------------------------------------------------------------------------------------------------#
```

```
#-----------------------------------------------------------------------------------------------------#
### Example datasets
#
# URL:  https://ecosis.org/package/dried-leaf-spectra-to-estimate-foliar-functional-traits-over-neon-do
#
#-----------------------------------------------------------------------------------------------------#
```

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

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

```
ecosis_id <- "87fbbced-0ccb-4b4f-99d7-b3b4c81bc151"  # NEON dried and ground data
ecosis_file <- sprintf(
  "https://ecosis.org/api/package/%s/export?metadata=true",
  ecosis_id
)
message("Downloading data...")
```

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

```
dat_raw <- read_csv(ecosis_file)
```

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

```
## See spec(...) for full column specifications.
```

```
message("Download complete!")
```

```
## Download complete!
```

```
head(dat_raw)
```

```
## # A tibble: 6 x 2,185
##   Affiliation Aluminum  Boron Calcium Carbon Cellulose `Common Name`
##   <chr>          <dbl>  <dbl>   <dbl>  <dbl>     <dbl> <chr>
## 1 University~   0.0207 0.0152     2.3   533.      108. <NA>
## 2 University~   0.0271 0.0142    11.3   550.      128. woodland mad~
## 3 University~   0.0061 0.0108     3.5   468.      121. black willow
## 4 University~   0.0176 0.0131     8.5   387       176. giant-trumpe~
## 5 University~   0.0078 0.0143     9.1   880.      234. giant-trumpe~
## 6 University~   0.108  0.0148     5.6   437.      194. sourwood
## # ... with 2,178 more variables: Copper <dbl>, Fiber <dbl>,
## #   Flavonoids <dbl>, Iron <dbl>, `Latin Genus` <chr>, `Latin
## #   Species` <chr>, Lignin <dbl>, Magnesium <dbl>, Manganese <dbl>,
## #   NSC <dbl>, Nitrogen <dbl>, PI <chr>, Phenolics <dbl>,
## #   Phosphorus <dbl>, Potassium <dbl>, Project <chr>, Sample_Date <dbl>,
## #   Sample_ID <chr>, Site <chr>, Site_Name <chr>, Starch <dbl>,
## #   Sugar <dbl>, Sulfur <dbl>, `USDA Symbol` <chr>, Zinc <dbl>,
## #   d13C <dbl>, d15N <dbl>, `350` <dbl>, `351` <dbl>, `352` <dbl>,
## #   `353` <dbl>, `354` <dbl>, `355` <dbl>, `356` <dbl>, `357` <dbl>,
## #   `358` <dbl>, `359` <dbl>, `360` <dbl>, `361` <dbl>, `362` <dbl>,
## #   `363` <dbl>, `364` <dbl>, `365` <dbl>, `366` <dbl>, `367` <dbl>,
## #   `368` <dbl>, `369` <dbl>, `370` <dbl>, `371` <dbl>, `372` <dbl>,
## #   `373` <dbl>, `374` <dbl>, `375` <dbl>, `376` <dbl>, `377` <dbl>,
## #   `378` <dbl>, `379` <dbl>, `380` <dbl>, `381` <dbl>, `382` <dbl>,
## #   `383` <dbl>, `384` <dbl>, `385` <dbl>, `386` <dbl>, `387` <dbl>,
## #   `388` <dbl>, `389` <dbl>, `390` <dbl>, `391` <dbl>, `392` <dbl>,
## #   `393` <dbl>, `394` <dbl>, `395` <dbl>, `396` <dbl>, `397` <dbl>,
## #   `398` <dbl>, `399` <dbl>, `400` <dbl>, `401` <dbl>, `402` <dbl>,
## #   `403` <dbl>, `404` <dbl>, `405` <dbl>, `406` <dbl>, `407` <dbl>,
## #   `408` <dbl>, `409` <dbl>, `410` <dbl>, `411` <dbl>, `412` <dbl>,
## #   `413` <dbl>, `414` <dbl>, `415` <dbl>, `416` <dbl>, `417` <dbl>,
## #   `418` <dbl>, `419` <dbl>, `420` <dbl>, `421` <dbl>, `422` <dbl>, ...
```

```
names(dat_raw)[1:40]
```

```
##  [1] "Affiliation"   "Aluminum"      "Boron"         "Calcium"
##  [5] "Carbon"        "Cellulose"     "Common Name"   "Copper"
##  [9] "Fiber"         "Flavonoids"    "Iron"          "Latin Genus"
## [13] "Latin Species" "Lignin"        "Magnesium"     "Manganese"
## [17] "NSC"           "Nitrogen"      "PI"            "Phenolics"
## [21] "Phosphorus"    "Potassium"     "Project"       "Sample_Date"
## [25] "Sample_ID"     "Site"          "Site_Name"     "Starch"
## [29] "Sugar"         "Sulfur"        "USDA Symbol"   "Zinc"
## [33] "d13C"          "d15N"          "350"           "351"
## [37] "352"           "353"           "354"           "355"
```

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

```
#-------------------------------------------------------------------------------------------------------#
## Create validation dataset
Start.wave <- 500
End.wave <- 2400
wv <- seq(Start.wave,End.wave,1)

spectra <- dat_raw[,names(dat_raw)[match(seq(Start.wave,End.wave,1),names(dat_raw))]]
sample_info <- data.frame(Sample_ID=dat_raw$`Sample_ID`, Sample_Date=dat_raw$`Sample_Date`,
                          USDA_Species_Code=dat_raw$`USDA Symbol`,
                          Common_Species_Name=dat_raw$`Common Name`,
                          Nitrogen=(dat_raw$`Nitrogen`)*0.1) # convert N from mg/g to g/g
head(sample_info)
```

```
##   Sample_ID Sample_Date USDA_Species_Code Common_Species_Name Nitrogen
## 1      L001    20170525              QUCO                <NA>     2.09
## 2      L002    20170526              MAMA      woodland madia     2.68
## 3      L003    20170527              SANI        black willow     2.49
## 4      L004    20170527              MAVI      giant-trumpets     1.72
## 5      L005    20170527              MAVI      giant-trumpets     2.26
## 6      L006    20170528              OXAR            sourwood     1.58
```

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

```
#-------------------------------------------------------------------------------------------------------#
## Plot data
waves <- seq(500,2400,1)
cexaxis <- 1.5
cexlab <- 1.8
ylim <- 100
ylim2 <- 100

mean_spec <- colMeans(spectra[,which(names(spectra) %in% seq(Start.wave,End.wave,1))])
spectra_quantiles <- apply(spectra[,which(names(spectra) %in% seq(Start.wave,End.wave,1))],
                           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 ****")
```

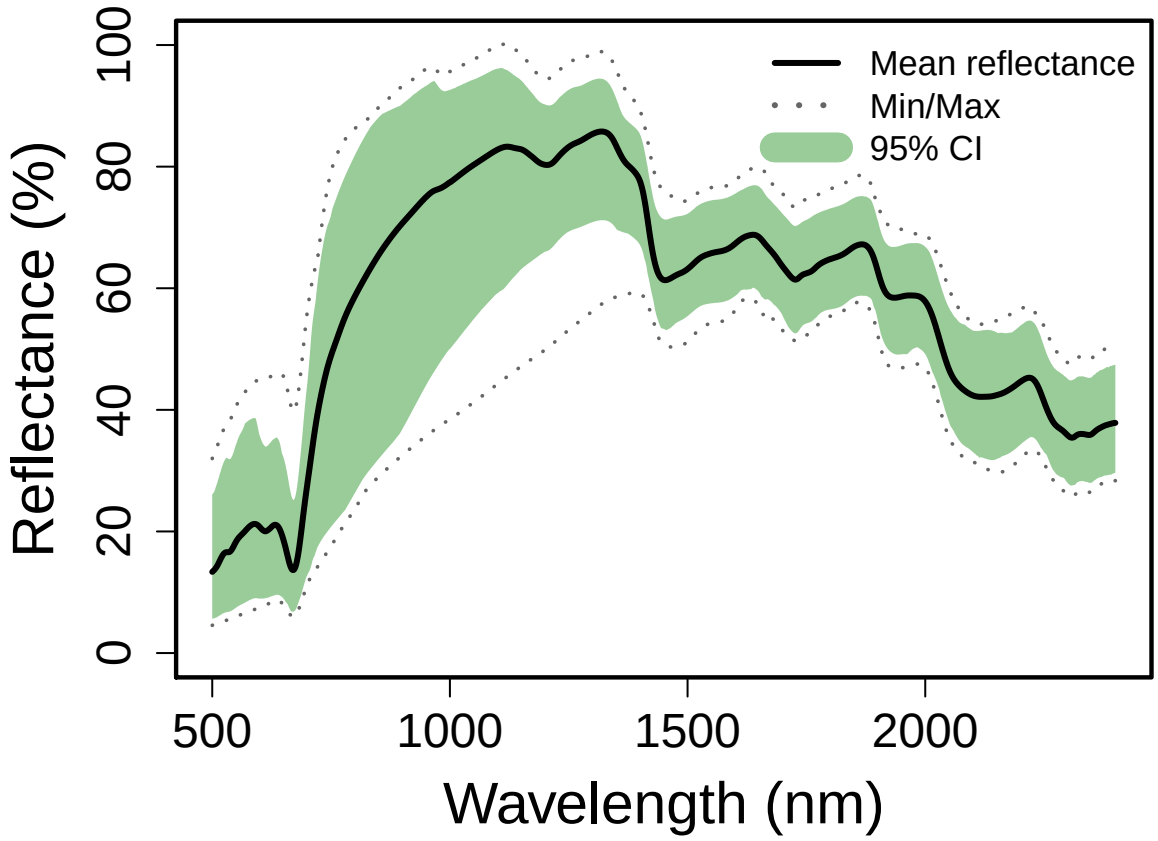
```
## [1] "**** Plotting Ecosis data. Writing to scratch space ****"
```

```
png(file=file.path(output_dir,'NEON_dried_and_ground_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(waves,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(waves ,rev(waves)),c(spectra_quantiles[6,]*100, rev(spectra_quantiles[2,]*100)),
        col="#99CC99",border=NA)
lines(waves,mean_spec*100,lwd=3, lty=1, col="black")
lines(waves,spectra_quantiles[1,]*100,lwd=1.85, lty=3, col="grey40")
lines(waves,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
```

```
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(waves,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(waves ,rev(waves)),c(spectra_quantiles[6,]*100, rev(spectra_quantiles[2,]*100)),
        col="#99CC99",border=NA)
lines(waves,mean_spec*100,lwd=3, lty=1, col="black")
lines(waves,spectra_quantiles[1,]*100,lwd=1.85, lty=3, col="grey40")
lines(waves,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=1.1)
box(lwd=2.2)
```



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

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

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

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

# estimate foliar N
Start.wave <- 1500
End.wave <- 2400
sub_spec <- as.matrix(droplevels(spectra[,which(names(spectra) %in% seq(Start.wave,End.wave,1))]))
temp <- as.matrix(sub_spec) %*% LeafN.plsr.coeffs.vec
leafN <- data.frame(rowSums(temp))+LeafN.plsr.intercept[,2]
leafN <- leafN[,1]  # convert to standard LMA units from sqrt(LMA)
names(leafN) <- "FS_PLSR_N_Perc"

# organize output
LeafN.PLSR.dataset <- data.frame(sample_info, FS_PLSR_N_Perc=leafN)
head(LeafN.PLSR.dataset)
```

```
##   Sample_ID Sample_Date USDA_Species_Code Common_Species_Name Nitrogen
## 1      L001    20170525              QUCO                <NA>     2.09
## 2      L002    20170526              MAMA      woodland madia     2.68
## 3      L003    20170527              SANI        black willow     2.49
## 4      L004    20170527              MAVI      giant-trumpets     1.72
## 5      L005    20170527              MAVI      giant-trumpets     2.26
## 6      L006    20170528              OXAR            sourwood     1.58
##   FS_PLSR_N_Perc
## 1       1.928792
## 2       2.448645
## 3       1.825797
## 4       1.516547
## 5       1.889162
## 6       2.308531
```

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

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

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

```
}

jk.leaf.n.est.quant <- apply(jk.leaf.n.est,1,quantile,probs=c(0.025,0.975))
jk.leaf.n.est.quant2 <- data.frame(t(jk.leaf.n.est.quant))
names(jk.leaf.n.est.quant2) <- c("FS_PLSR_Leaf_N_L5","FS_PLSR_Leaf_N_U95")
jk.leaf.n.est.sd <- apply(jk.leaf.n.est,1,sd)
names(jk.leaf.n.est.sd) <- "FS_PLSR_Leaf_N_Perc_Sdev"

## Combine into final dataset
stats <- data.frame(jk.leaf.n.est.sd,jk.leaf.n.est.quant2)
names(stats) <- c("FS_PLSR_Leaf_N_Perc_Sdev","FS_PLSR_Leaf_N_L5","FS_PLSR_Leaf_N_U95")
LeafN.PLSR.dataset.out <- data.frame(LeafN.PLSR.dataset,stats,
                                     residual=(LeafN.PLSR.dataset$FS_PLSR_N_Perc-LeafN.PLSR.dataset$N
head(LeafN.PLSR.dataset.out)
```

```
##   Sample_ID Sample_Date USDA_Species_Code Common_Species_Name Nitrogen
## 1      L001    20170525              QUCO                <NA>     2.09
## 2      L002    20170526              MAMA      woodland madia     2.68
## 3      L003    20170527              SANI        black willow     2.49
## 4      L004    20170527              MAVI      giant-trumpets     1.72
## 5      L005    20170527              MAVI      giant-trumpets     2.26
## 6      L006    20170528              OXAR            sourwood     1.58
##   FS_PLSR_N_Perc FS_PLSR_Leaf_N_Perc_Sdev FS_PLSR_Leaf_N_L5
## 1       1.928792               0.11501658          1.797976
## 2       2.448645               0.09137626          2.346434
## 3       1.825797               0.13983181          1.661016
## 4       1.516547               0.10016391          1.410133
## 5       1.889162               0.10957715          1.785171
## 6       2.308531               0.11613262          2.187824
##   FS_PLSR_Leaf_N_U95   residual
## 1           2.250608 -0.1612083
## 2           2.706357 -0.2313555
## 3           2.213883 -0.6642027
## 4           1.803098 -0.2034534
## 5           2.220676 -0.3708377
## 6           2.650508  0.7285307
```

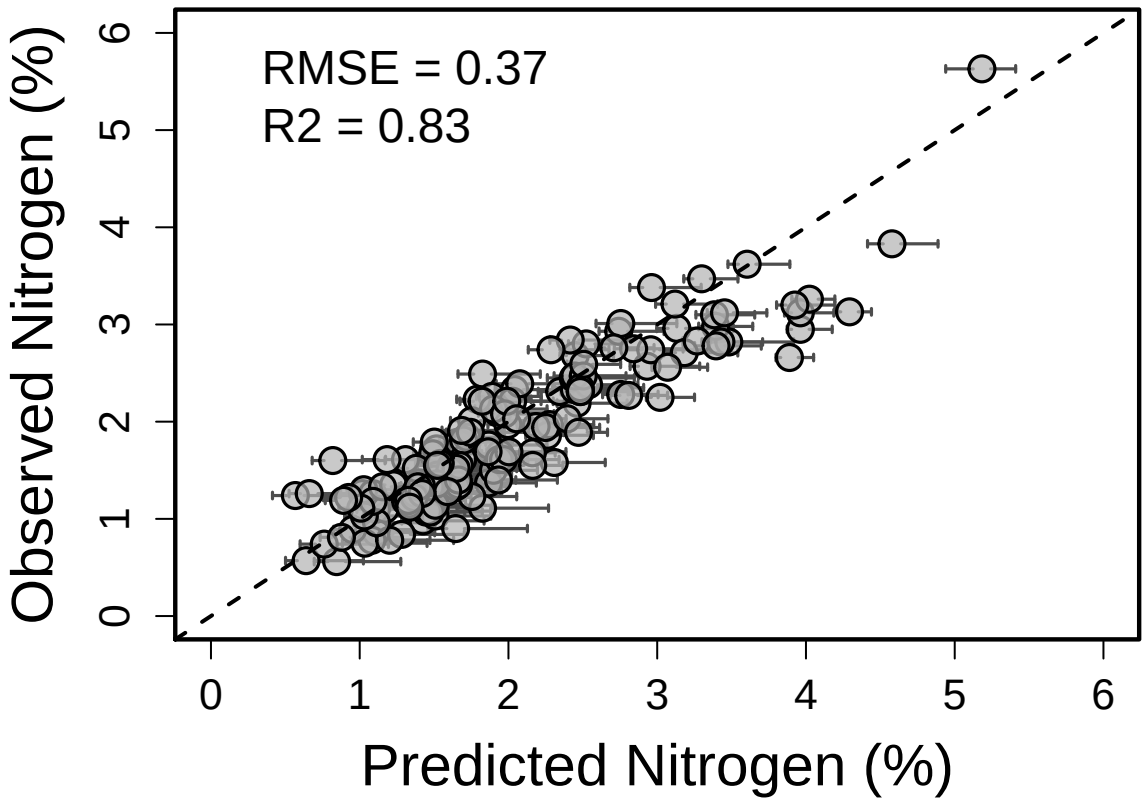
```
# output results
write.csv(x = LeafN.PLSR.dataset.out, file = file.path(output_dir,"NEON_PLSR_estimated_foliar_nitrogen_
          row.names = F)
# calculate error stats
rmse <- sqrt(mean(LeafN.PLSR.dataset.out$residual^2, na.rm=T))
# calculate fit stats
reg <- lm(LeafN.PLSR.dataset.out$FS_PLSR_N_Perc~LeafN.PLSR.dataset.out$Nitrogen)
#-------------------------------------------------------------------------------------------------#

#-------------------------------------------------------------------------------------------------#
## Plot up results
ptcex <- 1.8
cexaxis <- 1.3
cexlab <- 1.8
print("**** Plotting NEON PLSR estimated foliar nitrogen validation plot. Writing to scratch space ****
```

```
## [1] "**** Plotting NEON PLSR estimated foliar nitrogen validation plot. Writing to scratch space ***
```

```
png(file=file.path(output_dir,'NEON_PLSR_estimated_foliar_nitrogen_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(LeafN.PLSR.dataset.out$FS_PLSR_N_Perc,LeafN.PLSR.dataset.out$Nitrogen,
       li=LeafN.PLSR.dataset.out$FS_PLSR_Leaf_N_L5,gap=0.009,sfrac=0.004,lwd=1.6,
       ui=LeafN.PLSR.dataset.out$FS_PLSR_Leaf_N_U95,err="x",pch=21,col="black",
       pt.bg=alpha("grey70",0.7),scol="grey30",xlim=c(0,6),cex=ptcex,
       ylim=c(0,6),xlab="Predicted Nitrogen (%)",
       ylab="Observed Nitrogen (%)",main="",
       cex.axis=cexaxis,cex.lab=cexlab)
abline(0,1,lty=2,lw=2)
legend("topleft",legend = c(paste0("RMSE = ",round(rmse,2)),
                            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(LeafN.PLSR.dataset.out$FS_PLSR_N_Perc,LeafN.PLSR.dataset.out$Nitrogen,
       li=LeafN.PLSR.dataset.out$FS_PLSR_Leaf_N_L5,gap=0.009,sfrac=0.004,lwd=1.6,
       ui=LeafN.PLSR.dataset.out$FS_PLSR_Leaf_N_U95,err="x",pch=21,col="black",
       pt.bg=alpha("grey70",0.7),scol="grey30",xlim=c(0,6),cex=ptcex,
       ylim=c(0,6),xlab="Predicted Nitrogen (%)",
       ylab="Observed Nitrogen (%)",main="",
       cex.axis=cexaxis,cex.lab=cexlab)
abline(0,1,lty=2,lw=2)
legend("topleft",legend = c(paste0("RMSE = ",round(rmse,2)),
                            paste0("R2 = ",round(summary(reg)$r.squared,2))), bty="n", cex=1.5)
box(lwd=2.2)
```




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

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