Soi mineralogy estimates from lab spectra

Nélida E. Q. Silvero

26/01/2021

This tutorial is a step by step on how to estimate the relative amounts of iron oxides - hematite and goethite - from laboratory spectra and various methods. For this tutorial, we have 79 spectra collected in the Piracicaba region.

Install and load packages

```
if(!require(prospectr)) install.packages(prospectr)
if(!require(corrplot)) install.packages(corrplot)
if(!require(caret)) install.packages(caret)
require(prospectr)
require(raster)
require(corrplot)
```

Set working directory

```
setwd("C:/Users/neliq/Google Drive/dados")
list.files()
```

Load the data and the satellite image (SYSI)

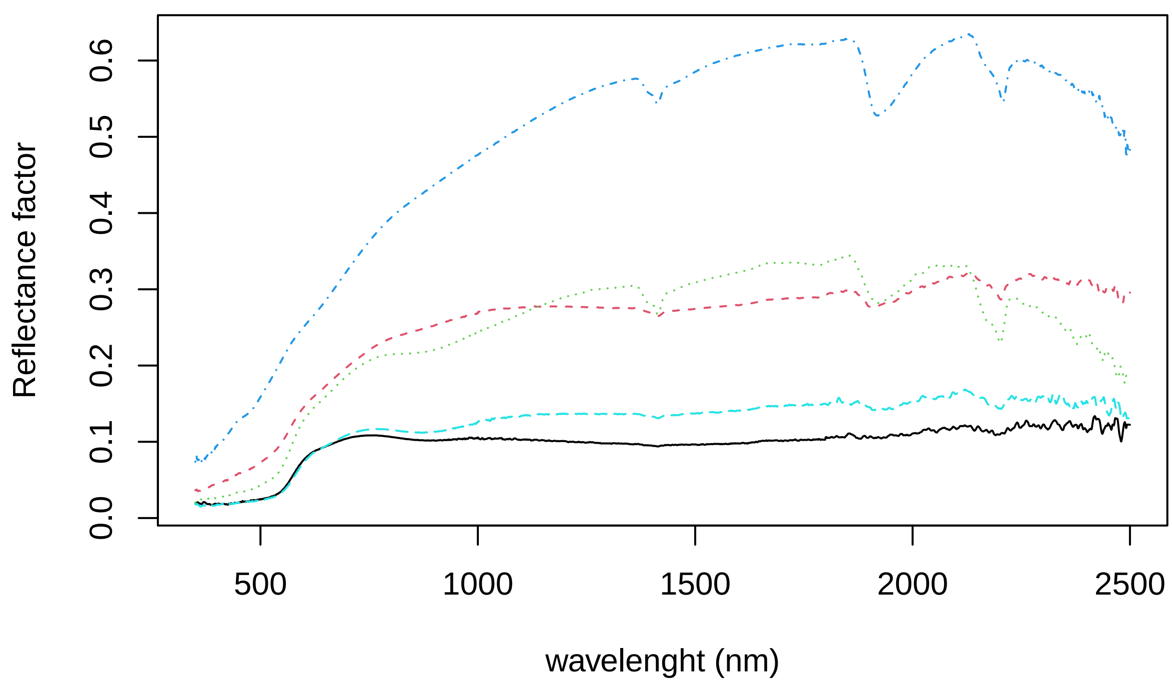
```
data <- read.csv("dados_spectra.csv", h=TRUE, sep=";")

rownames(data) <- data[,1]

spectra <- data[,-1:-3]
colnames(spectra) <- seq(from = 350, to = 2500, by=1)

SYSI <- stack("SYSI.tif")
SYSI[SYSI == 0] <- NA
```

Pre-processing by Savitzky and Golay (1968)

The method introduced by Savitzky and Golay (1968) is used both for smoothing and differentiating spectral curves. By utilizing higher order derivative spectra, subtle spectral features are detected and overlapping absorption characteristics resolved, which might not be possible from analysis of the original spectrum. Two reference are recommended for understanding of this tutorial **here** and **here**

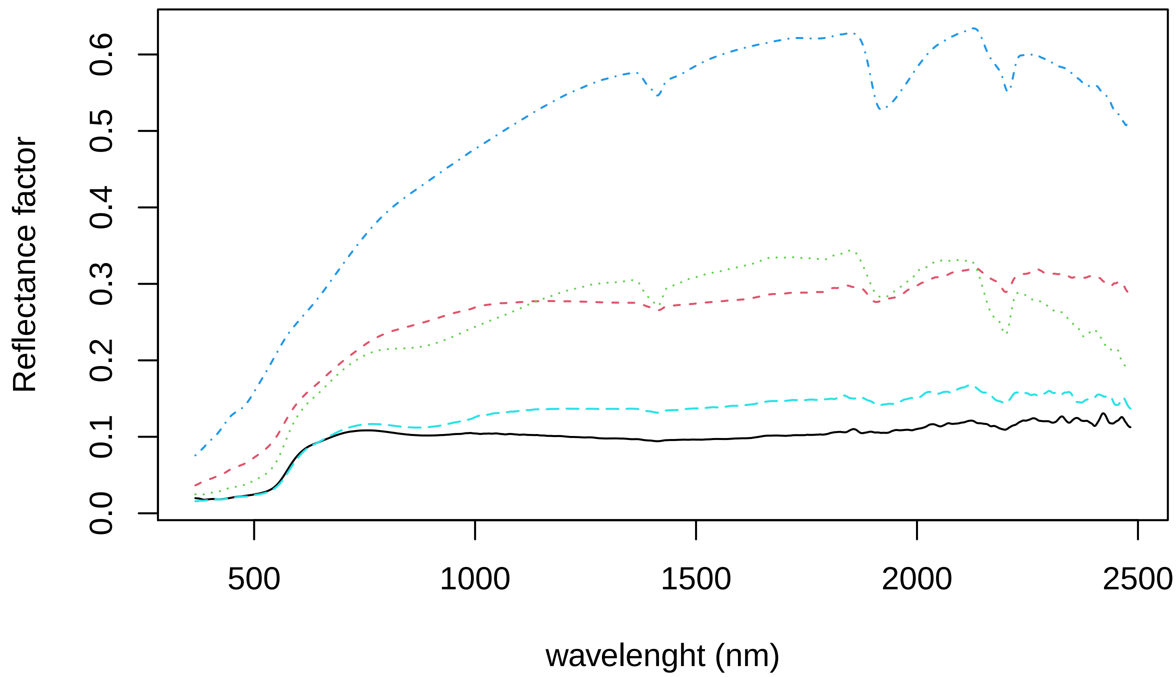
Plot raw spectra

```
matplot(as.numeric(colnames(spectra)), t(spectra[1:5,]), type = "l",
        xlab = "wavelenght (nm)",
        ylab = "Reflectance factor")
```



Smooth spectrum without deriving

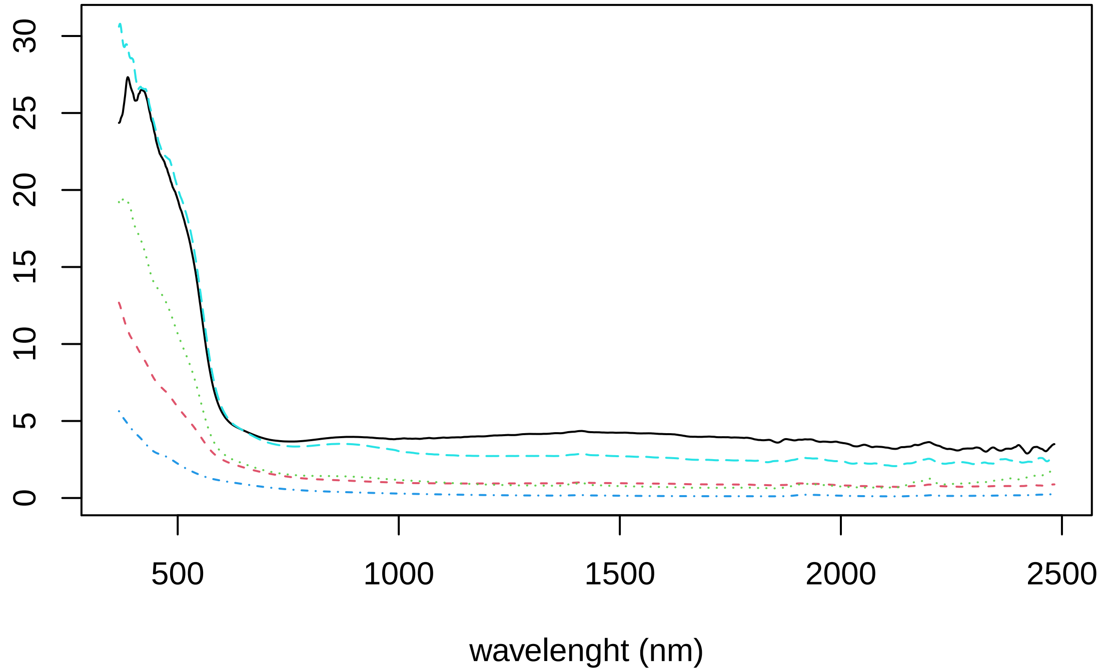
```
spectra.process1 <- savitzkyGolay(spectra, 0, 3, 35)
matplot(as.numeric(colnames(spectra.process1)), t(spectra.process1[1:5,]), type = "l",
        xlab = "wavelenght (nm)",
        ylab = "Reflectance factor")
```



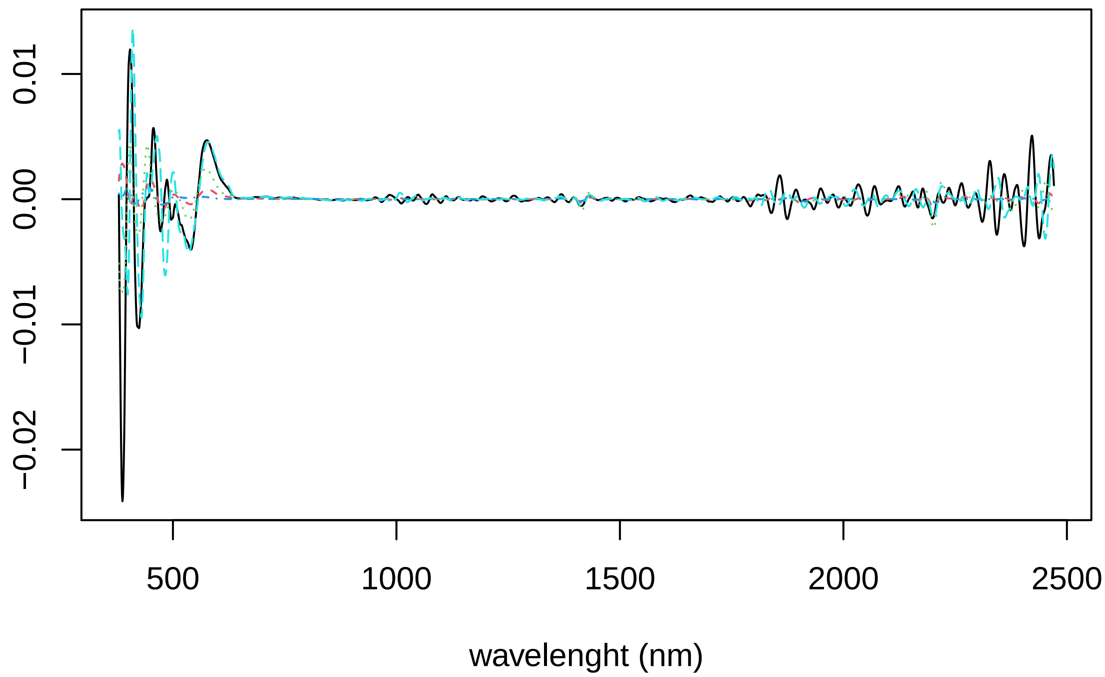
Kubelka-Munk transformation and second derivative

The Kubelka-Munk theory was developed for pigment mixtures but it can be applied to soils. Basically, the theory establishes that the behavior of a pigment mixture with respect to incident light of wavelength can be characterized by two coefficients: K (absorption) and S (scattering). The solution of the Kubelka-Munk equation for thick ("opaque") layers indicates that the K and S values of a pigment mixture can be obtained from the reflectance values of two or more mixtures of this pigment mixture with a white standard. It has been shown, in addition, that the K and S values of a pigment mixture are simply additive functions of the K and S values of the constituent pigments weighted in accordance with the proportion of each pigment. Thus, the theory permits one to predict the effect of a known soil pigment (e.g., hematite) on soil color, or vice versa, to calculate the proportion or the color of a soil pigment from soil reflectance data.

```
KM = ((1-spectra.process1)^2)/(2*spectra.process1) # from smooth spectra
matplot(as.numeric(colnames(KM)), t(KM[1:5,]), type = "l",
        xlab = "wavelenght (nm)",
        ylab = " ")
```




```r
sec.deriv = savitzkyGolay(KM, m = 2, p = 3, w = 25)
matplot(as.numeric(colnames(sec.deriv)), t(sec.deriv[1:5,]), type = "l",
        xlab = "wavelenght (nm)",
        ylab = " ")
```



Estimates of AHm and AGm

AHmt and AGt corresponds to the amplitude of Hematite and Goethita proposed by **Scheinost et al. (1998)**. In this paper, the authors used the amplitude between the ~415 nm minimum and the ~445 nm maximum for goethite (denoted as Y1), and between the ~535 nm minimum and the ~580 nm maximum for hematite (denoted as Y2).

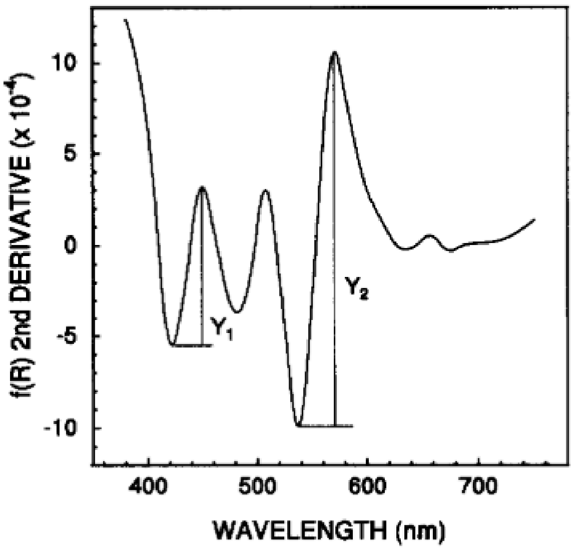


Figure 6. Position and amplitude of the bands selected for the quantification of goethite (Y_1) and hematite (Y_2) in soil samples.

AHm

```r
Hm.max = apply(data.frame(sec.deriv[ ,157:202]), MARGIN = 1, FUN = max)  #The sec.deriv[,157:202]
#corresponds to the amplitude between 415 and 445 nm

Hm.min = apply(data.frame(sec.deriv[ ,157:202]), MARGIN = 1, FUN = min)

AHm = as.data.frame((Hm.max^2)^(1/2))+((Hm.min^2)^(1/2))
colnames(AHm) <- "AHm"
head(AHm)
```

```
##                    AHm
## PIRT124 0.0087326937
## PIRT129 0.0012029425
## PIRT197 0.0040785691
## DI_ 146 0.0002988281
## PIRT123 0.0088903625
## PIRT125 0.0062529448
```

AGt

```r
Gt.max = apply(data.frame(sec.deriv[ ,37:67]), MARGIN = 1, FUN = max) #The sec.deriv[,37:67]
#corresponds to the amplitude between 415 and 445 nm

Gt.min = apply(data.frame(sec.deriv[ ,37:67]), MARGIN = 1, FUN = min)

AGt = as.data.frame((Gt.max^2)^(1/2))+((Gt.min^2)^(1/2))
```

```r
colnames(AGt) <- "AGt"
head(AGt)
```

```
##                 AGt
## PIRT124 0.010490066
## PIRT129 0.002463870
## PIRT197 0.007144992
## DI_ 146 0.001279398
## PIRT123 0.016374005
## PIRT125 0.010491674
```

The values provided here for AHm and AGt are considered in relative quantities only, as they do not have units of measure (for example, g kg). The higher the value, the greater the amount of the mineral in the soil. These values can be used to calculate the ratio Hm / (Hm + Gt) to get an idea of which of these minerals is in greater quantity in the soil under study.

```r
Hm.Gt <- cbind(AHm, AGt)
head(Hm.Gt)
```

```
##                  AHm         AGt
## PIRT124 0.0087326937 0.010490066
## PIRT129 0.0012029425 0.002463870
## PIRT197 0.0040785691 0.007144992
## DI_ 146 0.0002988281 0.001279398
## PIRT123 0.0088903625 0.016374005
## PIRT125 0.0062529448 0.010491674
```

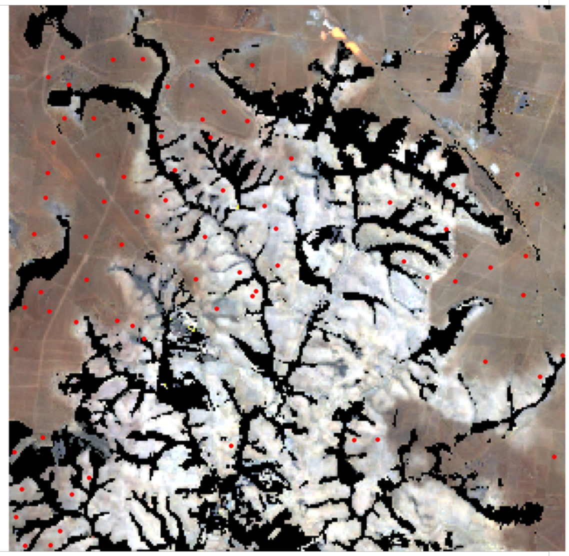
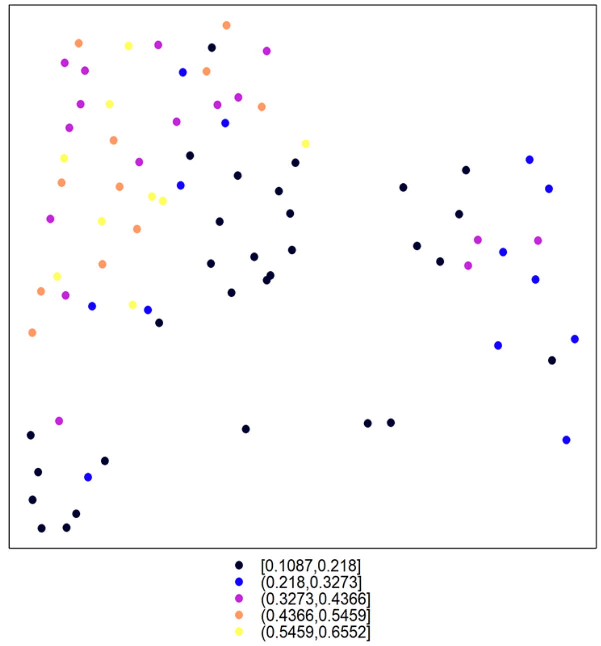
```r
ratio <- as.data.frame(Hm.Gt$AH / (Hm.Gt$AH + Hm.Gt$AG))
colnames(ratio) <- "ratio"
rownames(ratio) <- rownames(AHm)
head(ratio)
```

```
##             ratio
## PIRT124 0.4542893
## PIRT129 0.3280621
## PIRT197 0.3633935
## DI_ 146 0.1893443
## PIRT123 0.3518933
## PIRT125 0.3734301
```

Plot the SYSI and the sampling points

In the first figure, the SYSI (Syntehtic Soil Image) shows an area with highly variable soil texture, as demosntrated by its colors. Light areas corresponds to sandy soils while darker areas are related with clayey soils. In the second figure we can show that the highest values of the ratio Hm(Hm + Gt) are on clayey soils, which is coherent with the literature.

```r
dat2 <- cbind(data[,2:3], ratio)
coordinates(dat2) <- ~X+Y
#plotRGB(SYSI, r=3, g=2, b=1, stretch = "lin")
#plot(ratio2, add=T, col = "red", pch = 20)
#spplot(data2)
```

In their article, Scheinost et al. (1998) stated that a significant correlation was found between Y1 and goethite content and between Y2 and hematite content. They describe two equationos as follows:

$$Goethite(gkg^-1) = -0.06 + 268(Y_1)(R^2 = 0.86)$$

$$Hematite(gkg^-1) = -0.09 + 402(Y_2)(R^2 = 0.85)$$

This equations can be used to obtain a more quantitative estimate of Hm and Gt contents in the soils, but we should aware about the accuracy of these estimates.

Niomi Index proposed by Viscarra Rossel et al. (2010)

This is another way to obtain mineralogical information from lab spectra. This method was proposed by Viscarra Rossel et al. (2010) for mapping oxides in Australia, named NIODI and can be found here:

According to the authors "Positive values of the index indicate the presence of goethite, negative values the presence of hematite, and values approaching zero indicate measurements that are increasingly uncertain or where both minerals may be present. Note tha tthe NIODI indicates the relative proportion of hematite and goethite without being an estimate of their actual abundance in the soil sample.

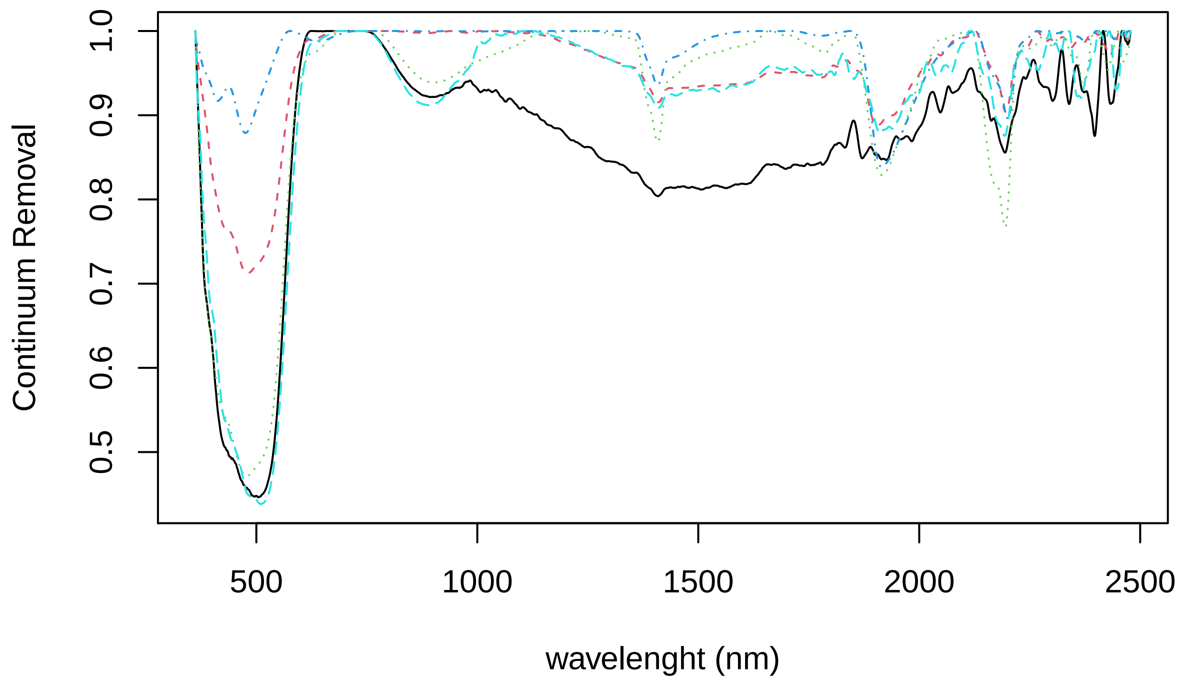
The first step consist in calculating the continuum removal of the lab spectra and find the main absorption features. The authors used the 880 and 920 nm, which correspond to the wavelengths at which major absorptions due to hematite and goethite occur, respectively. More details can be found **here**

Calculate the continuum removal

The “lenght” argument corresponds to the lenght of the first spectra pre-processed by Savitznky-Golay, that is the smooth spectra

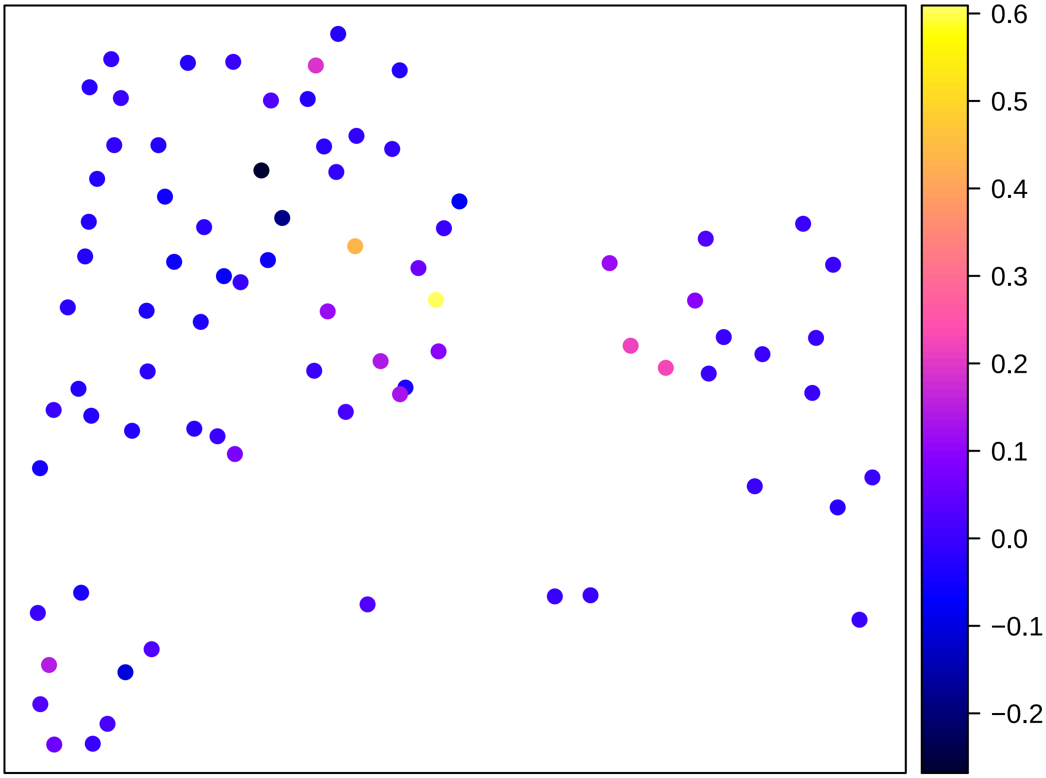
```
lenght = 362:2478
CR = continuumRemoval(spectra.process1, wav = lenght)
matplot(as.numeric(colnames(CR)), t(CR[1:5,]), type = "l",
```

```
        xlab = "wavelenght (nm)",
        ylab = "Continuum Removal")
```



After the CR was obtained, the absorption band depth (D) at the particular wavelength was calculated by substracting the CR reflectance from 1. According to the authors, "this provides an objective and physically based measurement of the abundance of hematite and goethite in soil".

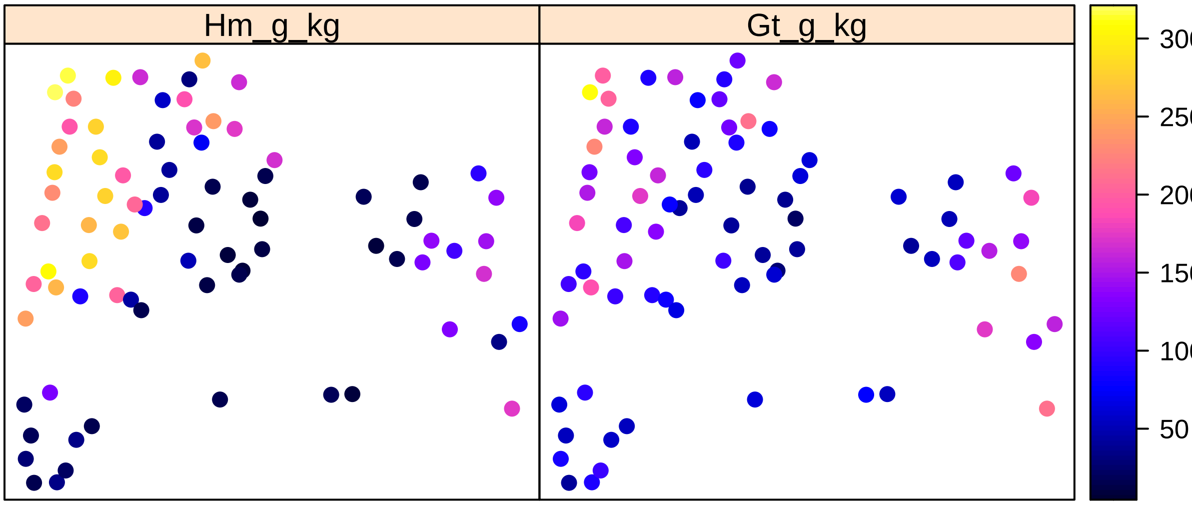
```
D = 1 - CR
D880 = apply(data.frame(D[ ,499:569]), MARGIN = 1, FUN = max)
D920 = apply(data.frame(D[ ,549:609]), MARGIN = 1, FUN = max)

NIODI = as.data.frame((D920 - D880)/(D880 + D920))
colnames(NIODI) <- "NIODI"


dat3 <- cbind(data[,2:3], NIODI)
coordinates(dat3) <- ~X+Y
spplot(dat3, colorkey = TRUE)
```



Hm and Gt according to Fernandes et al. (2004)

Fernandes et al. (2004) studied thirteen Ferralsols (Latossolos) from three Brazilian States (São Paulo, Minas Gerais and Espirito Santo). They followed the same procedure described in Scheinost to obtain the AHm and AGt amplitudes and then related these to the Hm and Gt content measured in laboratory. They proposed two equations to obtain Hm and Gt from the AHm and AGt amplitudes (420 to 450 nm and 530 to 570 nm).

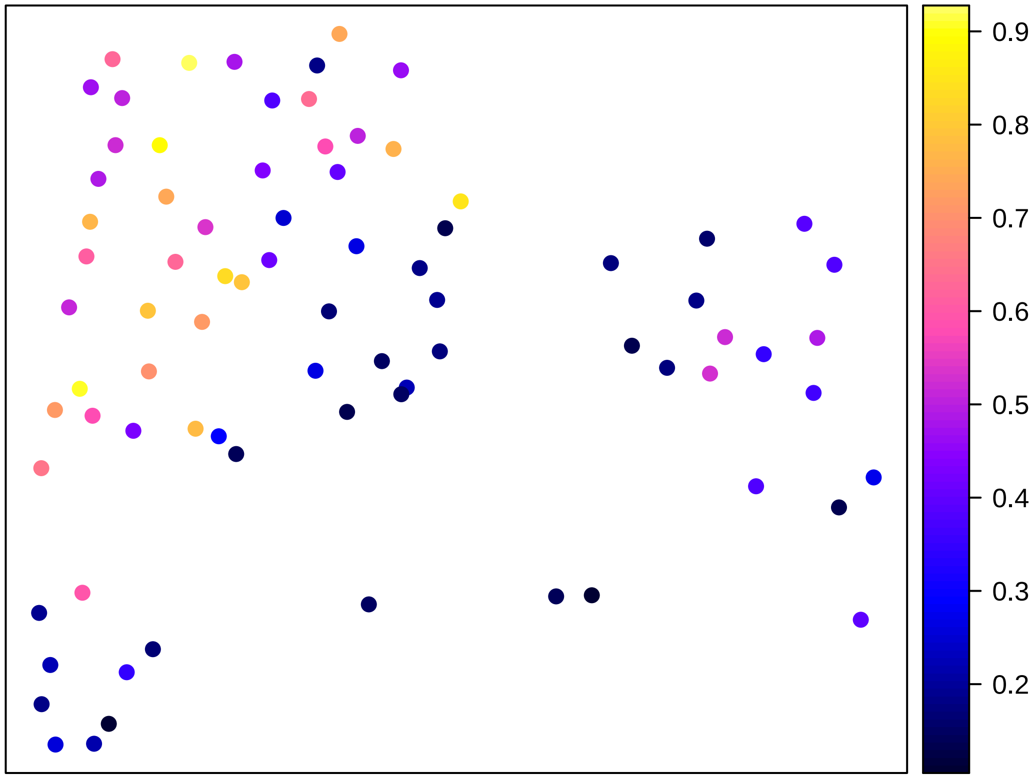
```
Hm_gkg = -1.6 + (36320*AHm)   #(R2 = 0.94)
Gt_gkg = 5.7 + (18607*AGt)    #(R2 = 0.63)

dat_Hm_Gt <- as.data.frame(cbind(data[,2:3], Hm_gkg, Gt_gkg))
colnames(dat_Hm_Gt) <- c("X", "Y", "Hm_g_kg", "Gt_g_kg")
coordinates(dat_Hm_Gt)<- ~X+Y
spplot(dat_Hm_Gt, colorkey = TRUE)
```



Hematite and goethite ratio (Hm/(Hm+Gt)) proposed by Fernandes et al. (2004)

The authors also proposed the ratio Hm(Hm+Gt), based on the following equation

```
Hm_Gt_calculated = -0.059 + (1.506*AHm)/(AHm+AGt)
Hm.Gt.ratio <- cbind(data[,2:3], Hm_Gt_calculated)
colnames(Hm.Gt.ratio) <- c("X", "Y", "ratioHm_Gt")
coordinates(Hm.Gt.ratio) <- ~X+Y
spplot(Hm.Gt.ratio, colorkey=TRUE)
```

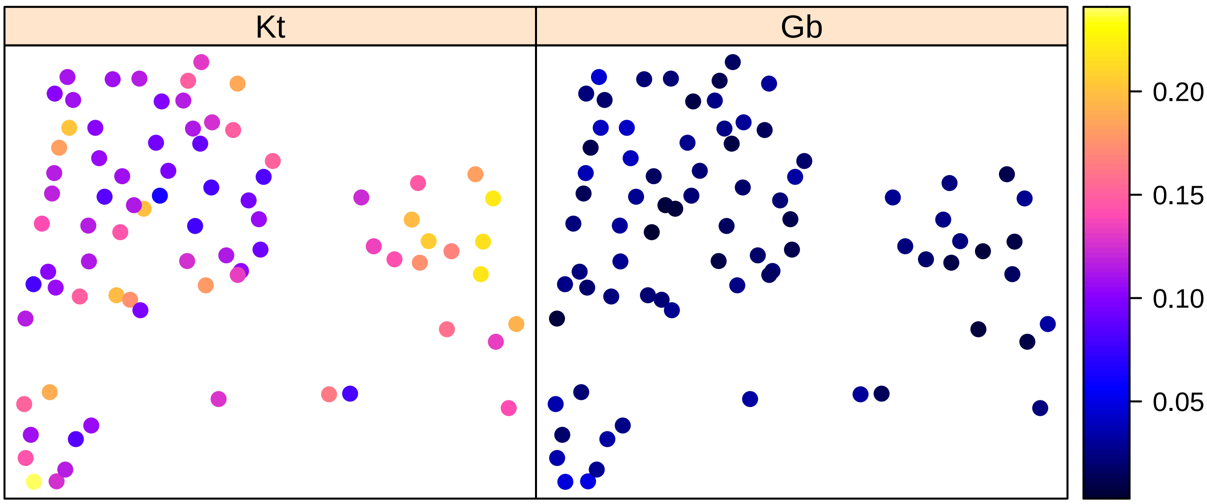


Kaolinite and Gibbsite from Contiuum removal

Calculate minumum and maximum reflectance between and 2293 nm

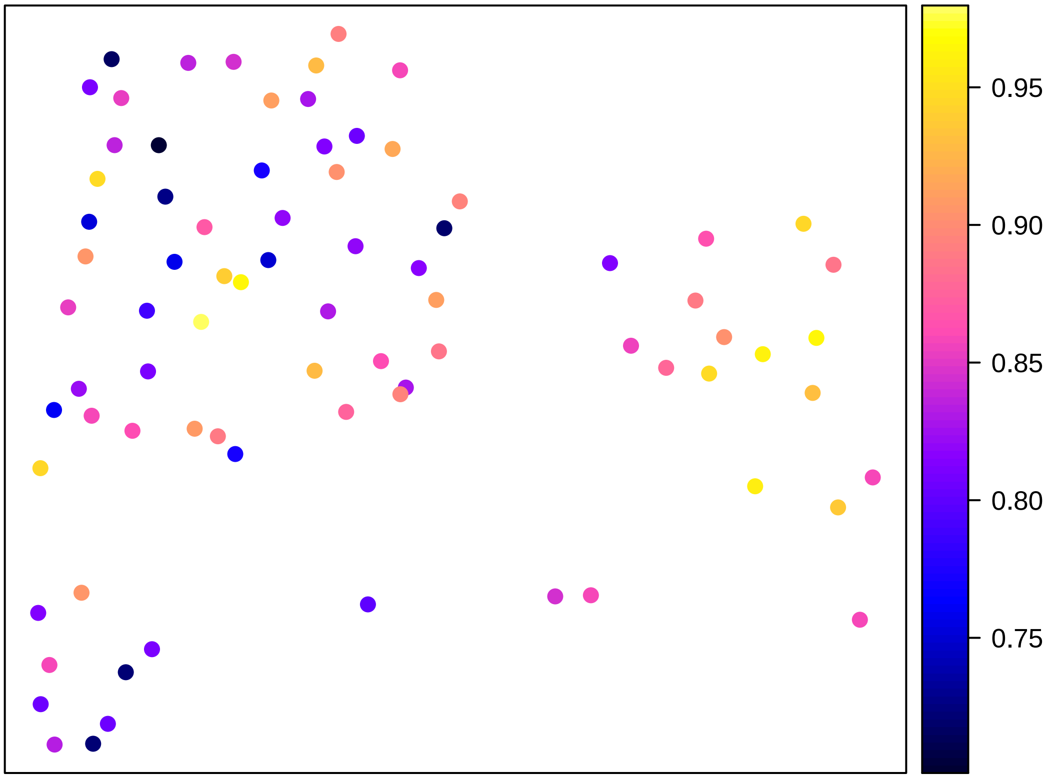
More details about these calculations canbe found in **Poppiel et al. (2020)**

```
CR_min = apply(data.frame(D[ ,1840:1903]), MARGIN = 1, FUN = min)
KT_max = apply(data.frame(D[ ,1771:1863]), MARGIN = 1, FUN = max)
GB_max = apply(data.frame(D[ ,1863:1932]), MARGIN = 1, FUN = max)

Kt_index = as.data.frame(KT_max - CR_min)
Gb_index = as.data.frame(GB_max - CR_min)

Kt_Gb <- cbind(data[,2:3], Kt_index, Gb_index)
colnames(Kt_Gb) <- c("X", "Y", "Kt", "Gb")
coordinates(Kt_Gb) <- ~X+Y
spplot(Kt_Gb, colorkey = TRUE)
```



From these values we can obtain the ratio Kt(Kt+Gb), where high values indicates higher contents of Kt

```
ratioKt_Gb <- (Kt_Gb$Kt/(Kt_Gb$Kt+Kt_Gb$Gb))
ratioKt_Gb_withXY <- cbind(data[,2:3],ratioKt_Gb)
colnames(ratioKt_Gb_withXY) <- c("X", "Y", "ratioKt_Gb")
coordinates(ratioKt_Gb_withXY) <- ~X+Y
spplot(ratioKt_Gb_withXY, colorkey = TRUE)
```



Join all estimated values together

```
data.mineralogy <- cbind(data[,2:3], ratio, NIODI, Hm_gkg, Gt_gkg, Hm_Gt_calculated, ratioKt_Gb)
colnames(data.mineralogy) <- c("X", "Y", "ratioHm_Gt", "NIODI", "Hm_g_kg", "Gt_g_kg",
                               "ratioHm_Gt_calculated", "ratioKt_Gb")
head(data.mineralogy)
```

```
##                 X         Y ratioHm_Gt         NIODI    Hm_g_kg   Gt_g_kg
## PIRT124 -47.59046 -22.86183  0.4542893 -0.0058790675 315.571437 200.88866
## PIRT129 -47.57588 -22.87263  0.3280621 -0.2682005701  42.090873  51.54524
## PIRT197 -47.52205 -22.88890  0.3633935 -0.0007642681 146.533631 138.64687
## DI_ 146 -47.56190 -22.89372  0.1893443 -0.0332127766   9.253438  29.50577
## PIRT123 -47.59256 -22.86455  0.3518933 -0.0185146565 321.297967 310.37112
## PIRT125 -47.58952 -22.86560  0.3734301 -0.0028002530 225.506956 200.91859
##         ratioHm_Gt_calculated ratioKt_Gb
## PIRT124             0.6251597  0.7157697
## PIRT129             0.4350616  0.7717712
## PIRT197             0.4882706  0.9646024
## DI_ 146             0.2261524  0.8271043
## PIRT123             0.4709514  0.8106064
## PIRT125             0.5033857  0.8534178
```

Spatial prediction of the estimates mienralogical indexes and contents

```r
coordinates(data.mineralogy) <- ~X+Y  #make spatial object

SYSI.df <- extract(SYSI, data.mineralogy)  #estract SYSI values for each soil sample
head(SYSI.df)
```

```
##      SYSI.1 SYSI.2 SYSI.3 SYSI.4 SYSI.5 SYSI.6
## [1,]    421    693   1041   1572   2053   1767
## [2,]    587    937   1300   1866   2236   2178
## [3,]    501    912   1262   1807   2255   1984
## [4,]   1018   1551   1977   2834   4045   3538
## [5,]    423    678   1006   1366   1640   1477
## [6,]    373    639    936   1261   1466   1379
```

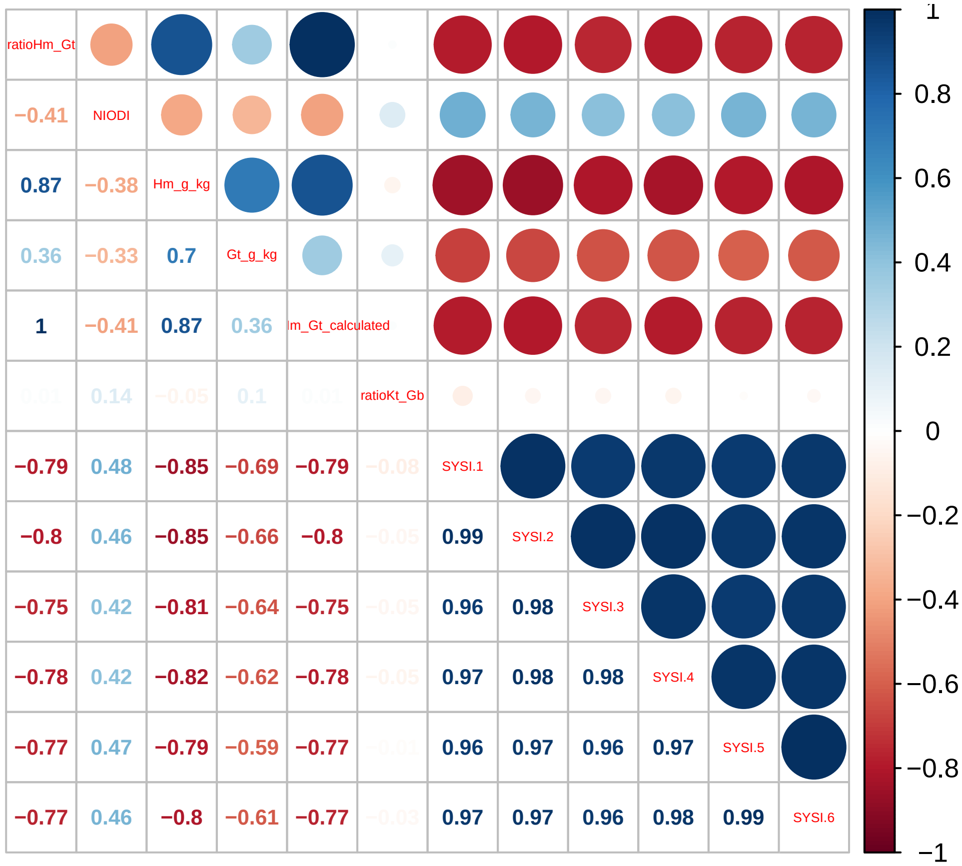
```r
dat1 <- cbind(as.data.frame(data.mineralogy), SYSI.df) #join data and SYSI values in a single table
head(dat1)
```

```
##                 X         Y ratioHm_Gt         NIODI    Hm_g_kg   Gt_g_kg
## PIRT124 -47.59046 -22.86183  0.4542893 -0.0058790675 315.571437 200.88866
## PIRT129 -47.57588 -22.87263  0.3280621 -0.2682005701  42.090873  51.54524
## PIRT197 -47.52205 -22.88890  0.3633935 -0.0007642681 146.533631 138.64687
## DI_ 146 -47.56190 -22.89372  0.1893443 -0.0332127766   9.253438  29.50577
## PIRT123 -47.59256 -22.86455  0.3518933 -0.0185146565 321.297967 310.37112
## PIRT125 -47.58952 -22.86560  0.3734301 -0.0028002530 225.506956 200.91859
##         ratioHm_Gt_calculated ratioKt_Gb SYSI.1 SYSI.2 SYSI.3 SYSI.4 SYSI.5
## PIRT124             0.6251597  0.7157697    421    693   1041   1572   2053
## PIRT129             0.4350616  0.7717712    587    937   1300   1866   2236
## PIRT197             0.4882706  0.9646024    501    912   1262   1807   2255
## DI_ 146             0.2261524  0.8271043   1018   1551   1977   2834   4045
## PIRT123             0.4709514  0.8106064    423    678   1006   1366   1640
## PIRT125             0.5033857  0.8534178    373    639    936   1261   1466
##         SYSI.6
## PIRT124   1767
## PIRT129   2178
## PIRT197   1984
## DI_ 146   3538
## PIRT123   1477
## PIRT125   1379
```

Pearson´s correlation between mineralogical index and SYSI bands

By the graph we can conclude that only the ratio Kt/(Kt+Gb) does not have a good correlation with the satellite image bands. The NIODI index, although showed good correlations, these were < 0.5

```r
corrplot.mixed(cor(dat1[,-1:-2]), number.cex = 0.7, #correlation between SYSI
               tl.cex = 0.4)  #and mineralogical values
```



Mapping mineralogical indexes and amounts using satellite images

To obtain spatial maps of the mineralogical indexes and amount, we are going to use the estimated values as dependent variables and the SYSI bands as independent variables. The Cubist algorithm with the default parameters of commitees and neighbours from the caret package will be used. It is worth noting that this is a simple example, and a combinations of hyperparameters should be tested to find out which one can provide the best estimates. Here we will build a model to predict the ratio Hm/(Hm+Gt) via 10-fold-cross validation, using all samples from our dataset. Therefore, there will be no external validation.

Construct the regression three model via Cubist for Hm/(Hm+Gt) ratio

```
ratioHm.Gt <- train(dat1[,9:14], dat1$ratioHm_Gt, method = "cubist",
                    trControl = trainControl(method = "cv", number = 10))
ratioHm.Gt
```

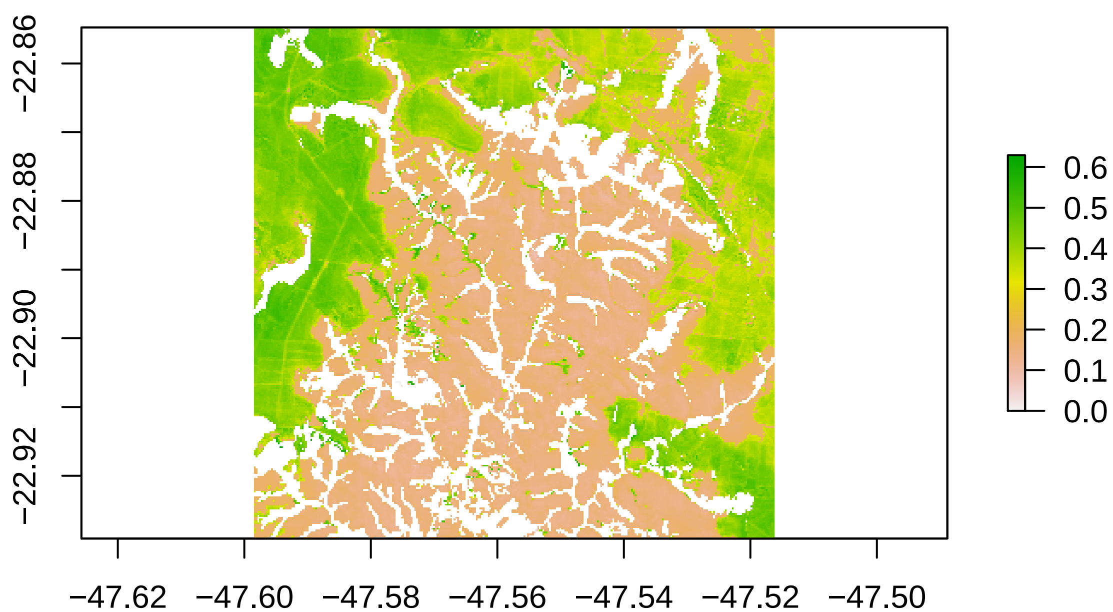
```
## Cubist
##
## 79 samples
##  6 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 71, 71, 72, 71, 71, 71, ...
```

```
## Resampling results across tuning parameters:
##
##   committees  neighbors  RMSE        Rsquared   MAE
##    1          0          0.08779620  0.7205493  0.06893807
##    1          5          0.09429927  0.6909859  0.07574463
##    1          9          0.09122913  0.7041319  0.07392163
##   10          0          0.08793922  0.7320058  0.06750103
##   10          5          0.09177508  0.7043105  0.07300331
##   10          9          0.08969310  0.7150850  0.07170417
##   20          0          0.08808290  0.7298901  0.06719810
##   20          5          0.09220725  0.7018546  0.07303827
##   20          9          0.08990692  0.7134131  0.07157089
##
## RMSE was used to select the optimal model using the smallest value.
## The final values used for the model were committees = 1 and neighbors = 0.
```

Predict on the satellite

We used the satellite image to obtain a spatial map of the ratio Hm/(Hm+Gt), which vary from 0 to 1. Values aproaching 1 indicate higher Hm contents while values approaching 0 indicate higher amounts of Gt.

```
ratioHm.Gt.map <- raster::predict(SYSI, ratioHm.Gt)
plot(ratioHm.Gt.map)
```



References

FERNANDES, R.B.A., BARRÓN, V., TORRENT, J., FONTES, M.P.F., 2004. Quantificação de óxidos de ferro de latossolos brasileiros por espectroscopia de refletância difusa. Rev. Bras. Ciênc. Solo 28, 245–257

Fernandes, K., Marques Júnior, J., Bahia, A.S.R. de S., Demattê, J.A.M., Ribon, A.A., 2020. Landscape-scale spatial variability of kaolinite-gibbsite ratio in tropical soils detected by diffuse reflectance spectroscopy. Catena 195, 104795. https://doi.org/10.1016/j.catena.2020.104795

Mendes, W. de S., Demattê, J.A.M., Bonfatti, B.R., de Resende, M.E.B., Campos, L.R., Saraiva da Costa, A.C., 2021. A novel framework to estimate soil mineralogy using soil spectroscopy. Appl. Geochemistry 127, 104909. https://doi.org/10.1016/j.apgeochem.2021.104909

POPPIEL, R. R. , LACERDA, M. P. C., RIZZO R., SAFANELLI J. L., BONFATTI B. R. SILVERO N. E. Q., DEMATTÊ J. A. M.. Soil Color and Mineralogy Mapping Using Proximal and Remote Sensing in Midwest Brazil. Remote Sens. 2020, 12, 1197; doi:10.3390/rs12071197

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36

SCHEINOST, A.C., CHAVERNAS, A., BARRÓN, V., TORRENT, J., 1998. Use and limitations of second derivative diffuse reflectance spectroscopy in the visible to near infrared range to identify and quantify Fe oxides in soils. Clays Clay Miner. 46, 528–536.

Silva, L.S., Marques Júnior, J., Barrón, V., Gomes, R.P., Teixeira, D.D.B., Siqueira, D.S., Vasconcelos, V., 2020. Spatial variability of iron oxides in soils from Brazilian sandstone and basalt. Catena 185, 104258. https://doi.org/10.1016/j.catena.2019.104258

VISCARRA ROSSEL, R.A., BUI, E.N., CARITAT, P., MCKENZIE, N.J., 2010. Mapping iron oxides and the color of Australian soil using visible-near-infrared reflectance spectra. J. Geophys. Earth Surface 115 F04031.

Viscarra Rossel, R.A., 2011. Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. J. Geophys. Res. 116, F04023. https://doi.org/10.1029/2011JF001977

An introduction to the prospectr package