An expanded PLSR example using leaf-level specrta and leaf mass per area (LMA) data from several CONUS NEON sites

Shawn P. Serbin

Overview

his is an R Markdown Notebook to illustrate how to conduct a basic model fit. This example shows you how to retrieve a dataset from the EcoSIS spectral database, choose the "optimal" number of plsr components, and fit a plsr model for leaf-mass area

When you click the **Knit** button in Rstudio a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

Getting Started

Installation

```
list.of.packages <- c("devtools","readr","RCurl","httr","pls","dplyr","reshape2",
                      "ggplot2","gridExtra")  # 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

```
#library(httr) #!! may not actually need this package
library(pls)
library(readr)
library(dplyr)
library(reshape2)
library(ggplot2)
library(gridExtra)
```

Setup other functions and options

```
# Source helper functions from GitHub
devtools::source_url("https://raw.githubusercontent.com/TESTgroup-BNL/How_to_PLSR/master/R_Scripts/funct

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

# Script options
```

```
pls.options(plsralg = "oscorespls")
pls.options("plsralg")
```

$plsralg [1] “oscorespls”

```
pls.options()$parallel
```

NULL

```
# NULL

# What is the target variable?
inVar <- "LMA_gDW_m2"
```

Set working directory (scratch space)

```
outdir <- tempdir()
setwd(outdir) # set working directory
print(paste0("Output directory: ",getwd()))  # check wd
```

[1] “Output directory: /private/var/folders/xp/h3k9vf3n2jx181ts786_yjrn9c2gjq/T/RtmpZRrvTT”

Grab data from EcoSIS

```
print("**** Downloading Ecosis data ****")
```

URL: https://ecosis.org/package/fresh-leaf-spectra-to-estimate-lma-over-neon-domains-in-eastern-united-states [1] “**** Downloading Ecosis data ****”

```
ecosis_id <- "5617da17-c925-49fb-b395-45a51291bd2d"  # NEON dataset
ecosis_file <- sprintf(
  "https://ecosis.org/api/package/%s/export?metadata=true",
  ecosis_id
)
message("Downloading data...")
dat_raw <- read_csv(ecosis_file)
message("Download complete!")
head(dat_raw)
```

A tibble: 6 x 2,162

Affiliation `Common Name` Domain Functional_type LMA `Latin Genus`
1 University~ black walnut D02 broadleaf 72.9 Juglans
2 University~ black walnut D02 broadleaf 72.9 Juglans
3 University~ black walnut D02 broadleaf 60.8 Juglans
4 University~ black walnut D02 broadleaf 60.8 Juglans
5 University~ black walnut D02 broadleaf 85.9 Juglans
6 University~ black walnut D02 broadleaf 85.9 Juglans
... with 2,156 more variables: `Latin Species` , PI , # Project , Sample_ID , `USDA Symbol` , 350 , # 351 , 352 , 353 , 354 , 355 , # 356 , 357 , 358 , 359 , 360 , # 361 , 362 , 363 , 364 , 365 , # 366 , 367 , 368 , 369 , 370 , # 371 , 372 , 373 , 374 , 375 , # 376 , 377 , 378 , 379 , 380 , # 381 , 382 , 383 , 384 ,

385 , # 386 , 387 , 388 , 389 , 390 , # 391 , 392 , 393 , 394 , 395 , # 396 , 397 , 398 , 399 , 400 , # 401 , 402 , 403 , 404 , 405 , # 406 , 407 , 408 , 409 , 410 , # 411 , 412 , 413 , 414 , 415 , # 416 , 417 , 418 , 419 , 420 , # 421 , 422 , 423 , 424 , 425 , # 426 , 427 , 428 , 429 , 430 , # 431 , 432 , 433 , 434 , 435 , # 436 , 437 , 438 , 439 , 440 , # 441 , 442 , 443 , 444 , ...

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

[1] “Affiliation” “Common Name” “Domain” “Functional_type” [5] “LMA” “Latin Genus” “Latin Species” “PI”
[9] “Project” “Sample_ID” “USDA Symbol” “350”
[13] “351” “352” “353” “354”
[17] “355” “356” “357” “358”
[21] “359” “360” “361” “362”
[25] “363” “364” “365” “366”
[29] “367” “368” “369” “370”
[33] “371” “372” “373” “374”
[37] “375” “376” “377” “378”