

Nota técnica

DESARROLLO DE MÉTODOS DE ANÁLISIS DE ESPECTROSCOPIA Y ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA LA EVALUACIÓN DE ALGUNAS PROPIEDADES DEL SUELO EN COSTA RICA

Johan Perret^{1/}, José Eduardo Villalobos-Leandro², Karim Abdalla-Bolaños³,
Carol Lucía Fuentes-Fallas⁴, Katherine Michelle Cuarezma-Espinoza⁵,
Esteban Nicolás Macas-Amaya⁶, María Teresa López-Maietta⁷, Darren Drewry⁸*

Palabras clave: Espectroradiometría; firmas hiperespectrales;
aprendizaje automático; modelado PLSR; agricultura de precisión.

Keywords: Spectroradiometry; hyperspectral signatures;
machine Learning; PLSR modeling; precision agriculture.

Recibido: 15/07/2019

Aceptado: 12/12/2019

RESUMEN


Introducción. Los métodos convencionales de laboratorio para analizar el contenido de nutrientes del suelo, generalmente requieren mucho tiempo y son costosos. En contraparte, la espectroradioscopia visible e infrarroja ofrece una técnica rápida para caracterizar el suelo en laboratorio. Las firmas espectrales codifican información sobre las características inherentes del suelo, como la composición mineral, el contenido de nutrientes, los compuestos orgánicos y

ABSTRACT


Development of spectroscopic methods and machine learning algorithms for evaluation of some soil properties in Costa Rica. Introduction. Conventional laboratory methods to analyze soil nutrients are usually time-consuming and costly. On the other hand, the visible and infrared spectroscopy offers a rapid technique to characterize soils. The spectral signatures encode information about the inherent composition of the soil, which

* Autor para correspondencia. Correo electrónico: jperret@earth.ac.cr


1 Universidad EARTH, Centro de Agricultura de Precisión, Costa Rica.

 0000-0002-3882-115X.


2 Universidad EARTH, Centro de Agricultura de Precisión, Costa Rica.

 0000-0002-9846-1723.


3 Universidad EARTH, Centro de Agricultura de Precisión, Costa Rica.

 0000-0003-1939-9536.


4 Universidad EARTH, Centro de Agricultura de Precisión, Costa Rica.

 0000-0002-3175-1523.


5 Universidad EARTH, Centro de Agricultura de Precisión, Costa Rica.

 0000-0002-4897-7254.


6 Universidad EARTH, Centro de Agricultura de Precisión, Costa Rica.

 0000-0001-7576-0272.

7 Universidad EARTH, Laboratorio de Suelos y Aguas, Costa Rica.

 0000-0002-9026-3570.

8 Ohio State University, Department of Food, Agricultural and Biological Engineering, USA.

 0000-0003-2593-7599.

el agua. **Objetivo.** El objetivo principal de este proyecto fue construir bibliotecas espectrales para los suelos tropicales de Costa Rica y determinar las bandas hiperespectrales óptimas en el rango espectral visible infrarrojo cercano e infrarrojo de onda corta para caracterizar propiedades de suelo. **Materiales y métodos.** Las mediciones hiperespectrales se llevaron a cabo con un espectroradiómetro ASD FieldSpec 4 para generar las firmas de reflectancia espectral de más de 1300 muestras de suelo de Costa Rica, pre-procesadas en el Laboratorio de Suelo de la Universidad EARTH. Se determinó el contenido de nutrientes de cada muestra de suelo, mediante plasma acoplado inductivamente. Además, se evaluaron el pH, la acidez extraíble, la saturación de bases, la saturación de acidez, la capacidad efectiva de intercambio catiónico, el carbono, la materia orgánica y la textura del suelo. Se desarrollaron modelos de regresión de mínimos cuadrados parciales (PLSR) en MATLAB para predecir esas propiedades de suelo a partir de firmas hiperespectrales. **Resultados.** Este enfoque espectroradioscópico, combinado con modelos de aprendizaje automático, permitió identificar bandas de ondas óptimas específicas en zonas espectrales en las que se puede predecir cada nutriente. Se logró una estimación precisa del contenido de diferentes componentes (Ca, Mg, Fe, C, N y CICE) con un R^2 superior a 0,8 y un error cuadrático medio (RMSE) inferior a 10%. **Conclusión.** Los análisis espectroscópicos combinados con el método Mínimo Cuadrático Parcial (PLS), pueden proporcionar una herramienta muy útil para la agricultura de precisión en los suelos tropicales de Costa Rica.

comprises mineral composition, nutrient content, organic compounds and water. **Objective.** The main objective of this study was to build spectral libraries for the tropical soils of Costa Rica and to determine the optimal hyperspectral wave bands in the visible and near infrared and shortwave infrared to characterize soil properties in laboratory. **Materials and methods.** The hyperspectral measurements were carried out with an ASD FieldSpec 4 spectroradiometer to generate spectral reflectance signatures of more than 1300 soil samples from Costa Rica preprocessed at EARTH University's soil laboratory. The nutrient content of each soil sample was determined by inductively coupled plasma. In addition, pH, exchangeable acidity, base saturation, acid saturation, effective cation exchange capacity, soil carbon, organic matter and soil texture were evaluated. Algorithms in MATLAB were developed to compile a hyperspectral data base and used a partial least squares regression (PLSR) methods to generate predictive models for these soil properties from hyperspectral signatures. **Results.** This spectroscopic approach combined with machine learning models allowed the identification of specific optimal wavebands in the spectral areas in which each nutrient can be predicted. An accurate estimation of the concentration of different components (Ca, Mg, Fe, C, N and CECe) was achieved with an R^2 greater than 0,8 and a mean square error (RMSE) lower than 10%. **Conclusion.** These spectroscopic techniques combined with the PLS regression, can provide a very useful tool for precision agriculture in tropical soils of Costa Rica.

INTRODUCCIÓN

La evaluación de la calidad del suelo, junto con su fertilidad, es sumamente importante para tomar decisiones que permitan optimizar

la producción de alimentos y así contribuir a la capacidad de adaptación al cambio climático. Los métodos de laboratorio convencionales permiten caracterizar la física, química y biología de los suelos. Sin embargo, la exploración en uso

de métodos novedosos permite, eventualmente, contar con métodos más rápidos, precisos y accesibles económicamente para evaluar la calidad del suelo. Actualmente se utilizan muestras de suelo y métodos de laboratorio convencionales como la Espectroscopia de Absorción Atómica o Espectroscopia de emisión por plasma de acoplamiento inductivo y colometría para cuantificar el contenido de nutrientes en suelo. Esos métodos convencionales involucran una serie de pasos e insumos necesarios para poder obtener resultados.

Sin embargo, la espectralradioscopia visible e infrarroja ofrece una técnica con menos pasos e insumos, por lo cual tiene potencial para ser menos costosa y más rápida en comparación con los métodos de análisis convencionales. Con frecuencia esos métodos convencionales requieren preparación de muestras que son propensas a la contaminación y errores durante el muestreo, la manipulación, la extracción, etc., aspectos que instan a la aplicación de métodos más rápidos y accesibles para evaluar la calidad del suelo.

La espectralradioscopia visible e infrarroja (Vis-NIR-SWIR) ha surgido como opción para caracterizar eficazmente varias propiedades del suelo como carbono orgánico, pH, textura, capacidad de intercambio catiónico, contenido de CaCO_3 , entre otros (Angelopoulou *et al.* 2020, Curcio *et al.* 2013, Demattê *et al.* 2016, Liu *et al.* 2016, Miloš y Bensa 2017, Silva *et al.* 2019, Steinberg *et al.* 2016, Viscarra Rossel *et al.* 2016).

Las características de absorción en el espectro visible están dominadas por las excitaciones electrónicas moleculares y las de la gama, NIR contienen una combinación de vibraciones moleculares de armónicos (Adeline *et al.* 2017). Sin embargo, caracterizar el efecto de cada una de las propiedades del suelo en los espectros de Vis-NIR, es una tarea difícil debido a la naturaleza compleja de la matriz del suelo con múltiples superposiciones de características espectrales y colinealidades espectrales fuertes entre las propiedades del suelo (Adeline *et al.* 2017). Las firmas espectrales codifican información sobre la composición inherente del suelo,

que comprende la composición mineral, el contenido de nutrientes, los compuestos orgánicos y el agua.

El objetivo principal de este proyecto fue construir bibliotecas espectrales para los suelos tropicales de Costa Rica, y determinar las bandas hiperespectrales óptimas en el rango espectral visible (400 nm – 700 nm), infrarrojo cercano (NIR, 701-1000 nm) e infrarrojo de onda corta (SWIR, 1001-2500 nm) para desarrollar modelos predictivos de propiedades de suelo mediante firmas hiperespectrales.

Rinnan *et al.* (2009) y Hadoux *et al.* (2014) dan una buena reseña de diversos métodos matemáticos para predecir las propiedades del suelo a partir de los espectros Vis-NIR del suelo, como la regresión lineal múltiple por pasos (Shibusawa *et al.* 2001), la regresión por componentes principales (Chang *et al.* 2001), el análisis de regresión múltiple (Ben-Dor y Banin 1995), los *splines* de regresión adaptativa multivariable (Shepherd y Walsh 2002) y la regresión por *support vector machine* (Stevens *et al.* 2010). El método de regresión de mínimos cuadrados parciales (Partial Least Square Regresión - PLSR) es la técnica estadística multivariada más común para la calibración espectral y la predicción de las propiedades del suelo (Viscarra Rossel *et al.* 2006). En este análisis se evaluó la capacidad de modelos PLSR para la predicción de 29 propiedades del suelo de muestras pre-procesadas en laboratorio a partir de firmas hiperespectrales Vis-NIR-SWIR generadas por un espectroradiómetro (ASD).

MATERIALES Y MÉTODOS

Ubicación y preparación de las muestras de suelo

Un total de 1375 muestras de suelo recolectadas en Costa Rica, principalmente en las provincias de Limón y Guanacaste, fueron analizadas en este estudio. Las muestras de suelo fueron secadas al horno por 48h a 60°C, molidas y homogeneizadas con un molino tipo Dyna-Crush y tamizadas a un tamaño de partícula de

2 mm. Del material se seleccionó una submuestra de 50 g de cada suelo, para conservarlas en la colección del Laboratorio de Suelo y Aguas de la Universidad EARTH para futuras verificaciones e investigaciones.

Análisis de laboratorio del suelo

El contenido de nutrientes de cada muestra de suelo se evaluó en el laboratorio de Suelo y Aguas de la Universidad EARTH con el método de extracción Mehlich III. Se analizaron los macronutrientes primarios y secundarios (P, K, S, Ca y Mg), los micronutrientes (Zn, Cu, Fe, Mn, B) y otros elementos (Si, Na), mediante un plasma de acoplamiento inductivo (ICP), junto con espectrofotómetro de emisión óptico (OES), marca PerkinElmer Optima 800. El carbono y el nitrógeno se estudiaron con un analizador CNS marca LecoCN Trumac. Además, se evaluó el pH, la acidez extractable, la saturación de bases, la saturación de acidez, la textura, la densidad aparente, la capacidad de intercambio catiónico efectiva (CICE) y la materia orgánica del suelo.

Recolección de firmas hiperespectrales

Para generar las firmas hiperespectrales, se utilizó un espectroradiómetro FieldSpec 4 Estándar-Res ASD (Analytical Spectral Devices Inc., Boulder, CO, EE. UU.). Ese equipo tiene un rango espectral de 350-2500 nm con resoluciones espectrales de 3 y 10 nm en los rangos de 350-1000 nm y 1000-2500 nm, respectivamente, y un intervalo de muestreo espectral de 1 nm. Antes de generar las firmas, se precalentó el instrumento por media hora. Durante la adquisición de las firmas, se recalibró el equipo cada 30 min con un panel Spectralon® de 12 × 12 cm² (Labsphere, North Sutton, EE. UU.) para evitar desviación de los sensores Vis-NIR-SWIR. Se colocó una lámpara de filamento de tungsteno del ciclo de halógeno-halógeno del 70W (12V dc regulado) (~ 3100 ° K de temperatura de color), en una caja de madera de 33 x 35,5 x 50 cm, para garantizar una fuente de luz constante y estandarizada que evitara contaminación por otras fuentes de luz. El interior de la

caja fue pintada de color negro para minimizar el efecto de reflexión de las paredes. La fibra óptica fue instalada con un ángulo de 55° y una distancia de 2 cm con respecto a la superficie de la muestra. Eso permitió tener un campo visual de 16 mm sobre la superficie de la muestra de suelo. Se estandarizó también la altura y profundidad de suelo de las muestra (10,2 mm) para evitar cualquier efecto del contenedor del suelo. Para reducir el efecto de la orientación de la muestra, se repitió la lectura hiperespectral en 4 orientaciones diferentes (0°, 90°, 180° y 270°). Se guardó el promedio de esas 4 firmas para las 1375 muestras de suelo en la librería espectral, por lo que se generaron 5500 firmas hiperespectrales en total (1375 muestras, con 4 orientaciones cada una).

Compilación de la librería hiperespectral

Los reportes de laboratorio para las propiedades de suelo (249 en total) se encontraban en diferentes formatos Excel, según el tipo de análisis. En otro orden, las 5500 firmas hiperespectrales generadas por el equipo FieldSpec 4, y a través del programa RS3, fueron guardadas en el formato binario (archivo “*.asd”), propietario de la compañía ASD. Se escribió un programa en MATLAB para extraer los resultados de los análisis de suelo de los archivos Excel y la depuración de los datos erróneos (problema de sintaxis, formato o a fuera de rangos) y luego se procedió a guardar cada resultado de forma sistemática en una base de datos. El programa también favoreció el cálculo de diferentes propiedades de suelo como capacidad de intercambio catiónico efectivo (CICE), saturación de bases (SB), saturación de acidez (SA), relaciones catiónicas (Ca/Mg, Ca/K, Mg/K, (Ca+Mg)/K) y relación de absorción de sodio (RAS). Cada muestra tuvo un código único, el cual permitió integrar las firmas hiperespectrales al asociar el promedio de las 4 firmas sacadas a 90° de orientación por cada muestra y al seguir este código único. El programa de MATLAB también permitió eliminar errores provenientes del cambio de sensores en el equipo FieldSpec 4 en las longitudes de onda de 1000 y 1800 nm (Figura 1).

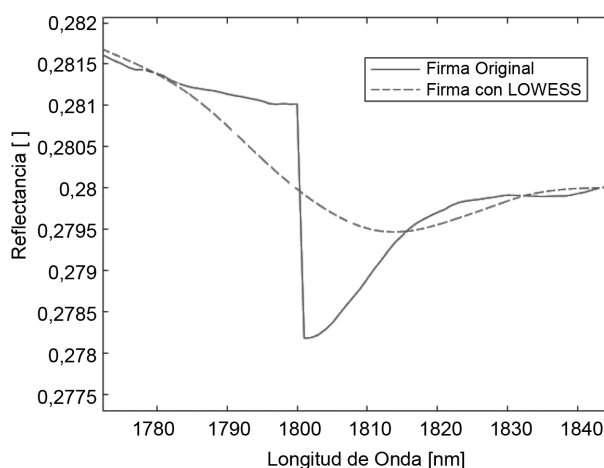


Figura 1. Efecto del algoritmo de LOWESS para eliminar errores provenientes del cambio de sensor en el equipo FieldSpec 4 a la longitud de onda de 1800 nm.

En estas longitudes de onda, las firmas presentan un cambio abrupto que fue suavizado con algoritmo de regresión local (también conocido por sus siglas en inglés, LOWESS - Locally Weighted Scatterplot Smoothing). Adicionalmente, el código de MATLAB logra eliminar las reflectancias de los primeros 50 nm (o sea de 350 a 400 nm), debido a que las firmas contienen mucho ruido para estas longitudes de onda. Finalmente, el programa compila la librería hiperespectral en formato MATLAB (archivo “*.MAT”) y Excel (Archivo “*.xlsx”) para el análisis PLSR.

Análisis PLSR y desarrollo de Modelos Predictivos

PLSR es una técnica de modelado popular utilizada en quimiometría y se utiliza comúnmente para análisis espectral cuantitativo; contribuye a la formulación de modelos predictivos cuando hay diversidad de variables predictoras que son altamente colineales. La técnica está estrechamente relacionada con regresión de componentes principales (PCR). El método PLSR se basa en los mínimos cuadrados parciales (PLS), ya que permite reducir el espacio dimensional al

maximizar la covarianza entre las puntuaciones de los espectros y cada propiedad del suelo. Este método realiza proyecciones rotativas iterativas para estimar matrices de variables latentes y puntuaciones asociadas tanto al espectro como a la variable de respuesta (propiedad de suelo). La regresión de mínimos cuadrados parciales es la asociación de una reducción de PLS con una regresión lineal multivariada clásica, que explica la correlación entre los espectros Vis-NIR y la propiedad del suelo. Los modelos predictivos tienen la forma de:

$$S = C_{400} R_{400} + C_{401} R_{401} + \dots + C_i R_i + m$$

Donde:

S = a la estimación de la propiedad de suelo.

C = al coeficiente de regresión PLSR para cada longitud de onda.

R = a la reflectancia generada por el espectroradiómetro (de 400 a 2500 nm).

m = a la intersección.

Antes del análisis estadístico, los espectros de reflectancia (R) se centraron en el promedio más/menos 2 desviaciones estándares para eliminar firmas atípicas (Figura 2). Luego,

para cada propiedad del suelo, la base de datos de cada configuración espectral se dividió en un conjunto de calibración de 80% de las muestras y un conjunto de validación de 20%

de las muestras, para predecir propiedades de suelo, con un juego de datos independientes a la calibración del modelo. Los algoritmos de PLSR fueron desarrollados en MATLAB.

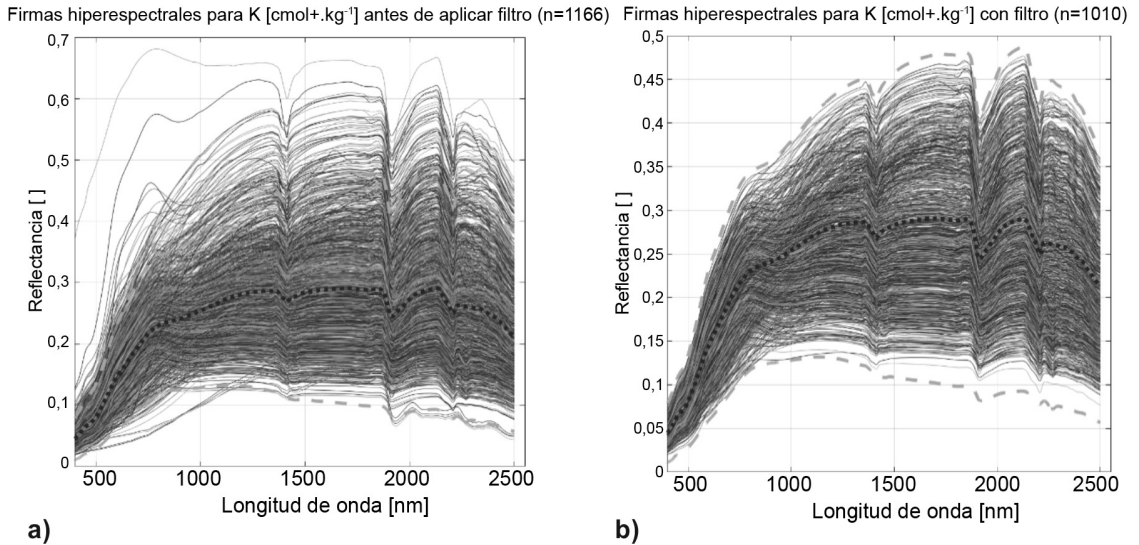


Figura 2. Eliminación de las firmas hiperespectrales atípicas. a) Firmas antes y b) después de aplicar el filtro.

Las líneas punteadas representan el promedio de las firmas y las discontinuadas representan más o menos 2 desviaciones estándares del promedio.

Capacidad de predictiva de los modelos PLSR

Se evaluó el coeficiente de determinación (R^2), así como el error cuadrático medio de predicción (RMSE) del inglés *Root Mean Square Error*, la predicción de la suma de cuadrados de los residuales (PRESS) del inglés *Predicted Residual Sum of Squares*, y el Criterio de información de Akaike (AIC), el Criterio de Wold R, el Criterio de Información de Akaike Normalizado y finalmente el Criterio de Información Bayesiana (BIC). Estas métricas de calidad de calibración y predicción sirvieron de base para indicar la cantidad óptima de componentes para luego proceder a seleccionar los modelos con mejor desempeño para cada una de las propiedades de suelo.

Evaluación de longitudes de onda significativas

Se desarrolló un algoritmo en MATLAB para graficar la ponderación de los coeficientes PLSR con respecto a la longitud de onda Vis-NIR-SWIR y luego se procedió a identificar los 10 máximos/mínimos locales para señalar las longitudes de ondas más relevantes para predicciones.

RESULTADOS Y DISCUSIÓN

Se predijeron 29 propiedades del suelo a partir de un conjunto de datos de 1375 muestras de suelo pre-procesadas en laboratorio recolectadas en Costa Rica; su pH, acidez intercambiable

(AE), K, Ca, Mg, P, Fe, Cu, Zn, Mn, Na, Si, B, S, C, N, Materia Orgánica (MO), textura (porcentaje de arena, limo arcilla), densidad aparente (DA), capacidad de intercambio catiónico (CICe), saturación de bases (SB), saturación de acidez (SA), relaciones catiónicas (Ca/Mg, Ca/K, Mg/K, (Ca+Mg)/K) y relación de absorción de sodio (RAS). Estas propiedades fueron seleccionadas por sus diferentes comportamientos espectroscópicos espectrales y predichos por PLSR.

La Figura 3 presenta los resultados de las diferentes métricas (R², RMSE, PRESS, AIC, Wold's R, AIC Normalizado y BIC) para la

variable Ca como ejemplo. Dichas métricas fueron utilizadas para evaluar la calidad de las predicciones PLSR y determinar la cantidad óptima de componentes en los modelos para cada una de las propiedades de suelo. Los mejores modelos presentaron valores de R², próximos a 1 (uno) tanto con la dinámica de datos de calibración (80%), como con los datos de validación (20%), que llegaron a minimizar los valores de PRESS, AIC y BIC. Las líneas verticales punteadas muestran la cantidad óptima de componentes que se emplearon en los modelos de predicción para cada propiedad de suelo.

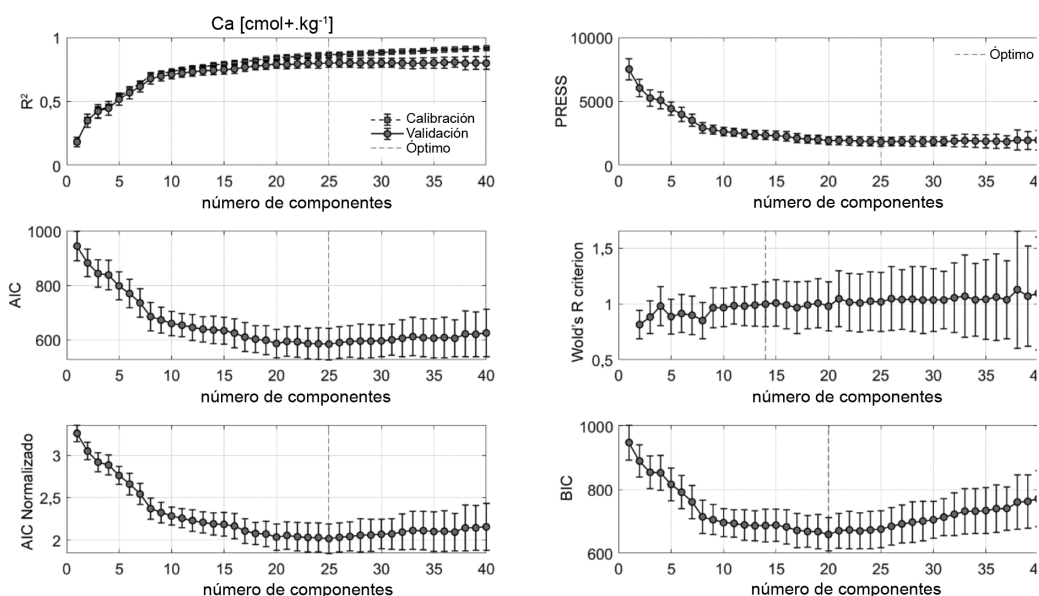


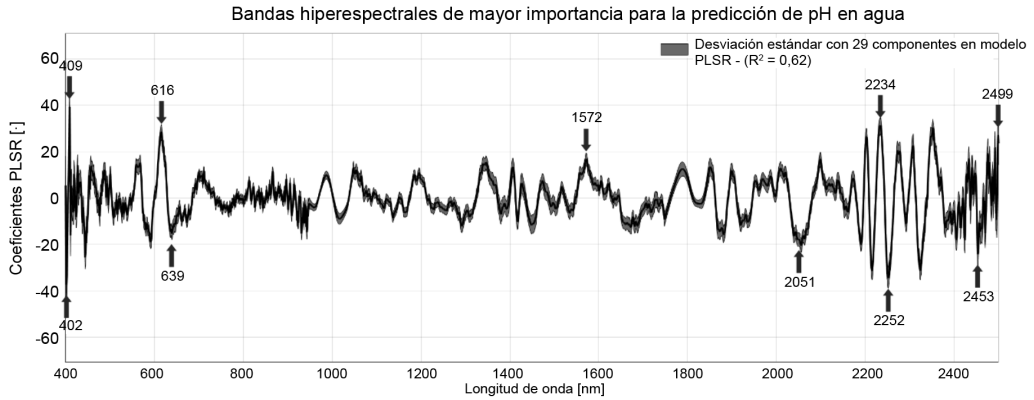
Figura 3. Resultados de las métricas para evaluar la calidad de las predicciones PLSR y determinar la cantidad óptima de componentes en los modelos para calcio.

La línea punteada corresponde al número de componentes óptimos para cada modelo.

Se determinó a qué longitud de onda se encontraban los coeficientes PLSR más sobresalientes para la descripción y la predicción de cada propiedad de suelo; posteriormente se eligieron los 10 picos con el valor absoluto más alto. La Figura 4 muestra resultados

para pH, Mg, y C. Las flechas indican los picos que tienen mayor peso para describir el parámetro de interés. Es interesante notar que ciertas longitudes de ondas muestran rasgos similares entre los parámetros que fueron analizados.

a)



b)



c)





Figura 4. Coeficientes PLSR versus longitud de onda para a) pH del agua, b) Magnesio y c) Carbono. Los coeficientes de regresión PLS pueden utilizarse para seleccionar predictores (longitud de onda) más relevantes de acuerdo con la magnitud de sus valores. Las longitudes de mayor importancia en los modelos PLSR están indicadas por las flechas.

La Tabla 1 presenta las 10 longitudes de onda de mayor relevancia para predecir cada propiedad de suelo. No fue posible reconocer un patrón claro, pero si se superpusieron los 10 picos más significativos para el pH, los macronutrientes y micronutrientes y CICE (Figura 5). Se puede

observar que las bandas se agrupan en 4 zonas del espectro del FieldSpec 4. Más precisamente, se puede observar que las bandas en el espectro visible tienen un papel importante en las predicciones PLSR así que 3 zonas de las longitudes de onda corta infrarroja (SWIR de 1001 a 2500 nm).

Tabla 1. Longitudes de onda más importantes para las predicciones PLSR para cada propiedad de suelo.

Propiedad de Suelo	Posición de los picos más importante para las predicciones PLSR [nm]									
	Mayor  Importancia en Modelos PLSR  Menor									
pH	409	402	2252	2234	616	2499	2453	2051	1572	639
AE	407	2215	2469	403	2231	1353	616	2477	1380	1881
K	2493	2497	406	584	1874	1428	2224	1413	2211	1860
Ca	1396	1432	1918	2214	2349	2425	421	407	2140	615
Mg	611	2214	2352	1393	2424	418	1917	407	1433	1952
P	2471	417	1433	2220	1872	1416	451	2493	2185	1885
Fe	401	409	2474	2484	1405	2239	2036	643	674	2255
Cu	2213	2227	2492	2436	1924	420	1375	1417	407	1898
Zn	440	406	2214	664	2494	629	2455	1943	1410	1905
Na	721	1818	1917	2494	2204	1378	451	1426	406	2426
Mn	2224	2173	2391	1420	1884	441	2497	412	733	665
Si	2492	2270	676	1906	1413	1055	1710	2060	1329	445
B	1812	1904	1358	544	1431	2197	2317	1026	983	464
S	1872	1812	2348	2244	1359	1390	674	445	1113	945
C	1352	2495	1848	1868	402	2459	453	2209	1381	681
N	1352	2495	1848	1868	402	2459	453	2209	1381	681
OM	401	2489	1385	2443	2225	1357	1019	2282	981	1869
Arena	2208	576	2301	979	1873	639	843	1134	1402	1893
Limo	1812	2236	2200	438	1871	1403	1376	1125	903	542
Arcilla	1837	2213	983	1416	2296	1363	1774	483	574	1032
DA	1410	2205	2289	1907	2499	2079	418	1858	789	653
CICE	407	2492	418	2425	2255	1863	611	917	1402	664
SB	1354	620	2214	1917	983	590	1784	902	2354	2469
SA	1354	620	1917	2323	983	517	1784	2492	2274	1456
Ca/Mg	405	454	2498	2332	1019	1361	1808	2294	981	1413
Ca/K	2180	1907	2223	1882	1429	403	2397	2456	1387	547
Mg/K	1905	2180	2223	1881	1431	2456	2395	1385	403	1016
(Ca+Mg).k ⁻¹	1906	2180	2222	1881	1429	403	2397	2456	1388	547
SAR	2270	2494	1382	1652	1942	748	401	562	947	2005

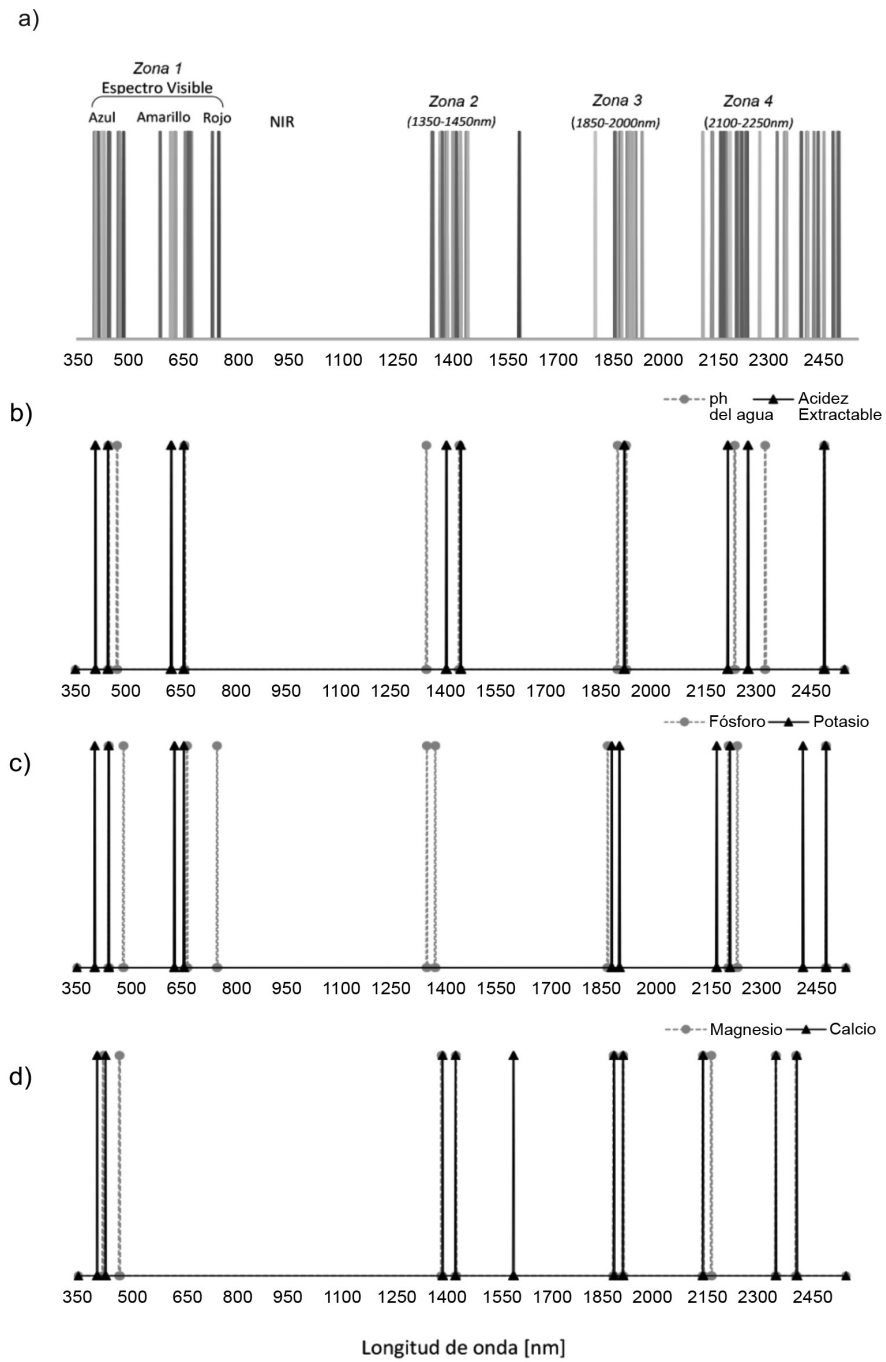


Figura 5. Longitudes de onda de mayor importancia para los modelos PLSR. a) todos los elementos confundidos, b) para solo pH del agua y Acidez Extractable (AE), c) para fósforo y potasio y d) para magnesio y calcio.

Como en el estudio de Adeline *et al.* (2017), se puede agrupar las bandas espectrales más sobresalientes para predecir propiedades de suelo en 4 rangos espectrales. La primera zona se ubica de 400 a 750 nm; según Adeline *et al.* (2017) esta zona responde a la presencia de óxidos de hierro en el suelo. La segunda zona incluye longitudes de onda de 1350 a 1450 nm; según Poppiel *et al.* (2018); ese rango responde a la presencia de minerales arcillosos tipo 2:1 (ej. caolinita) y agua adsorbida en la superficie de las partículas de suelo. La zona 3 de 1850 a 2000 nm estaba correlacionado con propiedades asociadas al contenido de agua (Adeline *et al.* 2017). La cuarta zona incluye bandas espectrales cercanas a 2200 nm y está relacionado con la presencia de CaCO_3 , minerales de arcilla en el suelo y materia orgánica.

La Figura 6 expone los gráficos de dispersión de valor predicho frente al valor

observado para algunas propiedades de suelo. La línea punteada negra en diagonal representa el eje de predicción perfecta. La punteada color gris obedece a la de ajuste del mejor modelo PLSR para esta variable. Entre más cerca se encuentren las 2 líneas punteadas, mejor ajuste predictivo. Las 2 continuas exponen una desviación estándar, aspecto que permite identificar los datos atípicos. El valor de R^2 sin paréntesis (esquina izquierda superior de cada gráfico) muestra el coeficiente de determinación para todos los datos disponibles para esa propiedad de suelo (“n” es la cantidad de muestras usadas para desarrollar el modelo PLSR), y el número entre paréntesis es el coeficiente de determinación para las predicciones con los datos apartados para la validación que correspondió al 20% de todos los datos. Por su parte el RMSE mostró el error cuadrático medio de predicción en porcentaje.

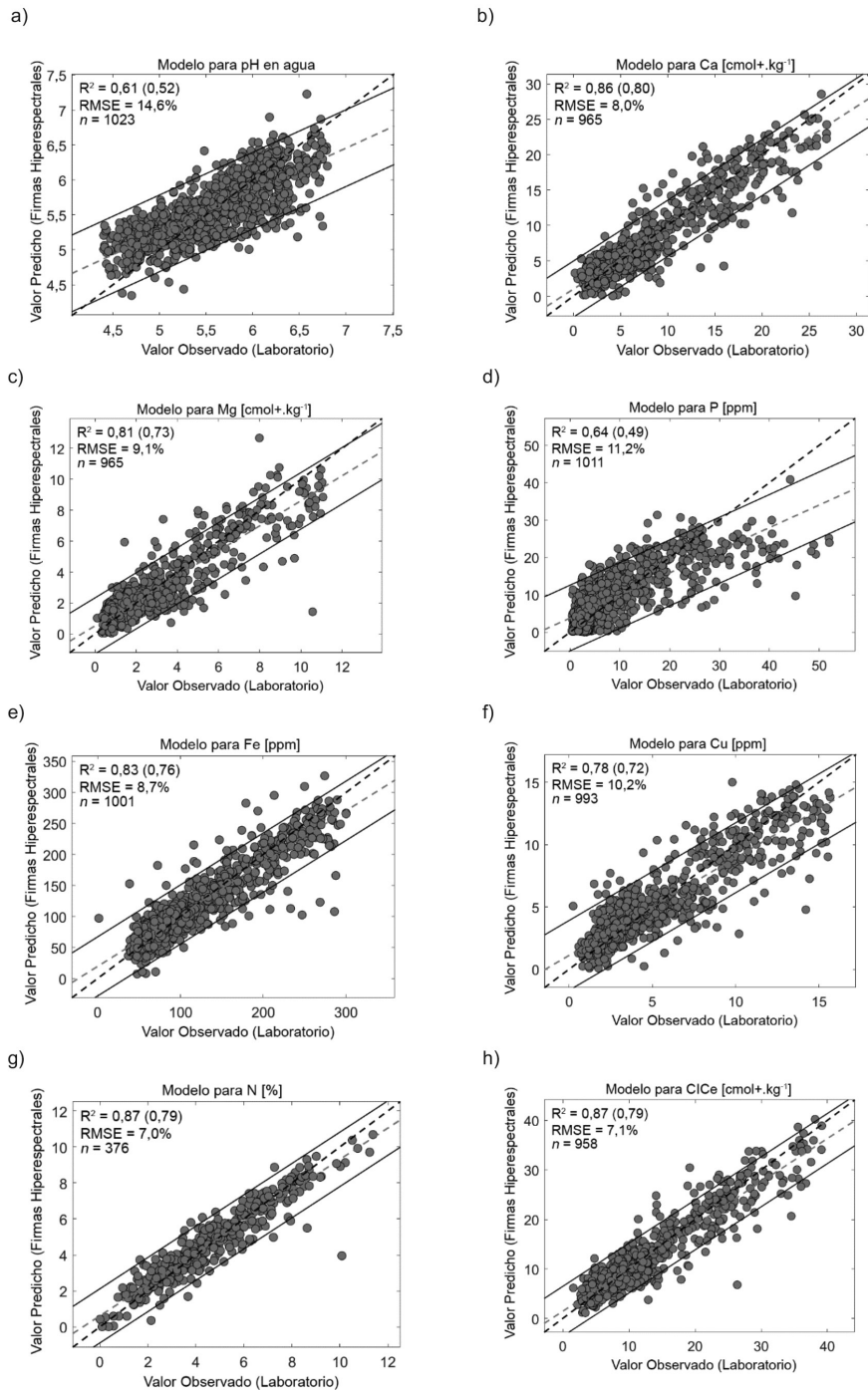


Figura 6. Gráficos de dispersión describiendo el ajuste de las predicciones con los resultados de laboratorio. a) pH del agua, b) calcio, c) magnesio, d) fósforo, e) hierro, f) cobre, g) nitrógeno y h) capacidad de intercambio catiónico efectivo para muestras de suelo pre-procesadas tomadas en Costa Rica mediante el modelo PLS.

La Tabla 2 aporta un resumen de la calidad de los modelos para la mayoría de las propiedades de suelo. El filtro aplicado a las muestras consistió en centrar el promedio a 2 desviaciones estándar, para eliminar firmas

irregulares. Algunos elementos y/o propiedades de suelo tenían menos análisis de laboratorio, por lo que se realizó la modelización con menos datos, que corresponden a la segunda columna de la Tabla 2.

Tabla 2. Estadísticas para la calibración y la validación externa de varios atributos de suelo mediante la aplicación de regresión de mínimos cuadrados parciales (PLSR).

Propiedad de suelo	Número de muestras después de aplicar filtro	R ² Calibración	R ₂ Validación	RMSE [%]	Número óptimo de componentes en Modelo PLSR
pH en agua	1023	0,61	0,52	14,57	29
AE [cmol+.kg ⁻¹]	965	0,53	0,41	15,86	27
K [cmol+.kg ⁻¹]	1010	0,54	0,37	14,85	33
Ca [cmol+.kg ⁻¹]	965	0,86	0,81	7,97	25
Mg [cmol+.kg ⁻¹]	965	0,81	0,73	9,14	28
P [ppm]	1011	0,64	0,49	11,2	33
Fe [ppm]	1001	0,83	0,76	8,66	31
Cu [ppm]	993	0,78	0,72	10,2	26
Zn [ppm]	1003	0,42	0,32	16,2	20
Na [cmol+.kg ⁻¹]	117	0,06	0,06	15,85	2
Mn [ppm]	992	0,65	0,58	14,87	20
Si [ppm]	27	0,56	0,4	12,32	3
B [ppm]	400	0,25	0,17	15,73	13
S [ppm]	416	0,50	0,36	11,1	17
C [%]	376	0,88	0,79	6,99	27
N [%]	376	0,87	0,79	7,00	27
Arena [%]	271	0,48	0,34	18,38	15
Limo [%]	266	0,49	0,34	15,17	17
Arcilla [%]	272	0,48	0,39	17,08	10
DA [g/cm ³]	67	0,36	0,21	18,21	6
CICe [cmol+.kg ⁻¹]	958	0,88	0,77	7,13	36
SB [%]	969	0,46	0,39	17,55	20
SA [%]	969	0,47	0,38	17,4	19

No se eliminaron datos de calibración para tratar de ajustar los modelos. Se generaron buenos modelos PLSR ($R^2 > 0,8$, RMSE < 10%) para Ca, Mg, Fe, C, N y CICE.

El número óptimo de componentes para el modelo PLSR se escogió basado en R^2 , esto debido a que es el más, comúnmente, utilizado en literatura, lo que permite realizar comparaciones entre los resultados obtenidos y otras investigaciones. Los resultados para los modelos de parámetros físicos de suelo como el porcentaje de arena, limo y arcilla, y densidad aparente indicaron una pobre capacidad de predictiva ($R^2 < 0,5$, RMSE entre 15-20%). Los resultados para carbono coincidieron con Reyna *et al.* (2017). Este análisis fue realizado con un número de 70 muestras, que aunque limitado se reportó que la modelización PLSR es útil para obtener información sobre el contenido de carbono. Kawamura *et al.* (2019) obtuvieron por su parte, predicciones para fósforo ($R^2 = 0,78$) al combinar un análisis PLS con un algoritmo genético con solamente 103 muestras. En el caso de esta investigación, se contó con una base de datos mucho más sólida con más de 1300 muestras.

La manipulación digital de cada propiedad de suelo, se identificó como limitante, ya que el tiempo promedio para evaluar los primeros 40 componentes de los modelos PLSR, con un procesador con 4 cores (Intel Core i7-4712HQ) y 16,0 GB de RAM, fue de 17 horas por cada propiedad de suelo. En ese sentido, Adeline *et al.* (2017) investigaron la posibilidad de reducir la resolución de las bandas hiperespectrales y obtuvieron buenos resultados. Ese proceso es conocido como degradación espectral y se presenta como una opción para llegar a reducir el tiempo de utilizado.

Con los modelos que se generaron, se pudieron predecir las variables de suelo analizado a partir de su firma hiperespectral con certidumbre, como fue en el caso de calcio, magnesio, hierro, carbono, nitrógeno y CICE, que reportaron R^2 de validación, superior al 0,8.

Se escribió un programa en MATLAB que usa los coeficientes PLSR generados para predecir cada propiedad de suelo con su firma espectral en menos de 5 minutos. Esta técnica facilitó una alternativa para aproximar rápidamente las deficiencias nutricionales que puede tener un suelo degradado a partir de muestras con un pre-procesamiento previo como secado y molido. A futuro, se considera incorporar técnicas de *machine learning* para cuantificar las propiedades de los suelos de Costa Rica y considerar así un mejor ajuste del modelo predictivo.

CONCLUSIONES

Los resultados obtenidos mostraron que la espectroscopía Vis-NIR-SWIR puede predecir de forma relativamente rápida en comparación a las técnicas convencionales de análisis de suelo con extracción en Mehlich III u Olsen y con espectroscopía de absorción atómica o espectroscopía de emisión por plasma de acoplamiento inductivo, una amplia gama de propiedades del suelo, independientemente de las diferentes ubicaciones geográficas en Costa Rica. A partir de los coeficientes PLSR encontrados en este estudio, se ha trabajado en un programa basado en MATLAB que permite predecir el valor de cada propiedad del suelo a partir de su firma hiperespectral. Se logró una estimación precisa del contenido de diferentes componentes (Ca, Mg, Fe, C, N y CICE) con un R^2 superior a 0,8 y un error cuadrático medio (RMSE) inferior a 10%.

Se identificó que los análisis espectroscópicos combinados con la regresión PLS pueden proporcionar una herramienta muy útil para la agricultura de precisión en los suelos tropicales de Costa Rica. A futuro, la capacidad predictiva se complementará con técnicas de pre-procesamiento, con el análisis de regresión PLS y otros algoritmos de aprendizaje automático, que permitan generar las bases para investigaciones en la predicción de las variables analizadas en campo.

AGRADECIMIENTOS

Las personas autoras desean agradecer el Sistema de Banca para el Desarrollo de Costa Rica por el financiamiento de este proyecto y el apoyo continuo al Centro de Agricultura de Precisión de la Universidad EARTH.

LITERATURA CITADA

- Adeline, KR; Gomez, C; Gorretta, N; Roger, JM. 2017. Predictive Ability of Soil Properties to Spectral Degradation from Laboratory Vis-NIR Spectroscopy Data. *Geoderma* 288:143-153.
- Angelopoulou, T; Balafoutis, A; Zalidis, G; Bochtis, D. 2020. From Laboratory to Proximal Sensing Spectroscopy for Soil Organic Carbon Estimation - A Review. *Sustainability* 12(2):443-467.
- Ben-Dor, E; Banin, A. 1995. Near-infrared analysis (NIRA) as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of American Journal* 59:364-372.
- Chang, CW; Laird, DA; Mausbach, MJ; Hurburgh, CR. 2001. Near-Infrared Reflectance Spectroscopy—Principal Components Regression Analyses of Soil Properties. *Soil Science Society of American Journal* 65:480-490.
- Curcio, D; Ciruolo, G; D'Asaro, F; Minacapilli, M. 2013. Prediction of Soil Texture Distributions Using VNIR-SWIR Reflectance Spectroscopy. *Procedia Environmental Sciences* 19:494-503.
- Demattê, JAM; Morgan, CLS; Chabrilat, S; Rizzo, R; Franceschini, MHD; Terra, FS; Vasques, GM; Wetterlind, J. 2016. Spectral Sensing from Ground to Space in Soil Science: State of the Art, Applications, Potential and Perspectives. *In* Thenkabail, PS (ed). *Remote Sensing Handbook*, vol. 2: Land Resources Monitoring, modeling and mapping with remote sensing. Florida, US, CRC press. p. 661-732.
- Hadoux, X; Gorretta, N; Roger, JM; Bendoula, R; Rabatel, G. 2014. Comparison of the efficacy of spectral pre-treatments for wheat and weed discrimination in outdoor conditions. *Journal of Computers and Electronics in Agriculture* 108:242-249.
- Kawamura, K; Tsujimoto, Y; Nishigaki, T; Andriamananjara, A; Rabenarivo, M; Asai, H; Rakotoson, T; Razafimbelo, T. 2019. Laboratory visible and near-infrared spectroscopy with genetic algorithm-based partial least squares regression for assessing the soil phosphorus content of upland and lowland rice fields in Madagascar. *Remote Sensing* 11(5):506-524.
- Liu, L; Ji, M; Dong, Y; Zhang, R; Buchroithner, M. 2016. Quantitative retrieval of organic soil properties from visible near-infrared shortwave infrared (Vis-NIR-SWIR) spectroscopy using fractal-based feature extraction. *Remote Sensing* 8(12):1-18.
- Miloš, B; Bensa, A. 2017. Prediction of soil organic carbon using VIS-NIR spectroscopy: Application to Red Mediterranean soils from Croatia. *Eurasian Journal of Soil Science* 6(4):365-373.
- Poppiel, RR; Lacerda, MPC; Junior, MP; Dematte, JAM; Romero, DJ; Júnior, LM; Sato, MV; Cassol, LFM. 2018. Surface Spectroscopy of Oxisols, Entisols and Inceptisol and Relationships with Selected Soil Properties. *Revista Brasileira de Ciência do Solo* 42:1-26.
- Reyna, L; Dube, F; Barrera J; Zagal, E. 2017. Potential model overfitting in predicting soil carbon content by visible and near-infrared spectroscopy. *Applied Sciences* 7:708-721.
- Rinnan, A; Van den Berg, F; Engelsen, SB. 2009. Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry* 28(10):1201-1222.
- Shepherd, KD; Walsh, MG. 2002. Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Science Society of America Journal* 66:988-998.
- Shibusawa, S; Imade Anom, SW; Sato, S; Sasao, A; Hirako, S. 2001. Soil mapping using the real-time soil spectrophotometer. *In* Grenier G; Blackmore, S (eds.). *Third European Conference on Precision Agriculture*. Montpellier, France, ECPA. p. 497-508.
- Silva, EB; Giasson, E; Dotto, AC; Caten, AT; Demattê, JAM; Bacic, ILZ; da Veiga, M. 2019. A regional legacy soil dataset for prediction of sand and clay content with VIS-NIR-SWIR, in southern Brazil. *Revista Brasileira de Ciência Do Solo* 43:1-20.
- Steinberg, A; Chabrilat, S; Stevens, A; Segl, K; Foerster, S. 2016. Prediction of Common Surface Soil Properties Based on Vis-NIR Airborne and Simulated EnMAP Imaging Spectroscopy Data: Prediction Accuracy and Influence of Spatial Resolution. *Remote Sensing* 8:613-633.
- Stevens, A; Udelhoven, T; Denis, A; Tychon, B; Liroy, R; Hoffmann, L; van Wesemael, B. 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma* 158:32-45.
- Viscarra Rossel, RA; Behrens, T; Ben-Dor, E; Brown, DJ; Demattê, JAM; Shepherd, KD; Shi, Z; Stenberg, B; Stevens, A; Adamchuk, V; Aichi, H; Barthès, BG; Bartholomeus, HM; Bayer, AD; Bernoux, M;

Böttcher, K; Brodský, L; Du, CW; Chappell, A; Fouad, Y; Genot, V; Gomez, C; Grunwald, S; Gubler, A; Guerrero, C; Hedley, CB; Knadel, M; Morrás, HJM; Nocita, M; Ramirez-Lopez, L; Roudier, P; Campos, EMR; Sanborn, P; Sellitto, VM; Sudduth, KA; Rawlins, BG; Walter, C; Winowiecki, LA; Hong, SY; Ji, W. 2016. A global spectral library to

characterize the world's soil. *EARTH Earth-Science Reviews* 155:198-230.

Viscarra Rossel, RA; Walvoort, DJ; McBratney, AB; Janik, LJ; Skjemstad, JO. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131:59-75.



Todos los derechos reservados. Universidad de Costa Rica. Este artículo se encuentra licenciado con Creative Commons Reconocimiento-NoComercial-SinObraDerivada 3.0 Costa Rica. Para mayor información escribir a rac.cia@ucr.ac.cr